Signal Processing in Space and Time - A Multidimensional Fourier Approach

THÈSE N^{O} 4871 (2010)

PRÉSENTÉE LE 3 DÉCEMBRE 2010

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

LABORATOIRE DE COMMUNICATIONS AUDIOVISUELLES 1

PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Francisco PEREIRA CORREIA PINTO

acceptée sur proposition du jury:

Prof. E. Telatar, président du jury
Prof. M. Vetterli, directeur de thèse
Prof. D. L. Jones, rapporteur
Prof. J. M. F. Moura, rapporteur
Prof. M. Unser, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2010

Abstract

Sound waves propagate through space and time by transference of energy between the particles in the medium, which vibrate according to the oscillation patterns of the waves. These vibrations can be captured by a microphone and translated into a digital signal, representing the amplitude of the sound pressure as a function of time. The signal obtained by the microphone characterizes the time-domain behavior of the acoustic wave field, but has no information related to the spatial domain. The spatial information can be obtained by measuring the vibrations with an array of microphones distributed at multiple locations in space. This allows the amplitude of the sound pressure to be represented not only as a function of time but also as a function of space.

The use of microphone arrays creates a new class of signals that is somewhat unfamiliar to Fourier analysis. Current paradigms try to circumvent the problem by treating the microphone signals as multiple "cooperating" signals, and applying the Fourier analysis to each signal individually. Conceptually, however, this is not faithful to the mathematics of the wave equation, which expresses the acoustic wave field as a single function of space and time, and not as multiple functions of time.

The goal of this thesis is to provide a formulation of Fourier theory that treats the wave field as a single function of space and time, and allows it to be processed as a multidimensional signal using the theory of digital signal processing (DSP). We base this on a physical principle known as the Huygens principle, which essentially says that the wave field can be sampled at the surface of a given region in space and subsequently reconstructed in the same region, using only the samples obtained at the surface. To translate this into DSP language, we show that the Huygens principle can be expressed as a linear system that is both space- and time-invariant, and can be formulated as a convolution operation. If the input signal is transformed into the spatio-temporal Fourier domain, the system can also be analyzed according to its frequency response.

In the first half of the thesis, we derive theoretical results that express the 4-D Fourier transform of the wave field as a function of the parameters of the scene, such as the number of sources and their locations, the source signals, and the geometry of the microphone array. We also show that the wave field can be effectively analyzed on a small scale using what we call the space/time–frequency representation space, consisting of a Gabor representation across the spatio-temporal manifold defined by the microphone array. These results are obtained by treating the signals as continuous functions of space and time.

The second half of the thesis is dedicated to processing the wave field in discrete space and time, using Nyquist sampling theory and multidimensional filter banks theory. In particular, we show examples of orthogonal filter banks that effectively represent the wave field in terms of its elementary components while satisfying the requirements of critical sampling

and perfect reconstruction of the input. We discuss the architecture of such filter banks, and demonstrate their applicability in the context of real applications, such as spatial filtering and wave field coding.

Keywords: Acoustic wave fields, array signal processing, multidimensional Fourier analysis, space/time–frequency analysis, multidimensional filter banks, directional filter banks, spatial filtering, wave field coding.

Résumé

Les ondes sonores se propagent sur l'espace et le temps à travers de transfert d'énergie entre les particules du milieu, qui vibrent selon les caractéristiques des ondes. Ces vibrations peuvent être captées par un microphone et traduites dans un signal numérique, ce qui représente l'amplitude de la pression acoustique en fonction du temps. Le signal obtenu par le microphone décrit la variation temporelle du champ d'ondes acoustiques, mais il n'a pas d'informations liées au domaine spatial. L'information spatiale peut être obtenue en mesurant les vibrations avec un réseau de microphones distribués par divers endroits dans l'espace. Cela permet à l'amplitude de la pression acoustique de se faire représenter non seulement en fonction du temps mais aussi en fonction de l'espace.

L'utilisation des réseaux de microphones crée une nouvelle classe de signaux qui sont peu familiers à l'analyse de Fourier. Les paradigmes actuels essayent de contourner le problème en interprétant les signaux des microphones ainsi que les signaux "coopérants", et en appliquant l'analyse de Fourier à chaque signal individuel. Théoriquement, toutefois, ce n'est pas fidèle aux mathématiques de l'équation des ondes, qui exprime le champ d'ondes acoustiques en utilisant une fonction unique de l'espace et du temps, et pas comme des multiples fonctions du temps.

L'objectif de cette thèse est présenter une formulation de la théorie de Fourier qui manipule le champ d'ondes en utilisant une fonction unique de l'espace et du temps, et lui permet d'être manipulé comme un signal multidimensionnel à l'aide de la théorie du traitement des signaux numériques (TSN). Nous nous basons sur un principe physique appelé le principe de Huygens, qui dit essentiellement que le champ d'ondes peut être mesuré à la surface d'une région dans l'espace, et puis reconstruit dans la même région, en utilisant seulement des mesurages obtenus à la surface. Pour traduire en langage TSN, nous montrons que le principe de Huygens peut être exprimé dans un système linéaire qui est invariant dans l'espace et dans le temps, et peut être formulé comme une opération de convolution. Si le signal est transformé dans le domaine spatio-temporel de Fourier, le système peut également être analysé en fonction de sa réponse en fréquence.

Dans la première moitié de la thèse, nous obtenons des résultats théoriques qui expriment la transformée de Fourier 4-D du champ d'ondes en fonction des paramètres de la scène, tels que le nombre de sources et de leur emplacement, les signaux des sources, et le géométrie du réseau de microphones. Nous montrons aussi que le champ d'ondes peut être analysé efficacement sur une petite échelle en utilisant ce que nous appelons la représentation espace/temps-fréquence, composé d'une représentation de Gabor au cours de la "manifold" spatio-temporelle définie par le réseau de microphones. Ces résultats sont obtenus en interprétant les signaux comme des fonctions continues de l'espace et du temps.

La seconde moitié de la thèse est consacrée au traitement du champ d'ondes dans l'espace

et dans le temps discrets, en utilisant la théorie de mesurage de Nyquist et la théorie des bancs de filtres multidimensionnels. En particulier, nous montrons des exemples de bancs de filtres orthogonaux qui représentent effectivement le champ d'ondes en fonction de ses composants élémentaires, tout en satisfaisant aux exigences de mesurage critique et reconstruction parfaite de l'entrée. Nous discutons l'architecture de ces bancs de filtres, et nous démontrons leur applicabilité dans le contexte d'applications réelles, telles que le filtrage spatial et la codification du champ d'ondes.

Mots-clés: Champs d'ondes acoustiques, traitement des signaux réseaux, analyse de Fourier multidimensionnelle, analyse espace/temps–fréquence, bancs de filtres multidimensionnels, bancs de filtres directionnels, filtrage spatial, codification du champ d'ondes.

Acknowledgements

This thesis is the crown-jewel of a 4-year collaboration with the Audiovisual Communications Laboratory under the supervision of Prof. Martin Vetterli. It has been a real honor to work all these years in such a vibrant and prolific research environment, surrounded by so many brilliant minds and inspiring personalities.

Needless to say, none of this would be possible without Martin's first-class supervision. He has influenced my way of thinking in countless ways, and greatly contributed to my personal growth. He is a visionary, a motivator, and has a contagious sense of humor. An outstanding human being on all possible levels. A special thanks to Christof Faller, who taught me everything relevant about audio coding, trained me in the art of being pragmatic about research, and who is simply an great friend. It was a fun experience to teach alongside with Andrea Ridolfi, who has always stunned me with his ability to climb up mountains all over the weekend and still passionately teach math on Monday morning to a crowd of sleepy students. I hope we keep doing our "bacalhau" dinners at the Portuguese restaurant in the future! Many thanks to Profs. Emre Telatar, Michael Unser, José Moura, and Doug Jones for kindly accepting to be part of my thesis committee, Prof. Moe Win for introducing me to Martin and for all the great advise on Ph.D. programs, and Luciano Sbaiz for his help with... well, random stuff! And of course, many thanks to Jocelyne and Jacqueline for their dedication to the lab and its people.

A very special thanks to all my friends in Portugal, Switzerland, and other corners in the world. My "wewofos" family Eva, Joana, and Ana, for 10 years of memorable moments and unshakable friendship. Tiago, for all the "jorda-jarda" and the monthly erudite conversations. Radha, Roberto, Paco, and Alexandra, for all the great moments we've spent together and all the tasty food we've shared. Ayush, for all the fun discussions, the dinners with randomly selected people, and simply for letting me mess up with your head. Kerstin, for being my loyal tango partner. Florence, Amina, Mihailo, and everyone else who has crossed my life during these four years.

Last but not least, a very special thanks to all my family. My parents, who have supported me unconditionally throughout my entire life. I hope I can keep making you proud. My brother, who is still my greatest friend and influence, and has always put me on the right track. My grand-parents, uncles, cousins, and everyone else who has watched me grow.

I would like to express my gratitude to the Portuguese Science and Technology Foundation (FCT) for supporting this research with the grant SFRH/BD/27257/2006, the Swiss National Science Foundation for the grant 200021-121935, and the ERC Support for Frontier Research for the grant 247006.

Glossary

Continuous variables

x, y, z	Spatial positions
t	Time
$\mathbf{r}$	Coordinate vector on a 3-D Euclidean space: $\mathbf{r} = (x, y, z)$
Φ_x, Φ_y, Φ_z	Spatial frequencies in rad/m
Ω	Temporal frequency in rad/s
$\mathbf{\Phi}$	Spatial frequency vector: $\mathbf{\Phi} = (\Phi_x, \Phi_y, \Phi_z)$

Discrete variables

n_x, n_y, n_z	Spatial sample indexes
n_t	Time sample index
$\mathbf{n}$	Sample index vector: $\mathbf{n} = (n_x, n_y, n_z, n_t)$
b_x, b_y, b_z	Spatial transform coefficient indexes
b_t	Temporal transform coefficient index
$\mathbf{b}$	Transform coefficient index vector: $\mathbf{b} = (b_x, b_y, b_z, b_t)$
N_x, N_y, N_z	Number of samples in space
N_t	Number of samples in time
$\mathbf{N}$	Diagonal matrix with number of samples N_x, N_y, N_z, and N_t

Spaces and regions

$\mathbb{R}$	Space of real numbers
$\mathbb{Z}$	Space of integer numbers
$\mathbb{A}$	Space of angles between 0 and π
$\mathbb{I}$	Space of transform block indexes
$\mathbb{K}$	Space of coset vectors
ℓ_1	Hilbert space of integrable functions
ℓ_2	Hilbert space of square-integrable functions
$\mathcal{V}$	Volume in the 3-D Euclidean space
$\mathcal{S}$	Surface in the 3-D Euclidean space
$\mathcal{A}$	Area in the 3-D Euclidean space
$\mathcal{L}$	Contour in the 3-D Euclidean space
$\mathcal{V}_\mathrm{a}$, $\mathcal{V}_\mathrm{s}$	Analysis and synthesis volumes
$\mathcal{S}_\mathrm{a}$, $\mathcal{S}_\mathrm{s}$	Analysis and synthesis surfaces
$\mathcal{L}_\mathrm{a}$, $\mathcal{L}_\mathrm{s}$	Analysis and synthesis contours
$\mathcal{U}$	Low-energy region of the spatio-temporal spectrum

Wave field analysis

Variables

α	Angle of arrival of a far-field wave front to the x-axis, with $\alpha \in [0, \pi]$
β	Angle of arrival of a far-field wave front to the y-axis, with $\beta \in [0, \pi]$
$\boldsymbol{\alpha}$	Vector of far-field arrival angles: $\boldsymbol{\alpha} = (\alpha, \beta)$
r	Radius
$\mathbf{r}_p$	Source position: $\mathbf{r}_p = (x_p, y_p, z_p)$
$\mathbf{r}_\mathrm{wall}$	Position where the wave front first hits a wall
$\mathbf{r}_\mathcal{S}$	Position at surface region $\mathcal{S}$: $\mathbf{r}_\mathcal{S} = (x_\mathcal{S}, y_\mathcal{S}, z_\mathcal{S})$
$\mathbf{r}_\mathcal{L}$	Position at contour region $\mathcal{L}$: $\mathbf{r}_\mathcal{L} = (x_\mathcal{L}, y_\mathcal{L}, z_\mathcal{L})$
$\boldsymbol{n}_\mathcal{S}$	Normal vector at surface region $\mathcal{S}$
$\boldsymbol{n}_\mathcal{L}$	Normal vector at contour region $\mathcal{L}$

$\boldsymbol{v}$	Coordinate vector on a 2-D Euclidean space: $\boldsymbol{v} = (v_0, v_1)$
$\mathbf{u}$	Wave front propagation vector: $\mathbf{u} = (u_x, u_y, u_z)$, where, in polar coordinates, $u_x = \cos\alpha \sin\beta$, $u_y = \cos\beta$, and $u_z = \sin\alpha \sin\beta$
$\alpha_{\mathrm{nf}}(x)$	Angle of incidence of a near-/intermediate-field wave front at point x: $\alpha_{\mathrm{nf}}(x) = \angle\big(x_p - x + jz_p\big)$

Functions

$p(r,t)$	Radial sound pressure
$g(x,t)$	2-D spatio-temporal Green's function
$g(\mathbf{r},t)$	4-D spatio-temporal Green's function
$g(r,\Omega)$	Temporal Fourier transform of radial Green's function
$\mathcal{G}(\mathbf{r},\Omega)$	Wave field synthesis operator

Surface convolution

$\varphi(\boldsymbol{v})$	Map of 2-D vector $\boldsymbol{v} = (v_0, v_1)$ to the 3-D coordinates of a surface $\mathcal{S}$
$\varphi_{\mathrm{a}}(\boldsymbol{v})$	Map of 2-D vector $\boldsymbol{v} = (v_0, v_1)$ to the 3-D coordinates of the analysis surface $\mathcal{S}_{\mathrm{a}}$
$\varphi_{\mathrm{s}}(\boldsymbol{v})$	Map of 2-D vector $\boldsymbol{v} = (v_0, v_1)$ to the 3-D coordinates of the synthesis surface $\mathcal{S}_{\mathrm{a}}$
$f_{\mathrm{a}}(\mathbf{r},\Omega)$	Analysis function in the surface convolution
$f_{\mathrm{s}}(\mathbf{r},\Omega)$	Synthesis function in the surface convolution
$f(\mathbf{r},\Omega)$	Output of the surface convolution
$\delta_{\mathcal{S}}(\mathbf{r})$	Sampling kernel on a surface $\mathcal{S}$ in the 3-D Euclidean space
$h(\boldsymbol{v},t)$	Spatio-temporal filter on a flat surface

Various parameters

c	Speed of sound: $c = 340\mathrm{m/s}$
m	Number of spatial dimensions
$\Phi_{S,x}$, $\Phi_{S,y}$, $\Phi_{S,z}$	Spatial sampling frequencies in rad/m
Ω_S	Temporal sampling frequency in rad/s
$\boldsymbol{\Phi}_S$	Vector of spatial sampling frequencies: $\boldsymbol{\Phi}_S = (\Phi_{S,x}, \Phi_{S,y}, \Phi_{S,z})$
Ω_M	Maximum temporal frequency

f_S	Temporal sampling frequency in Hz
T_x, T_y, T_z	Spatial sampling periods
T_t	Temporal sampling period
$\alpha_\mathrm{nf}^\mathrm{min}$	Minimum angle of incidence of near-/intermediate-field wave front with the microphone array (or observation contour)
$\alpha_\mathrm{nf}^\mathrm{max}$	Maximum angle of incidence of near-/intermediate-field wave front with the microphone array (or observation contour)
Φ_x^min	Minimum spatial frequency of high-energy spectral region
Φ_x^max	Maximum spatial frequency of high-energy spectral region
b_x^min	Minimum transform coefficient index of high-energy spectral region
b_x^max	Maximum transform coefficient index of high-energy spectral region

Signals and Sequences

Continuous

$s(t)$	Source signal
$p(x,t)$	Sound pressure on the x-axis
$p(x,y,t)$	Sound pressure on the xy-plane
$p(\mathbf{r},t)$	Sound pressure in the xyz-space
$p(\mathbf{r},\Omega)$	Temporal Fourier transform of sound pressure
$p(\mathbf{r}_\mathcal{S},t)$	Sound pressure on a surface $\mathcal{S}$
$p(\mathbf{r}_\mathcal{L},t)$	Sound pressure on a contour $\mathcal{L}$
$S(\Omega)$	Source spectrum
$P(\Phi_x,\Omega)$	Spatio-temporal spectrum on the x-axis
$P(\Phi_x,\Phi_y,\Omega)$	Spatio-temporal spectrum on the xy-plane
$P(\mathbf{\Phi},\Omega)$	Spatio-temporal spectrum in the xyz-space
$P(\alpha,\Omega)$	Directional spectrum on the α-axis
$P(\beta,\Omega)$	Directional spectrum on the β-axis
$P(\boldsymbol{\alpha},\Omega)$	Directional spectrum on the $\alpha\beta$-plane
$P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)$	Short space/time Fourier transform

Discrete

$s[n_t]$ Source sequence

$p[n_x, n_t]$ Sound pressure on the x-axis

$p[n_x, n_y, n_t]$ Sound pressure on the xy-plane

$p[\mathbf{n}]$ Sound pressure in the $xyzt$-space

$p[n_x, n_y, n_z, b_t]$ Temporal DFT of sound pressure

$S[b_t]$ Source transform

$P[b_x, b_t]$ Spatio-temporal transform on the x-axis

$P[b_x, b_y, b_t]$ Spatio-temporal transform on the xy-plane

$P[\mathbf{b}]$ Spatio-temporal transform in the $xyzt$-space

$P[\mathbf{n}_0, \mathbf{b}]$ Short space/time Fourier transform

Window functions

Continuous

$w_x(x)$, $w_y(y)$, $w_z(z)$ Window functions in space

$w_t(t)$ Window function in time

$w_{\mathbf{r}}(\mathbf{r})$ Spatial window function

$w(\mathbf{r}, t)$ Spatio-temporal window function

$W_x(\Phi_x)$, $W_y(\Phi_y)$, $W_z(\Phi_z)$ Window spectra in space

$W_t(\Omega)$ Window spectrum in time

$W_{\mathbf{r}}(\mathbf{\Phi})$ Spatial window spectrum

$W(\mathbf{\Phi}, \Omega)$ Spatio-temporal window spectrum

$M(\Phi_x, \Omega)$ Near-/intermediate-field spectral mask

Discrete

$w_x[n_x]$, $w_y[n_y]$, $w_z[n_z]$ Window sequences in space

$w_t[n_t]$ Window sequence in time

$w[\mathbf{n}]$ Spatio-temporal window sequence

$W_x[b_x]$, $W_y[b_y]$, $W_z[b_z]$ Window transforms in space

$W_t[b_t]$ Window transform in time

$W[\mathbf{b}]$	Spatio-temporal window transform
$L_x,\ L_y,\ L_z$	Number of non-zero samples in space
L_t	Number of non-zero samples in time
$\mathbf{L}$	Diagonal matrix with number of non-zero samples L_x, L_y, L_z, and L_t

Filtering

Continuous

$h_x(x),\ h_y(y),\ h_z(z)$	Impulse responses in space
$h_t(t)$	Impulse response in time
$h_{\mathbf{r}}(\mathbf{r})$	Spatial impulse response
$h(\mathbf{r}, t)$	Spatio-temporal impulse response
$H_x(\Phi_x),\ H_y(\Phi_y),\ H_z(\Phi_z)$	Frequency responses in space
$H_t(\Omega)$	Frequency response in time
$H_{\mathbf{r}}(\mathbf{\Phi})$	Spatial frequency response
$H(\mathbf{\Phi}, \Omega)$	Spatio-temporal frequency response

Discrete

$h_x[n_x],\ h_y[n_y],\ h_z[n_z]$	Impulse responses in space
$h_t[n_t]$	Impulse response in time
$h[\mathbf{n}]$	Spatio-temporal impulse response
$H_x[b_x],\ H_y[b_y],\ H_z[b_z]$	Frequency responses in space
$H_t[b_t]$	Frequency response in time
$H[\mathbf{b}]$	Spatio-temporal frequency response

Filter design

$h_{\mathrm{ideal}}[\mathbf{n}]$	Ideal impulse response
$H_{\mathrm{ideal}}[\mathbf{b}]$	Ideal frequency response
α_{Pass}	Cut-off angle of pass-band region
α_{Stop}	Cut-off angle of stop-band region
$\mathrm{A}_{\mathrm{Pass}}$	Maximum approximation error in pass-band region
$\mathrm{A}_{\mathrm{Stop}}$	Stop-band attenuation

Beamforming

α_{S}	Beam-steering angle
$\Delta_{\mathrm{FF}}^{\mathrm{Signal}}[\mathbf{b}]$	Spectral sampling kernel for extracting a far-field target source
$\Delta_{\mathrm{FF}}^{\mathrm{Noise}}[\mathbf{b}]$	Spectral sampling kernel for extracting a far-field interferer
$\Delta_{\mathrm{NF}}^{\mathrm{Signal}}[\mathbf{b}]$	Spectral sampling kernel for extracting a near-field target source
$\Delta_{\mathrm{NF}}^{\mathrm{Noise}}[\mathbf{b}]$	Spectral sampling kernel for extracting a near-field interferer

Bases and matrices

υ_{b_x,n_x}	Spatial orthogonal basis
ψ_{b_t,n_t}	Temporal orthogonal basis
$\varphi[\mathbf{b},\mathbf{n}]$	Separable spatio-temporal orthogonal basis: $\varphi[\mathbf{b},\mathbf{n}] = \upsilon_{b_x,n_x}\psi_{b_t,n_t}$
$\mathbf{P}$	Matrix expansion of $p[\mathbf{n}]$
$\mathbf{Y}$	Matrix expansion of $P[\mathbf{b}]$
$\boldsymbol{\Upsilon}$	Matrix expansion of υ_{b_x,n_x}
$\boldsymbol{\Psi}$	Matrix expansion of ψ_{b_t,n_t}
$\boldsymbol{\Upsilon}_L,\ \boldsymbol{\Upsilon}_R$	Left and right halves of $\boldsymbol{\Upsilon}$
$\boldsymbol{\Psi}_L,\ \boldsymbol{\Psi}_R$	Left and right halves of $\boldsymbol{\Psi}$

Filter banks

$z_x,\ z_t$	z-transform variables in space and time
$\mathbf{z}$	Vector with z-transform variables: $\mathbf{z} = (z_x, z_t)$
$H(z)$	1-D analysis filter
$F(z)$	1-D synthesis filter
$H(\mathbf{z})$	Multidimensional analysis filter
$F(\mathbf{z})$	Multidimensional synthesis filter
$\mathbf{H}$	1-D analysis polyphase matrix
$\mathbf{F}$	1-D synthesis polyphase matrix
$\prod_m \mathbf{H}_m$	Separable product of analysis polyphase matrices
$\prod_m \mathbf{F}_m$	Separable product of synthesis polyphase matrices

O_x, O_t	Number of overlapping samples in space and time
$\mathbf{O}$	Diagonal matrix with number of overlapping samples in space and time, O_x and O_t
$\mathbf{R}$	Parallelogram resampling matrix (non-diagonal)
$\mathbf{Q}$	Quincunx resampling matrix (non-diagonal)
$\mathbf{k}$	Spatio-temporal coset vector: $\mathbf{k} = \left[\begin{smallmatrix} k_x \\ k_t \end{smallmatrix}\right]$
$\mathbf{z}^{-\mathbf{k}}$	Spatio-temporal delay factor: $\mathbf{z}^{-\mathbf{k}} = z_x^{-k_x} z_t^{-k_t}$
$\mathbf{i}$	Spatio-temporal block index: $\mathbf{i} = \left[\begin{smallmatrix} i_x \\ i_t \end{smallmatrix}\right]$
$\mathbf{0}$, $\mathbf{1}$, ...	Vector integers: $\mathbf{0} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, $\mathbf{1} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, ...

Wave field coding

N_{bit}	Number of bits of amplitude quantizer
$p[\mathbf{n}]$	Input signal
$\hat{p}[\mathbf{n}]$	Decoded signal
$P[\mathbf{b}]$	Input transform
$\hat{P}[\mathbf{b}]$	Decoded transform
$P_{\mathrm{Q}}[\mathbf{b}]$	Quantized transform coefficients
$\mathrm{SF}[\mathbf{b}]$	Scale factors
$Q[\mathbf{b}]$	Maximum quantization noise power
$R[\mathbf{b}]$	Bit rate for each transform coefficient
R	Total bit rate
D	Total distortion
$D(R)$	Rate-distortion curve

Operators

$**$	Spatial convolution
$*$	Temporal convolution
$\circledast$	Surface convolution
$L_{(\mathbf{r}_0, t_0)}$	Delay operator

$\mathbb{E}$	Expectation
$((\cdot))_N$	Circular referencing: $((\cdot))_N = \cdot \bmod N$
$\det$	Matrix determinant
max	Maximum operator
sign	Sign operator
mean	Empirical mean
var	Empirical variance
probability	Empirical probability
Huffman	Map of amplitude value to Huffman code word

Contents

Chapter 1

Introduction

Signal processing is a vibrant field of engineering and mathematics that attracts a lot of attention from both the industrial and the academic world, causing the field to be in a permanent state of evolution, critical analysis, and rediscovery. In the last century, we have seen the emergence of revolutionary technologies and exciting new inventions that strongly influenced the course of technology. Inventions such as voice recording and reproduction, photography and motion picture, radio transmission, among others, have used the power of signal processing and changed the way we store and communicate information across the planet. The theory of signal processing has followed suit, and new algorithms and methods were developed by taking advantage of the technological advances.

In the literature, signal processing is loosely defined as the science that studies the representation, analysis, and manipulation of signals representing physical phenomena [68]. A signal is not a physical phenomenon in itself but rather a representation of the phenomenon provided by the measuring instruments. It is the printed record of physical events, and the instruments define what is the printing language.

1.1 Sound and audio signals

A signal may represent a multiplicity of physical measures, such as sound pressure, temperature, particle velocity, light intensity, among others. Audio signals, in particular, were one of the great catalysts in the development of signal processing. Since ancient times, we have depended on different forms of sound for communication, entertainment, and welfare. By mastering the art of audio signal processing, we have boosted our capability of using sounds to communicate at ever greater distances, to create innumerable forms of entertainment, and to bring societies closer together.

Modern audio signal processing came to life as a fusion between three major breakthroughs: (i) the advances in the study of harmonic analysis, (ii) the development of sound recording technology, and (iii) the advances in digital computing. The first two occurred almost independently, until the increase in computational power made it possible to apply the principles of harmonic analysis to the recorded signals.

1.1.1 The birth of harmonic analysis

In the late 18th century, and throughout the entire 19th century, a brilliant generation of physicists and mathematicians was working on the concepts that became the theoretical foundation of signal processing. One of the first important contributions was provided by the French mathematician Jean d'Alembert (1717–1783), who not only introduced the concept of "field" (in the context of fluid dynamics) but also derived the first known form of the wave equation, together with its solution [23]. Immediately after, the Swiss mathematician Leonhard Euler (1707–1783) expanded d'Alembert's results to a general framework that explained a multitude of wave-like phenomena, including acoustic wave propagation [31]. This work sparked a famous debate between d'Alembert and Euler about the nature of functions [44; 52]; d'Alembert argued that only analytical functions could be a solution to the wave equation, whereas Euler argued that any "hand drawn" function, including discontinuous functions, could be a solution to the wave equation in the physical sense. He also derived an important solution for a particular type of analytic functions, given by

$$f(x) = a_0 \sin \frac{\pi x}{l} + a_1 \sin \frac{2\pi x}{l} + a_2 \sin \frac{3\pi x}{l} + \cdots . \tag{1.1}$$

The debate went on for half a century, until the French mathematician Joseph Fourier (1768–1830) partially proved that any arbitrary function could be developed into a trigonometric series of the form given by (1.1). In essence, Fourier showed that the "hand drawn" functions that Euler believed had a physical significance could be decomposed into elementary analytical components (sinusoids), and therefore be a plausible solution to the wave equation. This marked the birth of harmonic analysis.

1.1.2 The invention of sound recording

While mathematicians kept on discovering and perfecting the theoretical concepts of harmonic analysis, another great discovery was about to take place. Before the 19th century, writing and painting were the only instruments available for keeping records of the world events; there was no instrument available for recording sounds such as voice and music. The sound of Mozart playing at the royal court, or of Alexander giving a speech to his people, are essentially lost in time.

Curiously, the first record ever found of human voice was of an amplitude wave form scribbled on a piece of paper [79], obtained by the French bookseller Édouard de Martinville (1817–1879) with a device he called phonautograph (shown in Figure 1.1-(a)). It was later discovered that the wave form represented an anonymous voice singing the French song *Au Clair de la Lune*[1].

The real invention, however, came 20 years later, when the American scientist Thomas Edison (1847–1931) introduced the phonograph [28]—a device capable not only of recording sound but also of reproducing it with an impressive quality. The phonograph (shown in Figure 1.1-(b)) used a tinfoil sheet cylinder to imprint the amplitude wave form such that it could be read by a copper needle.

This concept was later improved by other inventors, such as Oberlin Smith, who pioneered the concept of magnetic recording [29], and Emile Berliner, who initiated a 50-year

[1] The sound file can be downloaded from: http://en.wikipedia.org/wiki/Phonautograph/

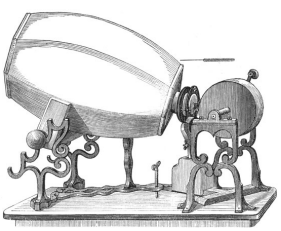


Image by F. Pisko

(a)

Image by N. Bruderhofer

(b)

Figure 1.1: First sound recording devices. (a) The phonautograph, invented by Édouard de Martinville in 1857, was capable of translating sound into an amplitude wave form printed on paper. (b) The phonograph, invented by Thomas Edison in 1877, recorded the amplitude wave form on a dedicated cylinder, and was capable of reproducing the recorded sound.

long transition from cylinders to discs as a recording medium [12]. These contributions culminated in the development of the Compact Cassette and the Long Playing (LP) recording media, which together dominated the market until the rise in popularity of the Compact Disc (CD) [66].

1.1.3 Digital computers and the DSP boom

Audio signal processing became well established as a scientific field when digital computers became fast enough to apply Fourier's principles to the recorded sound. The crown jewel of this period was the development of the Fast Fourier Transform (FFT), which made it possible to compute the Fourier transform with relatively low complexity using a digital computer. This allowed engineers to experiment with a variety of signal processing techniques, such as digital filters and spectral estimation, while manipulating the recorded audio signals. Digital signal processing (DSP) has since become an essential component of any electronic device that operates with audio signals, including cellphones, televisions, radios, and portable music players.

1.2 From mono to spatial audio

Almost as soon as Thomas Edison invented the technology for sound recording, people started to think of ways of recording (and communicating) spatial sound—*i.e.*, sound with preserved spatial impression. A curious comment made by an anonymous person to a British magazine goes as follows [83]

> «*When I sit in an acoustically perfect hall (full of people), in the best seat for hearing, and listen to an orchestra, I hear such and such sounds. I want to hear precisely this effect from a recording, if that be possible.*»—An anonymous music critic ("K. K.") writing in the 1928 July issue of The Gramophone.

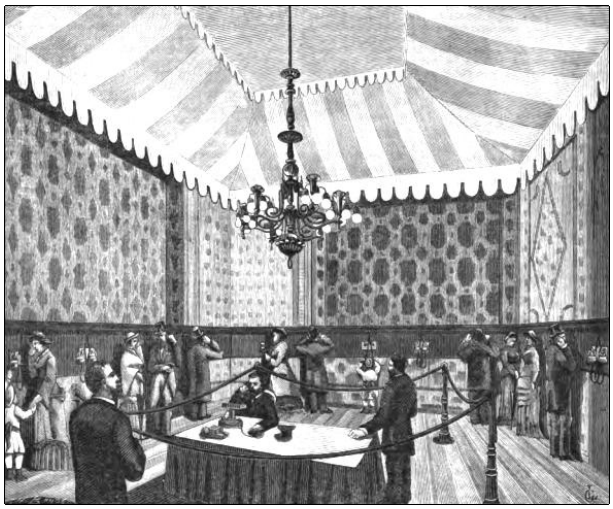


Image source: Nature, Oct. 20 Issue, 1881

Figure 1.2: First experience of a stereo audio transmission, at the Paris Opera in 1881. The transmission of the two channels was made with two telephones. The listener would experience a spatial impression by placing one speaker in each ear.

In his own words, "I want to hear precisely this effect", "K. K." was alluding to the listeners desire for a more realistic and immersive auditory experience. In fact, it was only four years after Edison filed his gramophone patent that the first demonstration of a stereo transmission was shown at the Paris Grand Opera, in 1881 [83]. The two audio channels were transmitted by two telephones, which together generated a spatial impression when the listener held one of the speakers to each ear (see Figure 1.2). The technique eventually evolved to the modern form of stereo that is still used today.

Many other techniques were experimented and introduced in the market, mostly with no success. The one that achieved the most popularity was the 5.1 surround format used in movie theaters, but it was not widely adopted by home users. Other recent techniques include the 7.1 surround format (and the variants 9.1, 10.2, and 22.2) [80], stereo based on HRTF (head-related transfer functions) [36], ambisonics [37], and wave field synthesis [9].

Regardless of all the technological differences between spatial audio techniques, the general trend is to increase the number of channels. In fact, there is a theoretical motivation for this increase: a physical principle known as the *Huygens principle*—discussed later in this thesis—suggests that the ideal playback setup is a continuous surface of loudspeakers (or a line, under certain conditions). In other words, the more channels are used, the better and more accurate is the spatial impression.

Yet, more than 80 years after "K. K." expressed his desire for "an orchestra in one recording", there is still a lot of uncertainty about which spatial audio format is going to succeed in replacing the well established stereo format. This creates uncertainty about the costs of implementation and compatibility issues, which makes the industry reluctant to adopt new formats. As a result, sound records are still mostly recorded using two channels, which is hardly enough to "store an orchestra".

1.3 Back to the origins

One of the questions that we raise in this thesis is: why not record the entire wave field first, and decide later on how to reproduce it? By getting too caught up in the discussion of how to reproduce spatial audio, one overlooks the implications that increasing the number of channels has on the recording side.

Throughout most of its history, digital signal processing has evolved as a technique for manipulating one-dimensional (1-D) signals, with only one independent variable. Typically, signals are characterized by an amplitude wave form expressed as a function of time. Music and speech signals, for example, are represented as the amplitude of the sound pressure as a function of time. Yet, there is no theoretical reason why *time* should be the independent variable of choice (or at least the only one).

Consider, for example, the equation derived by d'Alembert that describes the propagation of a wave front on a string, given by

$$\frac{\partial^2 p(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2}. \tag{1.2}$$

From (1.2) alone, there is no particular reason why the wave front is better characterized as a function of time, compared to, say, a function of space. The information given by $p(x_0, t)$ for a fixed x_0 can be equally given by $p(x, t_0)$ for a fixed t_0. In fact, in both cases, a huge amount of information is lost. The duality is illustrated in Figure 1.3.

Of course, there are practical reasons for using a function of time rather than a function of space. Audio sensors such as microphones are built to capture variations of the sound pressure as a function of time. To obtain an equivalent signal as a function of space would require an enormous number of pressure sensors placed along a line directed towards the source, with the microphones very tightly spaced. But this idea is not so far-fetched as it would seem a few years ago. While the processing power of digital computers is increasing, the size of the microphones is decreasing (as well as their cost). The only thing that has somewhat stagnated is the theory and application of multichannel Fourier analysis.

One of the consequences of focusing the development of signal processing on 1-D time-varying signals was that the theory of Fourier analysis became increasingly detached from its origins, which was the study of field equations. The temporal dimension became the only independent variable characterizing the acoustic wave field, while the three dimensions of space were mostly left in the background. This statement might sound contradictory, since the name "spatial audio" implies that the spatial dimensions are implicitly considered. But the fact is: in traditional spatial audio techniques, the multiple channels are treated as multiple functions of time, which is not justified in terms of the wave equation. This is particularly noticeable in coding formats for spatial audio, such as MPEG surround [48].

1.4 Arrays as spatial axes

The increasing use of microphone arrays in audio recording technology and consumer devices, as illustrated in the examples of Figure 1.4, has created an exciting new opportunity for taking the theory of audio DSP to a new level. An opportunity for applying Fourier analysis to the study of acoustic wave fields, and extending a breadth of existing DSP techniques to a four-dimensional (4-D) domain representing the three dimensions of space and the dimension of time.

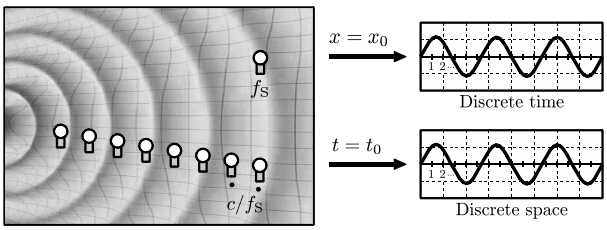


Figure 1.3: Duality between a time-varying signal and a space-varying signal. The signal obtained with the isolated microphone represents the amplitude of the sound pressure generated by the source as a function of discrete instances of time—sampled with the frequency f_S. The same signal can be obtained by placing multiple microphones on a straight line, uniformly spaced by c/f_S, where $c = 340\mathrm{m/s}$ is the speed of sound.

The main premise of this thesis can be summarized as follows: (i) microphone arrays, as they increase in size and number of microphones, approximate the spatial dimensions that define the wave equation; (ii) the wave equation has harmonic behavior along the spatial dimensions, the same way it has along the temporal dimension; (iii) therefore, it makes sense to extend the theory of Fourier analysis to the spatial dimensions.

In a recent work by Ajdler *et al.* [2], it was shown that, if the multiple channels are interpreted as spatial samples obtained with a certain "sampling frequency" (determined by the spacing between microphones) the spatio-temporal representation of the wave field is essentially band-limited and can be reduced to a closed-form solution. This means that, using Nyquist sampling theory, the acoustic wave field can be sampled across space with a finite sampling frequency, and yet be reconstructed with virtually no spatial aliasing. Thus, it is possible to record the acoustic wave field within a given area using a limited number of spatial samples (channels), and still retain all the information.

Reconstructing the acoustic wave field means, in physical terms, that an artificial wave front can be generated such that it is physically equivalent to the original wave front, using loudspeakers as a materialization of the spatial samples. This technique, introduced by Berkhout *et al.* [10; 14], is known as wave field synthesis (WFS), and it will be discussed in more detail later in the thesis.

In the context of DSP, Ajdler and Berkhout have essentially addressed the conversion of "analog wave fronts" into "discrete wave fronts"—a procedure known as *sampling*—as well as the reconstruction of the original "analog wave fronts" from the "discrete wave fronts"—a procedure known as *interpolation*. It has since become an open question how to *process* the wave field in discrete space and time—in particular, using multidimensional Fourier theory. That is the central question addressed in this thesis.

1.5 Contributions and thesis outline

The main contribution of this thesis is that we provide a framework of signal processing in space and time, formulated as a multidimensional extension of traditional 1-D signal processing. We address concepts such as linear time-invariant (LTI) systems, the Fourier transform, time–frequency analysis, and filter bank theory in the context of acoustic wave fields, and show how these concepts lead to powerful applications such as spatial filtering and sound

(a) (b) (c) (d)

Figure 1.4: Examples of the use of microphone arrays in consumer products and scientific experiments. (a) A typical Phonak hearing aid is equipped with two microphones for sound acquisition. (b) Recent Dell laptops have an integrated array with four microphones on the upper part of the screen. (c) High-end Sony camcorders have a square shaped array with four microphones for recording the entire sound field. (d) An enormous microphone array at the Denver International Airport was built by NASA in 2003 to study the wake vortex effects caused by the wing tip of airplanes.

field compression, all this based on simple DSP operations.

The structure of the thesis is inspired by the classic texts *Signals and Systems* by Oppenheim and Willsky [69] and *Discrete-Time Signal Processing* by Oppenheim and Schafer [68], going from the physics, to the Fourier analysis, the algorithms, the design of solutions and selected applications. Therefore, the thesis can also be used by graduate students with good knowledge of DSP but little knowledge of theoretical physics, or allow practitioners in array signal processing to broaden their perspective on acoustic wave fields using the tools we provide.

The outline consists of nine chapters separated by both scientific fields (physics and signal processing) and classes of signals (continuous and discrete), as listed in Table 1.1.

Table 1.1: List of chapters and topics.

Chapter	**Scientific Field**	**Class of Signals**
2 Fundamentals of Acoustics Theory	Physics	Continuous
3 Linear Space/Time Invariant Systems	Physics / Signal Processing	Continuous
4 The Spatio-Temporal Fourier Transform	Signal Processing	Continuous
5 Space/Time–Frequency Representation	Signal Processing	Continuous
6 Directional Representations	Signal Processing	Continuous
7 Filter-Bank Realizations of Transforms	Signal Processing	Discrete
8 Spatio-Temporal Filter Design	Signal Processing	Discrete
9 Acoustic Wave Field Coding	Signal Processing	Discrete

In **Chapter 2**, we review the basic theoretical concepts of acoustics theory that are necessary to understand the dynamics of wave propagation. These include the solution for spherical waves generated by point sources in space, the modes of wave propagation, and the Huygens principle, which explains why the wave field can be sampled and reconstructed like a regular signal. In **Chapter 3**, we introduce the concept of surface convolution, and prove that spatio-temporal systems based on the Huygens principle are both linear and space/time invariant. We also discuss different ways of representing the acoustic wave field as a signal. This chapter establishes the connection between the physical interpretation of acoustic wave fields and

the signal processing interpretation. In **Chapter 4**, we introduce the spatio-temporal Fourier transform, and derive the results for the most common type of acoustic scenes, such as point sources in the far-field, which are sparse in the Fourier domain, and point sources in the near-field. We also discuss the effects of sampling the acoustic wave field in space and time. In **Chapter 5**, we propose a generalization of Gabor's time–frequency representation such that it represents not only the frequency variations of sound along time but also the variations of the acoustic wave field across space. This consists of extending the local Fourier analysis to the three spatial dimensions, in addition to the temporal dimension, resulting in what we call the *space/time–frequency* representation space. In **Chapter 6**, we introduce an alternative to the spatio-temporal Fourier transform that we call *directional Fourier transform*, where the spectrum is represented as a function of the direction of propagation of the wave fronts. We prove that the acoustic wave field is equally sparse in this domain. In **Chapter 7**, we show how spatio-temporal transforms can be implemented in discrete space and time, and discuss filter bank structures that implement the desired transforms while satisfying the strict requirements of critical sampling and perfect reconstruction of the input. In **Chapter 8**, we show how point sources can be enhanced or eliminated from the acoustic scene by applying a non-separable filter in the spatio-temporal Fourier domain, and how such filters can be designed in a far more simple and intuitive way compared to existing spatial filtering techniques. We also introduce the concept of adaptive filtering in the spatio-temporal Fourier domain. In **Chapter 9**, we discuss the compression of acoustic wave fields based on plane-wave encoding, and prove that the spatio-temporal Fourier representation becomes increasingly efficient as the number of channels increases. We also propose a method for compressing acoustic wave fields using spatio-temporal perceptual criteria—a method we call *perceptual wave field coding*.

Chapter 2

Fundamentals of Acoustics Theory

2.1 Introduction

As mentioned in the introduction, one important aspect of this thesis is that we quickly step away from the world of theoretical physics and enter the world of DSP, where we avoid all the physics-related terminology and focus only on the relevant results—such as spherical radiation and modes of wave propagation. These results have a strong intuition behind all the mathematical formalism, and can be easily understood with the support of some basic theoretical background.

Many fundamental results of acoustics theory are derived from what is known as the *wave equation*—a partial differential equation that governs all the patterns of wave propagation. The wave equation, given in its simplest form by

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

determines how the sound pressure $p(x,t)$ at different points in space x (in this case, the 1-D x-axis) relates to the sound pressure at different points in time t, weighted by the speed of wave propagation c. We assume that the medium is homogeneous, continuous, adiabatic, isotropic, and perfectly elastic, as is the case with air [65]. The differential equation has an infinite number of solutions that can be expressed as $p(x,t) = s_1(t - \frac{x}{c}) + s_2(t + \frac{x}{c})$, where $s_1(t)$ and $s_2(t)$ are two arbitrary functions that propagate in the form of two waves with constant velocity c, moving in opposite directions along the x-axis. This solution is known as the d'Alembert solution, and is illustrated in Figure 2.1. It can be easily verified that $p(x,t)$ satisfies the wave equation, since

$$\frac{\partial^2 p(x,t)}{\partial t^2} = s_1''(t - \frac{x}{c}) + s_2''(t + \frac{x}{c})$$

and

$$\begin{aligned}\frac{\partial^2 p(x,t)}{\partial x^2} &= \frac{1}{c^2}s_1''(t - \frac{x}{c}) + \frac{1}{c^2}s_2''(t + \frac{x}{c}) \\ &= \frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2}.\end{aligned}$$

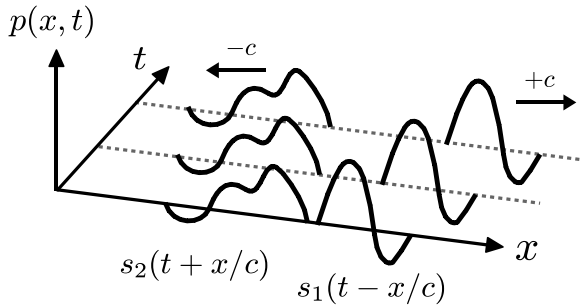


Figure 2.1: The d'Alembert solution to the wave equation consists of two waves traveling in opposite directions along the x-axis, with amplitudes given by $s_1(t-\frac{x}{c})$ and $s_2(t+\frac{x}{c})$. The two waves have a constant speed given by c.

To obtain a unique solution, we have to specify the initial values relative to t and the boundary conditions relative to x (the boundary conditions are the "initial values" in space). These conditions depend on both the characteristics of the excitation function and the physical setup where the wave propagates.

One important solution to the wave equation is the solution for spherical radiation caused by a point source in free space. This case is particularly important for two reasons: (i) in acoustic scenes, most types of sound sources can be approximated by point sources, and (ii) reflections and reverberation in a closed space can be modeled as virtual point sources (image sources) in free space [3; 63]. The solution for spherical radiation requires the use of the wave equation in spherical coordinates, given in its simplest form by

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p(r,t)}{\partial r}\right) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(r,t) \tag{2.1}$$

where r is the distance from the point source to the point of observation. The solution in this case is of the form

$$p(r,t) = \frac{1}{r}s_1(t-\frac{r}{c}) + \frac{1}{r}s_2(t+\frac{r}{c}) \tag{2.2}$$

which, similarly to the d'Alembert solution, represents two waves with constant speed c moving in opposite directions along r (away from, and towards the point source); we will see that only the first wave is physically possible, and hence the solution is simply $p(r,t) = \frac{1}{r}s(t-\frac{r}{c})$. It can also be verified that this solution satisfies the wave equation for any $s_1(t)$ and $s_2(t)$. In Sections 2.2 and 2.3 of this chapter, we provide a detailed derivation of $p(r,t)$ from the wave equation, for both the general case and the far-field case, *i.e.*, when the sound pressure is observed far away from the source.

Spherical waves radiate throughout the medium in a concentric pattern, meaning that the advancing wave fronts are equally valued along circular profiles with a common center. A well known example of concentric waves is the result of dropping an object over still water, as shown in the picture of Figure 2.2. To an outside observer, the falling object generates a concentric wave that smoothly moves away from the origin until it fades out into infinity. This macroscopic view of the phenomenon is predicted by the wave equation, as in the case of spherical waves, except that on the water's surface the wave equation is expressed as

Image by D. Jovic

Figure 2.2: Example of a radial wave caused by the impact of a falling object over still water. The wave propagates on the surface in the form of concentric rings, moving away from the point of impact and decaying in amplitude until it fades out completely. In this case, since the surface is a 2-D plane, the wave is characterized by the 2-D wave equation.

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

A closer view of the phenomenon, however, will reveal mathematical details in the equations, as well as physical subtleties at the microscopic level, that paint a different picture of how and why the wave front propagates throughout the medium until it fades out completely. Two concepts in particular, discussed in Sections 2.4 and 2.5, are known as the *modes of wave propagation* and the *Huygens principle*.

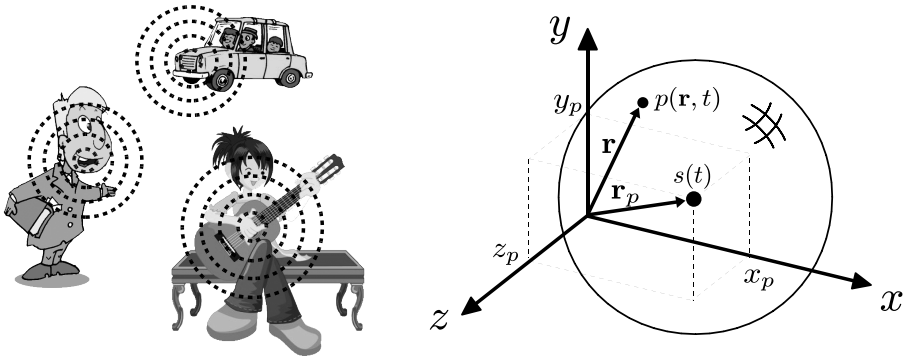
The modes of wave propagation [65] are related to how the wave behaves at different distances from the source. If the sound pressure is observed at a great distance, the wave front looks flat and harmonic; if the sound pressure is observed at a close distance, the wave front looks curved and damped. Such behavior is governed by the energy ratio between the two modes of wave propagation, known as the *propagating mode* and the *evanescent mode*. The propagating energy is responsible for the harmonic behavior of the wave front, whereas the evanescent energy is responsible for the amplitude decay.

The Huygens principle [50; 7] is a concept named after the Dutch physicist Christiaan Huygens (1629–1695), and it states that, at a microscopic level, each particle in the medium being excited by the advancing wave front acts itself as a source of excitation to the subsequent particles. This means that, at every point r_0 along the r-axis, the entire concentric wave front is a continuum of secondary point sources that generate the wave front at $r_0 + dr$, where dr is of the order of distance between particles. In other words, the particle vibrations observed at r_0 completely describe the wave field at $r > r_0$. We will see that this principle forms the basis of sampling and reconstruction of the acoustic wave field, and consequently the ability to analyze and process the wave field through DSP techniques.

2.2 Spherical radiation

2.2.1 Free space solution

Point sources are singularities in space that radiate sound equally in all directions, causing what is called a *spherical wave front*. Some examples are shown in Figure 2.3-(a). Given



(a) (b)

Figure 2.3: Point sources in a 3-dimensional space. (a) Many types of sound sources, such as people, machines, and musical instruments, are well approximated by point sources. (b) Each point source radiates a spherical wave front where the sound pressure $p(\mathbf{r}, t)$ is a function of the source signal $s(t)$ and the distance between the point of observation $\mathbf{r}$ and the point of origin $\mathbf{r}_p$, given by the ℓ_2-norm $\|\mathbf{r} - \mathbf{r}_p\|$.

a point source with source signal $s(t)$, the sound pressure at a distance r, denoted $p(r, t)$, is governed by the wave equation [65]

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) p(r, t) = -f(r, t), \tag{2.3}$$

where ∇^2 is the Laplacian operator ($\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)$ in spherical coordinates), c is the speed of sound, and $f(r, t)$ is a function with compact support around $r = 0$. The precise definition of compact support belongs to the field of measure theory, but it essentially means that the source occupies a compact region in space. In the case of a point source, $f(r, t) = \delta(r)s(t)$.

Taking the Fourier transform of (2.3) with respect to time, yields the Helmholtz equation [65]

$$\left(\nabla^2 + \left(\frac{\Omega}{c}\right)^2\right) p(r, \Omega) = -f(r, \Omega), \tag{2.4}$$

where Ω is the angular frequency in rad/s. The solution for (2.4) can be found using the Green's function method [39; 65], which consists of finding a solution $g(r, \Omega)$ to the Helmholtz equation when $f(r, t) = \delta(r)\delta(t)$, such that

$$\left(\nabla^2 + \left(\frac{\Omega}{c}\right)^2\right) g(r, \Omega) = -\delta(r). \tag{2.5}$$

The solution $g(r, \Omega)$ is called the Green's function, and is given by

$$g(r,\Omega) = A\frac{e^{-j\frac{\Omega}{c}r}}{r} + B\frac{e^{j\frac{\Omega}{c}r}}{r}, \tag{2.6}$$

where A and B are arbitrary constants. The general solution for $f(r,t) = \delta(r)s(t)$ is then given by the convolution $p(r,\Omega) = \int_{\mathbb{R}} f(\rho,\Omega)g(r-\rho,\Omega)d\rho$, which equals

$$p(r,\Omega) = S(\Omega)\left(A\frac{e^{-j\frac{\Omega}{c}r}}{r} + B\frac{e^{j\frac{\Omega}{c}r}}{r}\right). \tag{2.7}$$

The result in (2.7) can be interpreted as an input signal $s(t)$ filtered by a synthesis filter $g(r,t)$ that translates the input signal into a spherical radiation pattern.

Solving the Helmholtz equation uniquely requires a boundary condition at infinity. In physical terms, (2.7) represents two spherical waves: one radiating from the source to infinity, the other radiating from infinity into the source. The Sommerfeld radiation condition [82] states that no source can be a sink of energy. This implies that $B = 0$ and $A = \frac{1}{4\pi}$, and thus

$$p(r,\Omega) = \frac{S(\Omega)e^{-j\frac{\Omega}{c}r}}{4\pi r}. \tag{2.8}$$

Taking the inverse Fourier transform, the sound pressure is given by

$$p(r,t) = \frac{s(t-\frac{r}{c})}{4\pi r}, \tag{2.9}$$

which is simply a delayed and attenuated version of the source signal $s(t)$. The result can be rewritten in Cartesian coordinates, such that

$$\boxed{p(\mathbf{r},t) = \frac{s\left(t-\frac{\|\mathbf{r}-\mathbf{r}_p\|}{c}\right)}{4\pi\|\mathbf{r}-\mathbf{r}_p\|},} \tag{2.10}$$

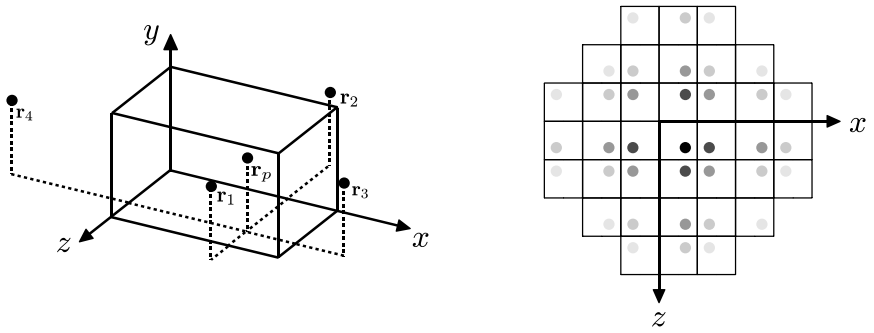
where $\mathbf{r} = (x,y,z)$ is the point of observation, $\mathbf{r}_p = (x_p,y_p,z_p)$ is the point of origin, and $\|.\|$ is the regular vector ℓ_2-norm. An illustration is shown in Figure 2.3-(b).

2.2.2 Image source model

In the case where the point source is not in free space but instead in a closed space, the result in (2.10) is only valid until the wave front hits a wall—then, the incident wave gets reflected and adds up to the original spherical wave. After the first reflection occurs, the sound pressure at any given location $\mathbf{r}$ is given by

$$p(\mathbf{r},t) = \underbrace{\frac{s\left(t-\frac{\|\mathbf{r}-\mathbf{r}_p\|}{c}\right)}{4\pi\|\mathbf{r}-\mathbf{r}_p\|}}_{\text{Direct wave}} + \underbrace{\frac{s\left(t-\frac{\|\mathbf{r}-(2\mathbf{r}_{\text{wall}}-\mathbf{r}_p)\|}{c}\right)}{4\pi\|\mathbf{r}-(2\mathbf{r}_{\text{wall}}-\mathbf{r}_p)\|}}_{\text{Reflected wave}} \tag{2.11}$$

where $\mathbf{r}_{\text{wall}}$ are the coordinates of where the incident wave hits the wall. The first and second terms in (2.11) are called the direct wave (or direct sound) and the reflected wave (or reflection). As the direct and reflected waves start hitting the other walls, the reflections start adding up until they fade out to zero due to the $\frac{1}{r}$ attenuation and the attenuation caused by the walls.



(a) (b)

Figure 2.4: Reflections and reverberation using the mirror source model. (a) Reflections are virtual point sources at every possible symmetric location, and driven by the same source signal $s(t)$. (b) Reverberation is caused by the accumulation of spherical waves originating from the virtual sources. The fading color represents the amplitude contribution of each source given the attenuation factor of the walls.

The reflected waves are equivalent to spherical waves generated by point sources centered at the mirror positions relative to the walls. Such interpretation is known as the image model [3] and can be generalized for rectangular spaces:

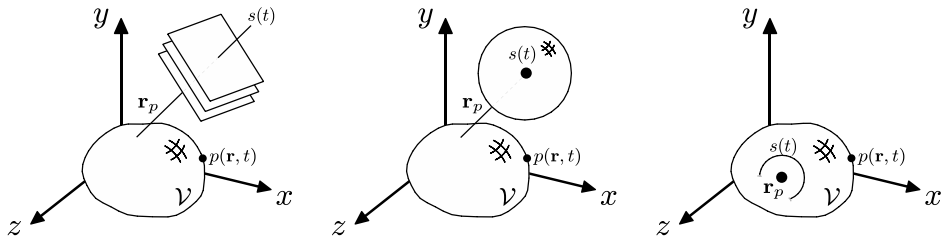
$$p(\mathbf{r},t) = \frac{s\left(t - \frac{\|\mathbf{r}-\mathbf{r}_p\|}{c}\right)}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} + \sum_{i=1}^{\infty} \frac{s\left(t - \frac{\|\mathbf{r}-\mathbf{r}_i\|}{c}\right)}{4\pi\|\mathbf{r}-\mathbf{r}_i\|}, \tag{2.12}$$

where $\mathbf{r}_i$ are the coordinates of the image sources. Similar models exist for non-rectangular spaces (*e.g.*, Martin *et al.* [63]), but they are not central in the context of this thesis. The important thing to retain is that a point source in a closed space can be modeled as an infinite number of point sources in free space, all with the same source signal but different delays and attenuations. The image model for rectangular spaces is illustrated in Figure 2.4.

2.3 Far-field radiation

A special case of spherical radiation is when the point source, located at $\mathbf{r}_p$, is very distant from the region of observation $\mathcal{V}$—a condition known as *far-field*. In practice, this means that the wave front reaching $\mathcal{V}$ from $\mathbf{r}_p$ has negligible curvature (*i.e.*, is nearly plane). The classification of a far-field source is always relative to the region of observation $\mathcal{V}$, and the condition is given by $\|\mathbf{r}_p\| \gg \|\mathbf{r}\|$. The opposite of far-field is known as *near-field*, and the condition is given by $\|\mathbf{r}-\mathbf{r}_p\| \ll \lambda$ [65], where $\lambda = \frac{c}{\Omega}$ is the wave length. Everything else in between will be referred to as *intermediate-field*. The three cases are illustrated in Figure 2.5.

Under the far-field assumption, $\|\mathbf{r}-\mathbf{r}_p\| = \|\mathbf{r}_p\| - \epsilon(\mathbf{r})$ where $\epsilon(\mathbf{r})$ is a residual term dependent on the point of observation $\mathbf{r}$. Replacing $\|\mathbf{r}-\mathbf{r}_p\|$ in (2.10) yields



(a) (b) (c)

Figure 2.5: Distance of the point source to region of observation. (a) In the far-field, where $\|\mathbf{r}_p\| \gg \|\mathbf{r}\|$, the wave front radiated by the distant point source is nearly plane. (b) In the intermediate-field, the general solution for spherical radiation applies. (c) In the near-field, where $\|\mathbf{r} - \mathbf{r}_p\| \ll \lambda$, the curvature of the wave front becomes more pronounced in the vicinity of $\mathcal{V}$.

$$p(\mathbf{r}, t) = \frac{s\left(t - \frac{\|\mathbf{r}_p\|}{c} + \frac{\epsilon(\mathbf{r})}{c}\right)}{4\pi \|\mathbf{r}_p\| - 4\pi\epsilon(\mathbf{r})} \tag{2.13}$$

$$\approx \frac{s\left(t - \frac{\|\mathbf{r}_p\|}{c} + \phi(\mathbf{r})\right)}{4\pi \|\mathbf{r}_p\|}, \tag{2.14}$$

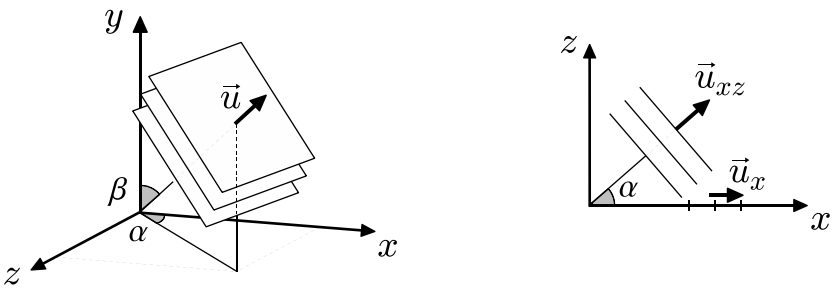
where $\phi(\mathbf{r}) = \frac{\epsilon(\mathbf{r})}{c}$ is the residual phase that depends on $\mathbf{r}$. This term can not be discarded, since it appears in the argument of $s(t)$ and it is generally unknown how sensitive the signal amplitude is to small differences in time, as opposed to the residual amplitude term $4\pi\epsilon(\mathbf{r})$ which is negligible regardless of $\mathbf{r}$. In addition, the terms $\frac{1}{4\pi\|\mathbf{r}_p\|}$ and $\frac{\|\mathbf{r}_p\|}{c}$ are fixed amplitude and phase values that we discard for simplicity, such that

$$p(\mathbf{r}, t) = s\left(t + \phi(\mathbf{r})\right). \tag{2.15}$$

To determine how $\phi(\mathbf{r})$ is expressed as a function of $\mathbf{r}$, recall that the wave front reaching $\mathcal{V}$ is plane, and therefore has a fixed direction of propagation (see Figure 2.6). This direction is given by the wave front normal, $\vec{u} = \mathbf{u} \cdot \vec{\mathbf{r}}$, where $\mathbf{u} = (u_x, u_y, u_z)$ contains the directional components in each axis and $\vec{\mathbf{r}} = (\vec{x}, \vec{y}, \vec{z})$ are the standard basis vectors $\vec{x}$, $\vec{y}$, and $\vec{z}$. Without loss of generality, consider that $\|\mathbf{u}\| = 1$. Since the wave front is plane, the expression $u_x x + u_y y + u_z z = c\phi(\mathbf{r})$ defines a profile of constant phase in the three-dimensional space. Thus,

$$\boxed{p(\mathbf{r}, t) = s\left(t + \frac{\mathbf{u} \cdot \mathbf{r}}{c}\right).} \tag{2.16}$$

The propagation vector $\mathbf{u}$ can also be expressed in polar coordinates, with $u_x = \cos\alpha \sin\beta$, $u_y = \cos\beta$, and $u_z = \sin\alpha \sin\beta$, where $\alpha, \beta \in [0, \pi]$ are the angles of arrival as defined in Figure 2.6.



(a) (b)

Figure 2.6: Wave front radiated by a wide-band point source located in the far-field: (a) in 3-D view and (b) in 2-D view. The wave front hits the x and y axes with angles $\alpha, \beta \in [0, \pi]$, and, in both cases, the wave front leaves a trail of constant-phase profiles (not to confuse with a sinusoidal wave front).

The most distinct aspect of this result is that the sound pressure is dependent only on the direction of propagation of the wave front given by $\mathbf{u}$. In the case where $s(t) = e^{j\Omega_0 t}$, the function $p(\mathbf{r}, t)$ is called a *plane wave* with frequency Ω_0 rad/s. We will see that the plane wave is the elementary component in the spatio-temporal Fourier analysis of the wave field, the same way complex frequencies are the elementary components in the traditional Fourier analysis.

2.4 Modes of wave propagation

Another distinct aspect of far-field radiation is that the amplitude of the sound pressure does not decay with the distance. In the strict sense, this is not possible under the Sommerfeld radiation condition [82], and it can be easily verified that (2.16) does not satisfy the wave equation. In fact, the far-field solution represents an idealized wave front.

The reason why a far-field acoustical wave front is not physically possible is because it contains only one mode of wave propagation, called the *propagating mode*. To satisfy the wave equation, the wave front must contain two modes of wave propagation: (i) the propagating mode, where the wave is perpetually harmonic, and (ii) the *evanescent mode*, where the wave decays exponentially. These two modes compete with each other at different directions and distances from the point source, starting with equal energy at the point of origin and tipping towards the propagating mode as the distance increases. The energy contribution of the evanescent mode becomes zero only asymptotically, when $\|\mathbf{r} - \mathbf{r}_p\| \to \infty$. This behavior is illustrated in Figure 2.7.

To understand the mathematics behind this concept, consider the Fourier transform of (2.16) with respect to time,

$$p(\mathbf{r}, \Omega) = S(\Omega) e^{j\frac{\Omega}{c}\mathbf{u}\cdot\mathbf{r}}. \tag{2.17}$$

The transition between modes can be observed when looking into a particular direction of

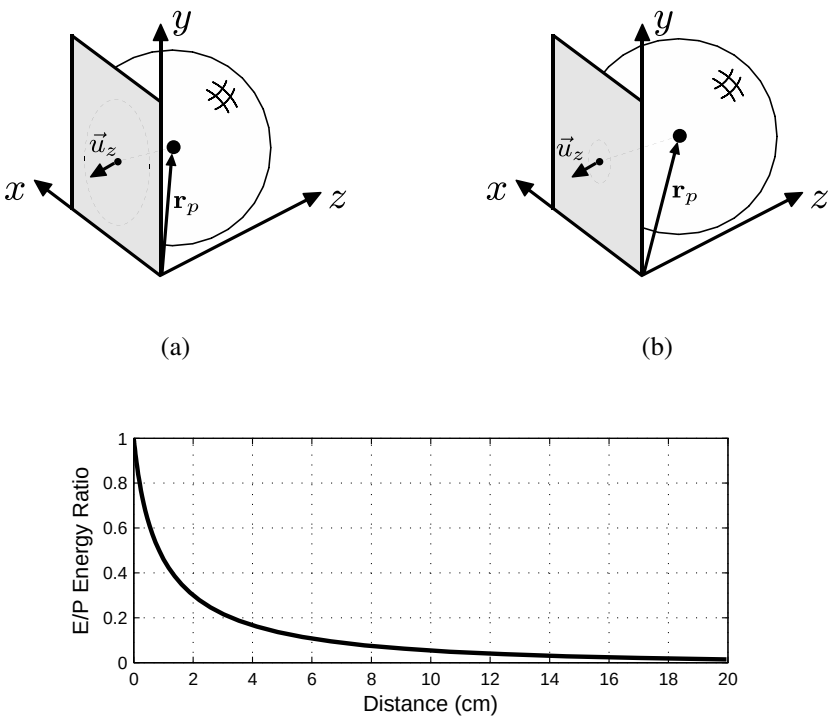


Figure 2.7: Modes of wave propagation in the vicinity of a point source. The total energy contribution of the propagating and evanescent modes in the spherical wave front is measured on the infinite xy-plane with respect to the z direction. The measurement starts with the xy-plane very close to the source in (a), and it moves away from the source in (b). The ratio between the two modes as a function of distance is shown in (c), where it is clear that the propagating mode becomes increasingly dominant over the evanescent mode as the distance increases, and the evanescent mode is asymptotically zero in the far-field. The method for obtaining the plot in (c) is explained in Chapter 4.

propagation (*i.e.*, any linear combination of the vectors u_x, u_y, and u_z). As an example, we consider the direction u_z only. For this purpose, let us separate the component in z, so that

$$p(\mathbf{r},\Omega) = S(\Omega)e^{j\frac{\Omega}{c}u_z z}e^{j\frac{\Omega}{c}(u_x x+u_y y)}. \tag{2.18}$$

Given the ℓ_2-norm $\|\mathbf{u}\|^2 = u_x^2 + u_y^2 + u_z^2$, the component in z can be expressed as $u_z = \pm\sqrt{\|\mathbf{u}\|^2 - (u_x^2+u_y^2)}$. Note that a transition occurs at $(u_x^2+u_y^2) = \|\mathbf{u}\|^2$, since $(u_x^2+u_y^2) < \|\mathbf{u}\|^2$ makes u_z real and $(u_x^2+u_y^2) > \|\mathbf{u}\|^2$ makes u_z complex. Accordingly, (2.18) can be written as

$$p(\mathbf{r},\Omega) = \begin{cases} S(\Omega)e^{j\frac{\Omega}{c}u_z z}e^{j\frac{\Omega}{c}(u_x x+u_y y)} & , (u_x^2+u_y^2) < \|\mathbf{u}\|^2 \\ S(\Omega)e^{\pm\frac{\Omega}{c}|u_z| z}e^{j\frac{\Omega}{c}(u_x x+u_y y)} & , (u_x^2+u_y^2) > \|\mathbf{u}\|^2 \end{cases}. \tag{2.19}$$

where the sign $\pm$ is chosen such that $\pm\Omega z$ is negative, thus providing a physical (non-divergent) solution [88]. In the first case, the wave propagates harmonically towards all directions, since all three exponential terms are complex. In the second case, the wave propagates harmonically towards x and y, and decays exponentially towards z.

At a first look, it may seem strange that a projection of the vector $\mathbf{u}$ onto the xy-plane can have a larger norm than $\mathbf{u}$ itself. And indeed, this never happens in the far-field case. In the near- and intermediate-field cases, however, the wave front is curved; to define a direction of propagation of a curved wave front requires a non-Euclidean space. In such a space, the projection of $\mathbf{u}$ onto a lower-dimensional subspace can in fact have a larger norm than $\mathbf{u}$ [61]. Thus, the more curved the wave front is in the region of observation, the more likely it is that $(u_x^2+u_y^2) > \|\mathbf{u}\|^2$ occurs.

2.5 Huygens principle

The propagation of acoustic waves throughout the medium is a process of transfer of energy between adjacent particles, that excite each other as the wave passes by. At a microscopic level, every time a particle is "pushed" by its immediate neighbor, it starts oscillating back and forth with decaying amplitude until it stops completely in its original position. This movement triggers the oscillation of subsequent particles—this time with less strength—and the process continues until the initial "push" is not strong enough to sustain the transfer of energy. This process is illustrated in Figure 2.8.

An important consequence of the decaying oscillation of particles is that, since they end up in the same position, there is no net displacement of mass in the medium. So, even though the waves travel in the medium, the medium itself does not follow the waves. This effectively turns every particle in the medium into a point source, driven by the same source signal of the original source. The particles are thus referred to as *secondary point sources*. At a macroscopic level, the combination of all the secondary point sources, and the spherical waves they generate, jointly build up the entire wave front generated by the original source. This is known as the Huygens principle [7], and is illustrated in Figure 2.9.

2.5.1 Mathematical formulation

The Huygens principle is formally described by the Kirchhoff-Helmholtz's equation [7], which expresses the sound pressure inside an arbitrary volume $\mathcal{V}$ as a function of the sound

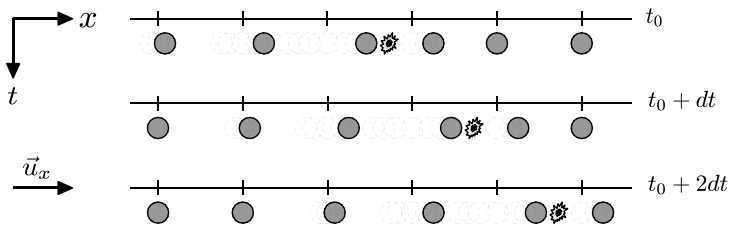


Figure 2.8: Energy transfer between particles in the medium. The traveling wave triggers an oscillating movement in the local particles, causing them to bump into their immediate neighbors and spread the oscillating movement. As the wave goes away, the oscillation decays and the particles go back to their original positions.

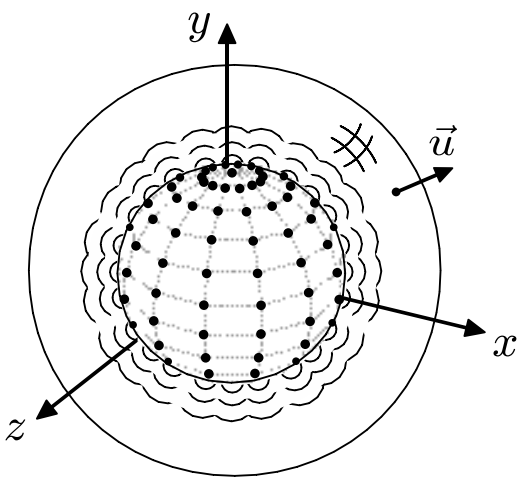


Figure 2.9: Illustration of the Huygens principle. At every "step" of the advancing wave front, the edge of the wave front is a continuum of secondary point sources that together build up the "next step" of the advancing wave front.

pressure on the boundary surface $\mathcal{S}$ (see Figure 2.10-(a)). The equation is given by

$$p(\mathbf{r},\Omega) = \oint_{\mathcal{S}} \left(\underbrace{\frac{\partial p(\mathbf{r}_{\mathcal{S}},\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}}}_{\text{Pressure gradient}} \underbrace{g(\mathbf{r}-\mathbf{r}_{\mathcal{S}},\Omega)}_{\text{Monopole}} - \underbrace{p(\mathbf{r}_{\mathcal{S}},\Omega)}_{\text{Pressure}} \underbrace{\frac{\partial g(\mathbf{r}-\mathbf{r}_{\mathcal{S}},\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}}}_{\text{Dipole}} \right) d\mathcal{S}, \tag{2.20}$$

for every $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{V}$, where $p(\mathbf{r}_{\mathcal{S}},\Omega)$ is the sound pressure on $\mathcal{S}$, $\boldsymbol{n}_{\mathcal{S}}$ is the surface normal pointing inwards, and $g(\mathbf{r}-\mathbf{r}_{\mathcal{S}},\Omega)$ is the 3-D Green's solution for spherical radiation centered at $\mathbf{r}_{\mathcal{S}}$,

$$g(\mathbf{r}-\mathbf{r}_{\mathcal{S}},\Omega) = \frac{e^{j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_{\mathcal{S}}\|}}{4\pi\|\mathbf{r}-\mathbf{r}_{\mathcal{S}}\|}. \tag{2.21}$$

The Kirchhoff-Helmholtz's equation in (2.20) represents a continuum of point sources g weighted by the pressure gradient $\frac{\partial p}{\partial \boldsymbol{n}_{\mathcal{S}}}$ and dipole sources $\frac{\partial g}{\partial \boldsymbol{n}_{\mathcal{S}}}$ weighted by the sound pressure p, covering the entire surface $\mathcal{S}$.

A simplified version of (2.20) is obtained by considering an area $\mathcal{A}$ enclosed by a boundary contour $\mathcal{L}$ (see Figure 2.10-(b)). The equation is then given by

$$p(\mathbf{r},\Omega) = \oint_{\mathcal{L}} \left(\underbrace{\frac{\partial p(\mathbf{r}_{\mathcal{L}},\Omega)}{\partial \boldsymbol{n}_{\mathcal{L}}}}_{\text{Pressure gradient}} \underbrace{g(\mathbf{r}-\mathbf{r}_{\mathcal{L}},\Omega)}_{\text{Monopole}} - \underbrace{p(\mathbf{r}_{\mathcal{L}},\Omega)}_{\text{Pressure}} \underbrace{\frac{\partial g(\mathbf{r}-\mathbf{r}_{\mathcal{L}},\Omega)}{\partial \boldsymbol{n}_{\mathcal{L}}}}_{\text{Dipole}} \right) d\mathcal{L}, \tag{2.22}$$

for every $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{A}$, where $g(\mathbf{r}-\mathbf{r}_{\mathcal{L}},\Omega)$ is the 2-D Green's function given by [88]

$$g(\mathbf{r}-\mathbf{r}_{\mathcal{L}},\Omega) = \frac{j}{4} H_0^{(1)}\left(\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_{\mathcal{L}}\|\right), \tag{2.23}$$

where $H_0^{(1)}$ is the zeroth-order Hankel function of the first kind. The results in (2.20) and (2.22) can be further simplified [8], when $\mathcal{V}$ "encloses" an entire source-free 3-D half-space (see Figure 2.10-(c)),

$$p(\mathbf{r},\Omega) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{p(\mathbf{r}_{\mathcal{S}},\Omega)}_{\text{Pressure}} \underbrace{\frac{\partial g(\mathbf{r}-\mathbf{r}_{\mathcal{S}},\Omega)}{\partial z}}_{\text{Dipole}} dxdy,\ z<0, \tag{2.24}$$

and when $\mathcal{A}$ "encloses" an entire source-free 2-D half-space (see Figure 2.10-(d)),

$$p(\mathbf{r},\Omega) = -\int_{-\infty}^{\infty} \underbrace{p(\mathbf{r}_{\mathcal{L}},\Omega)}_{\text{Pressure}} \underbrace{\frac{\partial g(\mathbf{r}-\mathbf{r}_{\mathcal{L}},\Omega)}{\partial z}}_{\text{Dipole}} dx,\ y=0 \text{ and } z<0. \tag{2.25}$$

The results in (2.24) and (2.25) are known as the Rayleigh equations. Note that, without loss of generality, we consider $\mathcal{S}$ to be the xy-plane and $\mathcal{L}$ the x-axis.

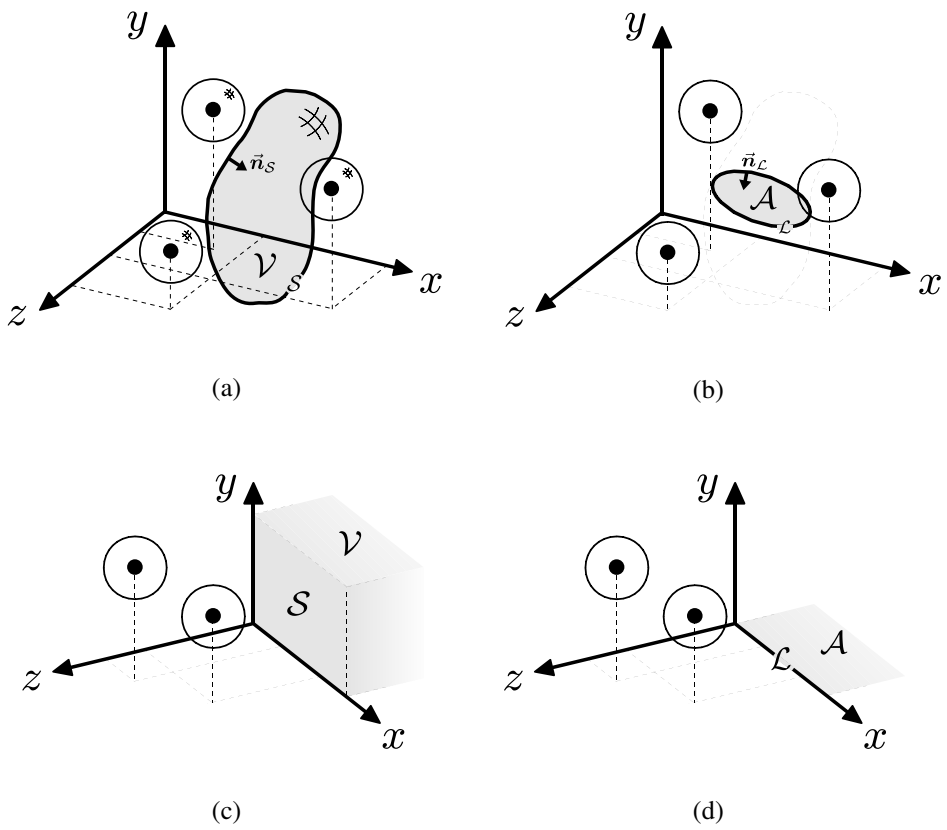


Figure 2.10: Illustration of the Kirchhoff-Helmholtz and Rayleigh equations. (a) General 3-D case, where $p(\mathbf{r}, t)$ for $\mathbf{r} \in \mathcal{V}$ is a function of $p(\mathbf{r}, t)$ for $\mathbf{r} \in \mathcal{S}$; (b) General 2-D case, where $p(\mathbf{r}, t)$ for $\mathbf{r} \in \mathcal{A}$ is a function of $p(\mathbf{r}, t)$ for $\mathbf{r} \in \mathcal{L}$; (c) 3-D half-space case, where $p(\mathbf{r}, t)$ for $z < 0$ is a function of $p(\mathbf{r}, t)$ for $z = 0$; (d) 2-D half-space case, where $p(\mathbf{r}, t)$ for $y = 0$ and $z < 0$ is a function of $p(\mathbf{r}, t)$ for $y = 0$ and $z = 0$.

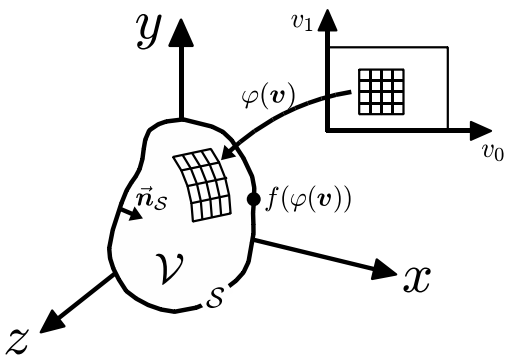


Figure 2.11: Expressing a surface integral as an integral in $\mathbb{R}^2$. The function $\varphi(\boldsymbol{v})$ maps a vector $\boldsymbol{v} \in \mathbb{R}^2$ to a vector $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{S}$, where $\mathcal{S}$ is a curved surface in $\mathbb{R}^3$. The normalizing factor is the ℓ_2-norm of the normal vector, $\|\boldsymbol{n}_{\mathcal{S}}\|$.

2.5.2 Convolution notation

An important consequence of the Huygens principle and its mathematical formulation is that any wave field can be observed on the surface of a given region in space and be resynthesized inside that same region, or potentially at a different region in space or at a different moment in time. These are the basic ingredients of a DSP system.

To make this relationship more apparent, it is convenient to express the surface integral in (2.20) as an integral in $\mathbb{R}^2$. This can be done using the following equality [59]

$$\oint_{\mathcal{S}} f(\mathbf{r}_{\mathcal{S}}) dS = \int_{\mathbb{R}^2} f(\varphi(\boldsymbol{v})) \left\| \frac{\partial \varphi(\boldsymbol{v})}{\partial v_0} \times \frac{\partial \varphi(\boldsymbol{v})}{\partial v_1} \right\| d\boldsymbol{v}, \tag{2.26}$$

where $\varphi(\boldsymbol{v}) : \mathbb{R}^2 \mapsto \mathbb{R}^3 \cap \mathcal{S}$ is a function that maps a vector $\boldsymbol{v} = (v_0, v_1)$ on a plane to a vector $\mathbf{r}_{\mathcal{S}} = (x_{\mathcal{S}}, y_{\mathcal{S}}, z_{\mathcal{S}})$ on the surface $\mathcal{S}$. The factor $\left\| \frac{\partial \varphi(\boldsymbol{v})}{\partial v_0} \times \frac{\partial \varphi(\boldsymbol{v})}{\partial v_1} \right\|$ is a normalization element, in this case given by the ℓ_2-norm of the normal vector, $\|\boldsymbol{n}_{\mathcal{S}}\|$. This transformation is illustrated in Figure 2.11.

The Kirchhoff-Helmholtz integral is then given by

$$p(\mathbf{r}, \Omega) = \int_{\mathbb{R}^2} \left(g(\mathbf{r} - \varphi(\boldsymbol{v}), \Omega) \frac{\partial p(\varphi(\boldsymbol{v}), \Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} - p(\varphi(\boldsymbol{v}), \Omega) \frac{\partial g(\mathbf{r} - \varphi(\boldsymbol{v}), \Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} \right) \|\boldsymbol{n}_{\mathcal{S}}\| \, d\boldsymbol{v}, \tag{2.27}$$

for every $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{V}$. To further simplify the result, define an operator $\mathcal{G}(\mathbf{r}, \Omega)$ such that

$$\mathcal{G}(\mathbf{r}, \Omega) = g(\mathbf{r}, \Omega) \, \|\boldsymbol{n}_{\mathcal{S}}\| \, \frac{\partial}{\partial \boldsymbol{n}_{\mathcal{S}}} - \|\boldsymbol{n}_{\mathcal{S}}\| \, \frac{\partial g(\mathbf{r}, \Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}}. \tag{2.28}$$

Then, (2.27) is given by

$$p(\mathbf{r}, \Omega) = \int_{\mathbb{R}^2} p(\varphi(\boldsymbol{v}), \Omega) \mathcal{G}(\mathbf{r} - \varphi(\boldsymbol{v}), \Omega) d\boldsymbol{v} \tag{2.29}$$

for every $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{V}$. We call the result in (2.29) the *surface convolution* of the sound pressure $p(\mathbf{r}, \Omega)$ with the *synthesis operator* $\mathcal{G}(\mathbf{r}, \Omega)$, and denote it with the symbol $\circledast$. The compact expression is given by

$$\boxed{p(\mathbf{r}, \Omega) = \begin{cases} p(\mathbf{r}, \Omega) \circledast \mathcal{G}(\mathbf{r}, \Omega) & , \mathbf{r} \in \mathbb{R}^3 \cap \mathcal{V} \\ \text{undefined} & , \mathbf{r} \notin \mathbb{R}^3 \cap \mathcal{V}. \end{cases}} \tag{2.30}$$

In accordance with the Huygens principle, the surface convolution of the sound pressure with the synthesis operator is the sound pressure itself inside the volume $\mathcal{V}$—by convention, it is undefined on the outside. In the next chapter, we will see that the surface convolution satisfies some of the properties of a regular convolution integral.

2.5.3 Wave field synthesis

The surface convolution expressed in (2.30) can be implemented in practice through a technique known as *wave field synthesis* (WFS), developed by Berkhout *et al.* [11] in the early 1990's. The method has been a subject of intense research among the audio engineering community, and it consists of using a large number of loudspeakers as secondary point sources representing the Green's solution $g(\mathbf{r}, \Omega)$ at various locations along $\mathcal{S}$—or, more commonly, along $\mathcal{L}$.

The Kirchhoff-Helmholtz equations present several challenges on a practical level—namely: (i) the number of secondary sources must infinite; (ii) the secondary sources have to cover the entire surface that encloses the listening region, which itself is infinitely large in the case of (2.24); (iii) both (2.20) and (2.24) require the use of dipoles as secondary sources, which are not easily implemented in hardware (loudspeakers typically act as monopoles). Some techniques for relaxing these theoretical requirements include: (i) discretizing the array of secondary sources, at the expense of adding spatial aliasing; (ii) limiting the listening region to a plane, using the 2-D solutions in (2.22) and (2.25), and performing segmented WFS; (iii) measuring the pressure gradient $\frac{\partial p}{\partial \boldsymbol{n}_{\mathcal{S}}}$ instead of the regular pressure p, so that the secondary sources can be implemented through monopoles instead of dipoles. Some real-world examples of WFS systems are shown in Figure 2.12.

Concepts such as spatial aliasing and local spatial analysis will become more clear as we turn into the discussion of spatio-temporal DSP systems.

Image source: dancetracksdigital.com

(a)

Image source: eaw.com

(b)

Image source: gizmag.com

(c)

Image source: soundbar.com

(d)

Figure 2.12: Real-world examples of WFS systems. (a) The Tresor Club is a popular dancing club in Berlin, Germany, and the first one to be equipped with a WFS sound system of 200 loudspeakers. (b) The Philadelphia Eagles Stadium is equipped with a gigantic loudspeaker array with over 700 loudspeakers surrounding the entire stadium. (c) The new Audi Q7 prototype has a WFS sound system with 62 loudspeakers surrounding the interior of the car. (d) An increasing number of home theater sound systems utilize linear arrays of loudspeakers instead of 5.1 setups, due to their ability to project sound into the walls.

2.6 Summary notes

- Acoustic wave fields are governed by a partial differential equation known as the wave equation;
- Point sources are the building blocks of every acoustic scene, whether the scene is in a free space or a reverberant room;
- Each point source radiates sound across space and time in a spherical wave pattern according to $p(\mathbf{r},t) = \frac{1}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} s\big(t - \frac{\|\mathbf{r}-\mathbf{r}_p\|}{c}\big)$ or $p(\mathbf{r},t) = s\big(t + \frac{\mathbf{u}\cdot\mathbf{r}}{c}\big)$, depending on the distance at which the sound pressure is observed;
- Spherical waves carry two types of energy—propagating and evanescent—which tend to manifest equally in the vicinity of the source, and converge to propagating energy at greater distances;
- The Huygens principle states that vibrating air particles are equivalent to secondary point sources, jointly sustaining the advancing wave front;
- The mathematical formulation of the Huygens principle—the Kirchhoff-Helmholtz equation—states that the wave field in any given source-free region in space is completely defined by the wave field at the boundary surface, and can be resynthesized using a technique known as wave field synthesis (WFS);
- The Kirchhoff-Helmholtz equation can be conveniently expressed in a convolution notation that we call surface convolution.

Chapter 3

Linear Space/Time Invariant Systems

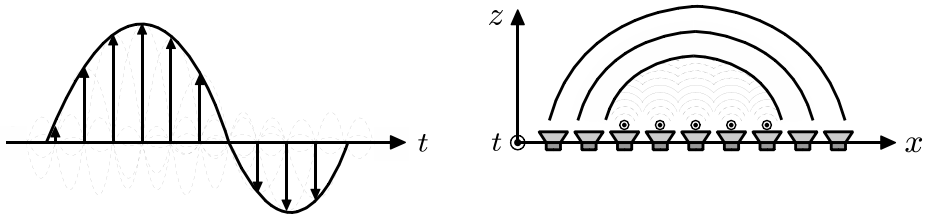
3.1 Introduction

The widespread adoption of digital signal processing (DSP) techniques in modern devices is possible thanks to two main technological pillars: the analog-to-digital (A/D) converter and the digital-to-analog (D/A) converter. The A/D converter is a device that collects samples of the input signal's amplitude at periodic instants in time; it translates physical signals—which are analog (or continuous-time) by nature—into a sequence of samples that can be processed by a computer. The D/A converter performs the opposite operation: it translates the sequence of samples into an analog signal that can be sent back to the physical world. It is essentially a low-pass interpolation filter that removes all the spectral images created by the sampling operation. In every DSP system, the A/D and D/A converters are the two building blocks that make it possible to process signals in the discrete domain, and without which most DSP theory would be useless. This is particularly true in applications related to sound processing.

When Christiaan Huygens developed his theory in the late 17th century [50], he could have hardly imagined the extent to which his ideas would affect today's technology. The Huygens principle was, in fact, the first comprehensive step in the creation of a D/A converter for acoustic wave fields—ultimately made possible by Berkhout's work on wave field synthesis (WFS) [9]. The main idea of a WFS algorithm is to pre-filter the signals that go into the loudspeakers such that the superposition of the wave fronts generated by each loudspeaker interpolate naturally in the air (or some general medium), resulting in the desired analog wave front. This is illustrated in Figure 3.1. Note, however, that this is not a "post-filter" in the traditional sense, since analog filters can not be implemented in space.

An A/D converter for acoustic wave fields can also be implemented with an array of microphones, where each microphone represents one spatial sample. The array can be seen as a device that collects samples of the input sound pressure at periodic locations in space (more specifically, on the boundary surface $\mathcal{S}$), thus translating the physical wave fronts into a sequence of samples that can be processed by a computer. Similarly, an analog "pre-filter" can not be implemented in space.

The availability of A/D and D/A technologies for sampling and reconstructing acoustic



(a) (b)

Figure 3.1: Digital-to-analog (D/A) conversion based on interpolation of elementary signals. (a) In traditional signal processing, the analog signal is reconstructed from its discrete counterpart through a sum of sinc functions centered at and wighted by each sample of the signal. (b) With the Huygens principle, the analog wave front is reconstructed from its samples through a sum of spherical waves centered at and weighted by each field value observed at the same position.

wave fields raises the following question: how can the samples be processed before being transmitted to a WFS system for playback? The development of DSP theory was mainly built upon a concept known as *linear time-invariant* (LTI) systems. According to the theory (see, *e.g.*, Oppenheim *et al.* [69]), an LTI system is one that responds to a linear combination (or superposition) of inputs with a linear combination of the individual outputs, and responds to a time shift on the input with an identical time shift on the output. For instance, if the response of the system to the inputs $s_1(t)$ and $s_2(t)$ is given by $y_1(t)$ and $y_2(t)$ respectively, then, due to linearity, the response to $a_1 s_1(t) + a_2 s_2(t)$ is given by $a_1 y_1(t) + a_2 y_2(t)$, where a_k is any constant such that $a_k \in \mathbb{C}$. Similarly, if the response of the system to the input $s(t)$ is given by $y(t)$, then, due to time invariance, the response to $s(t - t_0)$ is given by $y(t - t_0)$. In this chapter, we will show that the same properties are valid for a system of the type following the Huygens principle, *i.e.*, that the Huygens principle can be expressed as a system that is not only linear and time-invariant but also space-invariant. In the same line of thought, if the response of the system to the inputs $p_1(\mathbf{r}, t)$ and $p_2(\mathbf{r}, t)$ is given by $y_1(\mathbf{r}, t)$ and $y_2(\mathbf{r}, t)$ respectively, then, due to linearity, the response to $a_1 p_1(\mathbf{r}, t) + a_2 p_2(\mathbf{r}, t)$ is given by $a_1 y_1(\mathbf{r}, t) + a_2 y_2(\mathbf{r}, t)$, where a_k is any constant such that $a_k \in \mathbb{C}$. Similarly, if the response of the system to the input $p(\mathbf{r}, t)$ is given by $y(\mathbf{r}, t)$, then, due to space- and time-invariance, the response to $p(\mathbf{r} + \mathbf{r}_0, t + t_0)$ is given by $y(\mathbf{r} + \mathbf{r}_0, t + t_0)$. We call such a system *linear space/time invariant* (LSTI).

Expressing the Huygens principle as an LSTI system is a powerful conceptual tool that allows us to think of acoustic wave fields as *input signals* and physical principles as *impulse responses*. The bridge between the physical world and the DSP world is established by the surface convolution. For this reason, we dedicate most of this chapter to the discussion of the surface convolution and its properties, and to the particular cases that are relevant to the development of a comprehensive DSP theory for analyzing and processing acoustic wave fields.

3.2 Surface convolution

The surface convolution was introduced in the previous chapter as a compact formulation of the Kirchhoff-Helmholtz equation, which is the mathematical basis of the Huygens principle. The expression was defined as $p(\mathbf{r},\Omega) = p(\mathbf{r},\Omega) \circledast \mathcal{G}(\mathbf{r},\Omega)$ for every $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{V}$, where $\mathcal{V}$ is a source-free enclosed volume, and it states that the sound pressure inside $\mathcal{V}$ can be obtained by a convolution operation at the surface $\mathcal{S}$ between the sound pressure $p(\mathbf{r},\Omega)$ and the synthesis operator $\mathcal{G}(\mathbf{r},\Omega)$.

The result implies that the sound pressure inside $\mathcal{V}$ is completely defined by the sound pressure at the boundary surface $\mathcal{S}$—in agreement with the Huygens principle—meaning that any values of $p(\mathbf{r},\Omega)$ outside $\mathcal{S}$ are completely irrelevant to the outcome of the surface convolution. Thus, the expression can be written in the non-redundant form

$$p(\mathbf{r},\Omega) = (\delta_{\mathcal{S}}(\mathbf{r}) p(\mathbf{r},\Omega)) \circledast \mathcal{G}(\mathbf{r},\Omega)$$

where $\delta_{\mathcal{S}}(\mathbf{r})$ equals one for $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{S}$ and zero otherwise.

In this section, we discuss the definition and properties of the surface convolution in more detail, with special emphasis on the properties that are relevant to defining and constructing LSTI systems based on the Huygens principle.

3.2.1 General definition

The surface convolution is always between two functions: the *analysis function* $f_{\mathrm{a}}(\mathbf{r},\Omega)$ and the *synthesis function* $f_{\mathrm{s}}(\mathbf{r},\Omega)$. The analysis function represents the observed wave fronts that we intend to re-synthesize in the enclosed volume $\mathcal{V}$, whereas the synthesis function is the one that actually generates the wave fronts within $\mathcal{V}$ based on the observed values given by the analysis function.

The surface convolution between the two functions can be expressed as

$$f(\mathbf{r},\Omega) = f_{\mathrm{a}}(\mathbf{r},\Omega) \circledast f_{\mathrm{s}}(\mathbf{r},\Omega), \tag{3.1}$$

where

$$f_{\mathrm{a}}(\mathbf{r},\Omega) \circledast f_{\mathrm{s}}(\mathbf{r},\Omega) = \int_{\mathbb{R}^2} f_{\mathrm{a}}(\varphi_{\mathrm{a}}(\boldsymbol{v}),\Omega) f_{\mathrm{s}}(\mathbf{r} - \varphi_{\mathrm{s}}(\boldsymbol{v}),\Omega) d\boldsymbol{v} \tag{3.2}$$

for every $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{V}_{\mathrm{s}}$. The functions $\varphi_{\mathrm{a}}(\boldsymbol{v}) : \mathbb{R}^2 \mapsto \mathbb{R}^3 \cap \mathcal{S}_{\mathrm{a}}$ and $\varphi_{\mathrm{s}}(\boldsymbol{v}) : \mathbb{R}^2 \mapsto \mathbb{R}^3 \cap \mathcal{S}_{\mathrm{s}}$ are the functions that map a 2-D vector $\boldsymbol{v} = (v_0, v_1)$ to the 3-D coordinates of the analysis and synthesis surfaces, $\mathcal{S}_{\mathrm{a}}$ and $\mathcal{S}_{\mathrm{s}}$, which enclose the respective volumes $\mathcal{V}_{\mathrm{a}}, \mathcal{V}_{\mathrm{s}} \subset \mathbb{R}^3$. The geometrical shape of the analysis and synthesis surfaces is characterized by the functions $\varphi_{\mathrm{a}}(\boldsymbol{v})$ and $\varphi_{\mathrm{s}}(\boldsymbol{v})$, which are not necessarily equal. An illustration of the generalized surface convolution is given in Figure 3.2.

Note that the arguments in (3.1) are always the spatial coordinates $\mathbf{r}$ and the frequency Ω. Thus, for simplicity, we will additionally use the compact notation where the arguments are hidden:

$$f = f_{\mathrm{a}} \circledast f_{\mathrm{s}}. \tag{3.3}$$

This notation is particularly useful in the discussion of the properties of the surface convolution.

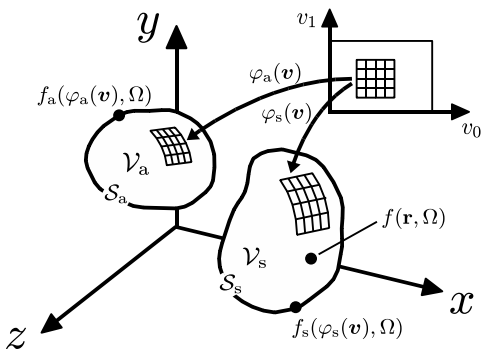


Figure 3.2: Generalized surface convolution. The analysis function observed at the analysis surface $\mathcal{S}_\mathrm{a}$ is convolved with the synthesis function defined at the synthesis surface $\mathcal{S}_\mathrm{s}$, generating the output function defined inside the source-free synthesis volume $\mathcal{V}_\mathrm{s}$, and undefined on the outside. The convolution integral is performed over a plane that maps its values to the analysis and synthesis surfaces through the functions $\varphi_\mathrm{a}(\boldsymbol{v})$ and $\varphi_\mathrm{s}(\boldsymbol{v})$, respectively.

3.2.2 Properties

3.2.2.1 Multiplicative identities

The surface convolution has two multiplicative identities, with direct implications on the physical and practical levels: the *analysis identity* and the *synthesis identity*. The analysis identity can be expressed as

$$(\delta_{\mathcal{S}_\mathrm{a}} f_\mathrm{a}) \circledast f_\mathrm{s} = f_\mathrm{a} \circledast f_\mathrm{s}, \tag{3.4}$$

for any $\mathcal{V}_\mathrm{a}, \mathcal{V}_\mathrm{s} \subset \mathbb{R}^3$, where $\delta_{\mathcal{S}_\mathrm{a}}$ is a delta function that samples the analysis surface $\mathcal{S}_\mathrm{a}$, defined as

$$\delta_{\mathcal{S}_\mathrm{a}}(\mathbf{r}) = \begin{cases} 1 & , \mathbf{r} \in \mathbb{R}^3 \cap \mathcal{S}_\mathrm{a} \\ 0 & , \mathbf{r} \notin \mathbb{R}^3 \cap \mathcal{S}_\mathrm{a}. \end{cases} \tag{3.5}$$

This property states that only the information at the analysis surface $\mathcal{S}_\mathrm{a}$ is relevant for the outcome of the surface convolution. We will see that, in practice, this reduces the number of spatial dimensions necessary for representing f_a.

The synthesis identity is given by

$$f_\mathrm{a} \circledast \mathcal{G} = f_\mathrm{a}, \tag{3.6}$$

provided that $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s}$, where $\mathcal{G}$ is the synthesis operator introduced in the previous chapter as

$$\mathcal{G} = g \left\| \boldsymbol{n}_\mathcal{S} \right\| \frac{\partial}{\partial \boldsymbol{n}_\mathcal{S}} - \left\| \boldsymbol{n}_\mathcal{S} \right\| \frac{\partial g}{\partial \boldsymbol{n}_\mathcal{S}}. \tag{3.7}$$

This property states that f_a is unaffected by the surface convolution when $f_\mathrm{s} = \mathcal{G}$. The proof of the analysis and synthesis identity properties is given in Theorem 4 and 5 in Section 3.6.

The two properties can also be combined such that

$$(\delta_{\mathcal{S}} f_{\mathrm{a}}) \circledast \mathcal{G} = f_{\mathrm{a}}, \tag{3.8}$$

where, due to the synthesis identity, $\mathcal{S}_{\mathrm{a}} = \mathcal{S}_{\mathrm{s}} = \mathcal{S}$. The result becomes more familiar when the analysis function is given by the sound pressure $p(\mathbf{r}, \Omega)$, such that

$$(\delta_{\mathcal{S}} p) \circledast \mathcal{G} = p. \tag{3.9}$$

The result is the Kirchhoff-Helmholtz equation, introduced in the previous chapter.

3.2.2.2 Linearity

One of the basic properties of the surface convolution is that it satisfies the superposition principle, *i.e.*, if the analysis function is a linear combination of individual functions $f_{\mathrm{a},k}$ such that $f_{\mathrm{a}} = \sum_k a_k f_{\mathrm{a},k}$, where $a_k \in \mathbb{C}$, then

$$\left(\sum_k a_k f_{\mathrm{a},k} \right) \circledast f_{\mathrm{s}} = \sum_k a_k \left(f_{\mathrm{a},k} \circledast f_{\mathrm{s}} \right) \tag{3.10}$$

$$= \sum_k a_k f_k. \tag{3.11}$$

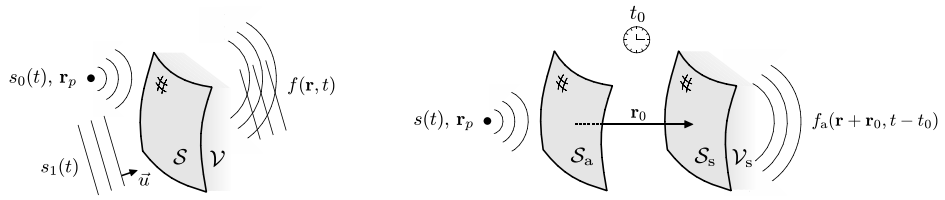
In other words, the response to a linear combination of individual inputs $f_{\mathrm{a},k}$ is a linear combination of individual outputs f_k. The proof of the linearity property is given in Theorem 6 in Section 3.6.

A direct consequence of the linearity property is that a null input always results in a null output for all $\mathbf{r}$ and t. In a physical context, this means that the air particles have to be excited by an advancing wave front in order to act as secondary sources. The following example shows how the linearity property applies to a real-world scenario.

Example 1. Consider a free space with two sound sources—one in the near-field and one in the far-field. The near-field source is located at $\mathbf{r}_p$ and, according to (2.10), has a radiation pattern given by $f_{\mathrm{a},0}(\mathbf{r}, \Omega) = \frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|}$. The far-field source has a radiation pattern given by $f_{\mathrm{a},1}(\mathbf{r}, \Omega) = e^{j\frac{\Omega}{c}\mathbf{u}\cdot\mathbf{r}}$, in accordance with (2.17), where $\mathbf{u}$ is the direction of propagation. Suppose the sources are driven by the signals $s_0(t)$ and $s_1(t)$, respectively. The resulting wave front is a linear combination of the two individual radiation patterns weighted by the respective source signals,

$$\begin{aligned} f_{\mathrm{a}} &= a_0 f_{\mathrm{a},0} + a_1 f_{\mathrm{a},1} \\ &= S_0(\Omega) \frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} + S_1(\Omega) e^{j\frac{\Omega}{c}\mathbf{u}\cdot\mathbf{r}}. \end{aligned}$$

The goal is to use the Huygens principle to re-synthesize the two wave fronts in an enclosed volume $\mathcal{V}$, by using the synthesis operator $f_{\mathrm{s}} = \mathcal{G}$. Due to the linearity property of the surface convolution, as well as the multiplicative identity, it follows that



(a) (b)

Figure 3.3: Examples of the linearity and shift invariance properties of the surface convolution. (a) The two radiation patterns generated on the outside of the enclosed volume $\mathcal{V}$—one in the near-field, other in the far-field—are re-synthesized and superposed inside $\mathcal{V}$, in accordance with the linearity property. (b) The near-field radiation pattern is observed at the analysis surface $\mathcal{S}_\mathrm{a}$ and re-synthesized t_0 seconds later at a remote synthesis surface $\mathcal{S}_\mathrm{s}$, shifted by $\mathbf{r}_0$ with respect to $\mathcal{S}_\mathrm{a}$. The result is an unaltered output, affected only by the spatial shift and the temporal delay.

$$
\begin{aligned}
f &= f_\mathrm{a} \circledast \mathcal{G} \\
&= \left(S_0(\Omega) \frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} + S_1(\Omega) e^{j\frac{\Omega}{c}\mathbf{u}\cdot\mathbf{r}} \right) \circledast \mathcal{G}(\mathbf{r},\Omega) \\
&= S_0(\Omega) \left(\frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} \circledast \mathcal{G}(\mathbf{r},\Omega) \right) + S_1(\Omega) \left(e^{j\frac{\Omega}{c}\mathbf{u}\cdot\mathbf{r}} \circledast \mathcal{G}(\mathbf{r},\Omega) \right) \\
&= S_0(\Omega) \frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} + S_1(\Omega) e^{j\frac{\Omega}{c}\mathbf{u}\cdot\mathbf{r}} \\
&= a_0 f_0 + a_1 f_1
\end{aligned}
$$

This example is illustrated in Figure 3.3-(a).

3.2.2.3 Shift invariance

Another important property of the surface convolution is that, under certain conditions, it is invariant to both spatial and temporal shifts of the analysis and synthesis functions. In the case of the synthesis function, if the output to a given f_s is given by $f = f_\mathrm{a} \circledast f_\mathrm{s}$, then a shifted version of f_s yields

$$f_\mathrm{a} \circledast L_{(\mathbf{r}_0,t_0)} f_\mathrm{s} = L_{(-\mathbf{r}_0,t_0)} f, \tag{3.12}$$

for any $\mathcal{V}_\mathrm{a}, \mathcal{V}_\mathrm{s} \subset \mathbb{R}^3$, where $L_{(\mathbf{r}_0,t_0)}$ is a delay operator such that $L_{(\mathbf{r}_0,t_0)} f(\mathbf{r},\Omega) = f(\mathbf{r} - \mathbf{r}_0, \Omega) e^{-j\Omega t_0}$. The same is true if the analysis function is shifted:

$$L_{(\mathbf{r}_0,t_0)} f_\mathrm{a} \circledast f_\mathrm{s} = L_{(\mathbf{r}_0,t_0)} f, \tag{3.13}$$

provided that the analysis and synthesis volumes are equally shaped, *i.e.*, $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s}$. Under this condition, the two cases combined yield

$$L_{(\mathbf{r}_0,t_0)} f_\mathrm{a} \circledast L_{(\mathbf{r}_1,t_1)} f_\mathrm{s} = L_{(\mathbf{r}_0-\mathbf{r}_1,t_0+t_1)} f. \tag{3.14}$$

In other words, the response to a shifted analysis or synthesis function, f_a and f_s, is a shifted output f, regardless of which variable is shifted. The proof of the shift invariance property is given in Theorem 7 in Section 3.6.

Note that the concept of causality does not exist in the spatial dimension as it exists in the temporal dimension. The difference between a plus and a minus sign in the components of the shift vectors $\mathbf{r}_0$ and $\mathbf{r}_1$ only determines the direction to which the surfaces S_a and S_s move in space. The following example shows an application of the shift invariance property.

Example 2. Consider an acoustic scene with one point source in the near-field located at $\mathbf{r}_p$ and driven by $s(t)$. From (2.10), it follows that

$$f_\mathrm{a}(\mathbf{r},\Omega) = S(\Omega)\frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|}.$$

The goal is to use the Huygens principle to re-synthesize the wave field *after* t_0 *seconds* in an enclosed volume $\mathcal{V}_\mathrm{s}$ that is *shifted in space by* $\mathbf{r}_0$ with respect to the original observation region. Using the synthesis operator $f_\mathrm{s} = \mathcal{G}$ and the shift invariance property, it follows that

$$\begin{aligned} f_\mathrm{a} \circledast L_{(\mathbf{r}_0,t_0)}\mathcal{G} &= S(\Omega)\left(\frac{e^{-j\frac{\Omega}{c}\|\mathbf{r}-\mathbf{r}_p\|}}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} \circledast \mathcal{G}(\mathbf{r}-\mathbf{r}_0,\Omega)\right) e^{-j\frac{\Omega}{c}t_0} \\ &= S(\Omega)\frac{e^{-j\frac{\Omega}{c}\|(\mathbf{r}+\mathbf{r}_0)-\mathbf{r}_p\|}}{4\pi\|(\mathbf{r}+\mathbf{r}_0)-\mathbf{r}_p\|} e^{-j\frac{\Omega}{c}t_0} \\ &= f_\mathrm{a}(\mathbf{r}+\mathbf{r}_0,\Omega)e^{-j\frac{\Omega}{c}t_0} \\ &= L_{(-\mathbf{r}_0,t_0)} f_\mathrm{a} \end{aligned}$$

This example is illustrated in Figure 3.3-(b).

3.2.2.4 Half-space property

The difference between a surface convolution between two functions f_a and f_s and a regular 2-D convolution integral is that in the surface convolution only the values within the range of $\varphi_\mathrm{a}(\boldsymbol{v})$ and $\varphi_\mathrm{s}(\boldsymbol{v})$ are used in the calculation (recall that these two functions map a plane region in $\mathbb{R}^2$ to a curved region in $\mathbb{R}^3$). This means that, even though the surface convolution is linear and shift-invariant, it does not satisfy other properties such as commutativity and associativity, *i.e.*, $f_\mathrm{a} \circledast f_\mathrm{s} \neq f_\mathrm{s} \circledast f_\mathrm{a}$ and $(f_0 \circledast f_1) \circledast f_2 \neq f_0 \circledast (f_1 \circledast f_2)$.

The scenario changes when the functions $\varphi_\mathrm{a}(\boldsymbol{v})$ and $\varphi_\mathrm{s}(\boldsymbol{v})$ are of the type $\varphi(\boldsymbol{v}) = \mathbf{a}\cdot\boldsymbol{v}+b$, where $\mathbf{a} = (a_0, a_1)$ and b are real-valued constants. This is when the functions map a plane region in $\mathbb{R}^2$ to another plane region in $\mathbb{R}^3$; the two cases are compared in Figure 3.4. In this particular case, the surface convolution reduces to a regular 2-D convolution, where the relation is given by

$$f_\mathrm{a} \circledast f_\mathrm{s} = \frac{1}{a_0 a_1} f_\mathrm{a} * f_\mathrm{s}. \tag{3.15}$$

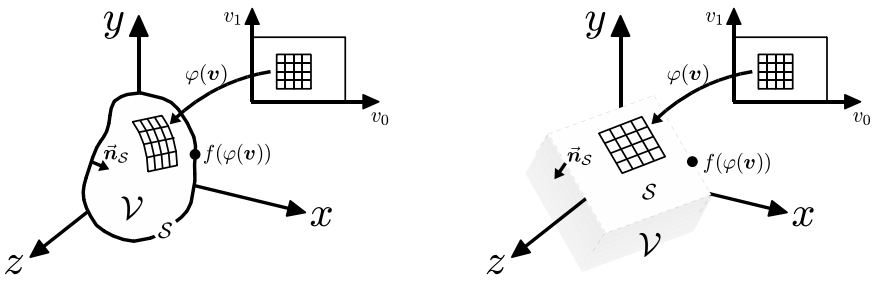


Figure 3.4: Comparison between mapping functions. The function $\varphi(\boldsymbol{v})$ maps a vector $\boldsymbol{v} \in \mathbb{R}^2$ to a vector $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{S}$, where $\mathcal{S}$ is: (a) a curved surface, or (b) a plane surface. In the case of (b), the plane surface has infinite size, and the volume it "encloses" is the entire half-space.

The proof of this property is given in Theorem 8 in Section 3.6. The expression can be further simplified if $\boldsymbol{v}$ represents the xy-plane, in which case $\mathbf{a} = (1, 1)$ and $b = 0$, and thus

$$f_{\mathrm{a}} \circledast f_{\mathrm{s}} = f_{\mathrm{a}} * f_{\mathrm{s}}. \tag{3.16}$$

By transforming the surface convolution into a regular 2-D convolution, the commutative and associative properties become valid, and thus $f_{\mathrm{a}} * f_{\mathrm{s}} = f_{\mathrm{s}} * f_{\mathrm{a}}$ and $(f_0 * f_1) * f_2 = f_0 * (f_1 * f_2)$. This also simplifies the analysis of acoustic wave fields in various ways, some of which we discuss later in the thesis.

3.3 LSTI systems based on the Huygens principle

The motivation for expressing and generalizing the Kirchhoff-Helmholtz equation into what we call the surface convolution is that it allows the use of the Huygens principle as a signal processing system. Looking at the properties of the surface convolution, summarized in Table 3.1, the Kirchhoff-Helmholtz equation is a particular case where the wave field is re-synthesized in the same region where it is observed, and at exactly the same time. The general definition, however, is much more flexible, in the sense that: (i) it allows the wave field to be re-synthesized at a remote region in space, at a later moment in time, and (ii) to be analyzed, processed, and stored before being re-synthesized.

The surface convolution has two properties that define a typical signal processing system: linearity and shift invariance. Since the Kirchhoff-Helmholtz equation is a particular case of the surface convolution, this means that any spatio-temporal system based on the Huygens principle is LSTI, as long as the filtering operations performed in between are also LSTI. Consider the following example.

Example 3. A famous orchestra is playing at the Montreux Jazz Festival in Switzerland, with the first test transmission in wave field synthesis (WFS) audio format. The recording

Table 3.1: List of properties of the surface convolution.

General definition	$f(\mathbf{r},\Omega) = f_\mathrm{a}(\mathbf{r},\Omega) \circledast f_\mathrm{s}(\mathbf{r},\Omega)$	, $\mathbf{r} \in \mathcal{V}_\mathrm{s}$
Analysis identity	$(\delta_{\mathcal{S}_\mathrm{a}} f_\mathrm{a}) \circledast f_\mathrm{s} = f_\mathrm{a} \circledast f_\mathrm{s}$	
Synthesis identity	$f_\mathrm{a} \circledast \mathcal{G} = f_\mathrm{a}$	, $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s}$
Kirchhoff-Helmholtz	$(\delta_\mathcal{S} p) \circledast \mathcal{G} = p$	, $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s}$
Linearity	$(\sum_k a_k f_{\mathrm{a},k}) \circledast f_\mathrm{s} = \sum_k a_k f_k$.	
Shift invariance	$f_\mathrm{a} \circledast L_{(\mathbf{r}_0,t_0)} f_\mathrm{s} = L_{(-\mathbf{r}_0,t_0)} f$	
	$L_{(\mathbf{r}_0,t_0)} f_\mathrm{a} \circledast L_{(\mathbf{r}_1,t_1)} f_\mathrm{s} = L_{(\mathbf{r}_0-\mathbf{r}_1,t_0+t_1)} f$	, $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s}$
Half-space property	$f_\mathrm{a} \circledast f_\mathrm{s} = f_\mathrm{a} * f_\mathrm{s}$	, $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s} = \mathbb{R}^3_{z<0}$

is made at the stage with a large array of microphones, which faithfully captures the entire sound field generated by the orchestra. The signal is then streamed to a research institute in Germany, where the sound field is reproduced in a listening room endowed with a WFS playback system. Our goal is to model this scenario as an LSTI system based on the Huygens principle.

From Table 3.1, the Kirchhoff-Helmholtz equation is given by

$$(\delta_\mathcal{S} p) \circledast \mathcal{G} = p. \tag{3.17}$$

In this case, we can consider that the scene is 2-dimensional, and both the microphone and loudspeaker arrays are placed on a contour $\mathcal{L}$ that borders the listeners area $\mathcal{A}$. Now, suppose the spatial shift between the concert hall in Switzerland and the listening room in Germany is $\mathbf{r}_0$ and the streaming delay is t_0. Then,

$$(\delta_\mathcal{L} p) \circledast L_{(\mathbf{r}_0,t_0)} \mathcal{G} = L_{(-\mathbf{r}_0,t_0)} p. \tag{3.18}$$

In addition, the recorded sound field is pre-processed in Switzerland before being streamed to Germany. This can be modeled as a filter h_a applied to the analysis surface, such that[1]

$$(\delta_\mathcal{L} p) \circledast (\delta_\mathcal{L} h_\mathrm{a}) \circledast L_{(\mathbf{r}_0,t_0)} \mathcal{G} = L_{(-\mathbf{r}_0,t_0)} y. \tag{3.19}$$

Note that the output is not p anymore, but some processed version y. Finally, we can assume that the listeners in Germany also have the means of processing the sound field before it is reproduced by the WFS playback system. This can be modeled as a filter h_s applied to the synthesis surface, such that

$$(\delta_\mathcal{L} p) \circledast (\delta_\mathcal{L} h_\mathrm{a}) \circledast (\delta_\mathcal{L} h_\mathrm{s}) \circledast L_{(\mathbf{r}_0,t_0)} \mathcal{G} = L_{(-\mathbf{r}_0,t_0)} y. \tag{3.20}$$

The example is illustrated in Figure 3.5.

This type of modeling, however, has a number of disadvantages: (i) it requires the use of a fixed Euclidean referential to account for the separation between the analysis and synthesis

[1]Here we assume a left-to-right convention, $f_0 \circledast f_1 \circledast \cdots \circledast f_{n-1} = (((f_0 \circledast f_1) \circledast \cdots) \circledast f_{n-1}$.

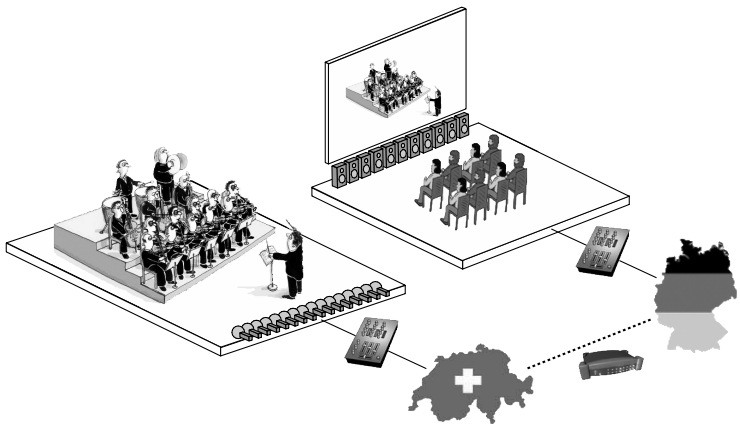


Figure 3.5: Example of a sound field streaming experiment modeled as an LSTI system based on the Huygens principle. The sound field generated by the orchestra in Switzerland is captured by a microphone array in the concert hall. The signal is then pre-processed and streamed to Germany, where the listening room is located. The received signal is processed again, and then fed to a loudspeaker array that re-synthesizes the sound field through WFS techniques. The equivalent LSTI system is given by a series of surface filtering operations on the input sound pressure $p(\mathbf{r}, t)$ and subsequent reconstruction by the synthesis operator $\mathcal{G}(\mathbf{r}, t)$, properly shifted to account for the distance between countries.

regions, and, most of all, (ii) it is highly redundant, since it uses a 3-dimensional space to describe operations that are essentially 2-dimensional, such as the surface integral. It is reasonable to say that there exists a more optimal, lower-dimensional representation of the analysis and synthesis regions that does the same job in terms of modeling.

In the next section, we will show how such a lower-dimensional representation can be constructed through the use of non-Euclidean spaces.

3.4 Signals and sequences

In digital signal processing (DSP), we tend to think of signals as unit-less representations of a given quantity, with no special regard for absolute measurements. A signal representing the sound pressure, for example, should ideally have its amplitude scaled in units of Pascal (Pa), which is the SI unit for pressure. In practice, however, the pressure measurements are performed by a transducer (*e.g.*, a microphone) that has its own output scale—typically voltage. It may also happen that the electrical signal goes through an A/D converter, which itself adjusts the amplitude scale to the dynamic range of the quantizer. Thus, by the time the signal is available for processing, all the absolute measurements related to the original metric are essentially lost. The differential measurements are the only information preserved.

Something similar happens with respect to the independent variables of the signal, such as time. When a continuous signal is discretized by an A/D converter, the result is always a timeless sequence of samples, regardless of the sampling frequency that is used. For this reason, discrete samples are represented as $s[0]$, $s[1]$, $s[2]$, *etc*, instead of $s(0)$, $s(T)$, $s(2T)$,

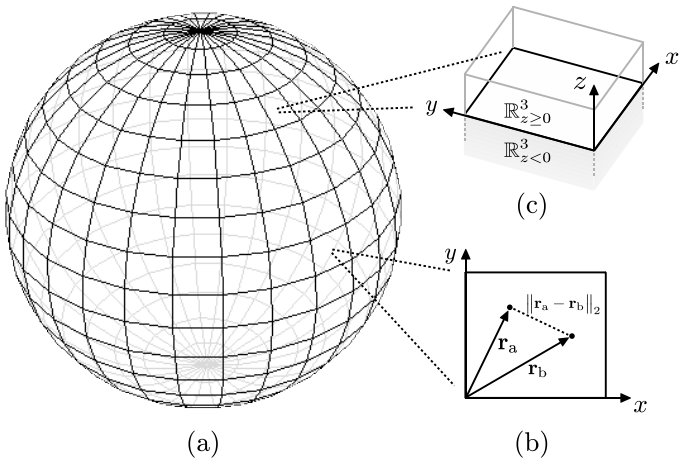


Figure 3.6: Example of a spherical manifold. (a) On a large scale, the topological space is curved and therefore non-Euclidean. (b) On a small scale, the space becomes Euclidean and the distance between vectors is given by the regular ℓ_2-norm. (c) If the local space is 3-dimensional—by considering the z variable—then the xy-plane separates the $\mathbb{R}^3$ space into two half-spaces, $\mathbb{R}^3_{z<0}$ and $\mathbb{R}^3_{z\geq 0}$.

etc. Another example is when pictures are captured by a digital camera: the result is always a matrix of amplitude values, regardless of the distance between pixel sensors.

This reasoning suggests that, if the sound pressure is measured at different points in space, the Cartesian coordinates with respect to an absolute referential will not be preserved in a practical application that involves some form of DSP. The only information preserved is the differential amplitude of the sound pressure from one point in space to the other. Therefore, it makes sense that the surface $\mathcal{S}$ delimiting the enclosed volume $\mathcal{V}$ is considered itself to be the main referential, because that is where all the information is obtained. The question is: how do we build such a referential, and how can a unit-less signal be obtained out of it?

3.4.1 Geometry of space/time

To transform the surface $\mathcal{S}$ into a referential for the representation of input signals, we use a concept known as a *manifold*. By definition, a manifold is a general topological space that is locally Euclidean [59], as the one illustrated in Figure 3.6. A classical example is the planet Earth, which is spherical when seen from space (global view) but flat when seen from the surface (local view). At the Earth's surface, it is reasonable to say that people move on a 2-D plane, and not on the tip of a vector connecting the center of Earth to them. Thus, for any person standing at a random place on Earth, the referential is always local and Euclidean. If the space is 3-dimensional ($\mathbb{R}^3$), then the sky above is one half-space ($\mathbb{R}^3_{z\geq 0}$) and the Earth bellow is the other half-space ($\mathbb{R}^3_{z<0}$).

In the context of LSTI systems, the conditions above invoke the half-space property of the

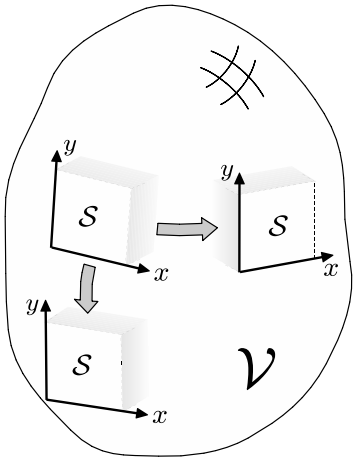


Figure 3.7: Local analysis of the surface manifold. The xy-plane moves tangentially along the surface, depending on the region where the sound pressure is observed. Due to the half-space property of the surface convolution, the Kirchhoff-Helmholtz equation is locally given by a regular 2-D convolution.

surface convolution, $f_\mathrm{a} \circledast f_\mathrm{s} = f_\mathrm{a} * f_\mathrm{s}$, which reduces the surface convolution to a regular 2-D convolution. For any surface $\mathcal{S}$ that is a smooth manifold, the Kirchhoff-Helmholtz equation is locally given by

$$p(\mathbf{r},t) = p(\mathbf{r},t) * \mathcal{G}(\mathbf{r},t) = \int_{\mathbb{R}^2} p(\boldsymbol{v},\Omega)\mathcal{G}(\mathbf{r}-\boldsymbol{v},\Omega)d\boldsymbol{v}, \tag{3.21}$$

where $\mathbf{r} = (x,y,z) \in \mathbb{R}^3_{z<0}$ and $\boldsymbol{v} = (x,y) \in \mathbb{R}^2$. This means that, when observing the different regions of $\mathcal{S}$, the xy-plane moves along the surface like a tangent sliding window, as illustrated in Figure 3.7. The process is analogous to the short-time analysis of signals, where a window function slides along the t-axis, and will be described in more detail in Chapter 5.

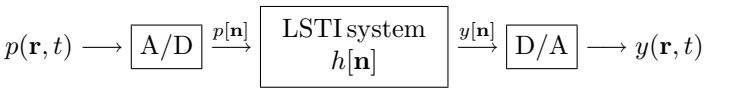
3.4.2 Continuous and discrete space/time signals

The series of mathematical simplifications and re-interpretation of the equations discussed until now places the entire problem of observing, processing, and re-synthesizing the acoustic wave field into a signal processing framework. All the physics-related terminology, such as Green's function and Kirchhoff-Helmholtz equation, as well as the involved calculations, such as line and surface integrals, can be replaced by a simple LSTI system endowed with a convolution operation:

$$p(\mathbf{r},t) \longrightarrow \boxed{\begin{array}{c}\text{LSTI system}\\ h(\boldsymbol{v},t)\end{array}} \longrightarrow y(\mathbf{r},t)$$

where $p(\mathbf{r},t)$ is the *input signal*, $y(\mathbf{r},t)$ is the *output signal*, and $h(\boldsymbol{v},t)$ is the *impulse response*. We call these signals the *continuous space/time signals*.

If the system is discrete, the A/D and D/A converters guarantee that the space/time signals are also discretized at the input and output of the system:



where $p[\mathbf{n}]$ is the *input signal*, $y[\mathbf{n}]$ is the *output signal*, and $h[\mathbf{n}]$ is the *impulse response*. We call these signals the *discrete space/time signals* (or sequences). The argument $\mathbf{n} = (n_x, n_y, n_t)$ contains the integer indexes of the spatial samples, n_x and n_y, and the temporal samples, n_t.

It is important to notice that, even though the input and output signals are 4-dimensional (*i.e.*, 3 spatial dimensions plus time), the LSTI system operates in a 3-dimensional space (2 spatial dimensions plus time); the z variable is not processed by the system. One intuitive way of looking at the problem is that the space/time signal is a projection of the wave field "image" onto the "screen" defined by the surface $\mathcal{S}$. Therefore, in practice, processing the acoustic wave field can be compared to processing gray-scale images. In the next chapter, we will show how such wave field "images" can be efficiently represented and processed in a Fourier-based domain.

3.5 Summary notes

- The microphone array acts as an A/D converter of acoustic wave fields, and the loudspeaker array acts as a D/A converter. Wave field synthesis (WFS) is one example of a D/A interpolation algorithm;
- The surface convolution expresses acoustic wave fields as input signals and physical principles (such as the Huygens principle) as impulse responses;
- The surface convolution is linear and space/time invariant (LSTI), and it yields the Kirchhoff-Helmholtz equation as a particular case. Therefore, the Huygens principle can be translated into an LSTI system;
- A long-range transmission of an entire acoustic wave field can be modeled by an LSTI system based on the Huygens principle, and thus conveniently expressed as a surface convolution;
- The pressure signal can be described by two spatial dimensions instead of three by treating the analysis surface as a manifold, locally approximated by a plane. In this case, the surface convolution becomes a regular 2-D convolution;
- Space/time signals are unit-less measurements of the sound pressure in space and time.

3.6 Theorems and proofs

Theorem 4 (Analysis identity)**.** *Given two arbitrary volumes $\mathcal{V}_\mathrm{a}, \mathcal{V}_\mathrm{s} \subset \mathbb{R}^3$ and a delta function $\delta_{\mathcal{S}_\mathrm{a}}(\mathbf{r})$ that equals one for $\mathbf{r} \in \mathbb{R}^3 \cap \mathcal{S}_\mathrm{a}$ and zero otherwise, for any analysis function $f_\mathrm{a}(\mathbf{r}, \Omega)$ and synthesis function $f_\mathrm{s}(\mathbf{r}, \Omega)$, it follows that*

$$(\delta_{\mathcal{S}_\mathrm{a}}(\mathbf{r}) f_\mathrm{a}(\mathbf{r},\Omega)) \circledast f_\mathrm{s}(\mathbf{r},\Omega) = f_\mathrm{a}(\mathbf{r},\Omega) \circledast f_\mathrm{s}(\mathbf{r},\Omega).$$

Proof. From the definition of the surface integral,

$$\begin{aligned}(\delta_{\mathcal{S}_\mathrm{a}}(\mathbf{r}) f_\mathrm{a}(\mathbf{r},\Omega)) \circledast f_\mathrm{s}(\mathbf{r},\Omega) &= \int_{\mathbb{R}^2} \delta_{\mathcal{S}_\mathrm{a}}(\varphi_\mathrm{a}(\boldsymbol{v})) f_\mathrm{a}(\varphi_\mathrm{a}(\boldsymbol{v}),\Omega) f_\mathrm{s}(\mathbf{r}-\varphi_\mathrm{s}(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\ &\overset{(a)}{=} \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi_\mathrm{a}(\boldsymbol{v}),\Omega) f_\mathrm{s}(\mathbf{r}-\varphi_\mathrm{s}(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\ &\overset{\mathrm{def}}{=} f_\mathrm{a}(\mathbf{r},\Omega) \circledast f_\mathrm{s}(\mathbf{r},\Omega)\end{aligned}$$

where (a) comes from the fact that $\varphi_\mathrm{a}(\boldsymbol{v}) : \mathbb{R}^2 \mapsto \mathbb{R}^3 \cap \mathcal{S}_\mathrm{a}$ and thus $\delta_{\mathcal{S}_\mathrm{a}}(\varphi_\mathrm{a}(\boldsymbol{v})) = 1$. Since there is no change in the synthesis function, this proof also holds for $f_\mathrm{s}(\mathbf{r},\Omega) = \mathcal{G}(\mathbf{r},\Omega)$. □

Theorem 5 (Synthesis identity)**.** *Given an arbitrary volume $\mathcal{V} = \mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s} \subset \mathbb{R}^3$ and a synthesis operator*

$$\mathcal{G}(\mathbf{r},\Omega) = g(\mathbf{r},\Omega) \left\|\boldsymbol{n}_\mathcal{S}\right\| \frac{\partial}{\partial \boldsymbol{n}_\mathcal{S}} - \left\|\boldsymbol{n}_\mathcal{S}\right\| \frac{\partial g(\mathbf{r},\Omega)}{\partial \boldsymbol{n}_\mathcal{S}}$$

for any analysis function $f_\mathrm{a}(\mathbf{r},\Omega)$, it follows that $f_\mathrm{a}(\mathbf{r},\Omega) \circledast \mathcal{G}(\mathbf{r},\Omega) = f_\mathrm{a}(\mathbf{r},\Omega)$.

Proof. From the definition of surface integral,

$$\begin{aligned}f_\mathrm{a}(\mathbf{r},\Omega) \circledast \mathcal{G}(\mathbf{r},\Omega) &= \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi_\mathrm{a}(\boldsymbol{v}),\Omega) \mathcal{G}(\mathbf{r}-\varphi_\mathrm{s}(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\ &\overset{(a)}{=} \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi(\boldsymbol{v}),\Omega) \mathcal{G}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\ &\overset{\mathrm{def}}{=} f_\mathrm{a}(\mathbf{r},\Omega)\end{aligned}$$

where (a) comes from the fact that $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s} = \mathcal{V}$ and thus $\varphi_\mathrm{a}(\boldsymbol{v}) = \varphi_\mathrm{s}(\boldsymbol{v}) = \varphi(\boldsymbol{v})$, where $\varphi(\boldsymbol{v}) : \mathbb{R}^2 \mapsto \mathbb{R}^3 \cap \mathcal{S}$. The result, by definition, is the Kirchhoff-Helmholtz equation. □

Theorem 6 (Distributivity and linearity)**.** *Given two arbitrary volumes $\mathcal{V}_\mathrm{a}, \mathcal{V}_\mathrm{s} \subset \mathbb{R}^3$, an arbitrary number of analysis functions $f_{\mathrm{a},k}(\mathbf{r},\Omega)$, a synthesis function $f_\mathrm{s}(\mathbf{r},\Omega)$, and a frequency dependent factor $a_k(\Omega) \in \mathbb{C}$, if $f_{\mathrm{a},k}(\mathbf{r},\Omega) \circledast f_\mathrm{s}(\mathbf{r},\Omega) = f_k(\mathbf{r},\Omega)$,* then it follows that

$$\left(\sum_k a_k(\Omega) f_{\mathrm{a},k}(\mathbf{r},\Omega)\right) \circledast f_\mathrm{s}(\mathbf{r},\Omega) = \sum_k a_k(\Omega) \left(f_{\mathrm{a},k}(\mathbf{r},\Omega) \circledast f_\mathrm{s}(\mathbf{r},\Omega)\right)$$

and thus

$$\left(\sum_k a_k(\Omega) f_{\mathrm{a},k}(\mathbf{r},\Omega)\right) \circledast f_\mathrm{s}(\mathbf{r},\Omega) = \sum_k a_k(\Omega) f_k(\mathbf{r},\Omega).$$

Proof. From the definition of the surface integral,

$$\begin{aligned}
&\left(\sum_k a_k(\Omega) f_{a,k}(\mathbf{r},\Omega)\right) \circledast f_{\mathrm{s}}(\mathbf{r},\Omega) \\
= &\int_{\mathbb{R}^2} \left(\sum_k a_k(\Omega) f_{a,k}(\varphi(\boldsymbol{v}),\Omega)\right) f_{\mathrm{s}}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\
\overset{(a)}{=} &\int_{\mathbb{R}^2} \sum_k a_k(\Omega) \left(f_{a,k}(\varphi(\boldsymbol{v}),\Omega) f_{\mathrm{s}}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega)\right) d\boldsymbol{v} \\
\overset{(b)}{=} &\sum_k a_k(\Omega) \int_{\mathbb{R}^2} f_{a,k}(\varphi(\boldsymbol{v}),\Omega) f_{\mathrm{s}}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\
\overset{(c)}{=} &\sum_k a_k(\Omega) \left(f_{a,k}(\mathbf{r},\Omega) \circledast f_{\mathrm{s}}(\mathbf{r},\Omega)\right) \\
\overset{(d)}{=} &\sum_k a_k(\Omega) f_k(\mathbf{r},\Omega)
\end{aligned}$$

In the case of step (a), where $f_{\mathrm{s}}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega)$ goes into the sum, the modification requires special care if $f_{\mathrm{s}}(\mathbf{r},\Omega) = \mathcal{G}(\mathbf{r},\Omega)$. The definition of $\mathcal{G}(\mathbf{r},\Omega)$ yields

$$\begin{aligned}
&\left(\sum_k a_k(\Omega) p_k(\varphi(\boldsymbol{v}),\Omega)\right) \mathcal{G}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) \\
=&g(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) \left\|\boldsymbol{n}_{\mathcal{S}}\right\| \frac{\partial}{\partial \boldsymbol{n}_{\mathcal{S}}} \sum_k a_k(\Omega) p_k(\varphi(\boldsymbol{v}),\Omega) - \left\|\boldsymbol{n}_{\mathcal{S}}\right\| \frac{\partial g(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} \sum_k a_k(\Omega) p_k(\varphi(\boldsymbol{v}),\Omega) \\
=&\sum_k a_k(\Omega) g(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) \left\|\boldsymbol{n}_{\mathcal{S}}\right\| \frac{\partial p_k(\varphi(\boldsymbol{v}),\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} - \sum_k a_k(\Omega) \left\|\boldsymbol{n}_{\mathcal{S}}\right\| \frac{\partial g(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} p_k(\varphi(\boldsymbol{v}),\Omega) \\
=&\sum_k a_k(\Omega) \left(g(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) \left\|\boldsymbol{n}_{\mathcal{S}}\right\| \frac{\partial p_k(\varphi(\boldsymbol{v}),\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} - \left\|\boldsymbol{n}_{\mathcal{S}}\right\| \frac{\partial g(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega)}{\partial \boldsymbol{n}_{\mathcal{S}}} p_k(\varphi(\boldsymbol{v}),\Omega)\right) \\
=&\sum_k a_k(\Omega) \left(p_k(\varphi(\boldsymbol{v}),\Omega) \mathcal{G}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega)\right)
\end{aligned}$$

Thus, the properties are also valid if $f_{\mathrm{s}}(\mathbf{r},\Omega) = \mathcal{G}(\mathbf{r},\Omega)$. Step (b) comes from the linearity of the integral. If the sum is infinite, then (b) requires that the Fubini theorem [16] is satisfied: $\int_{\mathbb{R}^2} \sum_k \left|a_k(\Omega) f_{a,k}(\varphi(\boldsymbol{v}),\Omega)\right| d\boldsymbol{v} < \infty$. The step (c) comes by definition of the surface convolution, and proves the distributive property. The step (d) proves the linearity property. □

Theorem 7 (Shift-invariance)**.** *Given two arbitrary volumes $\mathcal{V}_{\mathrm{a}}, \mathcal{V}_{\mathrm{s}} \subset \mathbb{R}^3$, if $L_{(\mathbf{r}_0,t_0)}$ is a delay operator such that*

$$L_{(\mathbf{r}_0,t_0)} f(\mathbf{r},\Omega) = f(\mathbf{r}-\mathbf{r}_0,\Omega) e^{-j\Omega t_0}$$

then, given any $\mathbf{r}_0, \mathbf{r}_1 \in \mathbb{R}^3$ and $t_0, t_1 \in \mathbb{R}$, for any analysis function $f_{\mathrm{a}}(\mathbf{r},\Omega)$ and synthesis function $f_{\mathrm{s}}(\mathbf{r},\Omega)$, it follows that

$$f_{\mathrm{a}}(\mathbf{r},\Omega) \circledast L_{(\mathbf{r}_0,t_0)} f_{\mathrm{s}}(\mathbf{r},\Omega) = L_{(-\mathbf{r}_0,t_0)} f(\mathbf{r},\Omega).$$

In particular, if $\mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s} = \mathcal{V}$, then

$$L_{(\mathbf{r}_0,t_0)} f_\mathrm{a}(\mathbf{r},\Omega) \circledast L_{(\mathbf{r}_1,t_1)} f_\mathrm{s}(\mathbf{r},\Omega) = L_{(\mathbf{r}_0-\mathbf{r}_1,t_0+t_1)} f(\mathbf{r},\Omega).$$

Proof. From the definition of the surface integral,

$$\begin{aligned}
& f_\mathrm{a}(\mathbf{r},\Omega) \circledast L_{(\mathbf{r}_0,t_0)} f_\mathrm{s}(\mathbf{r},\Omega) \\
= & \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi_\mathrm{a}(\boldsymbol{v}),\Omega) f_\mathrm{s}(\mathbf{r}-(\varphi_\mathrm{s}(\boldsymbol{v})-\mathbf{r}_0),\Omega) e^{-j\Omega t_0} d\boldsymbol{v} \\
= & f(\mathbf{r}+\mathbf{r}_0,\Omega) e^{-j\Omega t_0} \\
= & L_{(-\mathbf{r}_0,t_0)} f(\mathbf{r},\Omega)
\end{aligned}$$

$$\begin{aligned}
& L_{(\mathbf{r}_0,t_0)} f_\mathrm{a}(\mathbf{r},\Omega) \circledast L_{(\mathbf{r}_1,t_1)} f_\mathrm{s}(\mathbf{r},\Omega) \\
= & \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi(\boldsymbol{v})-\mathbf{r}_0,\Omega) e^{-j\Omega t_0} f_\mathrm{s}(\mathbf{r}-(\varphi(\boldsymbol{v})-\mathbf{r}_1),\Omega) e^{-j\Omega t_1} d\boldsymbol{v} \\
\overset{(a)}{=} & \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi_\mathrm{a}(\boldsymbol{v}),\Omega) f_\mathrm{s}(\mathbf{r}-(\varphi_\mathrm{a}(\boldsymbol{v})+\mathbf{r}_0-\mathbf{r}_1),\Omega) d\boldsymbol{v} e^{-j\Omega(t_0+t_1)} \\
= & f(\mathbf{r}-(\mathbf{r}_0-\mathbf{r}_1),\Omega) e^{-j\Omega(t_0+t_1)} \\
= & L_{(\mathbf{r}_0-\mathbf{r}_1,t_0+t_1)} f(\mathbf{r},\Omega)
\end{aligned}$$

where (a) comes from the change of variables $\varphi_\mathrm{a}(\boldsymbol{v}) = \varphi(\boldsymbol{v}) - \mathbf{r}_0$. Both cases are valid also for $f_\mathrm{s}(\mathbf{r},\Omega) = \mathcal{G}(\mathbf{r},\Omega)$, since the proof only involves a change of variables. □

Theorem 8 (Half-space property)**.** *Given an arbitrary volume $\mathcal{V} = \mathcal{V}_\mathrm{a} = \mathcal{V}_\mathrm{s} \subset \mathbb{R}^3$ with a surface mapped by a function of the type $\varphi(\boldsymbol{v}) = \mathbf{a}\cdot\boldsymbol{v} + b$, where $\mathbf{a} = (a_0, a_1) \in \mathbb{R}^2$ and $b \in \mathbb{R}$, for any analysis function $f_\mathrm{a}(\mathbf{r},\Omega)$ and synthesis function $f_\mathrm{s}(\mathbf{r},\Omega)$, it follows that $f_\mathrm{a} \circledast f_\mathrm{s} = \frac{1}{a_0 a_1} f_\mathrm{a} * f_\mathrm{s}$.*

Proof. From the definition of the surface integral,

$$\begin{aligned}
f_\mathrm{a}(\mathbf{r},\Omega) \circledast f_\mathrm{s}(\mathbf{r},\Omega) &= \int_{\mathbb{R}^2} f_\mathrm{a}(\varphi(\boldsymbol{v}),\Omega) f_\mathrm{s}(\mathbf{r}-\varphi(\boldsymbol{v}),\Omega) d\boldsymbol{v} \\
&= \int_{\mathbb{R}^2} f_\mathrm{a}(\mathbf{a}\cdot\boldsymbol{v}+b,\Omega) f_\mathrm{s}(\mathbf{r}-\mathbf{a}\cdot\boldsymbol{v}-b,\Omega) d\boldsymbol{v} \\
&\overset{(a)}{=} \frac{1}{a_0 a_1} \int_{\mathbb{R}^2} f_\mathrm{a}(\boldsymbol{\sigma},\Omega) f_\mathrm{s}(\mathbf{r}-\boldsymbol{\sigma},\Omega) d\boldsymbol{\sigma} \\
&= \frac{1}{a_0 a_1} f_\mathrm{a}(\mathbf{r},\Omega) * f_\mathrm{s}(\mathbf{r},\Omega)
\end{aligned}$$

where (a) comes from the change of variables $\boldsymbol{\sigma} = \mathbf{a}\cdot\boldsymbol{v} + b$ and $d\boldsymbol{\sigma} = a_0 a_1 d\boldsymbol{v}$. This is valid also for $f_\mathrm{s}(\mathbf{r},\Omega) = \mathcal{G}(\mathbf{r},\Omega)$, since the proof only involves a change of variables. □

Chapter 4

The Spatio-Temporal Fourier Transform

4.1 Introduction

The Fourier transform is arguably one of the most important discoveries in the history of physical sciences. It was first developed in the early 19th century by the French mathematician Joseph Fourier (1768–1830). In his famous paper [33], Fourier shows how to solve the heat equation in a solid medium for a general class of periodic excitation functions (at the time, only the solution for a few analytical functions was known). Fourier's ingenious approach, instead of trying to find a solution for each particular function, was to represent each function as a linear combination of sinusoids with different frequencies, and obtain the solution for each sinusoidal component. The result was then a linear combination of the sinusoidal solutions. The expression derived by Fourier is similar to what is known today as the Fourier series, and has progressed to its modern form due to the contribution of several mathematicians, such as Bernhard Riemann (1826–1866) and Johann Dirichlet (1805–1859) [24].

The Fourier transform had a deep impact in many fields of science. It has redefined fields such as classical mechanics, optics, and acoustics, and created or contributed to new ones such as harmonic analysis, quantum mechanics, and, of course, signal processing. Such a revolution was due to the fact that many physical phenomena occurring in nature have an inherent harmonic behavior that is efficiently represented by the Fourier transform. Physical forces and wave fields tend to manifest themselves in harmonic patterns, or to induce harmonic behavior in the surrounding medium. Such behavior can not always be directly observed through instrumentation, mostly due to technological limitations, and is therefore restricted to theoretical analysis. Notable examples include the harmonic behavior of gravity (see, *e.g.*, Ciufolini *et al.* on gravitational waves [20]) and the harmonic behavior of subatomic particles (see, *e.g.*, Becker *et al.* on string theory [5]). However, in many other cases—*e.g.*, sound waves and light waves—there are several measuring instruments that allow not only the phenomena to be directly observed but also to be manipulated—for example, through the use of filters and amplifiers. This is where harmonic analysis has proved to be a valuable tool in the development of applications and devices that have become part of our

(a) (b)

Figure 4.1: Silicon microphones based on MEMS (micro-electromechanical systems) technology. (a) The microphones can get as small as 3mm [21]—about 10% the diameter of a US quarter dollar coin. The maximum distance between microphones required to sample the acoustic wave field within the listening range without introducing spatial aliasing is given by $\frac{340\mathrm{m/s}}{44100\mathrm{Hz}} = 7.7\mathrm{mm}$, which can be easily satisfied with MEMS microphones. (b) The microphone operates through a pressure-sensitive diaphragm carved directly in the silicon chip, and it may contain other built-in components such as a pre-amplifier and an A/D converter.

everyday lives.

The recent progress in microphone technology poses an interesting scenario: as the number of microphones increases and the microphones themselves become smaller (see Figure 4.1), the sampling frequency in space becomes comparable to the sampling frequency in time, to the point where spatial aliasing can be virtually eliminated. The microphone array becomes a dimension in itself (the spatial dimension) and the signals observed at the array start manifesting harmonic behavior (recall that wave fields are characterized by functions of both space and time). Under these conditions, it makes sense to perform harmonic analysis not only on the temporal dimension but also on the spatial dimension.

As usual, the first step in performing harmonic analysis on a given signal is to calculate the Fourier transform of the signal, in order to obtain a spectral representation that expresses the amplitude and phase of each harmonic component as a function of frequency. The Fourier transform is taken over all the dimensions of the signal that exhibit harmonic behavior. In the case of the spatio-temporal signal $p(\mathbf{r}, t)$, which represents the sound pressure at a given point in space and time, this implies taking the Fourier transform over the three spatial dimensions as well as the temporal dimension. The purpose of calculating the Fourier transform of $p(\mathbf{r}, t)$ is to decompose the entire wave field into its elementary harmonic components, so that the wave field can be efficiently analyzed and manipulated for later reconstruction through wave field synthesis (WFS). We will see that the harmonic components of the wave field are generalizations of the traditional sinusoid, known as *plane waves*—these are sinusoids that propagate as a plane wave front. The plane waves are eigen-functions of the linear space/time invariant (LSTI) systems.

In this chapter, we derive and illustrate the Fourier transform of the most common types of wave fronts, and show how the spectral patterns change with the parameters of the scene. We address both the continuous version of the Fourier transform and the discrete version, as

well as the method through which one is obtained from the other—*i.e.*, through sampling and interpolation. We focus mostly on the theoretical aspects of Fourier analysis and its efficiency in the representation of acoustic wave fields. Spectral manipulation of acoustic wave fields is addressed in subsequent chapters.

4.2 Continuous Fourier transform

4.2.1 Definition

The spatio-temporal Fourier transform of the 4-D signal $p(\mathbf{r},t)$ can be obtained from the multi-dimensional formulation of the Fourier integral [27], and is given by

$$P(\mathbf{\Phi},\Omega) = \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\mathbf{r},t)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r} \tag{4.1}$$

where $\mathbf{\Phi} = (\Phi_x,\Phi_y,\Phi_z)$ are the spatial frequencies in rad/m and Ω is the temporal frequency in rad/s, and where $\int_{\mathbb{R}^3}d\mathbf{r} = \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}dxdydz$. The inverse transform is given by

$$p(\mathbf{r},t) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}} P(\mathbf{\Phi},\Omega)e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\mathbf{\Phi} \tag{4.2}$$

where $\int_{\mathbb{R}^3}d\mathbf{\Phi} = \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}d\Phi_x d\Phi_y d\Phi_z$. The Fourier integral can be interpreted as the projection of $p(\mathbf{r},t)$ onto the "orthogonal basis" $e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}$, which is the basic element of the continuous Fourier transform.

We will show several examples to help understand the spectral patterns generated by (4.1) given the most common inputs, and illustrate the results in the 3-D and 2-D spaces (which are the cases of interest in LSTI systems based on the Huygens principle). We focus only on the far-field and near-field cases, and leave the intermediate-field to the next chapter.

Example 9 (Far-field spectrum)**.** Consider a point source located in the far-field with source signal $s(t)$ and propagation vector $\mathbf{u}$, such that $p(\mathbf{r},t) = s\Big(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\Big)$. Plugging $p(\mathbf{r},t)$ into (4.1), yields

$$\begin{aligned} P(\mathbf{\Phi},\Omega) &= \int_{\mathbb{R}^m}\left(\int_{\mathbb{R}} s\Big(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\Big)e^{-j\Omega t}dt\right)e^{-j\mathbf{\Phi}\cdot\mathbf{r}}d\mathbf{r} \\ &= S(\Omega)\int_{\mathbb{R}^m} e^{-j\left(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\right)\cdot\mathbf{r}}d\mathbf{r}, \end{aligned}$$

and thus

$$\boxed{P(\mathbf{\Phi},\Omega) = S(\Omega)(2\pi)^m\delta\Big(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)} \tag{4.3}$$

where $\mathbf{\Phi}$,$\mathbf{u}\in\mathbb{R}^m$ and $m = 0,1,2,3$. The multi-dimensional Dirac function $\delta\Big(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)$ is non-zero for $\mathbf{\Phi} = \mathbf{u}\frac{\Omega}{c}$, and can be expressed as $\delta\Big(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big) = \delta\Big(\Phi_x - u_x\frac{\Omega}{c}\Big)\delta\Big(\Phi_y - u_y\frac{\Omega}{c}\Big)\delta\Big(\Phi_z - u_z\frac{\Omega}{c}\Big)$ for $m=3$. In the 3-D space of spatial frequencies, this corresponds to a point on a spherical surface defined by $||\mathbf{\Phi}||^2 = \big(\frac{\Omega}{c}\big)^2$, as illustrated in Figure 4.2-(a).

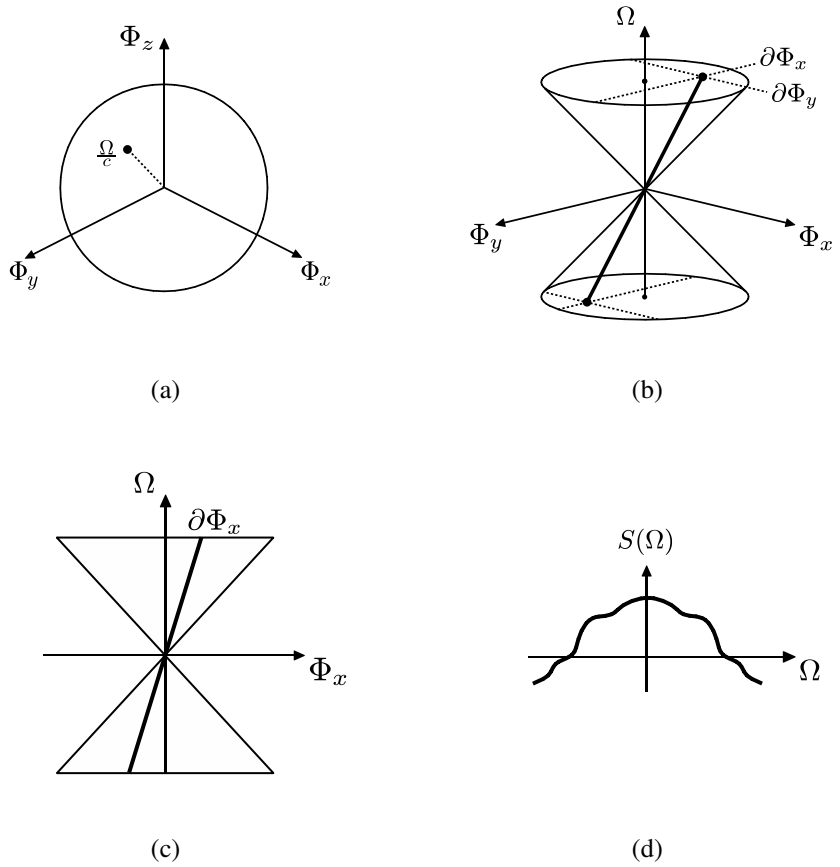


Figure 4.2: Spatio-temporal Fourier transform of a far-field source for different values of m (number of spatial dimensions). (a) For $m = 3$, and for each Ω, the spectrum is given by a Dirac point on the surface of a sphere of radius $\frac{\Omega}{c}$ (never inside the sphere), weighted by $S(\Omega)$. (b) For $m = 2$, the spectrum is given by a Dirac function defined on a line within a cone-shaped region, and weighted by $S(\Omega)$. The line always crosses the origin, and its slope across Φ_x and Φ_y is determined by the angle of arrival of the wave front with respect to the x and y-axes. (c) For $m = 1$, the spectrum is given by a Dirac function defined on a line within a triangular region, with slope determined by the angle of arrival, and weighted by $S(\Omega)$. (d) For $m = 0$, the spectrum has no spatial information, and therefore is equal to $S(\Omega)$.

According to (4.3), the spectral pattern generated by a far-field source is simply a Dirac function weighted by the Fourier transform of the source signal. The orientation of the Dirac function is given by the partial derivatives $\frac{\partial \mathbf{\Phi}}{\partial \Omega} = \frac{\mathbf{u}}{c}$, and, therefore, depends only on the direction and speed of wave propagation. In simple terms, the result in (4.3) is a product between the spatial behavior of the wave front and the temporal behavior.

The spectral pattern becomes more comprehensible by looking at the lower-dimensional cases, where $m = 2$ and $m = 1$. Consider also that $\mathbf{u}$ is expressed in polar coordinates in accordance to Figure 2.6, such that $u_x = \cos\alpha \sin\beta$, $u_y = \cos\beta$, and $u_z = \sin\alpha \sin\beta$, where $\alpha, \beta \in [0, \pi]$ are the angles with the x- and y-axes respectively.

Taking $p(\mathbf{r}, t)$ on the xy-plane, yields

$$P(\Phi_x, \Phi_y, \Omega) = S(\Omega)(2\pi)^2 \delta\Big(\Phi_x - \cos\alpha \sin\beta \frac{\Omega}{c}\Big) \delta\Big(\Phi_y - \cos\beta \frac{\Omega}{c}\Big), \tag{4.4}$$

which corresponds to a Dirac function defined along a line with partial derivatives $\frac{\partial \Phi_x}{\partial \Omega} = \frac{\cos\alpha \sin\beta}{c}$ and $\frac{\partial \Phi_y}{\partial \Omega} = \frac{\cos\beta}{c}$, and weighted by $S(\Omega)$. Given that $\alpha, \beta \in [0, \pi]$, the Dirac function is always within a cone-shaped region defined by $\Phi_x^2 + \Phi_y^2 \leq \left(\frac{\Omega}{c}\right)^2$. The result is illustrated in Figure 4.2-(b).

Taking $p(\mathbf{r}, t)$ on the x-axis, yields

$$P(\Phi_x, \Omega) = S(\Omega)(2\pi) \delta\Big(\Phi_x - \cos\alpha \frac{\Omega}{c}\Big), \tag{4.5}$$

which corresponds to a Dirac function that is non-zero along a line with slope $\frac{\partial \Phi_x}{\partial \Omega} = \frac{\cos\alpha}{c}$, and weighted by $S(\Omega)$. Given that $\alpha \in [0, \pi]$, the Dirac function is always within a triangular region defined by $\Phi_x^2 \leq \left(\frac{\Omega}{c}\right)^2$. The result is illustrated in Figure 4.2-(c).

If $m = 0$, the result reduces to the regular 1-D Fourier transform of $s(t)$, as shown in Figure 4.2-(d).

Example 10. Consider the case of Example 9 for $s(t) = e^{j\Omega_0 t}$, where Ω_0 is a fixed frequency. The Fourier transform of the complex exponential is given by $S(\Omega) = 2\pi\delta(\Omega - \Omega_0)$, and thus

$$P(\mathbf{\Phi}, \Omega) = (2\pi)^{m+1} \delta(\Omega - \Omega_0) \delta\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big). \tag{4.6}$$

This is illustrated in Figure 4.3-(a). The result is a single point in the 4-D spectrum located at $(\mathbf{\Phi}, \Omega) = \left(\mathbf{u}\frac{\Omega_0}{c}, \Omega_0\right)$, and is known as a *plane wave*[1]. The plane wave is the elementary component in the spatio-temporal Fourier analysis of the wave field, and it can be seen as a generalization of the 1-D complex frequency in the traditional context of Fourier analysis. The 1-D complex frequency is obtained as a particular case of (4.6) when $m = 0$, which means that the sound pressure is observed at one point in space.

[1]Note that any wide-band signal $s(t)$ generates a flat wave-front as long as $\|\mathbf{r}_p\| \gg \|\mathbf{r}\|$, though historically a plane wave refers to a single complex frequency [88].

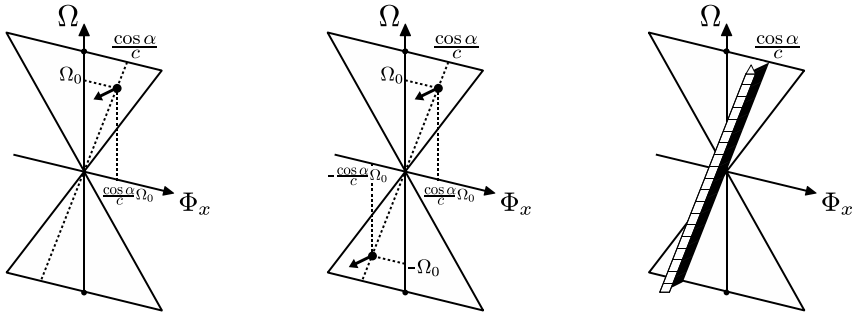


Figure 4.3: Examples of the spatio-temporal Fourier transform for $m = 1$ (one spatial dimension) and different source signals. (a) For $s(t) = e^{j\Omega_0 t}$, the spectrum is a single Dirac point located at $(\Phi_x, \Omega) = \Big(\cos\alpha\frac{\Omega_0}{c}, \Omega_0\Big)$. (b) For $s(t) = 2\cos(\Omega_0 t)$, the result is the same as in (a) plus the symmetric Dirac point at negative frequencies. (c) For $s(t) = \delta(t)$, the spectrum is a "flat" Dirac function defined on a line with slope $\frac{\partial \Phi_x}{\partial \Omega} = \frac{\cos\alpha}{c}$.

Example 11. Consider a variation of the previous example, where $s(t) = 2\cos(\Omega_0 t)$. Since $2\cos(\Omega_0 t) = e^{-j\Omega_0 t} + e^{j\Omega_0 t}$, the Fourier transform is given by $S(\Omega) = 2\pi\delta(\Omega + \Omega_0) + 2\pi\delta(\Omega - \Omega_0)$, and thus

$$P(\mathbf{\Phi}, \Omega) = (2\pi)^{m+1}\Big(\delta(\Omega + \Omega_0) + \delta(\Omega - \Omega_0)\Big)\delta\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big). \tag{4.7}$$

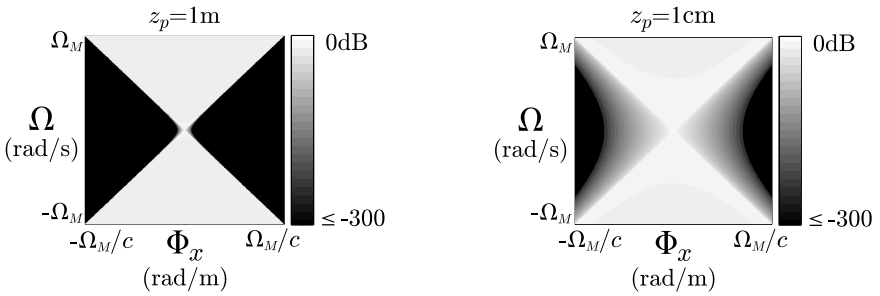
This is illustrated in Figure 4.3-(b). The result is the same as in the previous example, except for an additional point at the symmetric position $(\mathbf{\Phi}, \Omega) = \Big(-\mathbf{u}\frac{\Omega_0}{c}, -\Omega_0\Big)$. Similarly to the 1-D case, whenever $p(\mathbf{r}, t)$ is real, $P(\mathbf{\Phi}, \Omega)$ is symmetric with respect to the origin.

Example 12. If the source signal is a Dirac pulse, $s(t) = \delta(t)$, the Fourier transform is maximally flat and given by $S(\Omega) = 1$. The spatio-temporal Fourier transform is then given by

$$P(\mathbf{\Phi}, \Omega) = (2\pi)^m\delta\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big). \tag{4.8}$$

This is illustrated in Figure 4.3-(c).

The examples above show that when the source is in the far-field the spectral energy is always maximally concentrated in a Dirac function, regardless of the signal $s(t)$ that drives the source. This is not the case when the source is in the near-field. In the next examples, we will see that the proximity of the source to the observation region causes the spectral energy to spread across $\mathbf{\Phi}$.



(a) (b)

Figure 4.4: Spatio-temporal Fourier transform of a near-field source for $m = 1$ (one spatial dimension), $S(\Omega) = 1$, and different distances z_p. (a) For $z_p = 1\text{m}$, virtually all the energy is concentrated in the triangular region defined by $\Phi_x^2 \leq \left(\frac{\Omega}{c}\right)^2$. (b) For $z_p = 1\text{cm}$, the source is close to the x-axis, causing the energy to spread across Φ_x and intensifying the evanescent-wave content.

Example 13 (Near-field spectrum)**.** Consider a point source located in the near-field with source signal $s(t)$ and located at $\mathbf{r} = \mathbf{r}_p$, such that $p(\mathbf{r}, t) = \frac{1}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} s\Big(t - \frac{\|\mathbf{r}-\mathbf{r}_p\|}{c}\Big)$. Plugging $p(\mathbf{r}, t)$ into (4.1) for different values of m, yields

$$P(\mathbf{\Phi}, \Omega) = S(\Omega) \frac{e^{-j\mathbf{\Phi}\cdot\mathbf{r}_p}}{\|\mathbf{\Phi}\|^2 - \left(\frac{\Omega}{c}\right)^2}\Bigg|_{\mathbf{\Phi}=(\Phi_x,\Phi_y,\Phi_z)} \tag{4.9}$$

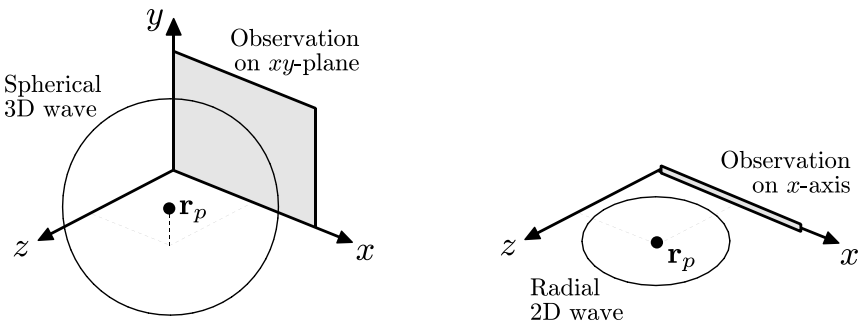
for $m = 3$,

$$P(\mathbf{\Phi}, \Omega) = S(\Omega)\frac{1}{2} \frac{e^{-|z_p|\sqrt{\|\mathbf{\Phi}\|^2 - \left(\frac{\Omega}{c}\right)^2}} e^{-j\mathbf{\Phi}\cdot(x_p,y_p)}}{\sqrt{\|\mathbf{\Phi}\|^2 - \left(\frac{\Omega}{c}\right)^2}}\Bigg|_{\mathbf{\Phi}=(\Phi_x,\Phi_y)} \tag{4.10}$$

for $m = 2$, and

$$P(\mathbf{\Phi}, \Omega) = S(\Omega)\frac{1}{4j} H_0^{(1)*}\left(\sqrt{y_p^2 + z_p^2}\sqrt{\|\mathbf{\Phi}\|^2 - \left(\frac{\Omega}{c}\right)^2}\right) e^{-j\mathbf{\Phi}\cdot x_p}\Bigg|_{\mathbf{\Phi}=\Phi_x} \tag{4.11}$$

for $m = 1$. The proof of these results is given in Ajdler *et al* [1], and the case for $m = 1$ is illustrated in Figure 4.4. The figure shows that the near-field spectrum contains most of the energy inside the triangular region $\Phi_x^2 \leq \left(\frac{\Omega}{c}\right)^2$, but it also contains some energy in the evanescent region, depending on the proximity of the source. For $\Phi_x^2 > \left(\frac{\Omega}{c}\right)^2$, assuming for simplicity that $y_p = 0$, the Hankel function is upper-bounded by $\frac{e^{-z_p\Phi_x}}{\sqrt{z_p\Phi_x}}$, and converges to the upper-bound for $\Phi_x^2 \gg \left(\frac{\Omega}{c}\right)^2$ [88; 2]. Thus, the farther away the source is from the x-axis, the faster is the amplitude decay and the less evanescent waves emerge. On the contrary,



(a) (b)

Figure 4.5: Equivalence between a spherical wave observed on a plane and a radial wave observed on a line. (a) The spherical wave caused by a point source located at $\mathbf{r}_p = (x_p, y_p, z_p)$ satisfies the 3-D wave equation. In this scenario, the sound pressure is observed on the xy-plane. (b) The radial wave caused by a point source located at $\mathbf{r}_p = (x_p, z_p)$ satisfies the 2-D wave equation. By observing the sound pressure on the x-axis, the respective Fourier transform is equivalent to the Fourier transform of (a) for $\Phi_y = 0$. The scenario in (b) is common in applications that make use of the propagation of sound on flat surfaces, such as tables and walls.

as the source moves closer to the x-axis, the balance between propagating and evanescent waves tips towards the evanescent waves, until the two modes become equally weighted (see Figure 2.7-(c)).

Note that the result for $m = 2$ is used to obtain the plot of Figure 2.7-(c), by computing the ratio $\int_{\frac{\Omega}{c}}^{\infty} \left|P(\mathbf{\Phi}, \Omega)\right|^2 d\mathbf{\Phi} / \int_0^{\frac{\Omega}{c}} \left|P(\mathbf{\Phi}, \Omega)\right|^2 d\mathbf{\Phi}$ for a random Ω.

Example 14. Another interesting case to consider is the spectral representation on the x-axis of the radial wave on the water surface, shown in the picture of Figure 2.2. The solution (see proof in Section 4.6) is given by

$$P(\Phi_x, \Omega) = S(\Omega) \frac{1}{2} \frac{e^{-|z_p|\sqrt{\Phi_x^2 - \left(\frac{\Omega}{c}\right)^2}} e^{-j\Phi_x x_p}}{\sqrt{\Phi_x^2 - \left(\frac{\Omega}{c}\right)^2}}. \tag{4.12}$$

Surprisingly, this expression is equivalent to one given by (4.10) for $\Phi_y = 0$. This suggests that the projection of a spherical wave on a flat plane is related to the projection of the respective spatio-temporal Fourier transform on the $\Phi_y = 0$ plane. The two situations are illustrated in Figure 4.5.

4.2.2 Properties

4.2.2.1 Convergence

The convergence of the Fourier transform has been a topic of intense debate since the very first paper that Joseph Fourier published on the topic [33]. The problem concerns whether the Fourier transform of a given class of signals is well defined, and, if so, whether the inverse transform of its spectrum returns the exact same signal. Eventually, mathematicians reached the conclusion that the Fourier transform can converge on many different levels—*e.g.*, uniformly, point-wise, or in norm [18].

Although some types of convergence are stronger than the others, the essential condition is that the Fourier integral is well defined in the region of integration. This typically means that the signal belongs to the ℓ_1 or ℓ_2 spaces, either locally or globally. If the region of integration in space is the volume V and the region of integration in time is the interval T, the condition for existence of the spatio-temporal Fourier transform is given by $p(\mathbf{r}, t) \in \ell_1(V, T)$ or $p(\mathbf{r}, t) \in \ell_2(V, T)$, and can be expressed respectively as

$$P(\mathbf{\Phi}, \Omega) = \int_V \int_T \big|p(\mathbf{r}, t)\big| dt d\mathbf{r} < \infty \tag{4.13}$$

and

$$P(\mathbf{\Phi}, \Omega) = \int_V \int_T \big|p(\mathbf{r}, t)\big|^2 dt d\mathbf{r} < \infty. \tag{4.14}$$

The region of integration can be, for example, a period of the signal if the signal is periodic, or the entire $\mathbb{R}$ space if the signal is non-periodic. This condition, however, is not fail-proof, since in some cases $p(\mathbf{r}, t)$ is not absolutely nor quadratically integrable in the strict sense and still has a Fourier transform—*e.g.*, (4.6) and (4.9). We will not analyze the topic of convergence in this thesis, given that in practice the input signals always have finite energy.

4.2.2.2 Convolution property

The convolution property of the Fourier transform is one of the most important consequences of the spatio-temporal Fourier analysis of acoustic wave fields, and it has fundamental implications on the way spatial audio applications may be addressed in the future. The importance of this property is comparable to the one the 1-D Fourier transform has in DSP applications that make use of filtering, as well as the filter design techniques developed in the context of Fourier theory. The impact of the convolution property will become clearer when we address the topic of spatio-temporal filter design in Chapter 8.

The definition of linear convolution between two functions $p(\mathbf{r}, t)$ and $h(\mathbf{r}, t)$ is given by

$$p(\mathbf{r}, t) \ast\ast h(\mathbf{r}, t) = \int_{\mathbb{R}^3} \int_{\mathbb{R}} p(\boldsymbol{\rho}, \tau) h(\mathbf{r} - \boldsymbol{\rho}, t - \tau) d\tau d\boldsymbol{\rho}, \tag{4.15}$$

where $\ast$ and $\ast$ denote convolution in space and time, respectively. A particular case of (4.15) is when the function $h(\mathbf{r}, t)$ is separable, *i.e.*, $h(\mathbf{r}, t) = h(\mathbf{r})h(t)$, which implies that

$$p(\mathbf{r}, t) \ast\ast h(\mathbf{r}, t) = p(\mathbf{r}, t) \ast h(\mathbf{r}) \ast h(t). \tag{4.16}$$

The convolution property states that the spatio-temporal Fourier transform maps the convolution of two functions into the product of their spatio-temporal Fourier transforms. This can be expressed as

$$p(\mathbf{r},t) * *h(\mathbf{r},t) \overset{\mathcal{F}}{\longleftrightarrow} P(\mathbf{\Phi},\Omega)H(\mathbf{\Phi},\Omega), \tag{4.17}$$

where $\mathcal{F}$ denotes Fourier transform. If the function $h(\mathbf{r},t)$ is separable, the property is simplified to

$$p(\mathbf{r},t) * h(\mathbf{r}) * h(t) \overset{\mathcal{F}}{\longleftrightarrow} P(\mathbf{\Phi},\Omega)H(\mathbf{\Phi})H(\Omega). \tag{4.18}$$

The proof of the continuous convolution property is given in Section 4.6.

4.2.2.3 Multiplication property

The frequency-domain counterpart of the convolution property is called the multiplication property. It states that the spatio-temporal Fourier transform maps the product of two functions into the convolution of their spatio-temporal Fourier transforms, and it can be expressed as

$$p(\mathbf{r},t)h(\mathbf{r},t) \overset{\mathcal{F}}{\longleftrightarrow} P(\mathbf{\Phi},\Omega) * *H(\mathbf{\Phi},\Omega), \tag{4.19}$$

where $*$ and $*$ denote convolution in $\mathbf{\Phi}$ and Ω, respectively. If the function $h(\mathbf{r},t)$ is separable, the property is simplified to

$$p(\mathbf{r},t)h(\mathbf{r})h(t) \overset{\mathcal{F}}{\longleftrightarrow} P(\mathbf{\Phi},\Omega) * H(\mathbf{\Phi}) * H(\Omega). \tag{4.20}$$

The proof of the continuous multiplication property is given in Section 4.6. This property will be useful in the next chapter, when we discuss the use of spatial windows in the Fourier analysis of the acoustic wave field.

4.3 Discrete Fourier transform

4.3.1 Definition

The discrete version of the spatio-temporal Fourier transform can be obtained from the uniformly sampled version of $p(\mathbf{r},t)$, denoted $p[\mathbf{n}]$. The index vector $\mathbf{n}$ is defined as $\mathbf{n} = (n_x, n_y, n_z, n_t)$, where n_x, n_y, and n_z are the integer indexes of the spatial samples and n_t is the integer index of the temporal samples. Denote the transform coefficients $P[\mathbf{b}]$ and the respective index vector $\mathbf{b} = (b_x, b_y, b_z, b_t)$, where b_x, b_y, and b_z are the integer indexes of the spatial transform coefficients and b_t is the integer index of the temporal transform coefficients[2]. The discrete Fourier transform of the 4-D sequence $p[\mathbf{n}]$ can be obtained from the multi-dimensional formulation of the Fourier sum [27], and is given by

$$P[\mathbf{b}] = \sum_{\mathbf{n}\in\mathbb{Z}^4} p[\mathbf{n}]e^{-j2\pi\mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}},\ \mathbf{b}\in\mathbb{Z}^4, \tag{4.21}$$

where $\mathbf{N}$ is a diagonal matrix with the number of samples in each dimension:

[2]The letter *b* stands for frequency "band".

$$\mathbf{N} = \begin{bmatrix} N_x & & & \\ & N_y & & \\ & & N_z & \\ & & & N_t \end{bmatrix}. \tag{4.22}$$

The inverse transform is given by

$$p[\mathbf{n}] = \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{b}\in\mathbb{Z}^4} P[\mathbf{b}] e^{j2\pi \mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}},\ \mathbf{n} \in \mathbb{Z}^4. \tag{4.23}$$

Similarly to the continuous case, the Fourier sum represents the projection of $p[\mathbf{n}]$ onto the orthogonal basis $e^{j2\pi \mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}}$, which is the basic element of the discrete Fourier transform.

4.3.2 Properties

4.3.2.1 Convergence

The essential condition for convergence of the discrete Fourier transform is that the Fourier sum is well defined within the summation bounds. Again, this typically means that the signal belongs to the ℓ_1 or ℓ_2 spaces, and thus

$$P[\mathbf{b}] = \sum_{\mathbf{n}} \big|p[\mathbf{n}]\big| < \infty \tag{4.24}$$

and

$$P[\mathbf{b}] = \sum_{\mathbf{n}} \big|p[\mathbf{n}]\big|^2 < \infty. \tag{4.25}$$

This condition is guaranteed in any practical scenario, where the number of samples and their respective amplitudes are always finite. To obtain an infinite number of samples in $p[\mathbf{n}]$, we would require either an infinite number of microphones or an infinite period of observation.

4.3.2.2 Convolution property

The convolution property in the discrete case is the same as in the continuous case, except that the discrete case is the one we actually use in a practical DSP application. We will often make use of the discrete convolution property in the discussion of spatio-temporal filtering in Chapter 8.

The definition of linear convolution between two sequences $p[\mathbf{n}]$ and $h[\mathbf{n}]$ is given by

$$p[\mathbf{n}] * *h[\mathbf{n}] = \sum_{\boldsymbol{\rho}\in\mathbb{Z}^3} \sum_{\tau\in\mathbb{Z}} p[\boldsymbol{\rho},\tau] h\Big[(n_x,n_y,n_z) - \boldsymbol{\rho}, n_t - \tau\Big], \tag{4.26}$$

where $*$ and $*$ denote convolution in space and time, respectively. The sequence $h[\mathbf{n}]$ can also be separable, in which case $h[\mathbf{n}] = h[n_x,n_y,n_z]h[n_t]$, and thus

$$p[\mathbf{n}] * *h[\mathbf{n}] = p[\mathbf{n}] * h[n_x,n_y,n_z] * h[n_t]. \tag{4.27}$$

The convolution property of the discrete Fourier transform can be expressed as

$$p[\mathbf{n}] * *h[\mathbf{n}] \overset{\mathcal{F}}{\longleftrightarrow} P[\mathbf{b}]H[\mathbf{b}], \tag{4.28}$$

or, if $h[\mathbf{n}]$ is separable,

$$p[\mathbf{n}] * h[n_x, n_y, n_z] * h[n_t] \overset{\mathcal{F}}{\longleftrightarrow} P[\mathbf{b}]H[b_x, b_y, b_z]H[b_t]. \tag{4.29}$$

The proof of the discrete convolution property is given in Section 4.6.

4.3.2.3 Multiplication property

The multiplication property of the discrete Fourier transform can be expressed as

$$p[\mathbf{n}]h[\mathbf{n}] \overset{\mathcal{F}}{\longleftrightarrow} P[\mathbf{b}] * *H[\mathbf{b}], \tag{4.30}$$

where $*$ and $*$ denote convolution in (b_x, b_y, b_z) and b_t, respectively. If the function $h[\mathbf{n}]$ is separable, the property is simplified to

$$p[\mathbf{n}]h[n_x, n_y, n_z]h[n_t] \overset{\mathcal{F}}{\longleftrightarrow} P[\mathbf{b}] * H[b_x, b_y, b_z] * H[b_t]. \tag{4.31}$$

The proof of the discrete multiplication property is given in Section 4.6. This property is most relevant in the context of block-wise processing of sequences, and will be analyzed in more detail in the discussion of spatio-temporal filter banks in Chapter 7.

4.4 Sampling and interpolation

In the previous two sections, the spatio-temporal Fourier transform was presented in two different worlds: the continuous and the discrete. The continuous Fourier transform deals with signals that are a function of continuous variables representing every possible coordinate in space and instant in time, whereas the discrete Fourier transform deals with signals that are a function of discrete variables representing a discrete selection of points in space and time. The two sides of the Fourier transform are connected by a theory known as Nyquist sampling (and interpolation) theory. In this section, we provide an overview of how this theory extends to the spatio-temporal domain.

In traditional signal processing, it is known that continuous-time systems are implemented through the use of analog circuits (*e.g.*, electronic, hydraulic, and pneumatic) that instantaneously modify the input and transmit it to the output. Discrete-time systems use a very different strategy: they first translate the input signal into a discrete set of samples, by obtaining periodical measurements of the signal amplitude—an operation known as *sampling*. The samples are then processed algebraically by a digital computer, and finally translated back into an analog output signal—an operation known as *interpolation*.

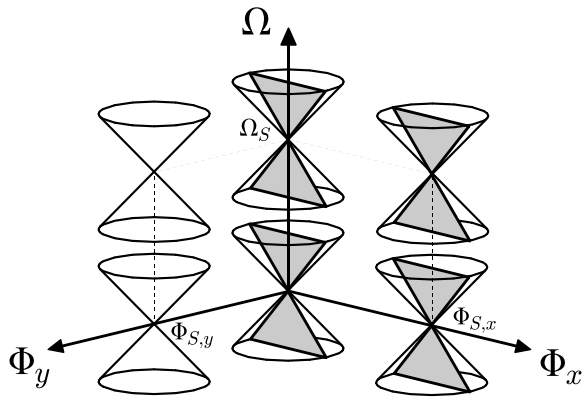
To the present day, there is no available technology for processing the acoustic wave field instantaneously in space and time in a systematic and separable way. An example of a non-separable spatio-temporal filter would be to place a wall in the path of the wave front, which could act both as a low-pass filter and as a curvature modifier, due to the building materials and density distribution across the wall. If the objective, however, is to amplify the sound pressure without modifying the wave front, or to flip the direction of propagation without changing the temporal behavior of the wave front, this is not possible with analog technology.

Therefore, in order to effectively apply DSP theory—and Fourier theory in particular—to acoustic wave fields, the sound pressure must be sampled in space.

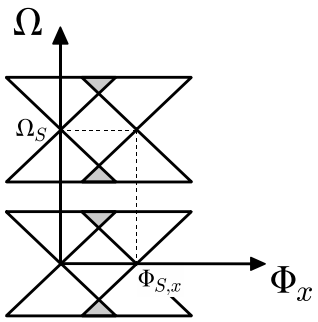
In order to process the wave field in discrete space, we have to address the problem of aliasing. When the sampling is uniform, spatial aliasing occurs when the distance between adjacent samples is larger than the Nyquist threshold, given by $\frac{\pi c}{\Omega_M}$ meters, where Ω_M is the maximum temporal frequency in rad/s. To understand how this value is obtained, consider the Fourier transform $P(\mathbf{\Phi}, \Omega)$ of the pressure signal $p(\mathbf{r}, t)$. According to Nyquist sampling theory [68], the sampling operation generates repetitions of $P(\mathbf{\Phi}, \Omega)$ centered in multiples of 2π, resulting in $\sum_{\boldsymbol{l}\in\mathbb{Z}^3} \sum_{l\in\mathbb{Z}} P(\mathbf{\Phi} - 2\pi\boldsymbol{l}, \Omega - 2\pi l)$. This means that the aliasing effects can occur in either of the four dimensions, if the respective Nyquist conditions are not satisfied. Conversely, if the four conditions are met, the signal can be perfectly reconstructed in all four dimensions. To determine these conditions, consider the case where the sources are in the far-field. We know from (4.3) that the spectral energy is contained at (or delimited by) the region $||\mathbf{\Phi}||^2 = \left(\frac{\Omega}{c}\right)^2$. Assuming the source signals are band-limited, the energy is either zero or negligible for $\left|\Omega\right| \geq \Omega_M$. Therefore, the Nyquist conditions for alias-free sampling are given by $\mathbf{\Phi}_S \geq 2\frac{\Omega_M}{c}$ and $\Omega_S \geq 2\Omega_M$, where $\mathbf{\Phi}_S$ and Ω_S are the spatial and temporal sampling frequencies. Examples of the different types of aliasing are illustrated in Figure 4.6.

In the near-field case, there is no condition that completely eliminates spatial aliasing, since $P(\mathbf{\Phi}, \Omega)$ is never band-limited across $\mathbf{\Phi}$ (see Figure 4.4). In the lower-dimensional cases, however, $P(\Phi_x, \Phi_y, \Omega)$ and $P(\Phi_x, \Omega)$ are approximately band-limited if the sources are not too close to the xy-plane and x-axis, respectively. In such cases, the Nyquist condition guarantees the reconstruction of the wave field with arbitrarily low spatial aliasing.

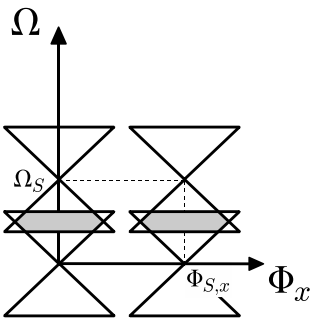
The Nyquist conditions demonstrate the remarkable fact that the acoustic wave field is essentially band-limited, and can be reconstructed from a few samples with minimum or no spatial aliasing. The reader may refer to the work of Ajdler *et al.* [2] on the plenacoustic function for a more detailed analysis on spatio-temporal sampling.



(a)



(b)



(c)

Figure 4.6: Effects of sampling the wave field in space and time. (a) When $p(\mathbf{r}, t)$ is sampled along x, y, and t with sampling frequencies $\Phi_{S,x}$, $\Phi_{S,y}$, and Ω_S respectively, the base spectrum is repeated at every combination of multiples of $\Phi_{S,x}$, $\Phi_{S,y}$, and Ω_S. For one spatial dimension, illustrated in the shaded region, the base spectrum is repeated at every combination of multiples of $\Phi_{S,x}$ and Ω_S. (b) If $\Phi_{S,x}$ is lower than the Nyquist threshold, the spectral images overlap and spatial aliasing occurs. (c) If Ω_S is lower than the Nyquist threshold, then temporal aliasing occurs.

4.5 Summary notes

- Microphone arrays with a large number of microphones reveal the harmonic behavior of the acoustic wave field across space. This behavior is efficiently represented by the Fourier transform;
- The acoustic wave field is composed of elementary wave fronts called plane waves, which are sparsely represented in the spatio-temporal Fourier domain in the form of Dirac points;
- A wide-band source in the far-field is represented in the spatio-temporal Fourier domain as a Dirac function defined on a line crossing the origin;
- If the source is in the near-field, the Fourier transform has a characteristic triangular shape, covering the entire propagating region of the spectrum. A percentage of the energy leaks into the evanescent region, but in most cases is negligible;
- Spherical wave fronts in the 3-D space and radial wave fronts in the 2-D space are related by the spatio-temporal Fourier transform;
- The convolution and multiplication properties of the regular Fourier transform can be extended to the spatio-temporal Fourier transform;
- Sampling the acoustic wave field in space generates an infinite number of spectral images, which overlap if the Nyquist sampling conditions are not satisfied;
- If all sources are band-limited in time and located in the far-field, then the acoustic wave field is band-limited in both space and time. If one or more sources are located in the near-field, then the acoustic wave field is approximately band-limited, and can be reconstructed from its samples with arbitrarily low spatial aliasing.

4.6 Theorems and proofs

Theorem 15 (Near-field spectrum on a line of 2-D radial wave)**.** *Consider a point source located on the xz-plane at $\mathbf{r}_p = (x_p, 0, z_p)$, driven with the source signal $s(t)$, and generating a radial wave front satisfying the 2-D wave equation. Consider also a point source located in space at $\mathbf{r}_p = (x_p, y_p, z_p)$, driven with the source signal $s(t)$, and generating a spherical wave front satisfying the 3-D wave equation. The Fourier transform of $p(\mathbf{r}, t)$ observed on the x-axis in the first case is equal to the Fourier transform of $p(\mathbf{r}, t)$ observed on the xy-plane in the second case for $\Phi_y = 0$, and it is given by*

$$P(\Phi_x, \Omega) = S(\Omega)\frac{1}{2}\Big(\Phi_x^2 - \big(\frac{\Omega}{c}\big)^2\Big)^{-\frac{1}{2}} e^{-\left|z_p\right|\sqrt{\Phi_x^2 - \left(\frac{\Omega}{c}\right)^2}}\, e^{-j\Phi_x x_p}.$$

Proof. From (2.23), the solution to the 2-D wave equation in the first case is given by

$$p(x, \Omega) = S(\Omega)\frac{j}{4}H_0^{(1)}\left(\frac{\Omega}{c}\sqrt{(x - x_p)^2 + z_p^2}\right).$$

The Fourier transform of $p(x, \Omega)$ with respect to x yields

$$
\begin{aligned}
P(\Phi,\Omega) &= \int_{\mathbb{R}} p(x,\Omega)e^{-j\Phi x}dx \\
&= S(\Omega)\frac{j}{4}\int_{\mathbb{R}} H_0^{(1)}\left(\frac{\Omega}{c}\sqrt{(x-x_p)^2+z_p^2}\right)e^{-j\Phi x}dx \\
&\overset{(a)}{=} S(\Omega)\frac{j}{4}e^{-j\Phi x_p}\int_{\mathbb{R}} H_0^{(1)}\left(\frac{\Omega}{c}\sqrt{l^2+z_p^2}\right)e^{-j\Phi l}dl \\
&= S(\Omega)\frac{j}{2}e^{-j\Phi x_p}\int_0^{\infty} H_0^{(1)}\left(\frac{\Omega}{c}\sqrt{l^2+z_p^2}\right)\cos(\Phi l)dl \\
&\overset{(b)}{=} S(\Omega)\frac{j}{2}\frac{e^{jz_p\sqrt{\left(\frac{\Omega}{c}\right)^2-\Phi^2}}e^{-j\Phi x_p}}{\sqrt{\left(\frac{\Omega}{c}\right)^2-\Phi^2}} \\
&= S(\Omega)\frac{1}{2}\frac{e^{-\left|z_p\right|\sqrt{\Phi^2-\left(\frac{\Omega}{c}\right)^2}}e^{-j\Phi x_p}}{\sqrt{\Phi^2-\left(\frac{\Omega}{c}\right)^2}},
\end{aligned}
$$

where (a) comes from the change of variable $l = x - x_p$ and (b) comes from a known integration result [38],

$$
\int_0^{\infty} H_0^{(1)}\left(\frac{\Omega}{c}\sqrt{l^2+z_p^2}\right)\cos(\Phi l)dl = \frac{e^{jz_p\sqrt{\left(\frac{\Omega}{c}\right)^2-\Phi^2}}}{\sqrt{\left(\frac{\Omega}{c}\right)^2-\Phi^2}}.
$$

The result is therefore equal to (4.10) for $\Phi_y = 0$. □

Theorem 16 (Convolution property)**.** *Given two spatio-temporal functions $p(\mathbf{r},t)$ and $h(\mathbf{r},t)$ with Fourier transforms $P(\mathbf{\Phi},\Omega)$ and $H(\mathbf{\Phi},\Omega)$ respectively, the Fourier transform of their convolution $p(\mathbf{r},t)**h(\mathbf{r},t)$ equals the product of the individual Fourier transforms, $P(\mathbf{\Phi},\Omega)H(\mathbf{\Phi},\Omega)$. The same is valid for the discrete Fourier transform case.*

Proof. Plugging (4.15) into (4.1), yields

$$
\begin{aligned}
& \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\mathbf{r},t) * *h(\mathbf{r},t)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r} \\
= & \int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)h(\boldsymbol{\rho}-\mathbf{r},\tau-t)d\tau d\boldsymbol{\rho}e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r} \\
= & \int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)h(\boldsymbol{\rho}-\mathbf{r},\tau-t)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}d\tau d\boldsymbol{\rho}dtd\mathbf{r} \\
= & \int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}h(\boldsymbol{\rho}-\mathbf{r},\tau-t)e^{-j(\mathbf{\Phi}\cdot(\boldsymbol{\rho}-\mathbf{r})+\Omega(\tau-t))}d\tau d\boldsymbol{\rho}dtd\mathbf{r} \\
\overset{(a)}{=} & \int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}h(\boldsymbol{\rho}-\mathbf{r},\tau-t)e^{-j(\mathbf{\Phi}\cdot(\boldsymbol{\rho}-\mathbf{r})+\Omega(\tau-t))}dtd\mathbf{r}d\tau d\boldsymbol{\rho} \\
\overset{(b)}{=} & \int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}h(\boldsymbol{\sigma},\varrho)e^{-j(\mathbf{\Phi}\cdot\boldsymbol{\sigma}+\Omega\varrho)}dtd\mathbf{r}d\varrho d\boldsymbol{\sigma} \\
\overset{(c)}{=} & \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r}\int_{\mathbb{R}^3}\int_{\mathbb{R}} h(\boldsymbol{\sigma},\varrho)e^{-j(\mathbf{\Phi}\cdot\boldsymbol{\sigma}+\Omega\varrho)}d\varrho d\boldsymbol{\sigma} \\
= & P(\mathbf{\Phi},\Omega)H(\mathbf{\Phi},\Omega),
\end{aligned}
$$

where (a) is valid as long as the Fubini theorem [16] is satisfied, *i.e.*,

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} \left|p(\boldsymbol{\rho},\tau)h(\boldsymbol{\rho}-\mathbf{r},\tau-t)\right|d\tau d\boldsymbol{\rho}dtd\mathbf{r} < \infty$$

and (b) comes from the change of variables $\boldsymbol{\sigma} = \boldsymbol{\rho} - \mathbf{r}$ and $\varrho = \tau - t$. Note that if $p(\mathbf{r},t)$ or $h(\mathbf{r},t)$ are Diracs functions, the Fubini theorem is still valid in the context of measure theory, since the Dirac functions can be translated into Dirac measures [16]. The same is true with the sums in the discrete Fourier transform case, since, according to measure theory, the sum can be translated into an integration with respect to the counting measure. □

Corollary 17 (Separability)**.** *If $h(\mathbf{r},t)$ is a separable product of a spatial function $h(\mathbf{r})$ and a temporal function $h(t)$ with Fourier transforms $H(\mathbf{\Phi})$ and $H(\Omega)$ respectively, such that $h(\mathbf{r},t) = h(\mathbf{r})h(t)$, the Fourier transform of the convolution $p(\mathbf{r},t) * h(\mathbf{r}) * h(t)$ equals the product of the individual Fourier transforms, $P(\mathbf{\Phi},\Omega)H(\mathbf{\Phi})H(\Omega)$. The same is valid for the discrete Fourier transform case.*

Proof. Directly from the equality (c) in Theorem 16, it follows that

$$
\begin{aligned}
& \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r}\int_{\mathbb{R}^3}\int_{\mathbb{R}} h(\boldsymbol{\sigma},\varrho)e^{-j(\mathbf{\Phi}\cdot\boldsymbol{\sigma}+\Omega\varrho)}d\varrho d\boldsymbol{\sigma} \\
= & \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r}\int_{\mathbb{R}^3} h(\boldsymbol{\sigma})e^{-j\mathbf{\Phi}\cdot\boldsymbol{\sigma}}d\boldsymbol{\sigma}\int_{\mathbb{R}} h(\varrho)e^{-j\Omega\varrho}d\varrho \\
= & P(\mathbf{\Phi},\Omega)H(\mathbf{\Phi})H(\Omega).
\end{aligned}
$$

□

Theorem 18 (Multiplication property)**.** *Given two spatio-temporal functions $p(\mathbf{r},t)$ and $h(\mathbf{r},t)$ with Fourier transforms $P(\mathbf{\Phi},\Omega)$ and $H(\mathbf{\Phi},\Omega)$ respectively, the Fourier transform of their product $p(\mathbf{r},t)h(\mathbf{r},t)$ equals the convolution of the individual Fourier transforms, $P(\mathbf{\Phi},\Omega) * *H(\mathbf{\Phi},\Omega)$. The same is valid for the discrete Fourier transform case.*

Proof. The proof is the same as in Theorem 16, requiring that, is accordance with the Fubini theorem [16],

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\left|P(\boldsymbol{\rho},\tau)H(\boldsymbol{\rho}-\boldsymbol{\Phi},\tau-\Omega)\right|d\tau d\boldsymbol{\rho}d\Omega d\boldsymbol{\Phi}<\infty$$

Again, if $P(\boldsymbol{\Phi},\Omega)$ or $H(\boldsymbol{\Phi},\Omega)$ are Diracs functions, the Fubini theorem is still valid in the context of measure theory. Also due to the representation of sums as integrals allowed by measure theory, the theorem is valid in the discrete Fourier transform case. □

Corollary 19 (Separability)**.** *If $h(\mathbf{r},t)$ is a separable product of a spatial function $h(\mathbf{r})$ and a temporal function $h(t)$ with Fourier transforms $H(\boldsymbol{\Phi})$ and $H(\Omega)$ respectively, such that $h(\mathbf{r},t)=h(\mathbf{r})h(t)$, the Fourier transform of the product $p(\mathbf{r},t)h(\mathbf{r})h(t)$ equals the convolution of the individual Fourier transforms, $P(\boldsymbol{\Phi},\Omega)*H(\boldsymbol{\Phi})*H(\Omega)$. The same is valid for the discrete Fourier transform case.*

Proof. The proof is the same as in Corollary 17. □

Chapter 5

Space/Time–Frequency Representation

5.1 Introduction

The modern definition of the Fourier transform consists of a projection of the signal $s(t)$ onto the orthogonal basis $e^{j\Omega t}$, with the direct and inverse transforms given by $S(\Omega) = \int_{\mathbb{R}} s(t)e^{-j\Omega t}dt$ and $s(t) = \int_{\mathbb{R}} S(\Omega)e^{j\Omega t}d\Omega$, respectively. This elegant formalism was developed on top of Fourier's original ideas during the 19th century, and perfected through the combined work of various mathematicians. The essential paradigm, however, as it was conceptualized by Fourier, remained unchanged. This is hardly surprising, since many of these mathematicians lived in a period when analytical rigour was of utmost importance to the community, and hence the main focus was on perfecting the mathematical formalism of the Fourier transform and providing a deeper understanding of the mathematical principles of the theory [52]. It was only later, in 1946, with the work of the Hungarian engineer Dennis Gabor (1900–1979), that a major theoretical breakthrough was observed. While trying to solve a concrete problem—how to optimally utilize the frequency bands in the transmission of audio signals—Gabor realized that the Fourier transform was not suited for representing the frequency variations that characterize non-stationary signals such as speech and music. He pointed out [35] that our everyday perception of sound is conditioned by the idea of "changing frequencies" (think, for example, of the sound of a police siren), which contradicts the strict definition of sinusoid as an infinitely long function with a fixed frequency. To circumvent this limitation, Gabor redefined the basic element of the Fourier transform as a time-limited harmonic function given by $w(t-t_0)e^{j\Omega t}$, where $w(t-t_0)$ is a window function with unit integral and decaying amplitude away from $t = t_0$. With Gabor's generalized basis, the Fourier transform becomes

$$S(t_0, \Omega) = \int_{\mathbb{R}} s(t)w(t - t_0)e^{-j\Omega t}dt$$

where $S(t_0, \Omega)$ is defined on what is commonly known as the *time–frequency representation space*. The regular Fourier transform $S(\Omega)$ is obtained as a particular case of $S(t_0, \Omega)$, when $w(t - t_0) = 1$. This new concept of time–frequency representation led to the development

of filter banks theory, and the widely used signal analysis techniques such as the short-time Fourier transform, the wavelet transform [42], and the modulated lapped transform [77], along with the many practical applications in the areas of audio, speech, and image processing.

In the spatio-temporal analysis of acoustic wave fields, the 4-D Fourier transform has a similar limitation to the one described by Gabor: it is unable to represent the frequency variations across space that characterize non-stationary wave fields. The wave field is non-stationary when the 4-D frequency content varies between different regions in space. This poses the problem of "changing plane waves", which contradicts the strict definition of plane wave as an infinitely long function in both space and time. In the same line of thought, we redefine the basic element of the Fourier transform as a space- and time-limited harmonic function given by $w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)w_t(t-t_0)e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}$, where $w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)$ and $w_t(t-t_0)$ are window functions with decaying amplitude away from $\mathbf{r}=\mathbf{r}_0$ and $t=t_0$, respectively. This generalizes the Fourier transform to

$$P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)=\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\mathbf{r},t)w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)w_t(t-t_0)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r}$$

where $P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)$ belongs to what we call the *space/time–frequency representation space*—an 8-dimensional space that spans all the variables related to space, time, and frequency. Similarly to what happens in Gabor's generalized definition, the regular 4-D Fourier transform $P(\mathbf{\Phi},\Omega)$ is obtained as a particular case of $P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)$, when $w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)w_t(t-t_0)=1$.

The space/time–frequency representation has fundamental implications in the Fourier analysis of the acoustic wave field. The limited length of the analysis window introduces uncertainty in the frequency domain, by spreading the support function across $\mathbf{\Phi}$ and Ω. In a practical scenario, this means that two sources that are close to each other in space may not be resolved in the frequency domain, due to the overlapping of their respective support functions. This uncertainty reflects the trade-off between spatial resolution and spatial-frequency resolution.

In this chapter, we analyze how the various parameters of the window functions and the acoustic scene affect the spectral patterns generated as a result. We will consider the same cases as in the examples of the previous chapter, such as point sources in the far-field and the near-field, driven by complex exponentials and Dirac pulses, and show how each spectral pattern transforms into the others by gradually changing the parameters. This analysis leads to the formulation of what we call the *short space/time Fourier transform*—a generalization of the short time Fourier transform that we discuss in detail in this chapter. An important foundation is also built for the discussion of spatio-temporal filter banks in Chapter 7.

5.2 Short space/time Fourier transform

5.2.1 Continuous case

The short space/time Fourier transform of the 4-D continuous signal $p(\mathbf{r},t)$ is given by an 8-D function $P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)$ defined as

$$P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)=\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\mathbf{r},t)\underbrace{w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)w_t(t-t_0)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}}_{\text{Generalized element}}dtd\mathbf{r}, \tag{5.1}$$

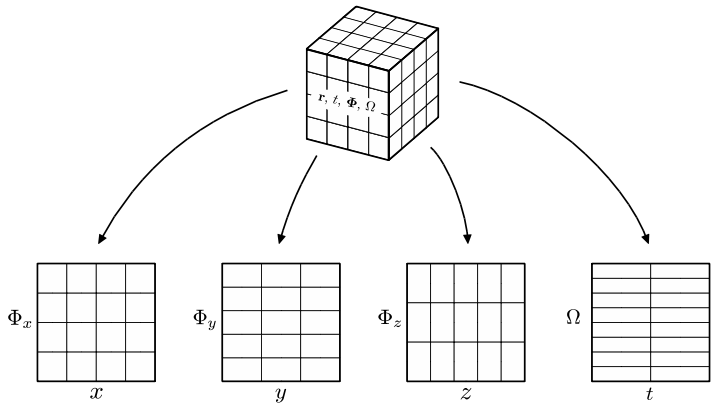


Figure 5.1: Space/time–frequency tiling that characterizes the short space/time Fourier transform, illustrated here as four separate 2-D tilings. The three spatial dimensions are mapped onto the three spatial frequency dimensions, and the temporal dimension is mapped onto the temporal frequency dimension. The resolution in each dimension can be arbitrarily chosen by varying the parameters of the four window functions.

where $w_{\mathbf{r}}(\mathbf{r})$ and $w_t(t)$ are the spatial and temporal window functions, and $\mathbf{r}_0$ and t_0 are the respective spatial and temporal shifts. The spatial window is a product between the window functions in x, y, and z, such that $w_{\mathbf{r}}(\mathbf{r}) = w_x(x)w_y(y)w_z(z)$.

The inverse transform is given by

$$p(\mathbf{r},t) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\mathbf{\Phi}dt_0d\mathbf{r}_0. \tag{5.2}$$

The inverse formula can be interpreted as the continuous sum of all the "signal blocks" indexed by $(\mathbf{r}_0, t_0)$, and it requires that $w_{\mathbf{r}}(\mathbf{r})$ and $w_t(t)$ have unit area. This guarantees that the distortion caused by the window functions is canceled by the inverse transform. A proof of the inversion formula is given in Section 5.5.

The definition in (5.1) is characterized by a partitioning of the space/time–frequency domain into uniform hypercubic regions spanning the variables $\mathbf{r}$, t, $\mathbf{\Phi}$, and Ω, which we call *space/time–frequency tiling*. A pseudo-illustration of the 8-D tiling is shown in Figure 5.1.

We will show several examples that help to understand the spectral patterns generated by (5.1) given the most common inputs, and illustrate the results in the 2-D space (one spatial dimension plus time). We cover the three cases of sources in the far-field, the near-field, and the intermediate-field.

Example 20 (Far-field spectrum)**.** Consider a point source located in the far-field with source signal $s(t)$ and propagation vector $\mathbf{u}$, such that $p(\mathbf{r},t) = s\big(t + \frac{\mathbf{u}\cdot\mathbf{r}}{c}\big)$. Plugging $p(\mathbf{r},t)$ into (5.1), yields

$$\boxed{P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega) = S(\Omega)W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) * W_t(\Omega)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}_0+\Omega t_0)}} \tag{5.3}$$

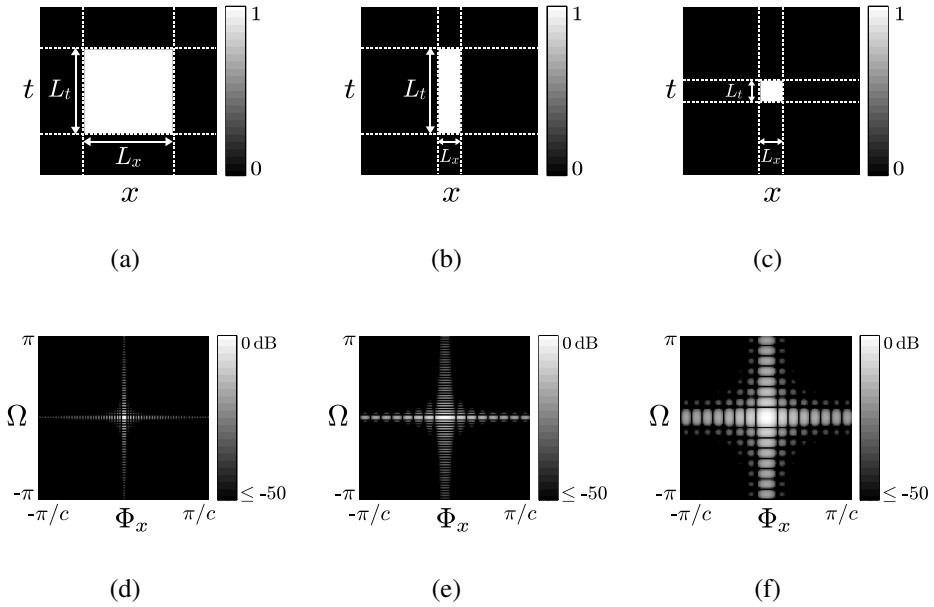


Figure 5.2: Space/time–frequency resolution trade-off for $m = 1$ (one spatial dimension) caused by a rectangular window. The spatio-temporal window in (a) has half of its size equal to one and the other half equal to zero, which is wide enough to generate a narrow sinc support function, as shown in (d). This corresponds to a low resolution in space and time, and a high resolution in frequency. In the case of (b), the window is narrow in space and wide in time, generating a sinc support function that is wide in Φ_x and narrow in Ω, as shown in (e). In the case of (c), the window is narrow in both space and time, generating a sinc support function that is wide in both Φ_x and Ω, as shown in (f). This corresponds to a high resolution in space and time, and a low resolution in frequency.

where $\mathbf{r}_0, \mathbf{\Phi}$,$\mathbf{u} \in \mathbb{R}^m$ for $m = 0, 1, 2, 3$. The functions $W_{\mathbf{r}}(\mathbf{\Phi})$ and $W_t(\Omega)$ are the Fourier transforms of $w_{\mathbf{r}}(\mathbf{r})$ and $w_t(t)$, and $W_{\mathbf{r}}\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big)$ is defined such that $W_{\mathbf{r}}\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big) = W_x\big(\Phi_x - u_x\frac{\Omega}{c}\big)W_y\big(\Phi_y - u_y\frac{\Omega}{c}\big)W_z\big(\Phi_z - u_z\frac{\Omega}{c}\big)$ for $m = 3$, where $W_x(\Phi_x)$, $W_y(\Phi_y)$, and $W_z(\Phi_z)$ are the Fourier transforms of $w_x(x)$, $w_y(y)$, and $w_z(z)$. The proof of this result is given in Section 5.5.

Comparing the result in (6.3) to the regular Fourier transform obtained in the previous chapter, given by $P(\mathbf{\Phi}, \Omega) = S(\Omega)(2\pi)^m\delta\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big)$, the effects of windowing can be interpreted as the Dirac function "opening up" into a smooth support function with the same orientation, $\frac{\partial\mathbf{\Phi}}{\partial\Omega} = \frac{\mathbf{u}}{c}$, and an additional smoothing effect caused by the temporal window. This support function has a variable resolution, governed by the parameters of the window functions, such as length and shape. Any variation of these parameters aimed at increasing the resolution in space and time necessarily decreases the resolution in frequency, and *vice versa*. This is a fundamental trade-off in the local Fourier analysis of signals. An illustration of the resolution trade-off is shown in Figure 5.2.

Example 21. Consider the case of Example 20 for $s(t) = e^{j\Omega_0 t}$, where Ω_0 is a fixed frequency. The Fourier transform of the complex exponential is given by $S(\Omega) = 2\pi\delta(\Omega - \Omega_0)$, and thus

$$P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) = 2\pi W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega_0}{c}\Big) W_t(\Omega - \Omega_0) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\Omega_0)t_0\right)}. \tag{5.4}$$

The result is proved in Section 5.5 and illustrated in Figure 5.3-(a). Comparing to the regular Fourier transform obtained in the previous chapter, given by $P(\mathbf{\Phi}, \Omega) = (2\pi)^{m+1}\delta\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega_0}{c}\big)\delta(\Omega - \Omega_0)$, we observe that the Dirac point centered at $(\mathbf{\Phi}, \Omega) = \big(\mathbf{u}\frac{\Omega_0}{c}, \Omega_0\big)$ "opens up" into a smooth support function centered at the same point. This implies that, by limiting the size of the analysis window, the plane waves become less concentrated in small regions of the spectrum, affecting the overall sparsity of the spatio-temporal Fourier transform.

Example 22. Consider a variation of the previous example, where $s(t) = 2\cos(\Omega_0 t)$. Since $2\cos(\Omega_0 t) = e^{-j\Omega_0 t} + e^{j\Omega_0 t}$, the Fourier transform is given by $S(\Omega) = 2\pi\delta(\Omega + \Omega_0) + 2\pi\delta(\Omega - \Omega_0)$, and thus

$$\begin{aligned} P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) = {} & 2\pi W_{\mathbf{r}}\Big(\mathbf{\Phi} + \mathbf{u}\frac{\Omega_0}{c}\Big) W_t(\Omega + \Omega_0) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega+\Omega_0)t_0\right)} + \\ & 2\pi W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega_0}{c}\Big) W_t(\Omega - \Omega_0) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\Omega_0)t_0\right)}. \end{aligned} \tag{5.5}$$

This result is illustrated in Figure 5.3-(b). Comparing to the previous example, the result is the same except for an additional support function centered at the symmetric position $(\mathbf{\Phi}, \Omega) = \big(-\mathbf{u}\frac{\Omega_0}{c}, -\Omega_0\big)$.

Example 23. If the source signal is a Dirac pulse, $s(t) = \delta(t)$, the Fourier transform is maximally flat and given by $S(\Omega) = 1$. The short space/time Fourier transform is then given by

$$P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) = W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) * W_t(\Omega) e^{-j(\mathbf{\Phi}\cdot\mathbf{r}_0 + \Omega t_0)}. \tag{5.6}$$

The result can be simplified if $W_t(\Omega)$ is a nascent Dirac function, *i.e.*, if $\lim_{a\to\infty} aW_t\,(a\Omega) = \delta\,(\Omega)$, in which case

$$P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) = W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) e^{-j\mathbf{\Phi}\cdot\mathbf{r}_0}. \tag{5.7}$$

Most window functions used in practice are nascent Dirac functions. Thus, for wide-band source signals and a long enough temporal window, the effects caused by $w_{\mathbf{r}}(\mathbf{r})$ tend to be dominant over the effects caused by $w_t(t)$.

The result in (5.7) is proved in Section 5.5 and illustrated in Figure 5.3-(c).

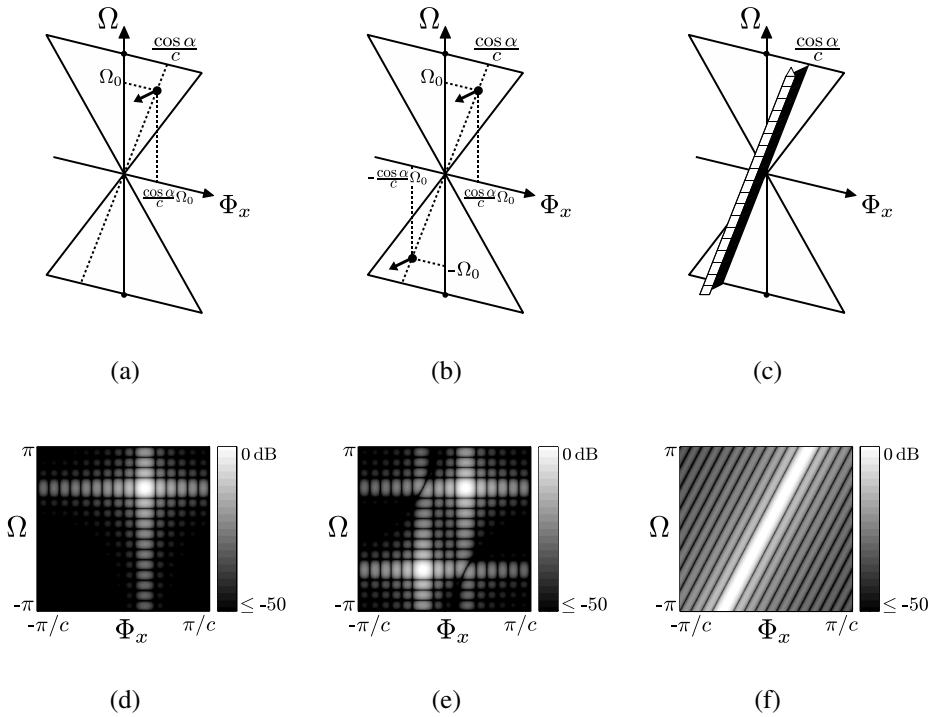


Figure 5.3: Effects of spatio-temporal windowing on the spectrum of a far-field source for $m = 1$ (one spatial dimension) and different source signals. For $s(t) = e^{j\Omega_0 t}$, the ideal (non-windowed) spectrum shown in (a) is a Dirac point located at $(\Phi_x, \Omega) = \big(\cos\alpha\frac{\Omega_0}{c}, \Omega_0\big)$. By applying a rectangular window function, the Dirac point "opens up" into a product of sinc functions with their main lobes centered at the same point, $(\Phi_x, \Omega) = \big(\cos\alpha\frac{\Omega_0}{c}, \Omega_0\big)$, as shown in (d). The extent to which the sinc functions spread the across Φ_x and Ω depends on the size of the spatio-temporal window. For $s(t) = 2\cos(\Omega_0 t)$, the ideal spectrum shown in (b) is a pair of symmetric Dirac points located at $(\Phi_x, \Omega) = \pm\big(\cos\alpha\frac{\Omega_0}{c}, \Omega_0\big)$. The rectangular window generates two support functions accordingly, as shown in (e). For $s(t) = \delta(t)$, the ideal spectrum shown in (c) is a Dirac function covering the entire line of slope $\frac{\partial\Phi_x}{\partial\Omega} = \frac{\cos\alpha}{c}$. Similarly, the rectangular window "opens up" the Dirac into a sinc function with its main lobe defined along the same line of slope $\frac{\partial\Phi_x}{\partial\Omega} = \frac{\cos\alpha}{c}$, as shown in (f). In this case, the sinc function is regulated by the spatial window alone, which suppresses the effect of the temporal window.

Example 24 (Intermediate-field spectrum)**.** Consider a point source located at $\mathbf{r} = \mathbf{r}_p$ with source signal $s(t)$, such that $p(\mathbf{r}, t) = \frac{1}{4\pi\|\mathbf{r}-\mathbf{r}_p\|} s\Big(t - \frac{\|\mathbf{r}-\mathbf{r}_p\|}{c}\Big)$. In this case, the spectral pattern generated as a result of windowing is difficult to express mathematically, since it is given by a convolution between one of the results in Example 13 and the Fourier transform of a shifted window function. Therefore, for simplicity, we will consider only the case for $m = 1$ (one spatial dimension) and take $\mathbf{r}_0 = 0$ and $t_0 = 0$. Under these conditions, the result can be approximated by the parametric expression

$$\boxed{P(\Phi_x, \Omega) = S(\Omega) \max\left\{W_x\left(\Phi_x - \cos\alpha\frac{\Omega}{c}\right), M(\Phi_x, \Omega)\right\},} \tag{5.8}$$

where $M(\Phi_x, \Omega)$ is a triangular mask given by

$$\boxed{M(\Phi_x, \Omega) = \begin{cases} W_x(0) & , (\Phi_x, \Omega) \notin \mathcal{U} \\ 0 & , (\Phi_x, \Omega) \in \mathcal{U} \end{cases},} \tag{5.9}$$

with $\mathcal{U} = \mathbb{R}^2 \backslash \big\{(\Phi_x, \Omega) : \Phi_x^{\min} \leq \Phi_x \leq \Phi_x^{\max},\ \Omega \geq 0\big\}$, and point-symmetric for $\Omega < 0$. The parameters α, $\Phi_x^{\min}$, and $\Phi_x^{\max}$ can be optimized for any given source.

Consider a spatial window of size L, and define $\alpha_{\text{nf}}(x) = \angle\big(x_p - x + jz_p\big)$ as the angle of incidence at point x, where $\alpha_{\text{nf}}^{\min} = \alpha_{\text{nf}}(0)$ is the smallest angle and $\alpha_{\text{nf}}^{\max} = \alpha_{\text{nf}}(L)$ is the largest angle, such that $\alpha_{\text{nf}}^{\min} \leq \alpha_{\text{nf}}(x) \leq \alpha_{\text{nf}}^{\max}$, in accordance to Figure 5.4-(a). The parameters are then given by $\cos\alpha = \mathbb{E}_x\left[\cos\alpha_{\text{nf}}(x)\right]$, where $\mathbb{E}_x$ denotes expectation over x, $\Phi_x^{\min} = \cos\alpha_{\text{nf}}^{\max}\frac{\Omega}{c}$, and $\Phi_x^{\max} = \cos\alpha_{\text{nf}}^{\min}\frac{\Omega}{c}$, in accordance to Figure 5.4-(b). The result is proved in Section 5.5.

An illustration of the windowed Fourier transform resulting from a near-field source and the respective approximation by (5.8) are shown in Figure 5.4-(c) and (d). Note that the triangular mask $M(\Phi_x, \Omega)$ opens and closes according to the distance of the source to the x-axis. If the distance is very small, the model approximates the Fourier transform of a near-field source, and the parameters converge to $\Phi_x^{\min} = -\frac{\Omega}{c}$ and $\Phi_x^{\max} = \frac{\Omega}{c}$. On the contrary, if the distance is very large, the model approximates the Fourier transform of a far-field source, and the parameters converge to $\alpha_{\text{nf}}(x) = \alpha$ and $\Phi_x^{\min} = \Phi_x^{\max} = \cos\alpha\frac{\Omega}{c}$. For any other scenario, we say that the source is in the intermediate-field. The convergence of (5.8) to the near-field and far-field cases is illustrated in Figure 5.5.

Example 25 (Co-linear sources)**.** In this example, we demonstrate the potential of the spatio-temporal Fourier representation to resolve multiple sources in the acoustic scene. For this purpose, we consider a worst case scenario that consists of having two co-linear sources, $s_1(t)$ and $s_2(t)$, with respect to the observation region, such that

$$\begin{aligned} P(\Phi_x, \Omega) = S_1(\Omega) \max\left\{W_x\left(\Phi_x - \cos\alpha_1\frac{\Omega}{c}\right), M_1(\Phi_x, \Omega)\right\} + \\ aS_2(\Omega) \max\left\{W_x\left(\Phi_x - \cos\alpha_2\frac{\Omega}{c}\right), M_2(\Phi_x, \Omega)\right\}, \end{aligned} \tag{5.10}$$

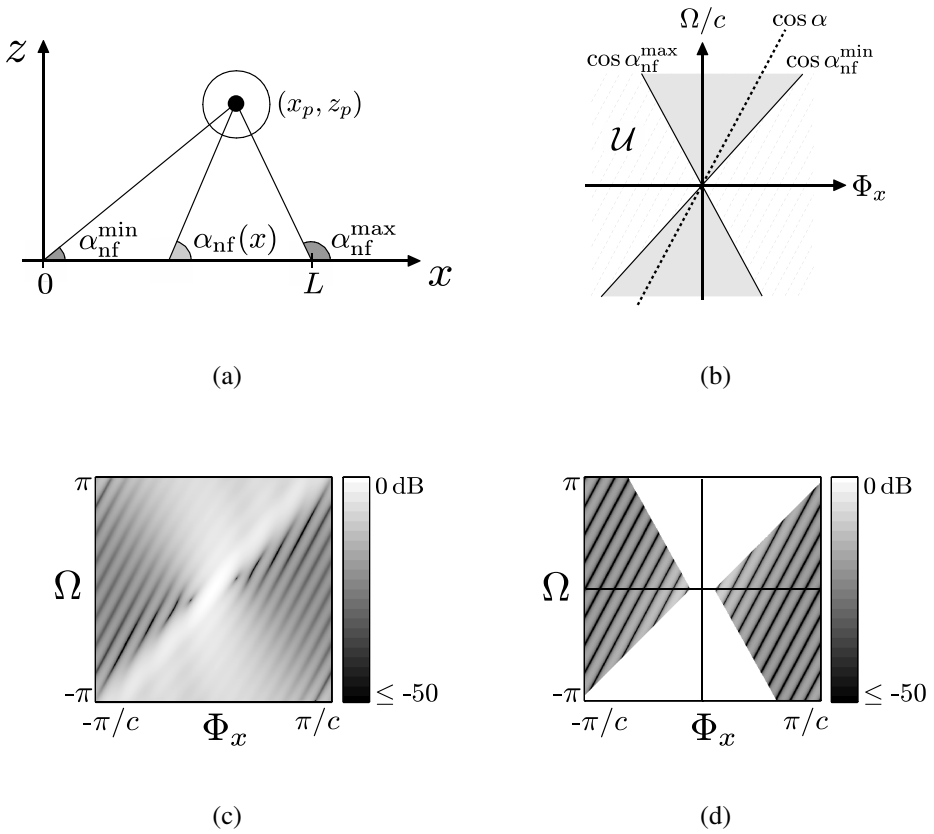


Figure 5.4: Effects of spatio-temporal windowing on the spectrum of an intermediate-field source for $m = 1$ (one spatial dimension) and $S(\Omega) = 1$. (a) The source is located at $\mathbf{r}_p = (x_p, z_p)$, and the resulting wave front arrives to the x-axis with an angle given by $\alpha_{\mathrm{nf}}(x)$. This angle is minimum for $x = 0$ and maximum for $x = L$, where L is the width of the rectangular spatial window. (b) The characteristic features of the resulting spectrum are mainly defined by the parameters $\cos\alpha = \mathbb{E}_x\left[\cos\alpha_{\mathrm{nf}}(x)\right]$, $\Phi_x^{\min} = \cos\alpha_{\mathrm{nf}}^{\max}\frac{\Omega}{c}$, and $\Phi_x^{\max} = \cos\alpha_{\mathrm{nf}}^{\min}\frac{\Omega}{c}$. The amplitude is maximum within the region $\Phi_x^{\min} \leq \Phi_x \leq \Phi_x^{\max}$, which can be interpreted as the main lobe of the support function, and the outside region (the region $\mathcal{U}$) consists of side-lobe ripples. The ripples are oriented towards $\frac{\partial\Phi_x}{\partial\Omega} = \frac{\cos\alpha}{c}$. The figure in (c) shows a Matlab simulation of an intermediate-field source, and the figure in (d) shows the respective approximation by the parametric model.

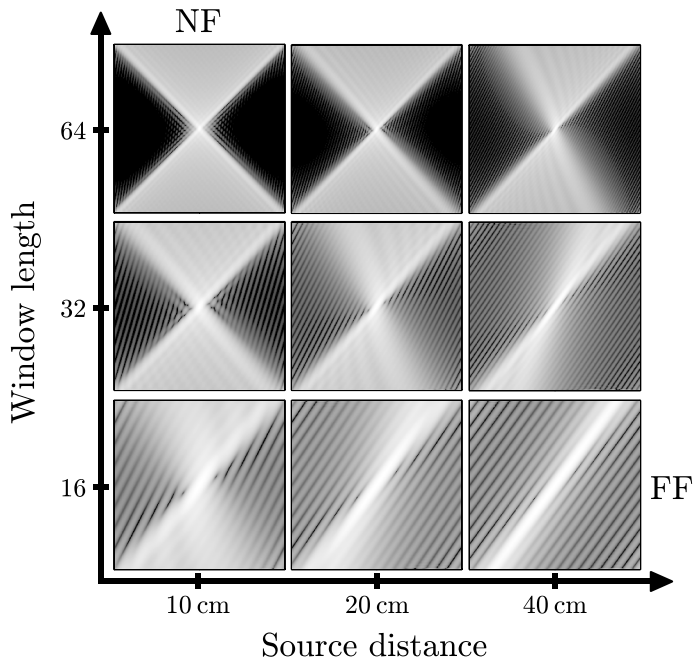
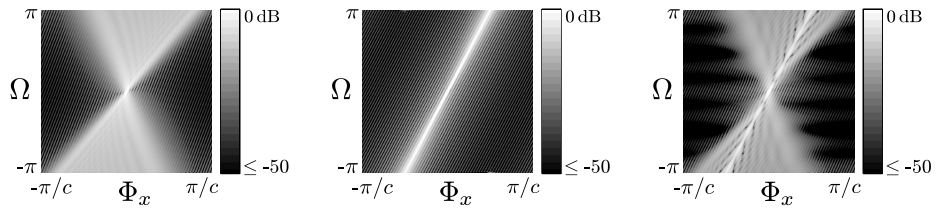


Figure 5.5: Convergence of the intermediate-field spectrum to the near-field and the far-field. The near-field scenario is obtained by either reducing the source distance or increasing the window size. The far-field scenario is obtained by increasing the source distance or decreasing the window size. Recall that a source is in the near-field when its distance to the observation region is small compared to the size of the region itself, and in the far-field when the opposite is true. Note also that the parameters $\cos\alpha$, $\Phi_x^{\min}$, and $\Phi_x^{\max}$ of the spectral pattern are preserved by increasing the source distance and the window size by the same factor.



(a) (b) (c)

Figure 5.6: Spatio-temporal spectrum generated by two co-linear sources at $\alpha = \frac{\pi}{3}$ with source signal $s(t) = \delta(t)$ and different distances. The spectrum of each source is shown in (a) and (b) respectively, and their sum in (c).

where a is the distance ratio between the two sources such that the signals have the same amplitude at the x-axis, and $\alpha_1 = \alpha_2$ due to co-linearity. An example is illustrated in Figure 5.6.

As the figure shows, the two sources can be easily distinguished in the Fourier spectrum, where the second source is clearly emphasized even though it is located behind the first source with respect to the observer.

5.2.2 Discrete case

The short space/time Fourier transform of the 4-D discrete signal $p[\mathbf{n}]$ is given by an 8-D function $P[\mathbf{n}_0, \mathbf{b}]$ defined as

$$P[\mathbf{n}_0, \mathbf{b}] = \sum_{\mathbf{n} \in \mathbb{Z}^4} p[\mathbf{n}] \underbrace{w[\mathbf{n} - \mathbf{n}_0] e^{-j2\pi \mathbf{b} \cdot \mathbf{N}^{-1} \mathbf{n}}}_{\text{Generalized element}}, \ \mathbf{n}_0 \in \mathbb{Z}^4 \text{ and } \mathbf{b} \in \mathbb{Z}^4 \tag{5.11}$$

where $\mathbf{N}$ is a diagonal matrix with the number of samples in each dimension,

$$\mathbf{N} = \begin{bmatrix} N_x & & & \\ & N_y & & \\ & & N_z & \\ & & & N_t \end{bmatrix}, \tag{5.12}$$

and $w[\mathbf{n} - \mathbf{n}_0]$ is the spatio-temporal window function, shifted by $\mathbf{n}_0$. The window function is a product between the three spatial windows and the temporal window.

The inverse transform is given by

$$p[\mathbf{n}] = \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{n}_0 \in \mathbb{Z}^4} \sum_{\mathbf{b} \in \mathbb{Z}^4} P[\mathbf{n}_0, \mathbf{b}] e^{j2\pi \mathbf{b} \cdot \mathbf{N}^{-1} \mathbf{n}}, \ \mathbf{n} \in \mathbb{Z}^4. \tag{5.13}$$

The inverse formula represents the discrete sum of all the signal blocks indexed by $\mathbf{n}_0$, and it requires that $w[\mathbf{n}]$ has unit sum. The proof is given in Section 5.5.

The discrete version of the short space/time Fourier transform is the simplest example of a block transform for acoustic wave fields, and it will serve as introduction to the topic of spatio-temporal filter banks discussed in Chapter 7.

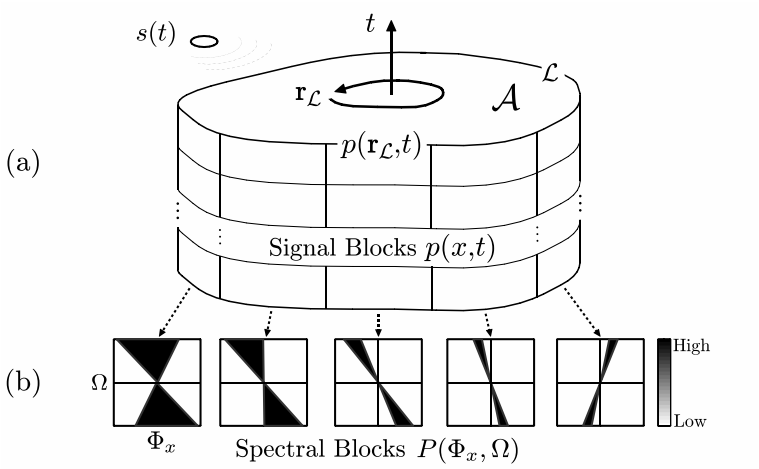


Figure 5.7: Local analysis of the wave field on a manifold. (a) In the spatio-temporal manifold generated by the smooth contour $\mathcal{L}$, the signal $p(\mathbf{r}_\mathcal{L}, t)$ is locally approximated by multiple spatio-temporal blocks $p(x, t)$. (b) The resulting spectral blocks $P(\Phi_x, \Omega)$ have different plane-wave content depending on the local properties of the wave field. In this example, the blocks on the left have more near-field characteristics than the blocks on the right, where the energy is more concentrated around the dominant direction.

5.3 Local analysis on a manifold

In Chapter 3, we introduced the concept of manifolds as a tool for reducing the number of spatial dimensions required to analyze the wave field on a surface $\mathcal{S}$ or a contour $\mathcal{L}$, such that only two or less spatial dimensions are needed. So, for instance, if the observation region is a curved contour $\mathcal{L}$, the analysis can be focused on small segments in space by sliding a tangential window along $\mathcal{L}$. Applying the same principle to the temporal dimension, the result is a spatio-temporal manifold along $\mathcal{L}$ similar to the one shown in Figure 5.7-(a). In the figure, each block is characterized by a straight spatial axis—for simplicity, the x-axis—in addition to the temporal axis.

As mentioned in the introduction, the use of local spatial analysis is motivated by the fact that the frequency content of the acoustic wave field varies between different regions in space. For instance, if the analysis window is closer to the source, the Fourier spectrum has near-field characteristics, whereas if the analysis window is further away from the source the Fourier spectrum turns into a far-field pattern. This is illustrated in Figure 5.7-(b).

Thus, by using local spatial analysis, we can take advantage of the stationary regions of the wave field, where the spectrum has mostly far-field characteristics and is therefore sparser and easier to analyze and manipulate. This advantage becomes apparent in operations such as filtering and coding, which we address in the last two chapters of this thesis.

The use of windowing on a curved manifold, however, comes with a cost: it reduces the sharpness of the spectrum. The effects of bending the spatial axis are somehow equivalent to increasing the curvature of the wave front, which happens when the source gets closer to the observation region. Thus, the more bended the spatial axis is at a certain region, the less sharp the spectrum will be, and *vice versa*. This can be compensated, for example, by varying the

window size according to the local smoothness of $\mathcal{L}$. The trade-off is illustrated in Figure 5.8.

5.4 Summary notes

- Non-stationary wave fields have different characteristics at different regions in space, which are not efficiently represented by a global Fourier analysis;
- A local analysis of the acoustic wave field requires a redefinition of the basic element of the Fourier transform, as a plane wave limited in space and time by a spatio-temporal window;
- A plane wave with finite size is represented in the spatio-temporal Fourier domain as a smooth support function, which converges to a Dirac as the size of the plane wave grows to infinity;
- A wide-band source in the far-field is represented in the spatio-temporal Fourier domain as a smooth support function with its main lobe defined on a line crossing the origin;
- If the source is in the intermediate-field, the spatio-temporal Fourier transform can be described by a parametric model that consists of a "max" operation between the Fourier transform of the spatial window and a triangular mask that opens and closes according to the source distance;
- The local Fourier analysis can be applied on a curved contour, at the expense of lower spectral sharpness.

5.5 Theorems and proofs

Theorem 26 (Continuous inversion formula)**.** *Given the continuous short space/time Fourier transform $P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega)$ of a signal $p(\mathbf{r}, t)$, the signal $p(\mathbf{r}, t)$ can be recovered through the inversion formula given by*

$$p(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int_{\mathbb{R}^3} \int_{\mathbb{R}} \int_{\mathbb{R}^3} \int_{\mathbb{R}} P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) e^{j(\mathbf{\Phi}\cdot\mathbf{r} + \Omega t)} d\Omega d\mathbf{\Phi} dt_0 d\mathbf{r}_0.$$

Proof. Plugging (5.1) into the inversion formula, yields

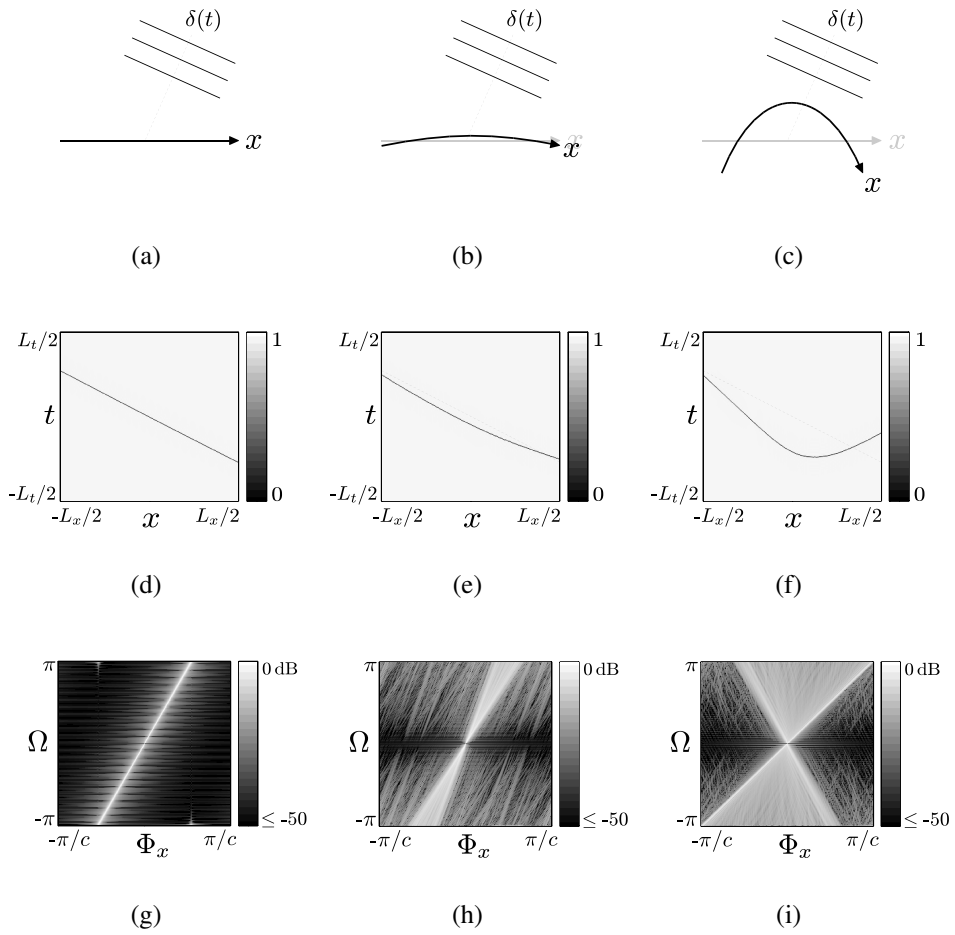


Figure 5.8: Resolution trade-off between the local curvature of $\mathcal{L}$ and the spatial frequency Φ_x, for the case of a rectangular window and a far-field source with source signal $s(t) = \delta(t)$. In (a), (b), and (c), the x-axis is progressively bent, causing the curvature of the wave front to increase accordingly, as shown in (d), (e), and (f). The resulting spectrum, shown in (g), (h), and (i) respectively, becomes increasingly blurred, reducing the sparsity of the Fourier representation.

$$
\begin{aligned}
& \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\left(\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)w_{\mathbf{r}}(\boldsymbol{\rho}-\mathbf{r}_0)w_t(\tau-t_0)e^{-j(\boldsymbol{\Phi}\cdot\boldsymbol{\rho}+\Omega\tau)}d\tau d\boldsymbol{\rho}\right) \\
& e^{j(\boldsymbol{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\boldsymbol{\Phi}dt_0 d\mathbf{r}_0 \\
\overset{(a)}{=}\ & \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)\left(\int_{\mathbb{R}^3} w_{\mathbf{r}}(\boldsymbol{\rho}-\mathbf{r}_0)d\mathbf{r}_0\right)\left(\int_{\mathbb{R}} w_t(\tau-t_0)dt_0\right) \\
& e^{j(\boldsymbol{\Phi}\cdot(\mathbf{r}-\boldsymbol{\rho})+\Omega(t-\tau))}d\tau d\boldsymbol{\rho}d\Omega d\boldsymbol{\Phi} \\
\overset{(b)}{=}\ & \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\boldsymbol{\rho},\tau)e^{-j(\boldsymbol{\Phi}\cdot\boldsymbol{\rho}+\Omega\tau)}d\tau d\boldsymbol{\rho}e^{j(\boldsymbol{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\boldsymbol{\Phi} \\
=\ & \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}} P(\boldsymbol{\Phi},\Omega)e^{j(\boldsymbol{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\boldsymbol{\Phi} \\
=\ & p(\mathbf{r},t),
\end{aligned}
$$

where (a) is valid as long as the Fubini theorem [16] is satisfied, *i.e.*, if the integral is well defined, and (b) comes from $\int_{\mathbb{R}^3} w_{\mathbf{r}}(\mathbf{r})d\mathbf{r} = 1$ and $\int_{\mathbb{R}} w_t(t)dt = 1$. □

Theorem 27 (Far-field spectrum)**.** *Given a far-field source with source signal $s(t)$ and propagation vector $\mathbf{u}$, and the spatial and temporal window functions $w_{\mathbf{r}}(\mathbf{r})$ and $w_t(t)$, with Fourier transforms $W_{\mathbf{r}}(\boldsymbol{\Phi})$ and $W_t(\Omega)$, the short space/time Fourier transform is given by*

$$
P(\mathbf{r}_0,t_0,\boldsymbol{\Phi},\Omega) = S(\Omega)W_{\mathbf{r}}\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big) * W_t(\Omega)e^{-j(\boldsymbol{\Phi}\cdot\mathbf{r}_0+\Omega t_0)}. \tag{5.14}
$$

Proof. Plugging $p(\mathbf{r},t) = s\big(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\big)$ into (5.1), yields

$$
\begin{aligned}
& P(\mathbf{r}_0,t_0,\boldsymbol{\Phi},\Omega) \\
=\ & \int_{\mathbb{R}^3}\int_{\mathbb{R}} s\Big(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\Big)w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)w_t(t-t_0)e^{-j(\boldsymbol{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r} \\
=\ & \int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}} s(\tau)\delta\Big(t-\tau+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\Big)d\tau w_t(t-t_0)e^{-j\Omega t}dtw_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r} \\
=\ & \int_{\mathbb{R}} s(\tau)\int_{\mathbb{R}^3}\left(e^{-j\Omega\left(\tau-\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)} *_{\Omega} W_t(\Omega)e^{-j\Omega t_0}\right)d\tau w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r} \\
=\ & \int_{\mathbb{R}} s(\tau)\int_{\mathbb{R}^3}\left(\int_{\mathbb{R}} e^{-j\Gamma\left(\tau-\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}W_t(\Omega-\Gamma)e^{-j(\Omega-\Gamma)t_0}d\Gamma\right)d\tau w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r} \\
\overset{(a)}{=}\ & \int_{\mathbb{R}}\left(\int_{\mathbb{R}} s(\tau)e^{-j\Gamma\tau}d\tau\right)\left(\int_{\mathbb{R}^3} e^{j\Gamma\frac{\mathbf{u}\cdot\mathbf{r}}{c}}w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r}\right)W_t(\Omega-\Gamma)e^{-j(\Omega-\Gamma)t_0}d\Gamma \\
=\ & \int_{\mathbb{R}} S(\Gamma)W_{\mathbf{r}}\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Gamma}{c}\Big)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}_0}W_t(\Omega-\Gamma)e^{-j(\Omega-\Gamma)t_0}d\Gamma \\
=\ & S(\Omega)W_{\mathbf{r}}\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}_0} * W_t(\Omega)e^{-j\Omega t_0} \\
=\ & S(\Omega)W_{\mathbf{r}}\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big) * W_t(\Omega)e^{-j(\boldsymbol{\Phi}\cdot\mathbf{r}_0+\Omega t_0)},
\end{aligned}
$$

where (a) is valid as long as the Fubini theorem [16] is satisfied, *i.e.*, if the integral is well defined. □

Corollary 28 (Complex-exponential source)**.** *The result of Theorem 27 for $s(t) = e^{j\Omega_0 t}$, where Ω_0 is a fixed frequency, is given by*

$$P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) = 2\pi W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega_0}{c}\Big) W_t(\Omega - \Omega_0) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\Omega_0)t_0\right)}.$$

Proof. The Fourier transform of $s(t) = e^{j\Omega_0 t}$ is given by $S(\Omega) = 2\pi\delta(\Omega - \Omega_0)$. Plugging this into (5.14), yields

$$\begin{aligned}
& P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) \\
= \;& 2\pi\delta(\Omega - \Omega_0) W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) * W_t(\Omega) e^{-j(\mathbf{\Phi}\cdot\mathbf{r}_0 + \Omega t_0)} \\
= \;& 2\pi \int_{\mathbb{R}} \delta(\tau - \Omega_0) W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\tau}{c}\Big) W_t(\Omega - \tau) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\tau)t_0\right)} d\tau \\
= \;& 2\pi W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega_0}{c}\Big) W_t(\Omega - \Omega_0) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\Omega_0)t_0\right)}.
\end{aligned}$$

□

Corollary 29 (Dirac source)**.** *The result of Theorem 27 for $s(t) = \delta(t)$, with $W_t(\Omega)$ defined as a nascent Dirac function such that $\lim_{a\to\infty} aW_t(a\Omega) = \delta(\Omega)$, is given by*

$$P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) = W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) e^{-j\mathbf{\Phi}\cdot\mathbf{r}_0}.$$

Proof. The Fourier transform of $s(t) = \delta(t)$ is given by $S(\Omega) = 1$. Plugging this into (5.14), yields

$$\begin{aligned}
& P(\mathbf{r}_0, t_0, \mathbf{\Phi}, \Omega) \\
= \;& W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) * W_t(\Omega) e^{-j(\mathbf{\Phi}\cdot\mathbf{r}_0 + \Omega t_0)} \\
= \;& \int_{\mathbb{R}} W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\tau}{c}\Big) W_t(\Omega - \tau) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\tau)t_0\right)} d\tau \\
= \;& \frac{\|\mathbf{u}\|}{c} \int_{\mathbb{R}} W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\tau}{c}\Big) \frac{c}{\|\mathbf{u}\|} W_t\left(\frac{c}{\|\mathbf{u}\|}\left(\|\mathbf{\Phi}\| - \frac{\|\mathbf{u}\|}{c}\tau\right)\right) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\tau)t_0\right)} d\tau \\
\overset{(a)}{=} \;& \frac{\|\mathbf{u}\|}{c} \int_{\mathbb{R}} W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\tau}{c}\Big) \delta\left(\|\mathbf{\Phi}\| - \frac{\|\mathbf{u}\|}{c}\tau\right) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\tau)t_0\right)} d\tau \\
= \;& \frac{\|\mathbf{u}\|}{c} \int_{\mathbb{R}} W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\tau}{c}\Big) \frac{c}{\|\mathbf{u}\|} \delta\left(\frac{c}{\|\mathbf{u}\|}\|\mathbf{\Phi}\| - \tau\right) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\tau)t_0\right)} d\tau \\
= \;& \int_{\mathbb{R}} W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\tau}{c}\Big) \delta(\Omega - \tau) e^{-j\left(\mathbf{\Phi}\cdot\mathbf{r}_0 + (\Omega-\tau)t_0\right)} d\tau \\
= \;& W_{\mathbf{r}}\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) e^{-j\mathbf{\Phi}\cdot\mathbf{r}_0},
\end{aligned}$$

where (a) comes from $\lim_{a\to\infty} aW_t(a\Omega) = \delta(\Omega)$ and from the fact that $\frac{c}{\|\mathbf{u}\|} \geq c$, where c is typically a large value ($c = 340\mathrm{m/s}$ in the case of sound). □

Theorem 30. *[Intermediate-field spectrum] Consider a point source with source signal $s(t)$ and position $\mathbf{r}_p = (x_p, z_p)$, and the approximation of $P(\Phi_x, \Omega)$ by the parametric model $P(\Phi_x, \Omega) = S(\Omega)\max\left\{W_x\left(\Phi_x - \cos\alpha\frac{\Omega}{c}\right), M(\Phi_x, \Omega)\right\}$, where $M(\Phi_x, \Omega)$ is given by (5.9) with parameters $\Phi_x^{\min}$ and $\Phi_x^{\max}$. Define $\alpha_{\mathrm{nf}}(x) = \angle\big(x_p - x + jz_p\big)$ as the angle of incidence at point x, where $\alpha_{\mathrm{nf}}^{\min} = \alpha_{\mathrm{nf}}(0)$ is the smallest angle and $\alpha_{\mathrm{nf}}^{\max} = \alpha_{\mathrm{nf}}(L)$ is the largest angle, such that $\alpha_{\mathrm{nf}}^{\min} \leq \alpha_{\mathrm{nf}}(x) \leq \alpha_{\mathrm{nf}}^{\max}$. The ripples within the region $\mathcal{U}$ are oriented towards $\cos\alpha = \mathbb{E}_x\big[\cos\alpha_{\mathrm{nf}}(x)\big]$, where $\mathbb{E}_x$ denotes expectation over x, and the triangular mask is delimited by $\Phi_x^{\min} = \cos\alpha_{\mathrm{nf}}^{\max}\frac{\Omega}{c}$ and $\Phi_x^{\max} = \cos\alpha_{\mathrm{nf}}^{\min}\frac{\Omega}{c}$.*

Proof. Divide the window function $w_x(x)$ of length L into M segments $w_x^{(m)}(x)$ of length $\frac{L}{M}$, such that $w_x(x) = \sum_{m=0}^{M-1} w_x^{(m)}\big(x - m\frac{L}{M}\big)$. For M large enough, the near-field wave fronts become increasingly far-field in the range of each segment (by definition, $\|\mathbf{r}_p\| \gg \|\mathbf{r}\|$). Thus,

$$P(\Phi_x, \Omega) = \sum_{m=0}^{M-1} \frac{2\pi}{r^{(m)}} S(\Omega)\delta\Big(\Phi_x - \cos\alpha^{(m)}\frac{\Omega}{c}\Big) *_{\Phi_x} \Big(W_x^{(m)}(\Phi_x)e^{-jm\frac{L}{M}\Phi_x}\Big),$$

where $\alpha^{(m)} = \alpha_{\mathrm{nf}}\big(m\frac{L}{M}\big)$ and $r^{(m)} = \big\|(m\frac{L}{M}, 0) - \mathbf{r}_p\big\|$. Note also that, as M increases, $W_x^{(m)}(\Phi_x)$ becomes increasingly flat with magnitude $\frac{L}{M}w_x\big(m\frac{L}{M}\big)$. This simplifies the result to

$$P(\Phi_x, \Omega) = S(\Omega)\frac{2\pi}{M}\sum_{m=0}^{M-1} \frac{Lw_x\big(m\frac{L}{M}\big)}{r^{(m)}} e^{-jm\frac{L}{M}\left(\Phi_x - \cos\alpha^{(m)}\frac{\Omega}{c}\right)}. \tag{5.15}$$

The notches in $\mathcal{U}$ occur when the sum in (5.15) is minimized, *i.e.*, when the exponential vectors are in maximum phase opposition. This requires a minimization of $\mathbb{E}\big[L\big(\Phi_x - \cos\alpha^{(m)}\frac{\Omega}{c}\big) - k2\pi\big]$ for $k \in \mathbb{Z}\backslash 0$, which occurs at $\Phi_x = \mathbb{E}\big[\cos\alpha^{(m)}\big]\frac{\Omega}{c} + \frac{k2\pi}{L}$; hence the orientation towards $\mathbb{E}\big[\cos\alpha^{(m)}\big]$. When $\Phi_x = \cos\alpha^{(m)}\frac{\Omega}{c}$, $\forall m$, at least one exponential vector in the sum equals 1. Since $\alpha_{\mathrm{nf}}(x)$ is a smooth function, the other $M-2$ vectors are also concentrated in the vicinity of 1. On the contrary, as $\big|\Phi_x - \cos\alpha^{(m)}\frac{\Omega}{c}\big|$ increases, the vectors become more dispersed in the complex plane, and the sum decreases in magnitude. This places the optimal limits at $\Phi_x^{\min} = \cos\alpha_{\mathrm{nf}}^{\max}\frac{\Omega}{c}$ and $\Phi_x^{\max} = \cos\alpha_{\mathrm{nf}}^{\min}\frac{\Omega}{c}$. □

Theorem 31 (Discrete inversion formula)**.** *Given the discrete short space/time Fourier transform $P[\mathbf{n}_0, \mathbf{b}]$ of a signal $p[\mathbf{n}]$, the signal $p[\mathbf{n}]$ can be recovered through the inversion formula given by*

$$p[\mathbf{n}] = \frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{n}_0\in\mathbb{N}^4}\sum_{\mathbf{b}\in\mathbb{N}^4} P[\mathbf{n}_0, \mathbf{b}]e^{j2\pi\mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}}.$$

Proof. The proof is the same as in Theorem 26, considering the fact that sums can be represented as integrals in the context of measure theory [16]. □

Chapter 6

Directional Representations

6.1 Introduction

The spatio-temporal Fourier transform was introduced as an efficient and sparse representation of the acoustic wave field that expresses the harmonic behavior across space as a function of the spatial frequency. The concept of spatial frequency can be understood as the frequency Φ_0 of a sinusoidal excitation pattern induced by the wave front on the microphones of the array axis. If the excitation pattern is characterized by a single frequency, it can either mean that there is only one source at the scene or there are multiple sources that coincidentally generate the same spatial frequency. Similarly, if the excitation pattern is characterized by more than one spatial frequency, it may be an indication that there is more than one source at the scene, or it may be that the same source is generating multiple temporal frequencies. The exact causes are not easy to determine unless we relate the spatial frequency to the angle of arrival of the wave front α_0, through the expression $\Phi_0 = \cos\alpha_0 \frac{\Omega}{c}$. Only then, by tracing a line from the origin to each spectral point, it becomes clear how many sources are present at the scene and what their individual directions are. Figure 6.1 illustrates a few examples of naturally occurring acoustic scenes, and shows how much information can be obtained from the spectrum by looking at the spatial frequency and the angle of arrival.

From a perceptual point of view, we, as human beings, are trained to abstract from the excitation patterns generated in our ears, focusing only on the subjective idea of direction induced by these patterns in our brains. If, for instance, we ask a friend "where is the sound coming from?", he will almost certainly point his arm in the direction of the source, indicating the angular nature of his spatial awareness. Indeed, the concept of direction has a much higher perceptual significance than the concept of spatial frequency.

The same is true from a mathematical point of view. In most of the results obtained in the previous two chapters, there is at least one parameter involved that is dependent on the angle of arrival of the wave front. In some cases (*e.g.*, far-field sources) it is the only parameter describing the spatial behavior of the wave front; in other cases (*e.g.*, intermediate-field sources) the spatial behavior can be expressed as a function of angle and distance (also a parameter with perceptual significance). Therefore, there is a strong motivation to express the spatio-temporal Fourier transform as a function of the angles of arrival α and β instead of the spatial frequencies Φ_x and Φ_y.

Defining a directional Fourier transform requires that the basic element of the Fourier

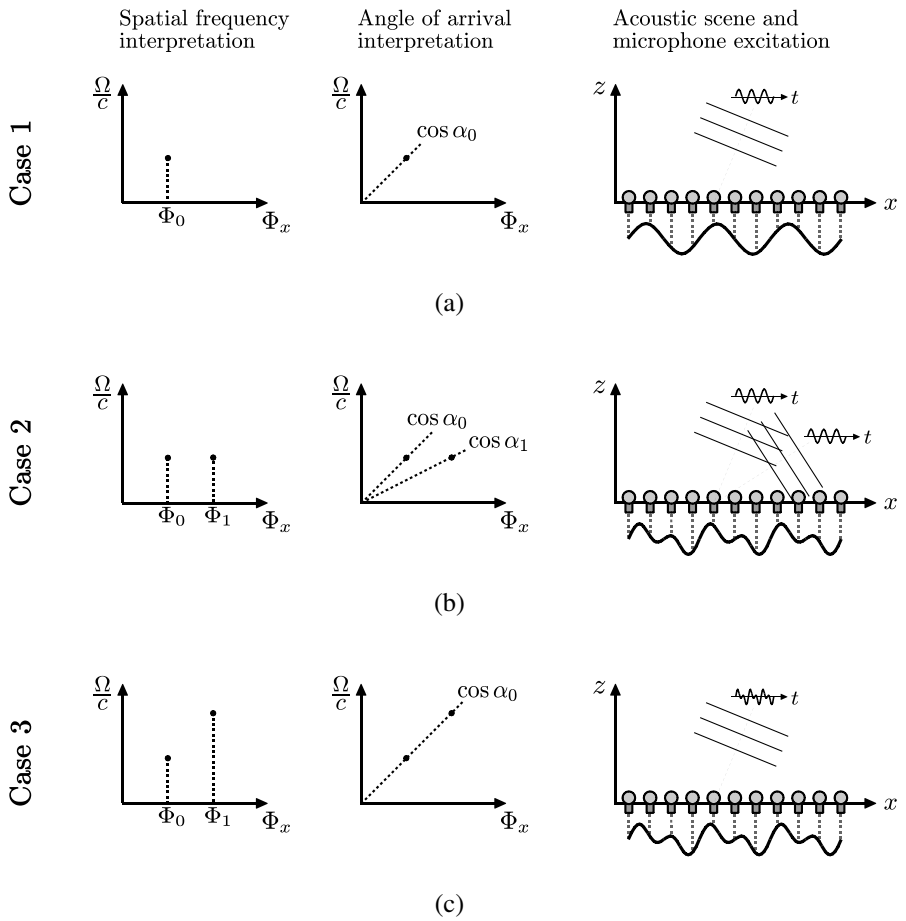


Figure 6.1: Is it possible to guess the content of the acoustic scene by looking at the spatial frequency? The figure shows three cases where the microphone array is excited in three different ways. The two spectral representations on the left are exactly the same, except that in the first we measure the spatial frequency of each point and in the second we measure the angle of arrival. On the right, the respective acoustic scene is shown together with the temporal excitation of the far-field sources and the spatial excitation of the microphone array (shown under the array). In (a), it is easy to guess that there is only one sinusoidal source at the scene. Yet, the spatial frequency Φ_0 does not intuitively tell where the source is located in space, as opposed to angle of arrival α_0. In (b), we can guess that there two sinusoidal sources at the scene, but again the respective spatial frequencies Φ_0 and Φ_1 do not intuitively tell where the sources are located in space, as opposed to angles of arrival α_0 and α_1. In (c), by looking at the spatial frequencies, one would be tempted to say that there are two sinusoidal sources at the scene. However, by looking at the angles of arrival, we realize that the two points in the spectrum actually belong to the same source, even though the spatial excitation of the microphone array is exactly the same as in (b). The true scene, in fact, contains only one source driven by a temporal signal with two frequencies.

integral, $e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}$, be represented as a function of α and β. For this purpose, we can use the relation $\mathbf{\Phi} = \mathbf{u}\frac{\Omega}{c}$ to redefine the spatio-temporal Fourier transform as

$$P\left(\mathbf{u}\frac{\Omega}{c},\Omega\right) = \int_{\mathbb{R}^2}\int_{\mathbb{R}} p(\mathbf{r},t)e^{-j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}dtd\mathbf{r},$$

where $\mathbf{r} = (x,y)$ and $\mathbf{u} = (u_x,u_y)$. This way, the basic element of the Fourier transform, $e^{j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}$, is characterized by a directional parameter $\mathbf{u}$ that replaces the spatial frequency $\mathbf{\Phi}$. The only question left is how to express $P(\mathbf{u}\frac{\Omega}{c},\Omega)$ as a function of α and β. Note that the concept of "angle of arrival" is defined as the angle between the normal vectors of a wave front and a flat surface in space. Therefore, it makes little sense to include the variable z in the Fourier integral.

The representation of the acoustic wave field in the directional Fourier domain has several effects in the type of spectral patterns obtained and the structure of the spectrum as a whole. The most important implication is that the spectrum becomes separable with respect to the defining parameters of the plane wave—angle and frequency. This means that changing the angle of the plane wave only shifts the spectral support function along the α or β axes, whereas changing the frequency of oscillation only shifts the support function along the Ω axis. This is an optimal representation in the directional sense, since, for every fixed direction α_0 and β_0, the spatio-temporal spectrum is completely described by the 1-D information along the Ω axis, plus the two parameters α_0 and β_0. We will see in Chapter 9 that, given the directional nature of acoustic wave fields, as well as of human auditory perception, this type of directional representation can result in higher coding gains.

The purpose of this chapter is to define a directional representation of the Fourier transform, and show how the spectral patterns obtained with the spatio-temporal Fourier transform translate into this new domain. The directional representation will give an alternative interpretation of the elementary components of the wave field, beyond that given by plane waves. We will introduce a new element called *far-field component*, which allows the spectrum of a near-field source to be expressed as a function of far-field patterns. This chapter will take us one step further into understanding what the basic elements of the acoustic wave field are.

6.2 Directional Fourier transform

The directional Fourier transform of the 3-D signal $p(\mathbf{r},t)$, with $\mathbf{r} = (x,y)$, is a 3-D function $P(\boldsymbol{\alpha},\Omega)$ given by

$$P(\boldsymbol{\alpha},\Omega) = \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2\int_{\mathbb{R}^2}\int_{\mathbb{R}} p(\mathbf{r},t)e^{-j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}dtd\mathbf{r} \tag{6.1}$$

where $\boldsymbol{\alpha} = (\alpha,\beta)$ are the angles of arrival of the wave fronts with respect to the x and y axes respectively, in accordance to Figure 2.6-(a), and $\mathbf{u} = (u_x,u_y)$ is the propagation vector—recall that, in polar coordinates, $u_x = \cos\alpha\sin\beta$ and $u_y = \cos\beta$. The normalization factor $\frac{d\mathbf{u}}{d\boldsymbol{\alpha}}$ is defined such that $\frac{d\mathbf{u}}{d\boldsymbol{\alpha}} = \frac{du_x}{d\alpha}\frac{du_y}{d\beta}$.

The inverse transform is given by

$$p(\mathbf{r},t) = \frac{1}{(2\pi)^3}\int_{\mathbb{A}^2}\int_{\mathbb{R}} P(\boldsymbol{\alpha},\Omega)e^{j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}d\Omega d\boldsymbol{\alpha} \tag{6.2}$$

where $\mathbb{A}$ is the space of angles between 0 and π, and $\int_{\mathbb{A}^2} d\boldsymbol{\alpha} = \int_0^\pi \int_0^\pi d\alpha d\beta$. The directional Fourier integral represents the projection of $p(\mathbf{r}, t)$ onto the basis $e^{j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}$, which is the basic element of the directional Fourier transform. This basis is notably similar to the sound pressure generated by a complex exponential in the far-field—also known as a *plane wave*—where $p(\mathbf{r}, t) = s\left(t + \frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)$ given $s(t) = e^{j\Omega_0 t}$. The directional Fourier transform can, therefore, be interpreted as a decomposition of the acoustic wave field into plane waves characterized by a direction of propagation $\mathbf{u}$—itself a function of $\boldsymbol{\alpha}$. The vector $\boldsymbol{\alpha}$ can take complex values as well, in case one or more components of $\mathbf{u}$ are complex. This is the case when $p(\mathbf{r}, t)$ contains evanescent-wave energy (*i.e.*, when the wave front is curved). A proof of the inversion formula is given in Section 6.5.

In line with the previous chapters, we will show examples that help to understand the spectral patterns generated by (6.1) given the most common inputs, and illustrate the results in the 3-D and 2-D spaces. The usual "test signals" are the complex exponential, the sinusoid, and the Dirac impulse, driving a point source in the far-field.

Example 32 (Far-field spectrum)**.** Consider a point source located in the far-field with source signal $s(t)$ and propagation vector $\mathbf{u}_0$ associated to a vector of arrival angles $\boldsymbol{\alpha}_0 = (\alpha_0, \beta_0)$, such that $p(\mathbf{r}, t) = s\left(t + \frac{\mathbf{u}_0\cdot\mathbf{r}}{c}\right)$. Plugging $p(\mathbf{r}, t)$ into (6.1), yields

$$\boxed{P(\boldsymbol{\alpha}, \Omega) = S(\Omega)(2\pi)^m \delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)} \tag{6.3}$$

where $\boldsymbol{\alpha}, \boldsymbol{\alpha}_0 \in \mathbb{R}^m$ for $m = 0, 1, 2$. The multi-dimensional Dirac function $\delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)$ is non-zero for $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, and can be expressed as $\delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) = \delta(\alpha - \alpha_0)\delta(\beta - \beta_0)$ for $m = 2$. Given that $\alpha, \beta \in [0, \pi]$, the Dirac function is always within a square prism defined by $0 \leq \alpha \leq \pi$ and $0 \leq \beta \leq \pi$, as illustrated in Figure 6.2. The proof of this result is given in Section 6.5.

Comparing the result in (6.3) to the regular Fourier transform obtained in the previous chapter, given by $P(\boldsymbol{\Phi}, \Omega) = S(\Omega)(2\pi)^m \delta\left(\boldsymbol{\Phi} - \mathbf{u}\frac{\Omega}{c}\right)$, the only change is the argument of the Dirac support function, which was translated to the new angular coordinate system.

Example 33. Consider the case of Example 32 for $s(t) = e^{j\Omega_0 t}$, where Ω_0 is a fixed frequency. The Fourier transform of the complex exponential is given by $S(\Omega) = 2\pi\delta(\Omega - \Omega_0)$, and thus

$$P(\boldsymbol{\alpha}, \Omega) = (2\pi)^{m+1}\delta(\Omega - \Omega_0)\delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0). \tag{6.4}$$

This is illustrated in Figure 6.3-(a). The result is a single point in the directional spectrum located at $(\boldsymbol{\alpha}, \Omega) = (\boldsymbol{\alpha}_0, \Omega_0)$, and, similarly to what happens in the regular Fourier transform, it is the elementary component in the directional Fourier analysis of the wave field, also known as *plane wave*.

Comparing to the regular Fourier transform obtained in the previous chapter, given by $P(\boldsymbol{\Phi}, \Omega) = (2\pi)^{m+1}\delta(\Omega - \Omega_0)\delta\left(\boldsymbol{\Phi} - \mathbf{u}\frac{\Omega_0}{c}\right)$, we observe that, in the directional Fourier representation, the Dirac point located at $(\boldsymbol{\Phi}, \Omega) = \left(\mathbf{u}\frac{\Omega_0}{c}, \Omega_0\right)$ is translated to a point located at $(\boldsymbol{\alpha}, \Omega) = (\boldsymbol{\alpha}_0, \Omega_0)$.

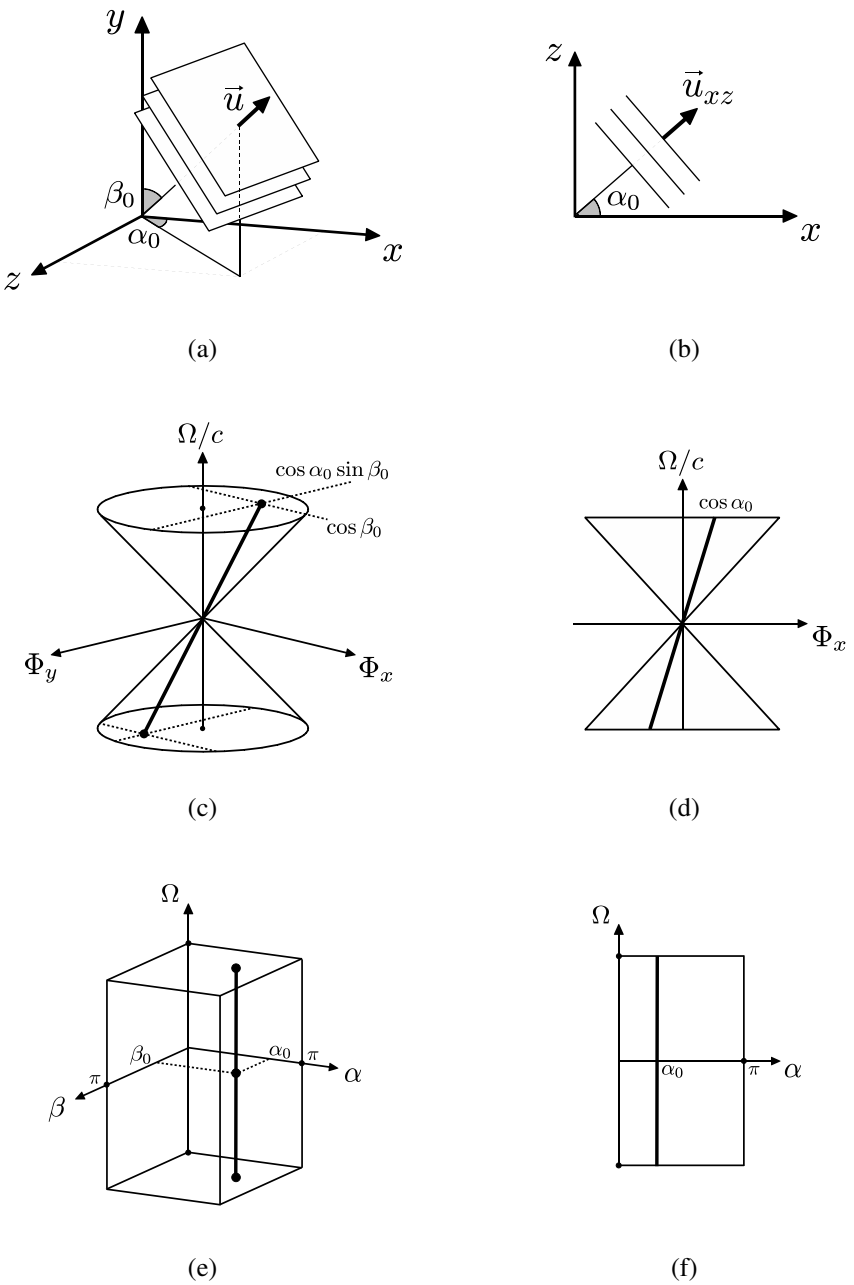


Figure 6.2: Directional Fourier transform of a far-field source for different values of m (number of spatial dimensions). (a) and (b) show the wave front generated by the far-field source for $m = 2$ and $m = 1$ respectively. (c) and (d) show the respective spatio-temporal Fourier transforms. (e) and (f) show the respective directional Fourier transforms. The result of translating the wave front a far-field source into a directional Fourier representation is that the Dirac function becomes a vertical line crossing the $\alpha\beta$-plane at the point $\boldsymbol{\alpha} = (\alpha_0, \beta_0)$, instead of a diagonal line crossing the origin. Similarly to the case of the spatio-temporal Fourier transform, the Dirac function in the directional representation is weighted by $S(\Omega)$.

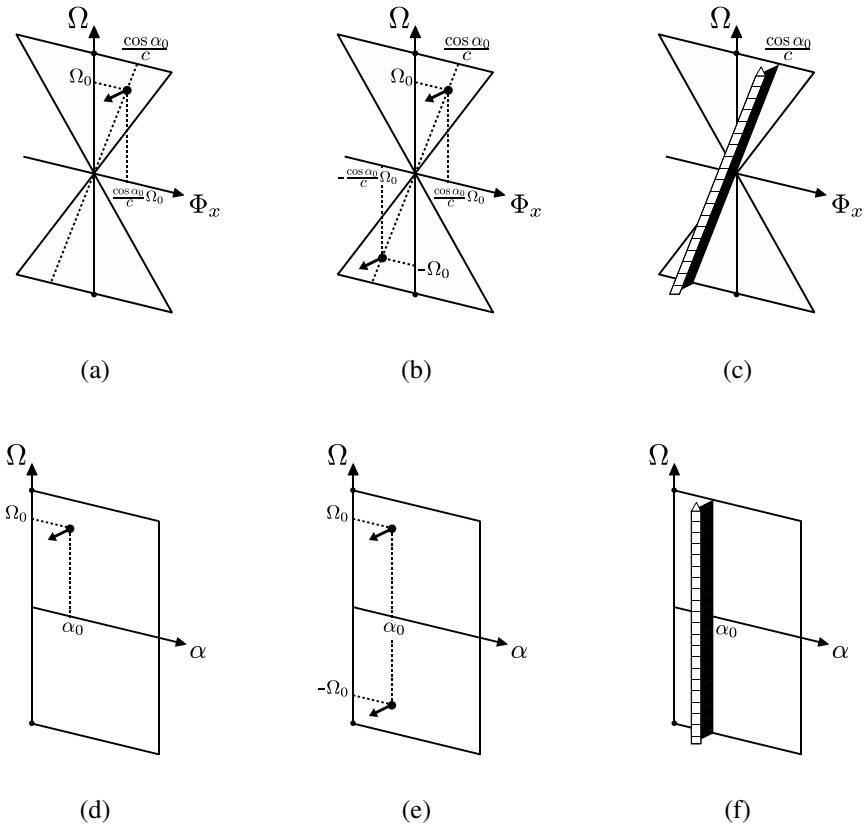


Figure 6.3: Examples of the directional Fourier transform for $m = 1$ (one spatial dimension) and different source signals. For $s(t) = e^{j\Omega_0 t}$, the spatio-temporal spectrum shown in (a) is a single Dirac point located at $(\Phi_x, \Omega) = \left(\cos\alpha_0 \frac{\Omega_0}{c}, \Omega_0\right)$; the directional spectrum shown in (d) is a single Dirac point located at $(\alpha, \Omega) = (\alpha_0, \Omega_0)$. The results for $s(t) = 2\cos(\Omega_0 t)$, shown in (b) and (e), are the same as in (a) and (d) except for the added symmetric point at negative frequencies. For $s(t) = \delta(t)$, the spatio-temporal spectrum shown in (c) is a Dirac function defined on a line with slope $\frac{\partial \Phi_x}{\partial \Omega} = \frac{\cos\alpha_0}{c}$; the directional spectrum shown in (f) is a Dirac function defined on a vertical line crossing $\alpha = \alpha_0$.

Example 34. Consider a variation of the previous example, where $s(t) = 2\cos(\Omega_0 t)$. Since $2\cos(\Omega_0 t) = e^{-j\Omega_0 t} + e^{j\Omega_0 t}$, the Fourier transform is given by $S(\Omega) = 2\pi\delta(\Omega + \Omega_0) + 2\pi\delta(\Omega - \Omega_0)$, and thus

$$P(\boldsymbol{\alpha}, \Omega) = (2\pi)^{m+1}\Big(\delta(\Omega + \Omega_0) + \delta(\Omega - \Omega_0)\Big)\delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0). \tag{6.5}$$

This is illustrated in Figure 6.3-(b). The result is the same as in the previous example, except for an additional point at the symmetric position $(\boldsymbol{\alpha}, \Omega) = (\boldsymbol{\alpha}_0, -\Omega_0)$. Note however that, as opposed to what happens in the regular Fourier transform $P(\boldsymbol{\Phi}, \Omega)$, the directional Fourier transform $P(\boldsymbol{\alpha}, \Omega)$ is not symmetric with respect to the origin, but with respect to the $\boldsymbol{\alpha}$-plane.

Example 35. If the source signal is a Dirac pulse, $s(t) = \delta(t)$, the Fourier transform is maximally flat and given by $S(\Omega) = 1$. The directional Fourier transform is then given by

$$P(\boldsymbol{\alpha}, \Omega) = (2\pi)^m \delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0). \tag{6.6}$$

This is illustrated in Figure 6.3-(c).

The examples given above show that the directional representation of the wave field represents the diagonal lines of the Fourier spectrum as vertical lines with the same amplitude, given by $S(\Omega)$. In the near-field and intermediate-field cases, for which we do not derive theoretical results, it can be expected that the triangular region containing the majority of the energy will be represented in the directional domain as a rectangular region, defined between $\alpha = \alpha_{\text{nf}}^{\min}$ and $\alpha = \alpha_{\text{nf}}^{\max}$. This result is demonstrated in the next section using a different interpretation of the directional Fourier transform.

6.3 Decomposition into far-field components

The directional Fourier transform defined in (6.1) and (6.2) represents $p(\mathbf{r}, t)$ as a function of its plane wave components, where each plane wave is characterized by a vector of angles $\boldsymbol{\alpha} = (\alpha, \beta)$ and a frequency of oscillation Ω. Such a transform can be seen as an alternative to the regular Fourier transform, better adapted to the directional nature of the acoustic wave field. It is possible, however, to obtain the directional Fourier transform $P(\boldsymbol{\alpha}, \Omega)$ from the regular Fourier transform $P(\boldsymbol{\Phi}, \Omega)$. For instance, if $P(\boldsymbol{\Phi}, \Omega)$ is already available at some point in the system, we may want to obtain $P(\boldsymbol{\alpha}, \Omega)$ directly without going back to computing $p(\mathbf{r}, t)$. This relation is defined as

$$P(\boldsymbol{\alpha}, \Omega) = \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} \left(\frac{\Omega}{c}\right)^2 \int_{\mathbb{R}^2} P(\boldsymbol{\Phi}, \Omega)\delta\Big(\boldsymbol{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\Phi}, \tag{6.7}$$

and, conversely,

$$P(\boldsymbol{\Phi}, \Omega) = \int_{\mathbb{A}^2} P(\boldsymbol{\alpha}, \Omega)\delta\Big(\boldsymbol{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\alpha}. \tag{6.8}$$

The results are proved in Section 6.5. This formulation of the directional Fourier transform represents the projection of $P(\boldsymbol{\Phi}, \Omega)$ onto the "space" formed by Dirac support functions

$\delta\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big)$. The result in (6.8) in particular can be interpreted as the continuous counterpart of a discrete superposition of far-field sources, with arrival angles ranging from 0 to π—recall from Section 4 that the spectrum of a far-field source is given by $P(\mathbf{\Phi}, \Omega) = S(\Omega)(2\pi)^m \delta\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big)$, and thus $P(\boldsymbol{\alpha}_0, \Omega)$ can be interpreted as the 1-D spectrum of a virtual source in the far-field with direction $\boldsymbol{\alpha}_0$. Therefore, we call (6.7) the decomposition of the wave field into *far-field components*. The concept is illustrated in Figure 6.4.

Note that we call the term $P(\boldsymbol{\alpha}, \Omega)\delta\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big)$ far-field "component" instead of a far-field "source". This is because $P(\boldsymbol{\alpha}, \Omega)$, for some given $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, does not necessarily represent a true source with direction $\boldsymbol{\alpha}_0$, but may represent for instance the energy contribution of a near-field source to that given direction, or the energy contribution of reverberation in the direction of $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$. The important idea is that any acoustic scene can be expressed as a function of far-field components, whether in open space or closed space, or whether it contains sources in the far-field or in the near-field.

It is also important to notice that, in the strict mathematical sense, the actual energy contribution of any given direction $\boldsymbol{\alpha}$ is zero. This can be easily understood by looking at $\delta\big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\big)$ as a probability density function, in which case any fixed $\boldsymbol{\alpha}$ yields a zero probability. To obtain a non-zero value, we have to consider a range of directions. For this purpose, it can be shown that, for any given $\boldsymbol{\alpha}_\mathrm{A}, \boldsymbol{\alpha}_\mathrm{B} \in \mathbb{A}^2$ such that $0 \leq \alpha_\mathrm{A} < \alpha_\mathrm{B} \leq \pi$ and $0 \leq \beta_\mathrm{A} < \beta_\mathrm{B} \leq \pi$, the accumulated spectral content within the interval $\boldsymbol{\alpha} \in [\boldsymbol{\alpha}_\mathrm{A}, \boldsymbol{\alpha}_\mathrm{B}]$ satisfies

$$\int_{\boldsymbol{\alpha}_\mathrm{A}}^{\boldsymbol{\alpha}_\mathrm{B}} P(\boldsymbol{\alpha}, \Omega) d\boldsymbol{\alpha} = \int_{\mathbf{u}_\mathrm{B}\frac{\Omega}{c}}^{\mathbf{u}_\mathrm{A}\frac{\Omega}{c}} P(\mathbf{\Phi}, \Omega) d\mathbf{\Phi}. \tag{6.9}$$

The proof is given in Section 6.5. In particular, if $\boldsymbol{\alpha}_\mathrm{A} = (0, 0)$ and $\boldsymbol{\alpha}_\mathrm{B} = (\pi, \pi)$,

$$\int_{\mathbb{A}^2} P(\boldsymbol{\alpha}, \Omega) d\boldsymbol{\alpha} = \int_{\mathbb{R}^2} P(\mathbf{\Phi}, \Omega) d\mathbf{\Phi}. \tag{6.10}$$

The result in (6.10) simply states that the total accumulated spectrum across $\boldsymbol{\alpha}$ in the directional domain (another way of saying: "the temporal information of the wave field") is equivalent to the total accumulated spectrum across $\mathbf{\Phi}$ in the spatio-temporal Fourier domain. The result in (6.9), however, has deeper implications, as it actually relates the spectral tiling of $P(\boldsymbol{\alpha}, \Omega)$ to a subband partitioning of $P(\mathbf{\Phi}, \Omega)$, as illustrated in Figure 6.5. Essentially, the result tells us how the spatio-temporal spectrum $P(\mathbf{\Phi}, \Omega)$ can be partitioned in order to obtain a discrete version of $P(\boldsymbol{\alpha}, \Omega)$, with a discrete number of "angular subbands". We will discuss the construction of a discrete directional transform in the next chapter.

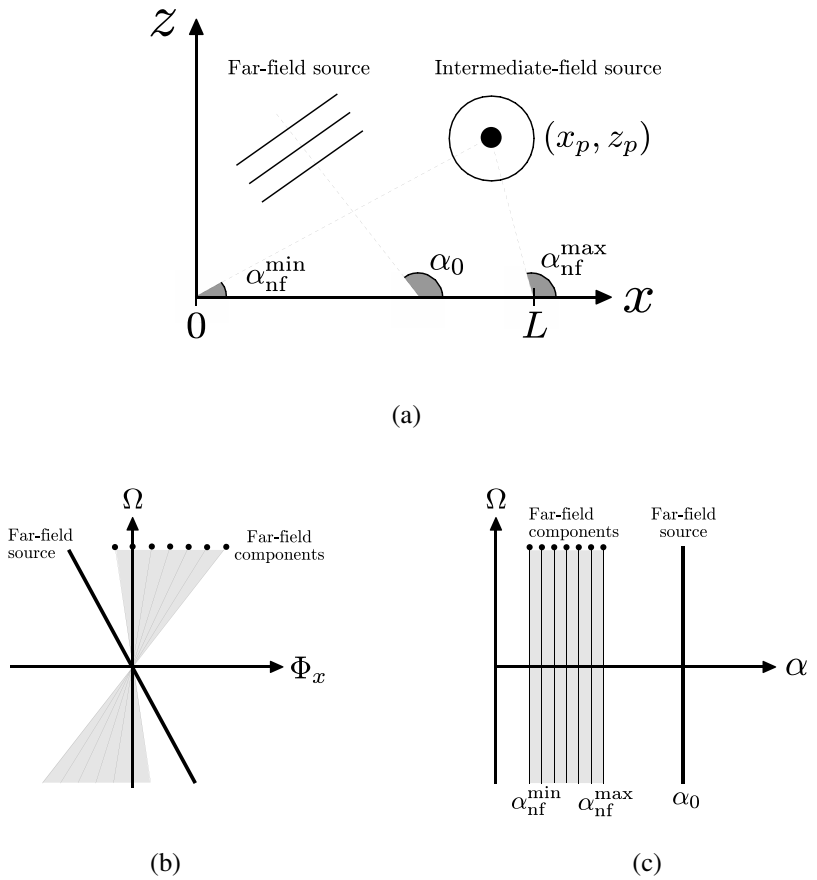
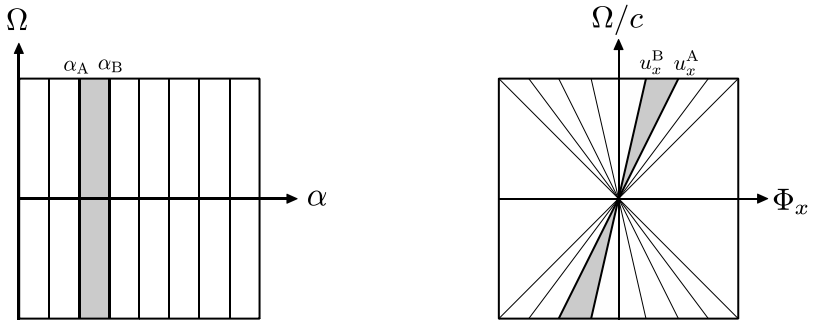


Figure 6.4: Decomposition of the wave field into far-field components. (a) In this example, the acoustic scene consists of two sources: one in the far-field with angle of arrival α_0 and one in the intermediate-field at position $(x, z) = (x_p, z_p)$. (b) In the spatio-temporal Fourier transform, the far-field source shows up as a Dirac function defined along a line crossing the origin with slope $\frac{\partial \Phi_x}{\partial \Omega} = \frac{\cos \alpha_0}{c}$, whereas the intermediate-field source shows up as a triangular region delimited by two lines of slope $\cos \alpha_{\mathrm{nf}}^{\mathrm{min}}/c$ and $\cos \alpha_{\mathrm{nf}}^{\mathrm{max}}/c$ (for simplicity, we discard the side-lobes outside the triangular region). This region can be interpreted as a continuum of zero-crossing lines with slopes ranging from $\cos \alpha_{\mathrm{nf}}^{\mathrm{min}}/c$ to $\cos \alpha_{\mathrm{nf}}^{\mathrm{max}}/c$, called far-field components. (c) In the directional Fourier transform, the far-field source shows up as a Dirac function defined along a vertical line crossing the α-axis at $\alpha = \alpha_0$, whereas the intermediate-field source shows up as a square region delimited by two vertical lines crossing the α-axis at $\alpha = \alpha_{\mathrm{nf}}^{\mathrm{min}}$ and $\alpha = \alpha_{\mathrm{nf}}^{\mathrm{max}}$. Similarly, the square region can be interpreted as a continuum of vertical lines crossing the α-axis at multiple points ranging from $\alpha_{\mathrm{nf}}^{\mathrm{min}}$ to $\alpha_{\mathrm{nf}}^{\mathrm{max}}$.



(a) (b)

Figure 6.5: Relation between the spectral tiling of $P(\alpha, \Omega)$ and the subband partitioning of $P(\Phi_x, \Omega)$. A rectangular-shaped region (a tile) in $P(\alpha, \Omega)$ ranging from $\alpha = \alpha_A$ to α_B, as shown in (a), has the same spectral content of a triangular-shaped region in $P(\Phi_x, \Omega)$ delimited by two lines with slope u_x^A/c and u_x^B/c, where $u_x = \cos\alpha$, as shown in (b).

6.4 Summary notes

- The spatial frequency is not the best intuitive measure of the spatial characteristics of the acoustic wave field;
- The spatial awareness of human beings is based on an angular localization of sources at the scene;
- A directional representation of the spatio-temporal Fourier transform can be obtained by expressing the Fourier basis as $e^{j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}$;
- The directional Fourier transform provides an equally sparse spectral representation where each far-field source is expressed as a vertical line perpendicular to the $\alpha\beta$-plane and weighted by the respective source spectrum $S(\Omega)$;
- In the directional Fourier domain, the spectral triangle generated by near-field and intermediate-field sources is expressed as a rectangular region defined between the minimum and maximum angles of incidence of the wave front;
- The directional Fourier transform can be redefined to express the wave field as a continuum of virtual sources in the far-field with different angles of arrival, called far-field components.

6.5 Theorems and proofs

Theorem 36 (Inversion formula)**.** *Given the directional Fourier transform $P(\boldsymbol{\alpha}, \Omega)$ of a signal $p(\mathbf{r}, t)$, the signal $p(\mathbf{r}, t)$ can be recovered through the inversion formula given by*

$$p(\mathbf{r}, t) = \frac{1}{(2\pi)^3} \int_{\mathbb{A}^2} \int_{\mathbb{R}} P(\boldsymbol{\alpha}, \Omega) e^{j\Omega\left(t + \frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)} d\Omega d\boldsymbol{\alpha}.$$

Proof. Plugging (5.1) into the inversion formula, yields

$$\begin{aligned}
& \frac{1}{(2\pi)^3} \int_{\mathbb{A}^2} \int_{\mathbb{R}} P(\boldsymbol{\alpha}, \Omega) e^{j\Omega\left(\frac{\mathbf{u}\cdot\mathbf{r}}{c} + t\right)} d\Omega d\boldsymbol{\alpha} \\
= \quad & \frac{1}{(2\pi)^3} \int_{\mathbb{A}^2} \int_{\mathbb{R}} \int_{\mathbb{R}^2} \int_{\mathbb{R}} p(\boldsymbol{\rho}, \tau) \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} \left(\frac{\Omega}{c}\right)^m e^{-j\Omega\left(\frac{\mathbf{u}\cdot\boldsymbol{\rho}}{c} + \tau\right)} d\tau d\boldsymbol{\rho} e^{j\Omega\left(\frac{\mathbf{u}\cdot\mathbf{r}}{c} + t\right)} d\Omega d\boldsymbol{\alpha} \\
\overset{(a)}{=} \quad & \frac{1}{(2\pi)^3} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \int_{\mathbb{A}^2} \int_{\mathbb{R}} p(\boldsymbol{\rho}, \tau) \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} \left(\frac{\Omega}{c}\right)^m e^{j\Omega \frac{\mathbf{u}}{c}\cdot(\mathbf{r}-\boldsymbol{\rho})} e^{j\Omega(t-\tau)} d\Omega d\boldsymbol{\alpha} d\tau d\boldsymbol{\rho} \\
= \quad & \frac{1}{(2\pi)^3} \int_{\mathbb{R}^2} \int_{\mathbb{R}} p(\boldsymbol{\rho}, \tau) \int_{\mathbb{A}^2} e^{j\frac{\Omega}{c}\mathbf{u}\cdot(\mathbf{r}-\boldsymbol{\rho})} d\left(\frac{\Omega}{c}\mathbf{u}\right) \int_{\mathbb{R}} e^{j\Omega(t-\tau)} d\Omega d\tau d\boldsymbol{\rho} \\
= \quad & \int_{\mathbb{R}^2} \int_{\mathbb{R}} p(\boldsymbol{\rho}, \tau) \delta(\mathbf{r} - \boldsymbol{\rho}) \delta(t - \tau) d\tau d\boldsymbol{\rho} \\
= \quad & p(\mathbf{r}, t)
\end{aligned}$$

where (a) is valid as long as the Fubini theorem [16] is satisfied, *i.e.*, if the integral is well defined (in the ℓ_1 sense). □

Theorem 37 (Far-field spectrum)**.** *Given a far-field source with source signal $s(t)$ and propagation vector $\mathbf{u}_0$ associated to an angle vector $\boldsymbol{\alpha}_0 = (\alpha_0, \beta_0)$, the directional Fourier transform is given by*

$$P(\boldsymbol{\alpha}, \Omega) = S(\Omega)(2\pi)^m \delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0).$$

Proof. Plugging $p(\mathbf{r}, t) = s\left(t + \frac{\mathbf{u}_0\cdot\mathbf{r}}{c}\right)$ into (6.1), yields

$$
\begin{aligned}
P(\boldsymbol{\alpha},\Omega) &= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m \int_{\mathbb{R}^m}\int_{\mathbb{R}} s\Big(t+\frac{\mathbf{u}_0\cdot\mathbf{r}}{c}\Big)e^{-j\Omega\left(\frac{\mathbf{u}\cdot\mathbf{r}}{c}+t\right)}dtd\mathbf{r}\\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m \int_{\mathbb{R}^m} S(\Omega)e^{j\Omega\frac{\mathbf{u}_0\cdot\mathbf{r}}{c}}e^{-j\Omega\frac{\mathbf{u}\cdot\mathbf{r}}{c}}d\mathbf{r}\\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m S(\Omega)\int_{\mathbb{R}^m} e^{j\Omega\frac{(\mathbf{u}_0-\mathbf{u})\cdot\mathbf{r}}{c}}d\mathbf{r}\\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m S(\Omega)(2\pi)^m\delta\Big(\Omega\frac{\mathbf{u}_0-\mathbf{u}}{c}\Big)\\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}S(\Omega)(2\pi)^m\delta\Big(\mathbf{u}-\mathbf{u}_0\Big)\\
&\overset{(a)}{=} \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}S(\Omega)(2\pi)^m\frac{\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)}{\left|\frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}}\\
&\overset{(b)}{=} \left|\frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} S(\Omega)(2\pi)^m\frac{\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)}{\left|\frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}}\\
&= S(\Omega)(2\pi)^m\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)
\end{aligned}
$$

where (a) comes from the equality $\delta(f(x)) = \delta(x-x_0)/\left|\frac{df(x)}{dx}\right|_{x=x_0}$ [41], and (b) comes from the fact that $\frac{du_x}{d\alpha}\frac{du_y}{d\beta} = \sin\alpha\sin^2\beta \geq 0$, $\forall\alpha,\beta\in[0,\pi]$, and $\left|\frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\right|\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) = \left|\frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)$. □

Theorem 38 (Decomposition into far-field components)**.** *Given the spatio-temporal Fourier transform $P(\boldsymbol{\Phi},\Omega)$ of a signal $p(\mathbf{r},t)$, the directional Fourier transform $P(\boldsymbol{\alpha},\Omega)$ is given by*

$$P(\boldsymbol{\alpha},\Omega) = \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2\int_{\mathbb{R}^2} P(\boldsymbol{\Phi},\Omega)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\Phi}.$$

Conversely,

$$P(\boldsymbol{\Phi},\Omega) = \int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\alpha}.$$

Proof. The first result can be obtained as follows:

$$
\begin{aligned}
P(\boldsymbol{\alpha},\Omega) &= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2\int_{\mathbb{R}^2}\int_{\mathbb{R}} p(\mathbf{r},t)e^{-j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}dtd\mathbf{r}\\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2 P\Big(\frac{\mathbf{u}\cdot\mathbf{r}}{c},\Omega\Big)\\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2\int_{\mathbb{R}^2} P(\boldsymbol{\Phi},\Omega)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\Phi}.
\end{aligned}
$$

The second result is obtained by plugging $p(\mathbf{r},\Omega) = \frac{1}{(2\pi)^2}\int_{\mathbb{A}^2}P(\boldsymbol{\alpha},\Omega)e^{j\Omega\frac{\mathbf{u}\cdot\mathbf{r}}{c}}d\boldsymbol{\alpha}$ into (4.1), such that

$$\begin{aligned}
P(\boldsymbol{\Phi},\Omega) &= \int_{\mathbb{R}^3} p(\mathbf{r},\Omega)e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r} \\
&= \int_{\mathbb{R}^3}\frac{1}{(2\pi)^2}\int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)e^{j\Omega\frac{\mathbf{u}\cdot\mathbf{r}}{c}}d\boldsymbol{\alpha}e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r} \\
&\overset{\text{(a)}}{=} \frac{1}{(2\pi)^2}\int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\int_{\mathbb{R}^3} e^{j\Omega\frac{\mathbf{u}\cdot\mathbf{r}}{c}}e^{-j\boldsymbol{\Phi}\cdot\mathbf{r}}d\mathbf{r}d\boldsymbol{\alpha} \\
&= \frac{1}{(2\pi)^2}\int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\int_{\mathbb{R}^3} e^{-j\left(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\right)\cdot\mathbf{r}}d\mathbf{r}d\boldsymbol{\alpha} \\
&= \int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\alpha}
\end{aligned}$$

where (a) is valid as long as the Fubini theorem [16] is satisfied, *i.e.*, if the integral is well defined. Note that the result of Theorem 37 can be obtained through a decomposition into far-field components. By plugging (4.3) into (6.7),

$$\begin{aligned}
P(\boldsymbol{\alpha},\Omega) &= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m\int_{\mathbb{R}^m} P(\boldsymbol{\Phi},\Omega)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\Phi} \\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m\int_{\mathbb{R}^m} S(\Omega)(2\pi)^m\delta\Big(\boldsymbol{\Phi}-\mathbf{u}_0\frac{\Omega}{c}\Big)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\Phi} \\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^m S(\Omega)(2\pi)^m\delta\Big(\mathbf{u}\frac{\Omega}{c}-\mathbf{u}_0\frac{\Omega}{c}\Big) \\
&= \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}S(\Omega)(2\pi)^m\delta\Big(\mathbf{u}-\mathbf{u}_0\Big) \\
&= S(\Omega)(2\pi)^m\delta(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0).
\end{aligned}$$

□

Theorem 39. *Given $\boldsymbol{\alpha}_\mathrm{A},\boldsymbol{\alpha}_\mathrm{B}\in\mathbb{A}^2$ such that $0\le\alpha_\mathrm{A}<\alpha_\mathrm{B}\le\pi$ and $0\le\beta_\mathrm{A}<\beta_\mathrm{B}\le\pi$, the accumulated spectral content within the interval $\boldsymbol{\alpha}\in[\boldsymbol{\alpha}_\mathrm{A},\boldsymbol{\alpha}_\mathrm{B}]$ satisfies*

$$\int_{\boldsymbol{\alpha}_\mathrm{A}}^{\boldsymbol{\alpha}_\mathrm{B}} P(\boldsymbol{\alpha},\Omega)d\boldsymbol{\alpha} = \int_{\mathbf{u}_\mathrm{B}\frac{\Omega}{c}}^{\mathbf{u}_\mathrm{A}\frac{\Omega}{c}} P(\boldsymbol{\Phi},\Omega)d\boldsymbol{\Phi}.$$

Proof. Plugging (6.8) yields

$$\begin{aligned}
\int_{\mathbf{u}_\mathrm{B}\frac{\Omega}{c}}^{\mathbf{u}_\mathrm{A}\frac{\Omega}{c}}\int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\alpha}d\boldsymbol{\Phi} &= \int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\int_{\mathbf{u}_\mathrm{B}\frac{\Omega}{c}}^{\mathbf{u}_\mathrm{A}\frac{\Omega}{c}}\delta\Big(\boldsymbol{\Phi}-\mathbf{u}\frac{\Omega}{c}\Big)d\boldsymbol{\Phi}d\boldsymbol{\alpha} \\
&= \int_{\boldsymbol{\alpha}_\mathrm{A}}^{\boldsymbol{\alpha}_\mathrm{B}} P(\boldsymbol{\alpha},\Omega)d\boldsymbol{\alpha},
\end{aligned}$$

The order of integration with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{\alpha}$ can be exchanged under the Fubini theorem [16] if $P(\boldsymbol{\alpha},\Omega)$ is absolutely integrable. The last equality is due to

$$\int_{\mathbf{u}_\mathrm{B}\frac{\Omega}{c}}^{\mathbf{u}_\mathrm{A}\frac{\Omega}{c}} \delta\Big(\mathbf{\Phi} - \mathbf{u}\frac{\Omega}{c}\Big) d\mathbf{\Phi} = \begin{cases} 1 & , \boldsymbol{\alpha} \in [\boldsymbol{\alpha}_\mathrm{A}, \boldsymbol{\alpha}_\mathrm{B}] \\ 0 & , \text{otherwise.} \end{cases}$$

□

Chapter 7

Filter-Bank Realizations of Spatio-Temporal Transforms

7.1 Introduction

The work carried out by Dennis Gabor in 1946 on the time–frequency representation of non-stationary signals [35] had an impact in the area of Fourier analysis well beyond that of the development of the short-time Fourier transform. The work essentially led to a generalized view of orthogonal transforms, based on the concept that different types of signals require a different partitioning of the time–frequency plane. Music signals, for instance, are better represented by a uniform partitioning of the spectrum, due to their harmonic nature. Electrocardiographic signals, on the contrary, are mostly characterized by low-frequency components generated by the heart beat plus the wide band noise generated by the surrounding muscles. For this type of signals, a dyadic partitioning of the spectrum—with higher resolution for lower frequencies—is a more appropriate representation. Such representations can be obtained through a class of discrete-time structures known as *filter banks*, which consist of a sequence of filters and rate converters organized in a tree structure (see, *e.g.*, Vaidyanathan [84] and Vetterli *et al.* [86]).

A typical filter bank structure, illustrated in Figure 7.1, consists of a signal decomposition stage, known as the *analysis stage*, and a signal reconstruction stage known as the *synthesis stage*. At every branch of the analysis stage, a given frequency band is filtered out and "stretched" by the rate converter in order to fully utilize the frequency range assigned by the initial sampling rate. The rate converters guarantee that the combination of all the filtered versions of the input has the desired number of samples in the end, which can be either: (i) less than the input (subsampling); (ii) more than the input (oversampling); (iii) or the same as the input (critically sampled). At the synthesis stage, the spectrum of the filtered signals is "squeezed" by the rate converters in order to occupy the original frequency range, and the branches are summed up. If the filters are properly designed, the aliasing caused by the rate converters is completely eliminated, due to aliasing compensation [84; 86].

Filter banks are a powerful tool used for modeling systems and obtaining efficient representations of a given class of signals through linear transforms that are invertible and critically sampled, and, in many cases, computationally efficient. In particular, filter banks can be used

Table 7.1: Summary of continuous spatio-temporal transforms.

Spatio-temporal Fourier transform
$P(\mathbf{\Phi},\Omega) = \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\mathbf{r},t)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r}$
$p(\mathbf{r},t) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}} P(\mathbf{\Phi},\Omega)e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\mathbf{\Phi}$
Short space/time Fourier transform
$P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega) = \int_{\mathbb{R}^3}\int_{\mathbb{R}} p(\mathbf{r},t)w_{\mathbf{r}}(\mathbf{r}-\mathbf{r}_0)w_t(t-t_0)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dtd\mathbf{r}$
$p(\mathbf{r},t) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^3}\int_{\mathbb{R}}\int_{\mathbb{R}^3}\int_{\mathbb{R}} P(\mathbf{r}_0,t_0,\mathbf{\Phi},\Omega)e^{j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}d\Omega d\mathbf{\Phi}dt_0d\mathbf{r}_0$
Directional Fourier transform
$P(\boldsymbol{\alpha},\Omega) = \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2\int_{\mathbb{R}^2}\int_{\mathbb{R}} p(\mathbf{r},t)e^{-j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}dtd\mathbf{r}$
$p(\mathbf{r},t) = \frac{1}{(2\pi)^3}\int_{\mathbb{A}^2}\int_{\mathbb{R}} P(\boldsymbol{\alpha},\Omega)e^{j\Omega\left(t+\frac{\mathbf{u}\cdot\mathbf{r}}{c}\right)}d\Omega d\boldsymbol{\alpha}$
$P(\boldsymbol{\alpha},\Omega) = \frac{d\mathbf{u}}{d\boldsymbol{\alpha}}\left(\frac{\Omega}{c}\right)^2\int_{\mathbb{R}^2} P(\mathbf{\Phi},\Omega)\delta\left(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\right)d\mathbf{\Phi}$
$P(\mathbf{\Phi},\Omega) = \int_{\mathbb{A}^2} P(\boldsymbol{\alpha},\Omega)\delta\left(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\right)d\boldsymbol{\alpha}$

to implement the discrete version of orthogonal transforms, such as the Fourier transform. For example, the frequency coefficients provided by the discrete Fourier transform (DFT) can be interpreted as the output of a uniform filter bank with as many bandpass filters as the number of coefficients. If the signal being transformed is multidimensional—say, a spatio-temporal signal—then the theory of multidimensional filter banks can be used instead, resulting in the typical filter bank structure shown in Figure 7.2. The generalization of filter banks theory (see, *e.g.*, Vaidyanathan [84]) consists of using multidimensional filters to obtain the different frequency bands from the input spectrum and using sampling lattices to regulate the spectral shaping prior to and after the filtering operations. Similarly to the one-dimensional case, the synthesis stage of the filter bank can be designed such that the output signal is a perfect reconstruction of the input.

In this chapter, we show how the theory of multidimensional filter banks can be used to translate the spatio-temporal transforms discussed in the previous chapters (summarized in Table 7.1) into realizable and computationally efficient discrete transforms that satisfy the strict requirements of perfect reconstruction and critical sampling. For simplicity, we discuss only the case where the spatio-temporal signals are two-dimensional, with the x-axis fully describing the observation region. We also show examples of different acoustic scenes and the respective signal representations they generate.

7.2 Realization of orthogonal transforms

Spatio-temporal orthogonal transforms can be obtained through any combination of orthogonal bases applied separately to the spatial and temporal dimensions of the input signal. Examples of transforms that can be used to exploit the temporal evolution of the sound field include the discrete Fourier transform (DFT), the discrete cosine transform (DCT), and the discrete wavelet transform (DWT). The DFT and the DCT are better suited for audio and

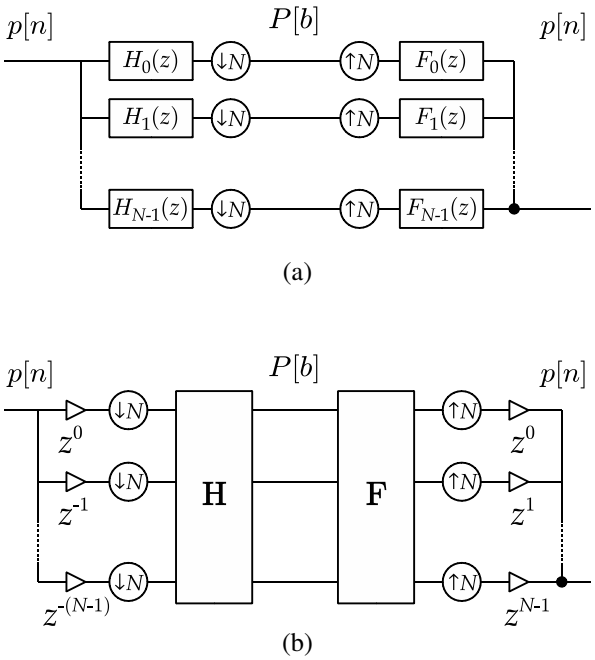


Figure 7.1: Typical structure of a one-dimensional filter bank. (a) The filter bank is composed of an analysis stage and a synthesis stage. The analysis stage splits up the input signal into N bandpass filtered versions, and downsamples each version by N in order to maintain the same sampling rate of the input. The synthesis stage then upsamples each bandpass signal by N and eliminates the aliasing components by filtering and summing up all the branches. If the filters $H(z)$ and $F(z)$ are properly designed, the output signal is a perfect reconstruction of the input. (b) The polyphase representation is an equivalent structure that expresses the filter bank as a computationally efficient block transform. The combination of delay factors and rate converters is known as "delay chain", and it basically generates a block of samples with size N from the input signal (in the analysis stage) or *vice versa* (in the synthesis stage). The filtering operations are then performed as a matrix product. Note that the delay chain on the right can be made causal by introducing a delay factor of $z^{-(N-1)}$.

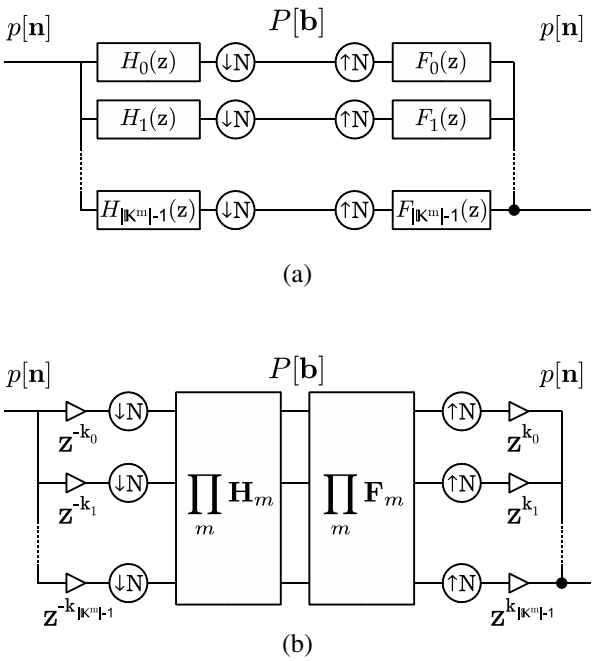


Figure 7.2: Typical structure of a multidimensional filter bank. (a) The filter bank structure is similar to the one-dimensional case, except that the filters and rate converters are multidimensional. The z-transform vector is defined such that, in the 2-D spatio-temporal domain, $\mathbf{z} = (z_x, z_t)$, and $\mathbf{N}$ is a diagonal resampling matrix given by $\mathbf{N} = \left[\begin{smallmatrix} N_x & 0 \\ 0 & N_t \end{smallmatrix}\right]$. The number of filters is determined by the size of the space of coset vectors $\mathbb{K}^2 \subset \mathbb{Z}^2$ (assuming $m = 2$ from the figure), which is essentially the space of all combinations of integer vectors $\mathbf{k} = \left[\begin{smallmatrix} k_x \\ k_t \end{smallmatrix}\right]$ from $\mathbf{k} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ to $\mathbf{k} = \left[\begin{smallmatrix} N_x - 1 \\ N_t - 1 \end{smallmatrix}\right]$. (b) The equivalent polyphase representation is characterized by a delay chain composed of vector delay factors $\mathbf{z}^{-\mathbf{k}} = z_x^{-k_x} z_t^{-k_t}$ and the resampling matrix $\mathbf{N}$, which generate 2-D sample blocks of size $N_x \times N_t$ from the input signal and *vice versa*. If the filter bank is separable, the filtering operations can be expressed as a product between transform matrices associated to each dimension.

speech sources, due to their harmonic nature, whereas the DWT is better suited for impulsive and transient-like sources (*e.g.*, shot events).

In the spatial domain, the choice of basis takes into account different factors, such as the position of the sources and the geometry of the acoustic environment (which influence the diffuseness of sound and the curvature of the wave field), as well as the geometry of the observation region. The Fourier transform, as we have shown, provides an efficient representation of the wave field on a straight line or a smoothly curved contour. However, under some particular conditions, a different choice of basis may prove to be more efficient. For instance, if the observation contour $\mathcal{L}$ is perfectly circular [2], a Fourier transform for circularly symmetric functions, such as the Hankel transform, can be used instead. In contrast, if $\mathcal{L}$ is sharply curved, the wavelet transform is more able to represent the sharp transitions across space in the wave field representation. In this chapter, however, we limit our analysis to the realization of traditional Fourier-based transforms.

The general formulation of a 2-D spatio-temporal orthogonal transform is given by

$$P[\mathbf{b}] = \sum_{\mathbf{n}\in\mathbb{Z}^2} p[\mathbf{n}]\upsilon^*_{b_x,n_x}\psi^*_{b_t,n_t},\ \mathbf{b}\in\mathbb{Z}^2 \tag{7.1}$$

and

$$p[\mathbf{n}] = \sum_{\mathbf{b}\in\mathbb{Z}^2} P[\mathbf{b}]\upsilon_{b_x,n_x}\psi_{b_t,n_t},\ \mathbf{n}\in\mathbb{Z}^2. \tag{7.2}$$

where υ_{b_x,n_x} and ψ_{b_t,n_t} are the spatial and temporal orthogonal bases, respectively. In matrix notation, (7.1) and (7.2) can be written as

$$\mathbf{Y} = \mathbf{\Upsilon}\mathbf{P}\mathbf{\Psi}^H \tag{7.3}$$

and

$$\mathbf{P} = \mathbf{\Upsilon}^H\mathbf{Y}\mathbf{\Psi}, \tag{7.4}$$

where $\mathbf{P}$, $\mathbf{Y}$, $\mathbf{\Upsilon}$, and $\mathbf{\Psi}$ are the matrix expansions of $p[\mathbf{n}]$, $P[\mathbf{b}]$, υ_{b_x,n_x}, and ψ_{b_t,n_t}, respectively. The results in (7.3) and (7.4) show that a spatio-temporal orthogonal transform is simply a matrix product between the input samples and the transformation matrices $\mathbf{\Upsilon}$ and $\mathbf{\Psi}$. The transform can thus be expressed as a multidimensional filter bank structure similar to the one shown in Figure 7.2, where the block $\prod_m \mathbf{H}_m$ represents a left product by $\mathbf{\Upsilon}$ and a right product by $\mathbf{\Psi}^H$, and $\prod_m \mathbf{F}_m$ represents a left product by $\mathbf{\Upsilon}^H$ and a right product by $\mathbf{\Psi}$. The input signal $p[\mathbf{n}]$ of size $N_x \times N_t$ is decomposed by the analysis stage of the filter bank into a transform matrix $P[\mathbf{b}]$ of equal size, and reconstructed back to $p[\mathbf{n}]$ by the synthesis stage.

7.2.1 Discrete Fourier transform

To perform a spatio-temporal discrete Fourier transform (DFT), the basis functions are defined as

$$\upsilon_{b_x,n_x} = \frac{1}{\sqrt{N_x}} e^{j\frac{2\pi}{N_x}b_x n_x},\ b_x = 0,\ldots,N_x-1 \text{ and } n_x = 0,\ldots,N_x-1 \tag{7.5}$$

$$\psi_{b_t,n_t} = \frac{1}{\sqrt{N_t}} e^{j\frac{2\pi}{N_t} b_t n_t},\ b_t = 0,\ldots,N_t-1 \text{ and } n_t = 0,\ldots,N_t-1. \tag{7.6}$$

This implies that $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are DFT matrices of size $N_x \times N_x$ and $N_t \times N_t$ respectively. An example of the output of the DFT filter bank is shown in Figure 7.3-(c).

7.2.2 Discrete cosine transform

To perform a spatio-temporal discrete cosine transform (DCT), the basis functions are defined as

$$\upsilon_{b_x,n_x} = \sqrt{\frac{2}{N_x}} \cos\left(\frac{\pi}{N_x}\left(n_x+\frac{1}{2}\right)\left(b_x+\frac{1}{2}\right)\right), \quad b_x = 0,\ldots,N_x-1 \text{ and } n_x = 0,\ldots,N_x-1$$

$$\upsilon_{b_t,n_t} = \sqrt{\frac{2}{N_t}} \cos\left(\frac{\pi}{N_t}\left(n_t+\frac{1}{2}\right)\left(b_t+\frac{1}{2}\right)\right), \quad b_t = 0,\ldots,N_t-1 \text{ and } n_t = 0,\ldots,N_t-1$$

Similarly, this implies that $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are DCT matrices of size $N_x \times N_x$ and $N_t \times N_t$ respectively. An example of the output of the DCT filter bank is shown in Figure 7.3-(d).

7.3 Realization of lapped orthogonal transforms

A lapped orthogonal transform (LOT) is a class of linear transforms where the input signal is split up into smaller overlapped blocks before each block is projected onto a given basis (and typically processed individually). A perfect reconstruction of the input signal is obtained by inverting the individual blocks and adding them through a technique known as overlap-and-add [62]. An example of a spatio-temporal LOT is the short space/time Fourier transform, introduced in Chapter 5.

The multidimensional filter bank structure of Figure 7.2 can be converted into a lapped transform simply by applying the resampling matrix $\mathbf{N}-\mathbf{O}$ instead of $\mathbf{N}$, where $\mathbf{O} = \left[\begin{smallmatrix} O_x & 0 \\ 0 & O_t \end{smallmatrix}\right]$ contains the number of overlapping samples O_x and O_t in each dimension. Since the resulting number of samples is greater than the number samples of the input signal, the filter bank becomes oversampled. Without loss of generality, we assume that $\mathbf{O} = \frac{1}{2}\mathbf{N}$, representing 50% of overlapping in both dimensions.

The decomposition of $p[\mathbf{n}]$ into overlapped blocks $p_{\mathbf{i}}[\mathbf{n}]$ can be written as

$$p_{\mathbf{i}}[\mathbf{n}] = p[\mathbf{n}]\,,\ \ \mathbf{n} = \frac{\mathbf{N}}{2}\mathbf{i},\ldots,\frac{\mathbf{N}}{2}(\mathbf{i}+\mathbf{2})-\mathbf{1}\,,\ \ \mathbf{i}\in\mathbb{I}^2, \tag{7.7}$$

where $\mathbf{i} = \left[\begin{smallmatrix} i_x \\ i_t \end{smallmatrix}\right]$ is the block index and $\mathbb{I}^2 \subset \mathbb{Z}^2$ is the respective set of block indexes. The notation $\mathbf{n} = \frac{\mathbf{N}}{2}\mathbf{i},\ldots,\frac{\mathbf{N}}{2}(\mathbf{i}+\mathbf{2})-\mathbf{1}$ means that $n_x = \frac{N_x}{2}i_x,\ldots,\frac{N_x}{2}(i_x+2)-1$ and $n_t = \frac{N_t}{2}i_t,\ldots,\frac{N_t}{2}(i_t+2)-1$. The vector integers are defined as $\mathbf{0} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, $\mathbf{1} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, and so on. Note also that, in order to handle the blocks that go outside the boundaries of $\mathbf{n}$, we

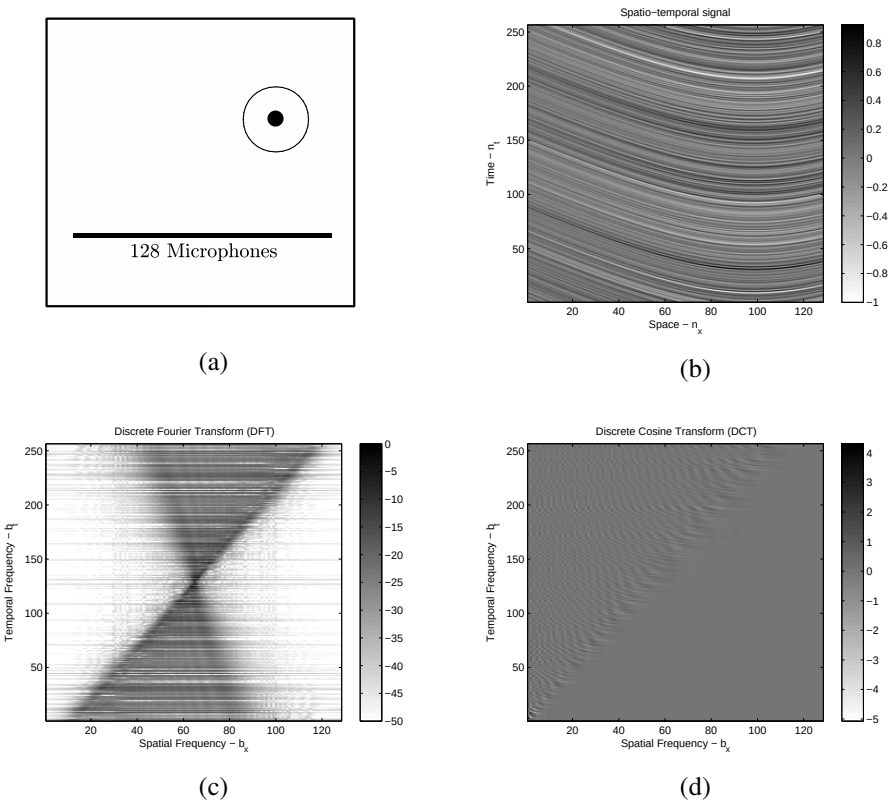


Figure 7.3: Example of discrete orthogonal transforms. (a) The acoustic scene consists of an intermediate-field source driven by white noise and a microphone array of size 128. (b) The spatio-temporal signal shows the curvature of the wave front in the region where the source is closer to the array. (c) The discrete Fourier transform (DFT) of the input signal shows a spectral pattern similar to the one predicted by the theory. (d) The discrete cosine transform (DCT) shows a similar spectral pattern to the DFT except that it only contains real-valued coefficients and is only defined for positive frequencies.

consider the signal to be circular (or periodic) in both dimensions. This presents an advantage over zero-padding, in particular, when $\mathcal{L}$ is closed.

Denoting $\varphi[\mathbf{b},\mathbf{n}] = \upsilon_{b_x,n_x}\psi_{b_t,n_t}$, the direct and inverse transforms for each block are given by

$$P_{\mathbf{i}}[\mathbf{b}] = \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N1}-\mathbf{1}} p_{\mathbf{i}}[\mathbf{n}]\varphi[\mathbf{b},\mathbf{n}]\,, \quad \mathbf{b} = \mathbf{0},\ldots,\mathbf{N1}-\mathbf{1} \tag{7.8}$$

and

$$\hat{p}_{\mathbf{i}}[\mathbf{n}] = \sum_{\mathbf{b}=\mathbf{0}}^{\mathbf{N1}-\mathbf{1}} P_{\mathbf{i}}[\mathbf{b}]\varphi[\mathbf{b},\mathbf{n}]\,, \quad \mathbf{n} = \mathbf{0},\ldots,\mathbf{N1}-\mathbf{1}. \tag{7.9}$$

Finally, the reconstruction of $p[\mathbf{n}]$ through overlap-and-add is given by

$$p[\mathbf{n}] = \sum_{\mathbf{i}\in\mathbb{I}^2} \hat{p}_{\mathbf{i}}\Big[\mathbf{n} - \frac{1}{2}\mathbf{N}\mathbf{i}\Big]\,, \quad \mathbf{n}\in\mathbb{Z}^2. \tag{7.10}$$

In matrix notation, the complete mechanism of the spatio-temporal LOT can be expressed as

$$\begin{bmatrix} \cdots\vdots & \vdots\cdots \\ \mathbf{Y_i} & \mathbf{Y}_{\mathbf{i}+\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]} \\ \mathbf{Y}_{\mathbf{i}+\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]} & \mathbf{Y_{i+1}} \\ \cdots\vdots & \vdots\cdots \end{bmatrix} = \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Upsilon}} & \underset{R}{\boldsymbol{\Upsilon}} & \\ & \underset{L}{\boldsymbol{\Upsilon}} & \underset{R}{\boldsymbol{\Upsilon}} \\ & & \ddots \end{bmatrix} \begin{bmatrix} \cdots\vdots & \vdots\cdots \\ \mathbf{P_i} & \mathbf{P}_{\mathbf{i}+\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]} \\ \mathbf{P}_{\mathbf{i}+\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]} & \mathbf{P_{i+1}} \\ \cdots\vdots & \vdots\cdots \end{bmatrix} \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Psi}} & \underset{R}{\boldsymbol{\Psi}} & \\ & \underset{L}{\boldsymbol{\Psi}} & \underset{R}{\boldsymbol{\Psi}} \\ & & \ddots \end{bmatrix}^H$$

and

$$\begin{bmatrix} \cdots\vdots & \vdots\cdots \\ \mathbf{P_i} & \mathbf{P}_{\mathbf{i}+\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]} \\ \mathbf{P}_{\mathbf{i}+\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]} & \mathbf{P_{i+1}} \\ \cdots\vdots & \vdots\cdots \end{bmatrix} = \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Upsilon}} & \underset{R}{\boldsymbol{\Upsilon}} & \\ & \underset{L}{\boldsymbol{\Upsilon}} & \underset{R}{\boldsymbol{\Upsilon}} \\ & & \ddots \end{bmatrix}^H \begin{bmatrix} \cdots\vdots & \vdots\cdots \\ \mathbf{Y_i} & \mathbf{Y}_{\mathbf{i}+\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]} \\ \mathbf{Y}_{\mathbf{i}+\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]} & \mathbf{Y_{i+1}} \\ \cdots\vdots & \vdots\cdots \end{bmatrix} \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Psi}} & \underset{R}{\boldsymbol{\Psi}} & \\ & \underset{L}{\boldsymbol{\Psi}} & \underset{R}{\boldsymbol{\Psi}} \\ & & \ddots \end{bmatrix}$$

where $\boldsymbol{\Upsilon} = \begin{bmatrix}\boldsymbol{\upsilon}_L & \boldsymbol{\upsilon}_R\end{bmatrix}$ and $\boldsymbol{\Psi} = \begin{bmatrix}\boldsymbol{\psi}_L & \boldsymbol{\psi}_R\end{bmatrix}$ are split into left and right halves in order to enforce overlapping directly in the transformation matrix. The resulting block-bidiagonal matrices are orthogonal.

7.3.1 Short space/time Fourier transform

The short space/time Fourier transform is obtained by defining the basis functions as

$$\upsilon_{b_x,n_x} = w_x[n_x]\frac{1}{\sqrt{N_x}}e^{j\frac{2\pi}{N_x}b_x n_x},\ b_x = 0,\ldots,N_x-1 \text{ and } n_x = 0,\ldots,N_x-1 \tag{7.11}$$

$$\psi_{b_t,n_t} = w_t[n_t]\frac{1}{\sqrt{N_t}}e^{j\frac{2\pi}{N_t}b_t n_t},\ b_t = 0,\ldots,N_t-1 \text{ and } n_t = 0,\ldots,N_t-1. \quad (7.12)$$

where $w_x[n_x]$ and $w_t[n_t]$ are the window functions in space and time, satisfying the conditions $w[n] = w[N-1-n]$ and $w^2[n] + w^2[n+\frac{N}{2}] = 1$ [77]. An example of the space/time Fourier transform is shown in Figure 7.4-(c).

7.3.2 Modified discrete cosine transform

The modified discrete cosine transform (MDCT) is an adaptation of the LOT introduced by Princen *et al.* [76], and used today as the base filter bank of most state-of-art audio coders (see, *e.g.*, Bosi *et al.* [15]). The main attribute of the MDCT is that it converts the LOT into a critically sampled filter bank while preserving the perfect reconstruction of the input. This is achieved through a technique known as time-domain aliasing cancelation (TDAC), which depends on a proper choice of the basis functions and can be extended to the 2-D case [60; 53]. An example of basis functions that satisfy the TDAC principle is given by

$$\upsilon_{b_x,n_x} = w_x[n_x]\sqrt{\frac{4}{N_x}}\cos\left(\frac{2\pi}{N_x}\left(n_x + \frac{N_x}{4} + \frac{1}{2}\right)\left(b_x + \frac{1}{2}\right)\right),$$
$$b_x = 0,\ldots,\frac{N_x}{2}-1 \text{ and } n_x = 0,\ldots,N_x-1, \quad (7.13)$$

and

$$\psi_{b_t,n_t} = w_t[n_t]\sqrt{\frac{4}{N_t}}\cos\left(\frac{2\pi}{N_t}\left(n_t + \frac{N_t}{4} + \frac{1}{2}\right)\left(b_t + \frac{1}{2}\right)\right),$$
$$b_t = 0,\ldots,\frac{N_t}{2}-1 \text{ and } n_t = 0,\ldots,N_t-1. \quad (7.14)$$

where $w_x[n_x]$ and $w_t[n_t]$ are the window functions in space and time, satisfying the conditions $w[n] = w[N-1-n]$ and $w^2[n] + w^2[n+\frac{N}{2}] = 1$ [77]. Note that the basis functions generate only half the number of coefficients, and hence result in a subsampled transform. This requires that we change the range of $\mathbf{b}$ in (7.8) and (7.9) from $\mathbf{b} = \mathbf{0},\ldots,\mathbf{N1}-\mathbf{1}$ to $\mathbf{b} = \mathbf{0},\ldots,\frac{\mathbf{N}}{2}\mathbf{1}-\mathbf{1}$. The TDAC principle guarantees that the aliasing component of a given block—caused by the subsampled transform—is canceled by the aliasing components of the adjacent blocks [76; 62].

An example of the spatio-temporal MDCT is shown in Figure 7.4-(d).

7.4 Realization of directional transforms

As discussed in the previous chapter, the directional Fourier transform can be obtained from the regular spatio-temporal Fourier transform by partitioning the spectrum into a discrete number of "angular subbands", such that each subband captures the far-field components arriving to the region of observation with a given range of angles. To obtain a discrete version

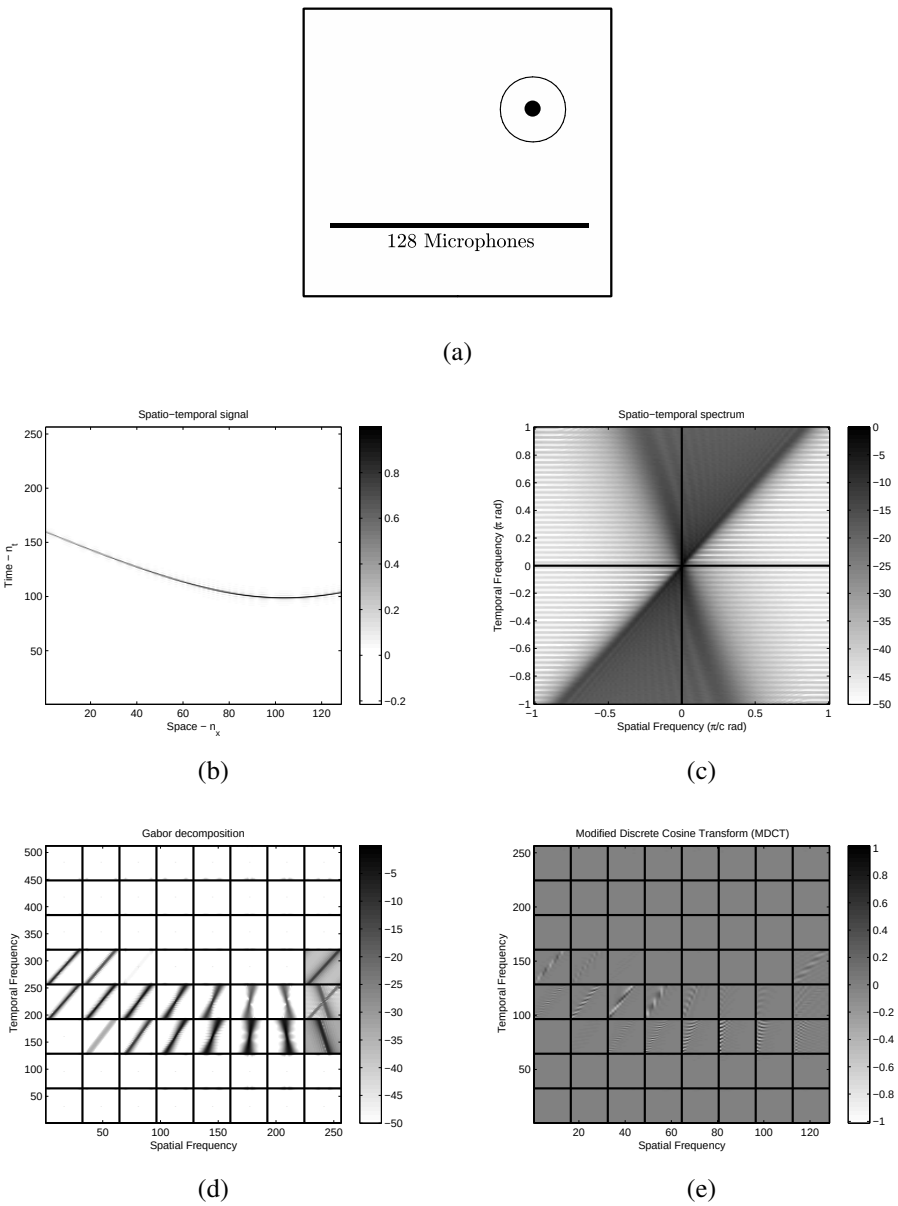


Figure 7.4: Example of lapped orthogonal transforms. (a) The acoustic scene consists of an intermediate-field source driven by a Dirac pulse and a microphone array of size 128. (b) Again, the spatio-temporal signal shows the curvature of the wave front in the region where the source is closer to the array. (c) The DFT of the signal is a triangular pattern as predicted by the theory, but with no local spatial information. (d) The short space/time Fourier transform (or Gabor decomposition) splits up the Fourier transform into blocks of size 32×64, resulting in a representation that discriminates the local curvature of the wave front. Notice how the blocks on the left have more far-field characteristics than the blocks on the right. (e) The modified discrete cosine transform (MDCT) is similar to the result in (d) except that, besides containing only real-valued coefficients at positive frequencies, it is critically sampled. Note that the number of samples in each axis of (e) is half the number of samples seen in the oversampled transform in (d).

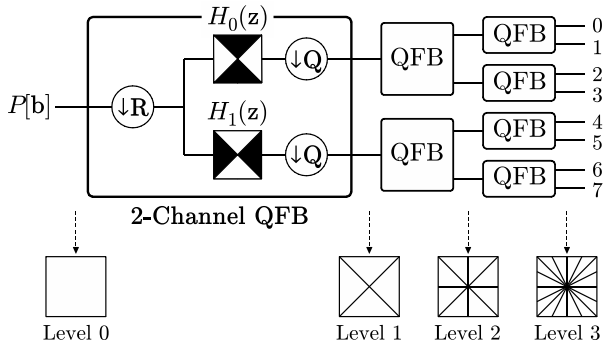


Figure 7.5: Three-level directional filter bank (DFB). This non-separable filter bank consists of an iterated tree-structure of quincunx filter banks (QFB), in this case performing a uniform decomposition of the input spectra into directional subbands. The matrices $\mathbf{R}$ and $\mathbf{Q}$ represent a parallelogram resampler followed by a quincunx resampler, and the filters $H_0(\mathbf{z})$ and $H_1(\mathbf{z})$ represent two diamond-shaped half-band filters.

of the directional Fourier transform, one can think of a filter bank that decomposes the spectrum into directional subbands defined in the range $\cos\big(\alpha + \frac{\Delta}{2}\big)\frac{\Omega}{c} \leq \Phi_x \leq \cos\big(\alpha - \frac{\Delta}{2}\big)\frac{\Omega}{c}$, for $\Omega \geq 0$ (and point-symmetric for $\Omega < 0$), where α is the central direction and Δ is the directional bandwidth. Such a filter bank is necessarily non-separable, since the subband partitioning of the spectrum does not correspond to a rectangular tiling.

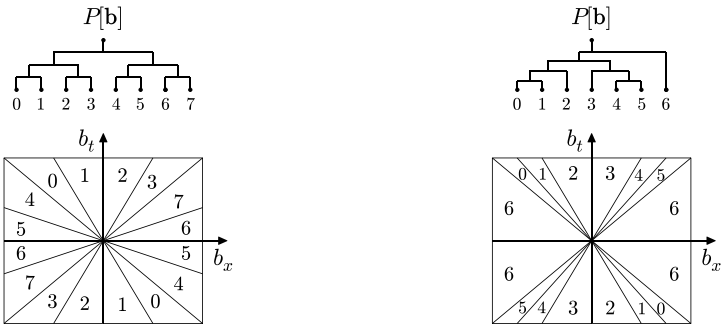
A directional filter bank (DFB) capable of achieving the desired subband partitioning can be obtained through an iterated tree-structure of 2-channel quincunx filter banks (QFB), illustrated in Figure 7.5. The QFB is a non-separable, critically sampled, perfect reconstruction filter bank defined by two diamond-shaped half-band filters, $H_0(\mathbf{z})$ and $H_1(\mathbf{z})$, preceded by a parallelogram resampler $\mathbf{R}$ and followed by a quincunx resampler $\mathbf{Q}$ [4; 70; 26; 19]. The number of levels of the DFB and the configuration of its tree branches determine the angular resolution of the filter bank, as well as the angular selectivity between 0 and π. Some examples of subband partitionings are illustrated in Figure 7.6.

There are seven variants of $\mathbf{R}$ and $\mathbf{Q}$, given by

$$\begin{aligned} \mathbf{Q}_0 &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{Q}_1 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \mathbf{R}_0 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ \mathbf{R}_1 &= \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \mathbf{R}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathbf{R}_3 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \end{aligned} \tag{7.15}$$

plus the identity matrix $\mathbf{R} = \mathbf{I}$, which have resampling densities given by $|\det(\mathbf{R})| = 1$ and $|\det(\mathbf{Q})| = 2$. The selection among these matrices depends on the tree level and the branch considered [26; 25]. The output of both channels is maximally decimated due to a global resampling density of $|\det(\mathbf{RQ})| = 2$.

The intuition behind the iterated QFB structure is that, instead of using different filters to obtain different subbands, we use the resampling matrices to rotate and skew the subbands before slicing them with a fixed filter. The design of the half-band filters $H_0(\mathbf{z})$ and $H_1(\mathbf{z})$, and the respective synthesis filters $F_0(\mathbf{z})$ and $F_1(\mathbf{z})$, can be done using two-dimensional filter



(a) (b)

Figure 7.6: Decomposition of the spatio-temporal spectrum into directional subbands using a tree-structured directional filter bank (DFB). The subband partitioning can be either: (a) uniform or (b) non-uniform. A partitioning similar to (b) is particularly suitable for spectral patterns generated by acoustic wave fields, since it gathers all the evanescent waves into a single subband and has variable directional resolution in the propagating-waves region.

design techniques. Notably, an efficient structure by Phoong *et al* [70] reduces the procedure to the design of an FIR all-pass filter $A(z) = z^{-\frac{1}{2}(2K-1)}$ of order K, such that

$$H_0(\mathbf{z}) = \frac{1}{2} z_x^{-2K} + \frac{1}{2} z_x^{-1} A(z_x z_t^{-1}) A(z_x z_t) \tag{7.16a}$$

$$H_1(\mathbf{z}) = -A(z_x z_t^{-1}) A(z_x z_t) H_0(\mathbf{z}) + z_x^{-4K+1} \tag{7.16b}$$

$$F_0(\mathbf{z}) = -H_1(-\mathbf{z}) \tag{7.16c}$$

$$F_1(\mathbf{z}) = H_0(-\mathbf{z}), \tag{7.16d}$$

where $\mathbf{z} = (z_x, z_t)$ represents the z-transform variables. The design of the all-pass filter $A(z)$ controls the frequency selectivity of the half-band filters, which has an influence in the final amount of cross-band artifacts.

An example of the DFB decomposition of an input wave field is shown in Figure 7.7.

7.5 Matlab examples

The figures shown in the previous sections were generated using a graphical user interface (GUI) programmed in Matlab v7.3. The development of this GUI was inspired by the concept of reproducible research [85], which states that the processes used for obtaining the results should be made available for verification, and the results should be reproducible by other researchers. The package is available in the repository of the Ecole Polytechnique Fédérale de Lausanne at `http://rr.epfl.ch/31/`. A snapshot of the GUI is shown in Figure 7.8.

The GUI is essentially a tool for simulating, analyzing, and processing acoustic wave fields using the theory of spatio-temporal Fourier analysis. The user can customize an acous-

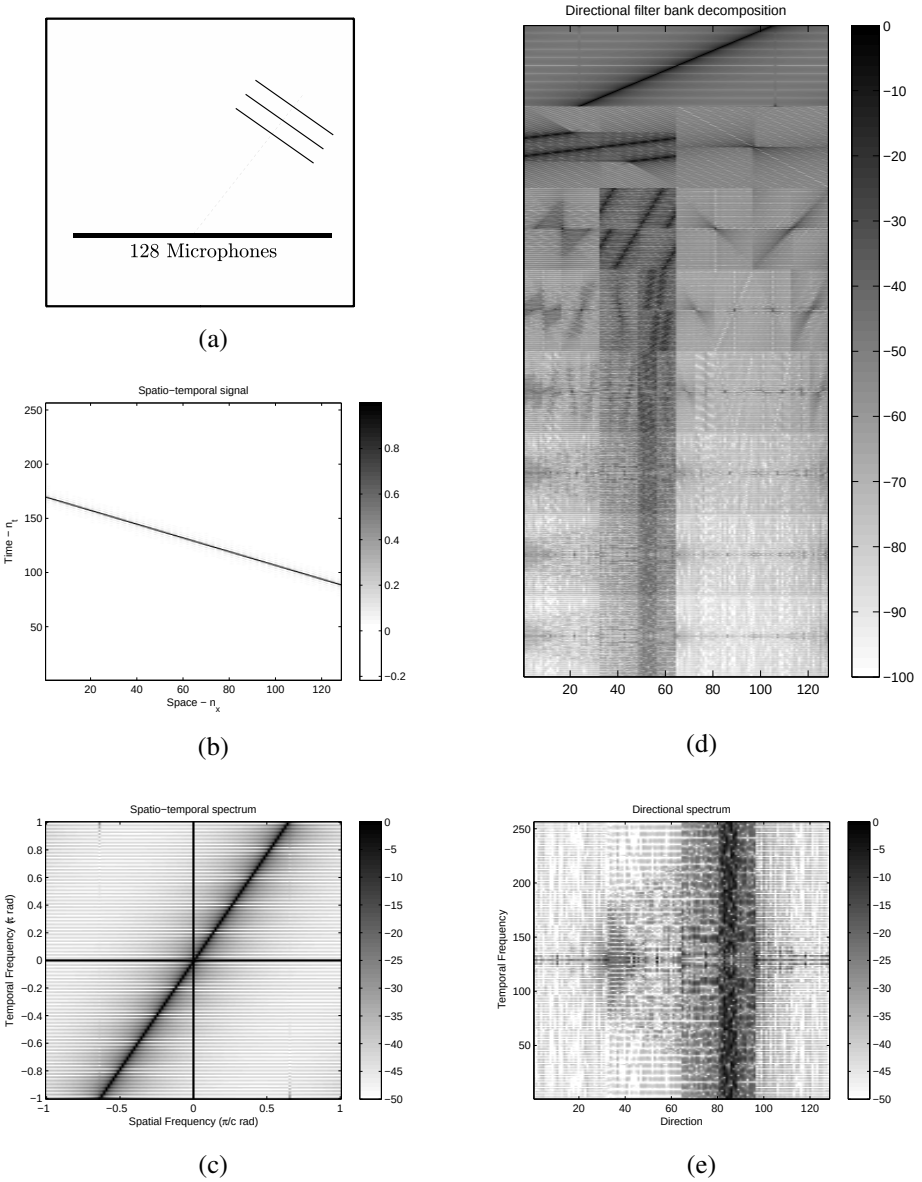


Figure 7.7: Example of a discrete directional transform. (a) The acoustic scene consists of a far-field source driven by a Dirac pulse and a microphone array of size 128. (b) The spatio-temporal signal shows that the wave front is completely plane. (c) The DFT of the signal shows that the spectral energy is concentrated on a single diagonal line crossing the origin, as predicted by the theory. (d) The directional filter bank (DFB) performs an iterated 2-channel quincunx decomposition until the number of samples of each subband in the directional axis is equal to 1. (e) The output of the DFB, after a proper rearrangement of the subbands, is the discrete directional transform. The result shows the spectral energy concentrated on a vertical line crossing the direction axis at the corresponding angle, as predicted by the theory. Note, however, that the directional transform is less sharp than the regular Fourier transform. This is a result of cross-band artifacts caused by the finite resolution of the half-band filters.

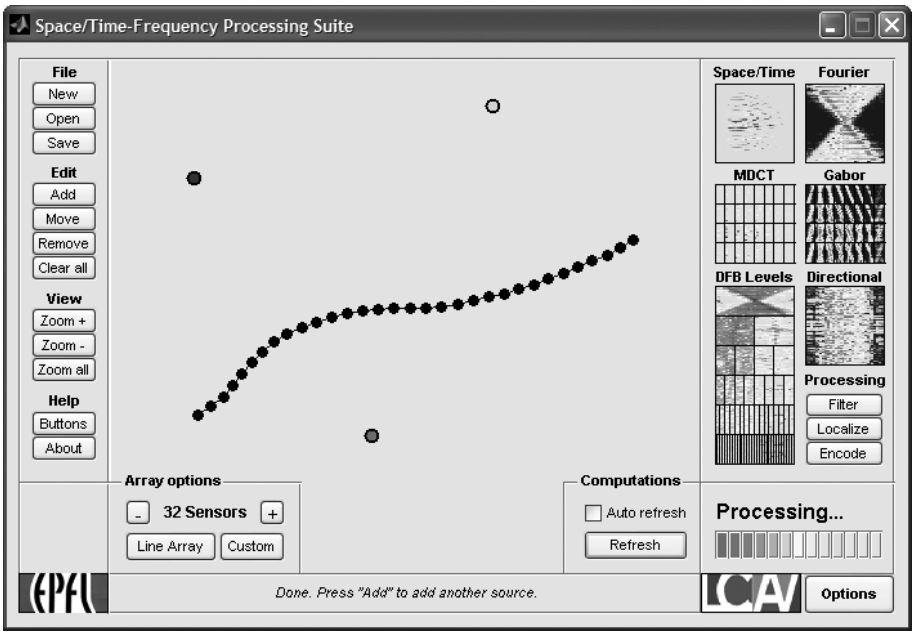


Figure 7.8: Snapshot of the Matlab graphical user interface (GUI).

tic scene with point sources and a sensor array, and observe the resulting wave field representations. Simple DSP operations can also be performed on the wave field, such as spatial filtering, localization, and wave field coding. These applications are discussed in the next two chapters.

Some examples of filter bank decompositions obtained using the GUI of Figure 7.8 are shown in Figure 7.9 and 7.10.

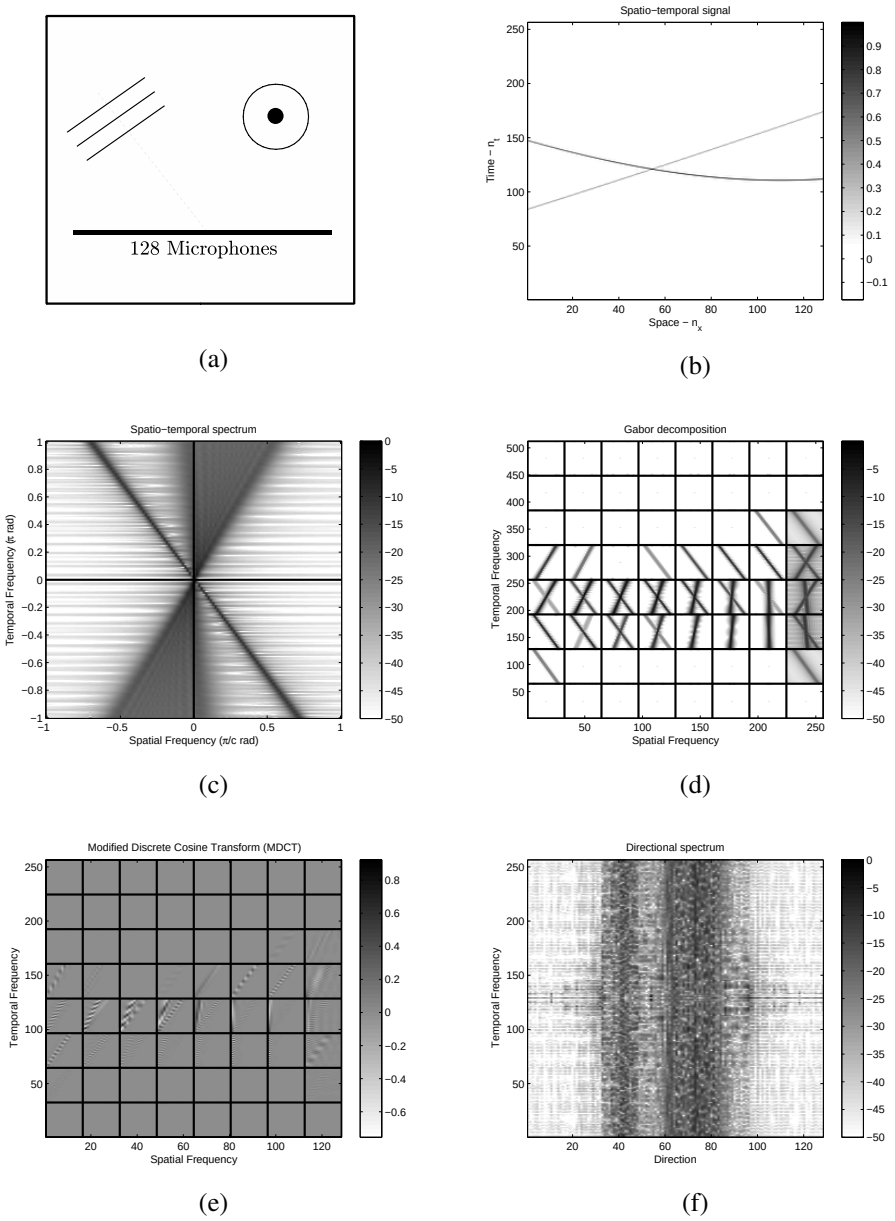


Figure 7.9: Far-field and intermediate-field sources driven by a Dirac pulse, observed on a linear microphone array. (a) Acoustic scene; (b) spatio-temporal signal; (c) DFT; (d) short space/time Fourier transform; (e) spatio-temporal MDCT; (f) discrete directional transform.

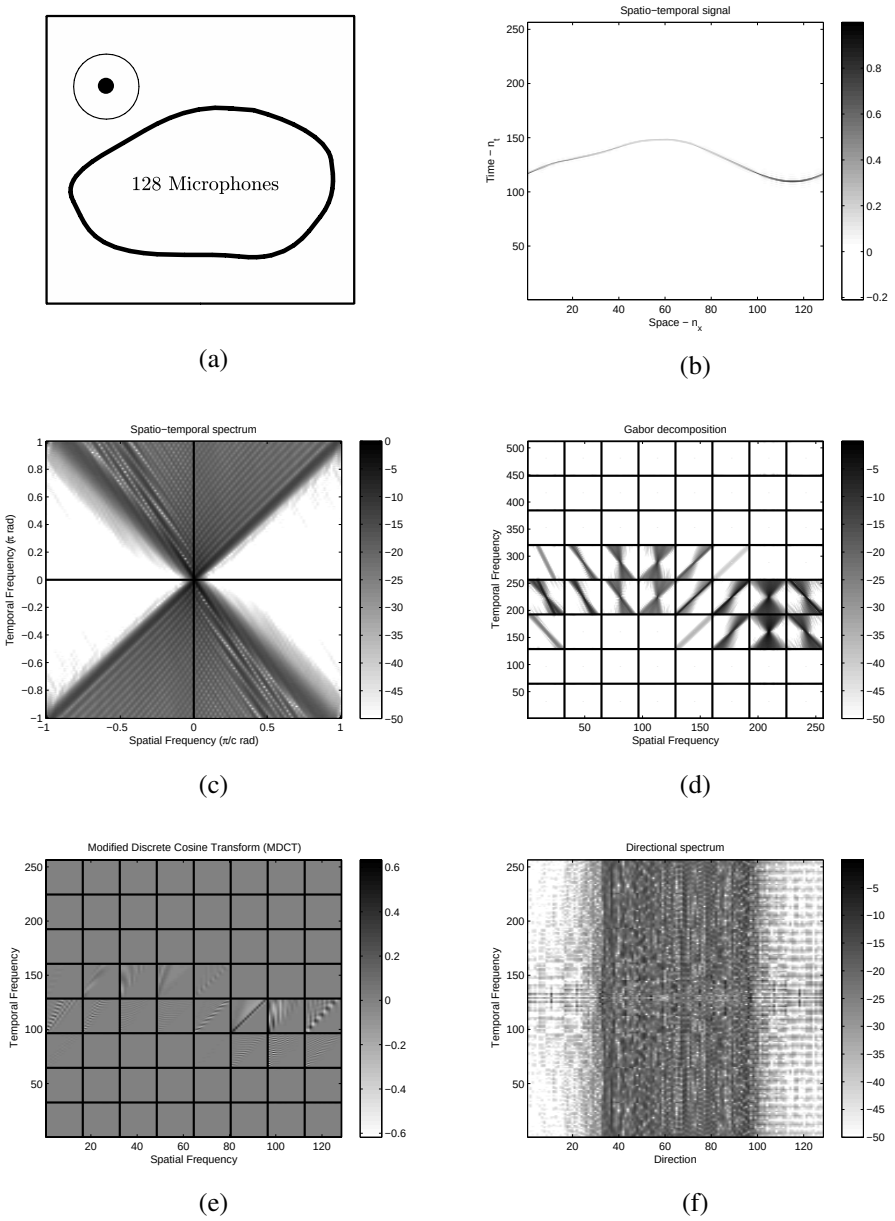


Figure 7.10: Intermediate-field source driven by a Dirac pulse and observed on a curved microphone array. (a) Acoustic scene; (b) spatio-temporal signal; (c) DFT; (d) short space/time Fourier transform; (e) spatio-temporal MDCT; (f) discrete directional transform.

7.6 Summary notes

- Filter banks are efficient discrete-domain structures that can be used for implementing the discrete version of orthogonal transforms;
- Spatio-temporal transforms can be implemented through the use of multidimensional filter banks theory;
- The realization of the discrete Fourier and cosine transforms can be obtained with a separable, critically sampled, and perfect reconstruction filter bank that performs a uniform partitioning of the spectrum;
- The realization of lapped orthogonal transforms such as the short space/time Fourier transform can be obtained with the same filter bank with a lower resampling ratio (oversampled). In the case of the modified discrete cosine transform, the filter bank is critically sampled;
- The discrete version of the directional Fourier transform requires the use of non-separable filter banks, and can be obtained with an iterated tree-structure of quincunx filter banks;
- A Matlab graphical user interface for the purpose of reproducing the results in this chapter is available in our repository at `http://rr.epfl.ch/31/`.

Chapter 8

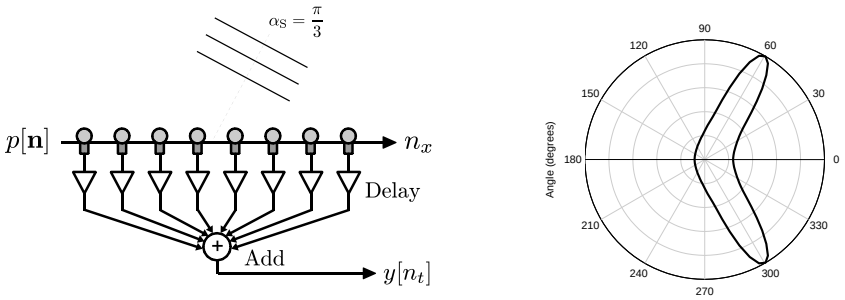
Spatio-Temporal Filter Design

8.1 Introduction

In digital signal processing (DSP), filtering is the cornerstone operation when it comes to manipulating signals, images, and other types of data. It is arguably the most used DSP technique in modern technology and electronic devices, as well as in the domain of Fourier analysis in general. The outstanding variety of applications of filtering go beyond the simple elimination of undesired frequencies in signals: it allows, for example, the elimination of several types of interferences, the cancelation of echoes in two-way communications, and the frequency multiplexing of radio signals. Moreover, the theory led to the invention of filter banks, and hence the development of new types of linear transforms and signal representations.

A linear filtering operation is defined as a convolution between the input signal $s[n_t]$ and the impulse response of the filter $h[n_t]$, given by $y[n_t] = \sum_{k\in\mathbb{Z}} s[k]h[k - n_t]$. The filter can have a finite impulse response (FIR), in which case it is non-recursive, always stable, and can easily be designed to have linear phase, or it can have an infinite impulse response (IIR), in which case it is recursive, can be either stable or unstable, and the linear phase property has to be approximated for a given range of frequencies. IIR filters typically require less computational power than FIR filters. Alternatively, the filtering operation can be performed in the Fourier domain. If $S(\omega)$ and $H(\omega)$ are the Fourier transforms of $s[n_t]$ and $h[n_t]$, where ω is the temporal frequency in rad, then, using the convolution property, $Y(\omega) = S(\omega)H(\omega)$.

In the field of study known as array signal processing (see, *e.g.*, Johnson *et al.* [54]), there exists a similar concept called spatial filtering (or beamforming). A spatial filter is a filter that favors a given range of directions in space, implemented directly through the array of sensors. The sensors are synchronized such that there is phase alignment for a desired angle of arrival and phase opposition for the other angles. Spatial filters have been used in many contexts throughout history with enormous success—most notably during warfare with the use of radars, and during the era of wireless communications with the use of antennas (see, *e.g.*, Brown *et al.* [17] for a history of radar development during world war II, and Sarkar *et al.* [81] for a history of wireless communications). Other applications include sonar, seismic wave monitoring, spatial audio, noise cancelation, and hearing aids technology. As long as more than one sensor is available, it is always possible to implement a spatial filter. The human auditory system, for example, uses an array of two sensors (the ears) to localize the sound sources in space.



(a) (b)

Figure 8.1: Typical implementation of a delay-and-add beamformer. (a) A linear array of 8 microphones, separated by the Nyquist distance $T_x = 340/44100\,\mathrm{m}$, captures the sound from the upper half-space. The signals received in each microphone at n_x are delayed by an appropriate factor, such that the signals are phase-aligned with respect to the steering angle α_S (in this case $\alpha_S = \frac{\pi}{3}$), and then added up to generate the output signal $y[n_t]$. (b) The directivity plot shows the normalized gain of the beamformer with respect to the arrival angle of Dirac sources in the far-field. The gain is clearly higher for the selected steering angle, and decreases as the angle moves away from α_S. The symmetry of the directivity plot is due to front-back ambiguity.

The problem with most beamforming techniques (*e.g.*, Frost [34], Widrow *et al.* [87] and Griffiths *et al.* [40]) is that they provide little flexibility in terms of the design of spatial filters. The basic technique consists of assigning different gains and delays to each sensor, and adding the results in order to obtain a higher acquisition gain (or reproduction gain) for a specific direction of arrival, as illustrated in Figure 8.1. This means that the spatial filtering operation is not performed on the signal samples directly but rather on the temporal DFT coefficients of each sensor, which is hardly an intuitive interpretation of convolutional filtering. One of the consequences is that we can not easily design the beamformer such that it has a flat response for a given range of directions, while rejecting all the others. The angular response of the beamformer is typically dependent of the spectral characteristics of the spatial window that regulates the gain of each sensor. We will show in this chapter that, in the spatio-temporal Fourier domain, the delay-and-add technique implemented by beamformers is equivalent to a 2-D sampling operation.

One of the greatest attributes of the Fourier transform is that it allows the interpretation of convolutional filtering in terms of intuitive parameters such as the cut-off frequencies, stop-band attenuation, and phase response. For instance, a filter can be sketched in the Fourier domain such that it has a unitary response for a given range of frequencies (pass-band) and a high attenuation for the remaining frequencies (stop-bands), plus an equiripple magnitude response and a linear phase. Using existing algorithms [68], the ideal filter can be translated into a realizable filter that optimally obtains the desired response. In the context of spatio-temporal Fourier analysis, the same reasoning can be used: we can sketch a spatial filter in the Fourier domain such that it has a unitary response for every plane wave within a given range of directions (pass-band) and a high attenuation for the remaining plane waves (stop-bands),

plus any additional magnitude and phase constraints. The ideal filter can be translated into a realizable filter by using two-dimensional filter design techniques. Once the spatio-temporal filter coefficients $h[\mathbf{n}]$ are obtained, the filtering operation can be performed either in the spatio-temporal domain, using the convolution formula

$$y[\mathbf{n}] = \sum_{\rho\in\mathbb{Z}}\sum_{\tau\in\mathbb{Z}} p[\rho,\tau]h[n_x-\rho, n_t-\tau],$$

or in the Fourier domain, using the convolution property $Y(\phi,\omega) = P(\phi,\omega)H(\phi,\omega)$, where $H(\phi,\omega)$ is the Fourier transform of the spatio-temporal filter $h[\mathbf{n}]$ and ϕ is the spatial frequency in rad.

The purpose of this chapter is to show how the ideal spatio-temporal filter can be designed as a function of the most relevant parameters, and how it can be translated into a realizable and well behaved filter with arbitrary order. We also show how the filter can be applied to a curved observation contour through the use of space/time–frequency analysis. Finally, we provide a brief study on adaptive filtering in the spatio-temporal Fourier domain.

8.2 Spatial filtering in the Fourier domain

The spatio-temporal Fourier transform is more than an efficient representation of acoustic wave fields: it is tool that generalizes several concepts of acoustics theory and array signal processing into a single framework. Concepts such as plane waves, near-field and far-field radiation, propagating and evanescent modes, spatial sampling, spatial windowing, *etc.*, all come together in the spatio-temporal Fourier domain. In this section, we show that beamforming techniques are no exception, and can be easily translated into operations in the spatio-temporal Fourier domain. More importantly, we show that beamforming is a very limiting type of spatial filtering, which can be characterized in a more generally way through 2-D convolutional filtering.

8.2.1 Beamforming

A standard beamforming algorithm operates by delaying the multiple signals obtained by each microphone in the array such that they are in phase alignment with respect to a given angle of arrival, as illustrated in Figure 8.1. This operation can be expressed as

$$y[b_t] = \sum_{n_x=0}^{N_x-1} e^{-j\frac{2\pi}{N_t}b_t\cos\alpha_{\mathrm{S}}\frac{n_xT_x}{c}} p[n_x, b_t] \tag{8.1}$$

where $p[n_x, b_t]$ is the temporal DFT of $p[\mathbf{n}]$, given by $p[n_x, b_t] = \sum_{n_t=0}^{N_t-1} p[n_x, n_t]e^{-j2\pi\frac{b_tn_t}{N_t}}$, T_x is the sampling period in space, and α_{S} is the beam-steering angle (*i.e.*, the center of the angular filter). Note that the output of the beamformer is a one-dimensional signal $y[b_t]$ with no spatial information.

The expression in (8.1) can be rewritten as

$$y[b_t] = \sum_{n_x=0}^{N_x-1} e^{-j\frac{2\pi}{N_x}\left(\cos\alpha_{\mathrm{S}}\frac{b_t}{c}\frac{T_xN_x}{N_t}\right)n_x} p[n_x, b_t]. \tag{8.2}$$

By identification, the result is equivalent to the spatial DFT of $p[n_x, b_t]$, where b_x is given by $b_x = \cos\alpha_{\mathrm{S}} \frac{T_x}{c} \frac{N_x}{N_t} b_t$. Hence,

$$y[b_t] = P\left[\cos\alpha_{\mathrm{S}} \frac{b_t}{c} \left(\frac{T_x N_x}{N_t}\right), b_t\right]. \tag{8.3}$$

Finally, using the discrete Dirac formalism,

$$\boxed{y[b_t] = \sum_{b_x=0}^{N_x-1} P[\mathbf{b}]\delta\left[b_x - \cos\alpha_{\mathrm{S}} \frac{b_t}{c} \left(\frac{T_x N_x}{N_t}\right)\right],} \tag{8.4}$$

where $\delta[b]$ is a Kronecker delta, defined as $\delta[b] = 1$ for $b = 0$ and $\delta[b] = 0$ otherwise.

The result in (8.4) essentially states that beamforming algorithms based on delay-and-add can be translated into a sampling operation in the spatio-temporal DFT domain, where the sampling kernel is given by the Kronecker delta $\delta\big[b_x - \cos\alpha_{\mathrm{S}} \frac{b_t}{c} \big(\frac{T_x N_x}{N_t}\big)\big]$. This operation is illustrated in Figure 8.2.

8.2.2 Convolutional filtering

A more general approach compared to sampling the Fourier spectrum along a diagonal line (beamforming) is to apply a properly designed 2-D filter in order to eliminate undesired plane waves. The pass-band region of the filter can be specified such that the plane waves within a given range of temporal frequencies, spatial frequencies, arrival angles, or any combination thereof, are preserved. The bandwidth of such a filter depends on which parameter is being measured—*e.g.*, we can measure the temporal frequency bandwidth, the spatial frequency bandwidth, or the angular bandwidth. Other important parameters include the cut-off and center frequencies (or angles), the ripple behavior of the pass-band region, the attenuation of the stop-band regions, and the phase response of the filter. An optimal and realizable filter satisfying the specifications of the ideal filter can be obtained through the use of 2-D filter design techniques, that we discuss later in the chapter. Some examples of spatio-temporal filters are shown in Figure 8.3.

A spatio-temporal filter is a 2-D discrete sequence defined as $h[\mathbf{n}] = h[n_x, n_t]$ with DFT coefficients $H[\mathbf{b}] = H[b_x, b_t]$. If the input signal is $p[\mathbf{n}]$, then that spatio-temporal filtering operation is given by a 2-D circular convolution [68],

$$y[\mathbf{n}] = \sum_{\rho=0}^{N_x-1} \sum_{\tau=0}^{N_t-1} p[\rho, \tau] h[((n_x - \rho))_{N_x}, ((n_t - \tau))_{N_t}], \tag{8.5}$$

where $((\cdot))_N = \cdot \bmod N$. Unlike beamforming, the output of (8.5) is the 2-D signal $y[\mathbf{n}]$ representing the entire wave field (*i.e.*, with spatial information).

Using the convolution property of the DFT, it follows that

$$\boxed{Y[\mathbf{b}] = P[\mathbf{b}]H[\mathbf{b}].} \tag{8.6}$$

To specify the parameters of the ideal filter, we first need to decide what is the purpose of the filter. A reasonable goal is to focus on the wave fronts originating from a particular point in space—perhaps the location of a target source—while suppressing every other wave front with a different origin. For this purpose, the expression of the spatio-temporal spectrum of

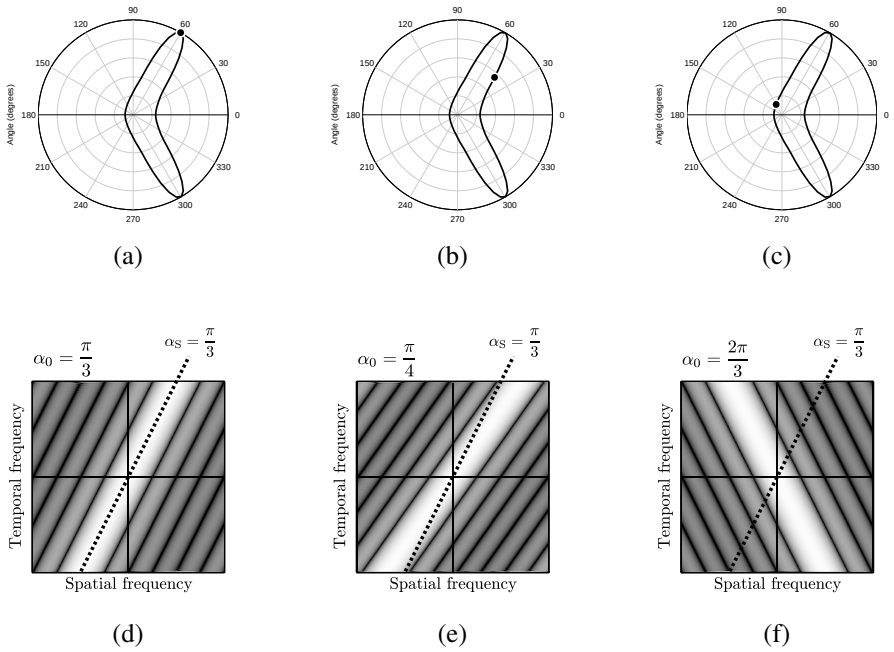


Figure 8.2: Delay-and-add beamformer as a 2-D sampling operation. In (a), (b), and (c), the directivity plot shows the normalized gain of the beamformer of Figure 8.1 with respect to the arrival angle of Dirac sources in the far-field. A black circle indicates the arrival angle of an actual Dirac source present at the scene, and the respective normalized gain. In (d), (e), and (f), the same beamforming operations are performed in the spatio-temporal Fourier domain. The spatio-temporal Fourier transform of the Dirac wave front on a linear array with 8 microphones is a sinc support function oriented towards the angle of arrival α_0. The dashed line represents the sampling kernel that directly obtains the output spectrum $y[b_t]$, and therefore the output signal $y[n_t]$. The slope of the line is determined by the steering angle $\alpha_S = \frac{\pi}{3}$.

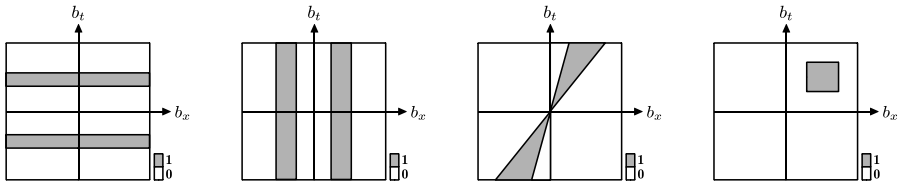


Figure 8.3: Examples of spatio-temporal filters. (a) A temporal filter $h[n_t]$ with symmetric 1-D Fourier transform. (b) A spatial filter $h[n_x]$ with symmetric 1-D Fourier transform. (c) A non-separable directional filter $h[\mathbf{n}] = h[n_x, n_t]$ with point-symmetric 2-D Fourier transform. (d) A separable spatio-temporal filter $h[\mathbf{n}] = h_x[n_x]h_t[n_t]$ with non-symmetric 2-D Fourier transform. The filters in (a), (b), and (c) have real-valued coefficients due to their symmetry, as opposed to the non-symmetric filter in (d), which has complex-valued coefficients.

an intermediate-field source can be used to specify the parameters of the ideal filter. Recall the expression from Chapter 5 in *continuous space and time*, given by

$$P(\Phi_x, \Omega) = S(\Omega) \max \left\{ W_x \left(\Phi_x - \cos\alpha \frac{\Omega}{c} \right), M(\Phi_x, \Omega) \right\}, \tag{8.7}$$

where $M(\Phi_x, \Omega)$ is a triangular mask given by

$$M(\Phi_x, \Omega) = \begin{cases} W_x(0) & , (\Phi_x, \Omega) \notin \mathcal{U} \\ 0 & , (\Phi_x, \Omega) \in \mathcal{U}, \end{cases} \tag{8.8}$$

with $\mathcal{U} = \mathbb{R}^2 \backslash \left\{ (\Phi_x, \Omega) : \Phi_x^{\min} \leq \Phi_x \leq \Phi_x^{\max}, \, \Omega \geq 0 \right\}$, and point-symmetric for $\Omega < 0$. The parameters α, $\Phi_x^{\min}$, and $\Phi_x^{\max}$ define the orientation and aperture of the triangular mask, and are given by $\cos\alpha = \mathbb{E}_x \left[\cos\alpha_{\text{nf}}(x) \right]$, $\Phi_x^{\min} = \cos\alpha_{\text{nf}}^{\max} \frac{\Omega}{c}$, and $\Phi_x^{\max} = \cos\alpha_{\text{nf}}^{\min} \frac{\Omega}{c}$.

According to (8.7) and (8.8), the maximum concentration of energy is contained within the region defined by the triangular mask (*i.e.*, for $(\Phi_x, \Omega) \notin \mathcal{U}$). Thus, the ideal filter can be defined in *discrete space and time* as

$$\boxed{H[\mathbf{b}] = \begin{cases} 1 & , \mathbf{b} \notin \mathcal{U} \\ 0 & , \mathbf{b} \in \mathcal{U}, \end{cases}} \tag{8.9}$$

where $\mathcal{U} = \mathbb{Z}^2 \backslash \left\{ \mathbf{b} : b_x^{\min} \leq b_x \leq b_x^{\max}, \, b_t \geq 0 \right\}$, and point-symmetric for $b_t < 0$. The parameters $b_x^{\min}$ and $b_x^{\max}$ are the discrete counterparts of $\Phi_x^{\min}$ and $\Phi_x^{\max}$, and are given by $b_x^{\min} = \cos\alpha_{\text{nf}}^{\max} \frac{b_t}{c} \left(\frac{T_x N_x}{N_t} \right)$ and $b_x^{\max} = \cos\alpha_{\text{nf}}^{\min} \frac{b_t}{c} \left(\frac{T_x N_x}{N_t} \right)$. The relation between the focus point and the filter specifications is illustrated in Figure 8.4.

The Matlab graphical user interface (GUI) used for obtaining the results in the previous chapter—available for download at the EPFL Reproducible Research repository [49]—contains a tool for filtering sources in a simulated acoustic scene. The filtering operation can be easily performed by selecting the type of filter (band-pass or band-stop) and clicking on

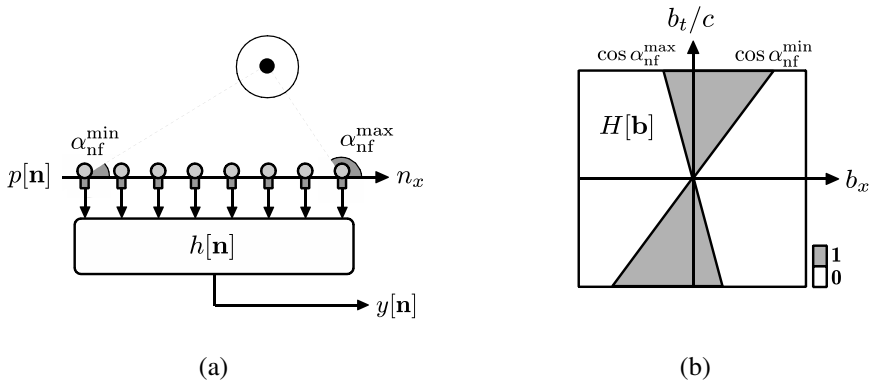


Figure 8.4: Filter specifications for a given target source. (a) The spatio-temporal filtering operation consists of a 2-D convolution between the input signal $p[\mathbf{n}]$ and the filter $h[\mathbf{n}]$, generating the output signal $y[\mathbf{n}]$. The target source is positioned in the intermediate-field such that the minimum and maximum angles of incidence are given by $\alpha_{\mathrm{nf}}^{\min}$ and $\alpha_{\mathrm{nf}}^{\max}$. (b) The ideal filter that retains most of the spectral content belonging to the source has a triangular shape that matches the shape of the spectral mask given by (8.8). Since the position of the source is completely defined by $\alpha_{\mathrm{nf}}^{\min}$ and $\alpha_{\mathrm{nf}}^{\max}$, as well as the respective spectral mask, the filter that targets the source is also completely defined by $\alpha_{\mathrm{nf}}^{\min}$ and $\alpha_{\mathrm{nf}}^{\max}$.

the target source. The algorithm takes care of designing the best filter for the selected source. A few simulation examples are shown at the end of this chapter.

8.2.3 Filtering in the short space/time Fourier domain

One of the limitations of filtering sources in the spatio-temporal Fourier domain is related to something we call the *shadow region* of the point source. To visualize this region, imagine that a beam of light is emitted from each point in the array axis, and directed towards the source. If the beam of light dies upon hitting the source, this causes the array to cast a shadow over an infinite triangular region behind the source. This region becomes larger the closer the source is to the array. As a consequence, any other source that falls within the shadow region can not be separated by applying a spatial filter (we will see that other techniques may be used instead), since their spectral content will be mixed with the spectral content of the source casting the shadow. This phenomenon is illustrated in Figure 8.5-(a).

In a less extreme scenario, the second source is not located in the shadow region of the first source, but their two individual shadows intersect at some point in space. In this case, a portion of their respective spectra will necessarily intersect in the Fourier domain, while the other portion will not. The parts of the spectrum that do not intersect can be extracted through spatial filtering, although the filtering operation only preserves a limited amount of spatial information related to the sources. This scenario is illustrated in Figure 8.5-(b).

An easy solution for reducing the influence of the shadow region on the spatial filtering operation is to use local Fourier analysis. Recall from Chapter 5 that the aperture of the spectral mask $M(\Phi_x, \Omega)$ given by (8.8) can be reduced either by moving the source away from

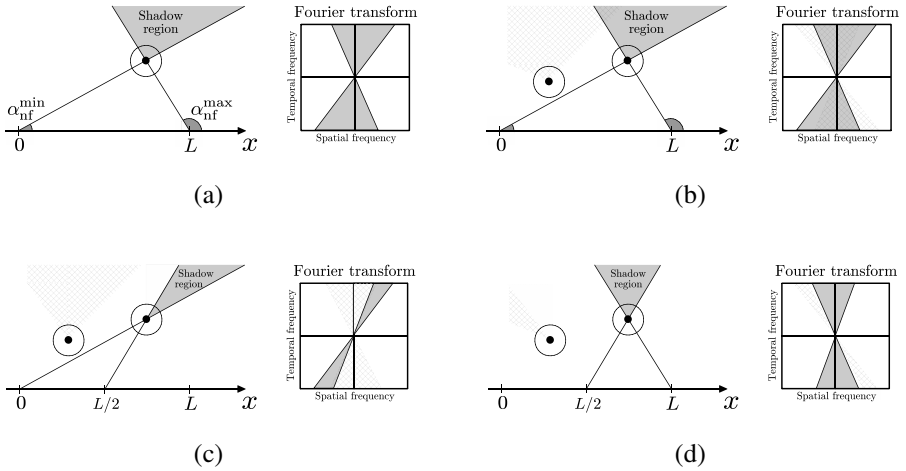


Figure 8.5: Shadow region of an intermediate-field source. (a) The array axis "casts a shadow" behind the source with a triangular shape defined by the minimum and maximum angles of incidence $\alpha_{\rm nf}^{\rm min}$ and $\alpha_{\rm nf}^{\rm max}$. Any other source that falls within the shadow region will have all its spectral content inside the shaded region of the Fourier transform, and therefore can not be recovered by a spatial filter alone. (b) If the second source is outside the shadow region of the first source, there are two possibilities: (i) the two shadows intersect at some point in space (as shown in the figure), in which case only a portion of their respective spectra will intersect; (ii) the two shadows never intersect, and consequently the "spectral triangles" never intersect in the Fourier domain (although the side lobes of $W_x(\Phi_x)$ do). The spectrum of the second source is represented by the hatched region in the figure. In (c) and (d), the array axis is divided into two blocks of equal size. This reduces the aperture of the shadow regions, and hence the aperture of the "spectral triangles". The axis can be divided into an arbitrary number of blocks, to the point where little or no intersection occurs between shadows.

the array axis or by reducing the spatial window length. By applying the short space/time Fourier transform to the input signal $p[\mathbf{n}]$, the window length can be arbitrarily reduced until the shadow regions of each source are not intersecting anymore, as illustrated in Figure 8.5-(c) and (d). On the downside, reducing the window length also results in a degradation of the spatial frequency resolution, due to the widening of the spectral support function $W_x(\Phi_x)$—the left argument of the $\max$ operation in (8.7).

Another situation where it is convenient to filter $p[\mathbf{n}]$ in the short space/time Fourier domain is when the array axis is curved. Going back to Figure 5.8, we see that the increased curvature of the axis causes the spectral mask to open up, even when the source is in the far-field. Reducing the length of the spatial window counters this effect, by reducing the aperture of the spectral mask. A few examples are shown in the end of this chapter.

8.3 2-D filter design techniques

The design of digital filters is typically performed in two steps: (i) the specification of the ideal filter, and (ii) the approximation by a realizable filter. The specifications include the cut-off frequencies of the pass-bands and stop-bands, as well as the transition bands, the maximum gain error in the pass-bands and stop-bands, and the order of the filter. Some of these parameters may be interdependent, and consequently the number of design parameters may be actually smaller. The approximation step in the design consists of deriving a realizable filter of arbitrary order that minimizes the error between its own impulse or frequency response and the impulse or frequency response of the ideal filter, with respect to a given error measure. Some examples of filter design techniques include the bilinear transformation method [68], the window method [68], and the widely used Parks-McClellan method [78].

Directional spatio-temporal filters of the type shown in Figure 8.4-(b) can be derived, under certain limitations, through the use of 2-D filter design techniques (see, *e.g.*, Dudgeon *et al.* [27] and Vaidyanathan *et al.* [84]). The main difficulty is that these filters are non-separable, and therefore can not be separately designed in space and time with 1-D filter design techniques. A technique known as the McClellan transformation, however, can be used to map an optimally designed 1-D filter to an equally optimal 2-D filter, using a suitable mapping function [64]. Generalizations of the Parks-McClellan method to the 2-D case also exist [57; 45], although these tend to be computationally expensive.

In this section, we show two simple techniques for designing 2-D FIR filters based on the window method and the Parks-McClellan method. These techniques are easy to implement, although they do not provide optimal results. The optimal spatio-temporal filter can be obtained through the generalized Parks-McClellan method.

8.3.1 Window method

The design of 2-D FIR filters with zero phase is straightforward, and can be obtained directly from the ideal specification of the filter,

$$H_{\text{ideal}}[\mathbf{b}] = \begin{cases} 1 & , \mathbf{b} \notin \mathcal{U} \\ 0 & , \mathbf{b} \in \mathcal{U}, \end{cases} \tag{8.10}$$

where $\mathcal{U}$ is the stop-band region of the spectrum. The impulse response of the ideal filter is infinite, due to the discontinuous transition between the pass and stop bands, and is given by

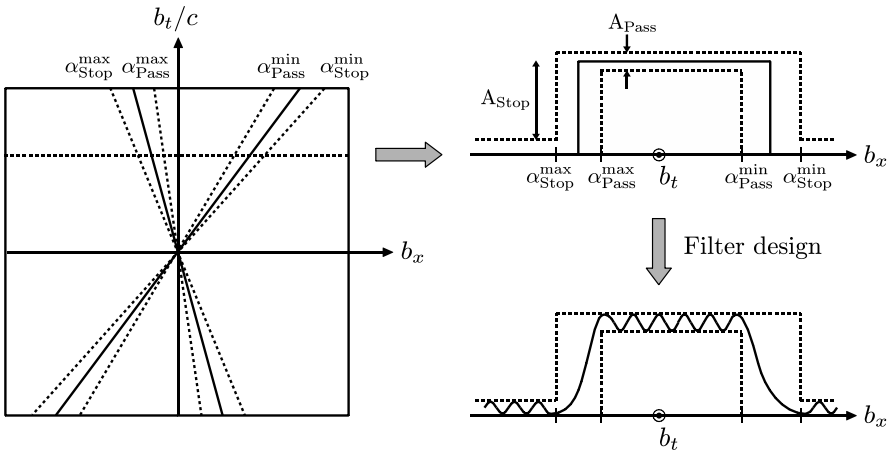


Figure 8.6: Spatio-temporal FIR filter design by the profiling method. The ideal filter is specified according to the desired pass-band and stop-band parameters, such as the cut-off angles $\alpha^{\min}$ and $\alpha^{\max}$ and the limits of the transition bands α_{Pass} and α_{Stop}. Additionally, we may want to specify the maximum approximation error in the pass-band, $\mathrm{A}_{\mathrm{Pass}}$, and the stop-band attenuation, $\mathrm{A}_{\mathrm{Stop}}$. Then, the ideal 2-D filter is profiled horizontally (or vertically) in order to obtain 1-D specifications. For each profile, the optimal 1-D filter is obtained by the Parks-McClellan algorithm, resulting in a filter profile with equiripple response.

$$h_{\mathrm{ideal}}[\mathbf{n}] = \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{b}\in\mathbb{Z}^2} H_{\mathrm{ideal}}[\mathbf{b}] e^{j2\pi \mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}},\ \mathbf{n}\in\mathbb{Z}^2. \tag{8.11}$$

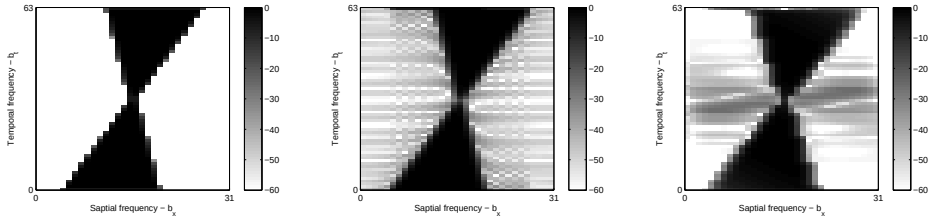
A finite approximation of the ideal filter can be obtained by multiplying $h_{\mathrm{ideal}}[\mathbf{n}]$ by a window function $w[\mathbf{n}]$ of size $L_x \times L_t$, such that

$$H[\mathbf{b}] = \sum_{\mathbf{n}=-\frac{1}{2}\mathbf{L1}}^{\frac{1}{2}\mathbf{L1}-\mathbf{1}} w[\mathbf{n}] h_{\mathrm{ideal}}[\mathbf{n}] e^{-j2\pi \mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}},\ \mathbf{b} = \mathbf{0},\ldots,\mathbf{L1}-\mathbf{1}, \tag{8.12}$$

where $\mathbf{L} = \left[\begin{smallmatrix} L_x & 0 \\ 0 & L_t \end{smallmatrix}\right]$ and $\mathbf{1} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. Using the convolution property of the spatio-temporal Fourier transform, it follows that

$$H[\mathbf{b}] = W[\mathbf{b}] * H_{\mathrm{ideal}}[\mathbf{b}]. \tag{8.13}$$

The disadvantage of using the window method is that the frequency response of the approximated filter is constrained by the Fourier transform of the window function, and therefore does not allow an individual control over the approximation errors in different bands. Nevertheless, the method is fast, general, and capable of approximating complicated filter specifications.



(a) (b) (c)

Figure 8.7: Comparison between filter design methods. (a) The ideal filter is a directional filter with zero phase and size 32×64. (b) The window method (using a Kaiser window) provides a good approximation of the desired magnitude response. (c) The profiling method results in a filter with better stop-band attenuation at higher temporal frequencies, but worse at lower temporal frequencies.

8.3.2 Profiling method

Another method for designing a 2-D FIR filter with zero phase is the profiling method—a somewhat crude approximation of the 2-D Parks-McClellan method. The strategy consists of tracing profiles across the ideal filter $H_{\text{ideal}}[\mathbf{b}]$ defined in (8.10) along horizontal lines of constant b_t and obtaining the optimal 1-D filter $h_{b_t}[b_x]$ for each profile, as illustrated in Figure 8.6. The approximation filter of size $L_x \times L_t$ is then given by

$$H[\mathbf{b}] = h_{b_t}[b_x],\ \mathbf{b} = \mathbf{0}, \ldots, \mathbf{L1} - \mathbf{1}.$$

The filter can alternatively be obtained by tracing vertical lines of constant b_x and obtaining the optimal 1-D filter $h_{b_x}[b_t]$ for each profile, such that

$$H[\mathbf{b}] = h_{b_x}[b_t],\ \mathbf{b} = \mathbf{0}, \ldots, \mathbf{L1} - \mathbf{1}.$$

This method is considerably slower compared to the window method, but allows more control over the various parameters of the ideal filter. A comparison between the two methods is shown in Figure 8.7.

8.4 Adaptive filtering in the Fourier domain

In the previous sections, we saw that when two sources are present at the scene—say, the target and the interferer—and their shadow regions intersect, it is impossible to fully recover the wave front generated by the target source through the use of spatial filtering. It is possible, however, to recover the target with a higher signal-to-noise ratio (SNR) by using adaptive filtering. The idea consists of steering the zeros of the directivity pattern towards the target direction, such that only the interferer noise is captured. This noise is then canceled out in the beamformer output by an adaptive filter, in order to maximize the output SNR. The technique was introduced by Griffiths *et al* [40], and is known as sidelobe canceling beamformer.

In this section, we show that the same operation can be performed in the spatio-temporal Fourier domain, and, similarly to the case of linear filtering, can be generalized in order to

account for sources in the near- and intermediate-field, and preserve the spatial information of the target source (*i.e.*, the entire wave front).

8.4.1 Adaptive beamforming

The result in (8.4) shows that the output signal of the beamformer can be obtained in the Fourier domain by sampling the center of the main lobe generated by a Dirac source in the far-field on the windowed array. The same strategy can be followed in order to obtain the output signal of the beamformer when the zeros of the directivity pattern are steered towards the target source [74]: in the spatio-temporal Fourier domain, zero-steering is equivalent to a 2-D sampling operation.

We have seen in Chapter 5 that a Dirac source in the far-field observed at a linear array of length L_x results in a spectrum with periodic notches at

$$\Phi_x = \cos\alpha\frac{\Omega}{c} + \frac{k2\pi}{L_x},\ (\Phi_x,\Omega)\in\mathbb{R}^2 \text{ and } k\in\mathbb{Z}\backslash 0. \tag{8.14}$$

If the Dirac source is in the near- or intermediate-field, then, according to Theorem 30, the spectral notches are located at

$$\Phi_x = \mathbb{E}_x\big[\cos\alpha_{\rm nf}(x)\big]\frac{\Omega}{c} + \frac{k2\pi}{L_x},\ (\Phi_x,\Omega)\in\mathcal{U} \text{ and } k\in\mathbb{Z}\backslash 0. \tag{8.15}$$

In both cases, the center of the main lobe is obtained by taking $k=0$. By defining the discrete sampling kernels as

$$\Delta_{\rm FF}^{\rm Signal}[\mathbf{b}] = \delta\left[b_x - \cos\alpha_{\rm S}\frac{b_t}{c}\left(\frac{T_xN_x}{N_t}\right)\right] \tag{8.16}$$

$$\Delta_{\rm FF}^{\rm Noise}[\mathbf{b}] = \delta\left[b_x - \left(\cos\alpha_{\rm S}\frac{b_t}{c} + \frac{k}{L_x}\frac{N_t}{c}\right)\left(\frac{T_xN_x}{N_t}\right)\right] \tag{8.17}$$

in the far-field case, and

$$\Delta_{\rm NF}^{\rm Signal}[\mathbf{b}] = \delta\left[b_x - \mathbb{E}_x\big[\cos\alpha_{\rm nf,S}(x)\big]\frac{b_t}{c}\left(\frac{T_xN_x}{N_t}\right)\right] \tag{8.18}$$

$$\Delta_{\rm NF}^{\rm Noise}[\mathbf{b}] = \delta\left[b_x - \left(\mathbb{E}_x\big[\cos\alpha_{\rm nf,S}(x)\big]\frac{b_t}{c} + \frac{k}{L_x}\frac{N_t}{c}\right)\left(\frac{T_xN_x}{N_t}\right)\right] \tag{8.19}$$

in the near-field case, then the desired output can be obtained through the 2-D sampling operation

$$\boxed{y[b_t] = \sum_{b_x=0}^{N_x-1} P[\mathbf{b}]\Delta[\mathbf{b}],} \tag{8.20}$$

by selecting the respective kernel $\Delta[\mathbf{b}]$. The sampling kernels given by (8.16)-(8.19) are illustrated in Figure 8.8.

The target signal obtained by sampling $P[\mathbf{b}]$ with the sampling kernel $\Delta_{\rm FF}^{\rm Signal}[\mathbf{b}]$ is not a clean version of the signal, but a noisy version due to the influence of the interferer. On the

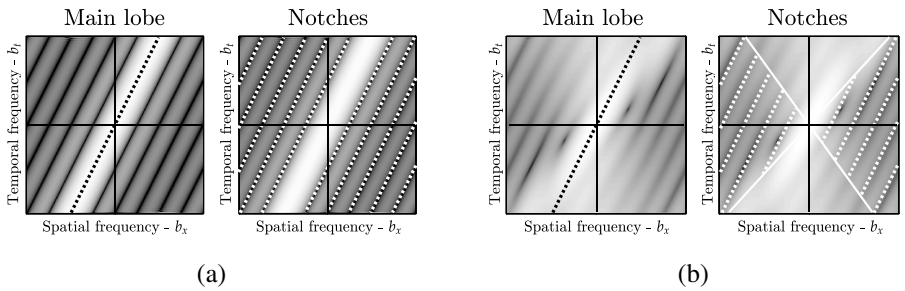


Figure 8.8: Implementation of a sidelobe canceling beamformer as a 2-D sampling operation in the spatio-temporal Fourier domain. (a) For sources in the far-field, the target signal is obtained by applying the sampling kernel $\Delta_{\mathrm{FF}}^{\mathrm{Signal}}[\mathbf{b}]$ (in dashed black), whereas the noise signal is obtained with the sampling kernel $\Delta_{\mathrm{FF}}^{\mathrm{Noise}}[\mathbf{b}]$ (in dashed white). (b) For sources in the near-field or intermediate-field, the target signal is obtained by applying the sampling kernel $\Delta_{\mathrm{NF}}^{\mathrm{Signal}}[\mathbf{b}]$ (in dashed black), whereas the noise signal is obtained with the sampling kernel $\Delta_{\mathrm{NF}}^{\mathrm{Noise}}[\mathbf{b}]$ (in dashed white). In the near-field case, special care must be taken not to sample the spectrum outside the region $\mathcal{U}$, where the high-amplitude area does not contain well defined notches. This also means that, for certain cases, there may be frequencies at which it is not possible to obtain an estimation of the background noise - typically at higher frequencies.

contrary, the output obtained by sampling $P[\mathbf{b}]$ with the sampling kernel $\Delta_{\mathrm{FF}}^{\mathrm{Noise}}[\mathbf{b}]$ is a clean version of the interferer, with no influence from the target signal. This results in a Wiener filtering problem, where the goal is to reduce the amount of noise in the target signal using an estimation of the noise signal. A possible solution is shown in Figure 8.9.

8.4.2 Spatio-temporal LMS

Another way of obtaining an estimation of the interferer is by applying a spatio-temporal filter to the input $p[\mathbf{n}]$, designed such that the pass bands contain little energy from the target source. This filter can be simply the complementary of the filter that obtains the target source, although any other suitable design is possible. Once the noisy target source is filtered out and the noise is estimated, a spatio-temporal adaptive filter can be used to obtain a clean version of the target source. The method is illustrated in Figure 8.10.

The advantage of using this strategy over the one based on 2-D sampling is that it is more general in the sense of spatio-temporal Fourier analysis. Since we are directly filtering in the spatio-temporal Fourier domain, the output of the adaptive filter is the entire filtered wave front, rather than a 1-D time-varying signal. Furthermore, the noise caused by the interferer can be easily estimated from any region of the spectrum, instead of only from diagonal lines. This implies that a larger amount of spatial information can be preserved from the interferer, as well as from the target source. If the upper filter in Figure 8.10 captures the entire triangular region where most of the target energy is contained, then the entire wave front generated by the target source is preserved. This attribute may be critical in audio applications, particularly those dependent on spatial perception (*e.g.*, surround sound, hearing aids, *etc.*). The spatio-

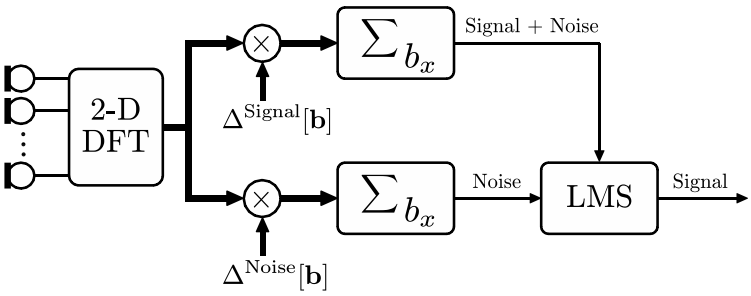


Figure 8.9: Implementation of a 1-D adaptive beamformer in the spatio-temporal Fourier domain. The noisy target signal is obtained by sampling the spectrum at the main lobe, while the noise estimation is obtained by sampling the spectrum at the notches. The noise is then filtered out from the target signal using an adaptive filtering algorithm, such as the least mean squares (LMS) filter [46].

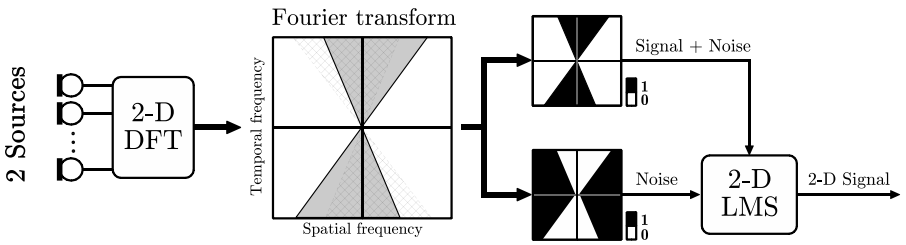


Figure 8.10: Implementation of a 2-D adaptive beamformer in the spatio-temporal Fourier domain. The noisy target signal is obtained by filtering the spectrum with a directional filter that captures its high-amplitude triangular region. The noise is estimated with the complementary filter, which captures the low-amplitude region $\mathcal{U}$ of the target source. This region is expected to contain far more energy from the interferer, specially if a non-rectangular window function is used in space. The 2-D noise wave front is then filtered out from 2-D target wave front using a 2-D version of the LMS adaptive filtering algorithm. The output 2-D signal represents the entire wave front generated by the target source, with no loss of spatial information.

temporal adaptive filtering algorithm can be implemented in the spatio-temporal domain or directly in the Fourier domain, using a 2-D version of the LMS algorithm (see, *e.g.*, Hadhoud *et al.* [43]).

It is important to mention that this section is not meant to be a critical analysis of Wiener filtering itself, or the optimality of our method compared to traditional methods, but rather to show that there is an intuitive motivation for performing adaptive filtering in the spatio-temporal Fourier domain, and that it makes sense to think in terms spatio-temporal Fourier analysis when the goal is to obtain the entire filtered wave front at the output.

8.5 Summary notes

- Beamforming is a spatial filtering technique that allows the wave field to be captured by favoring a given direction;
- The traditional delay-and-add beamformer can be translated into a 2-D sampling operation in the spatio-temporal Fourier domain;
- Spatial filtering can be performed through a 2-D convolution between the array inputs and a spatio-temporal filter, which may be separable or non-separable;
- The ideal spatio-temporal filter is one where the pass-band captures the triangular region of the spectrum where most of the energy of the target source is contained;
- The filtering can be performed in the short space/time Fourier domain in order to deal with the cases where the spectral support of multiple sources overlap;
- Ideal filters can be approximated by realizable filters with limited order, for example by using the window method or the profiling method;
- In adaptive beamforming, steering the zeros of the directivity pattern towards the target source can be translated into a 2-D sampling operation in the spatio-temporal Fourier domain;
- Adaptive beamforming can be performed directly in the Fourier domain through the use of spatio-temporal filtering and 2-D adaptive filtering.

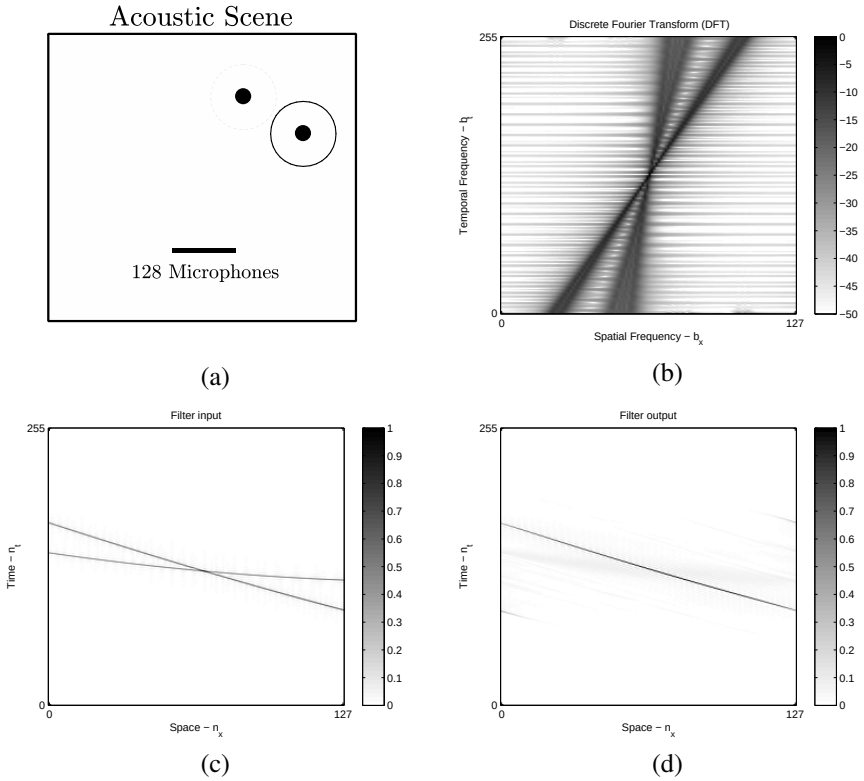


Figure 8.11: Example of filtering *directly* in the spatio-temporal Fourier domain (with no lapped transform). (a) The acoustic scene consists of two Dirac sources in the intermediate-field. The goal is to suppress the dashed source. (b) DFT along the entire spatial axis. (c) Filter input. (d) Filter output.

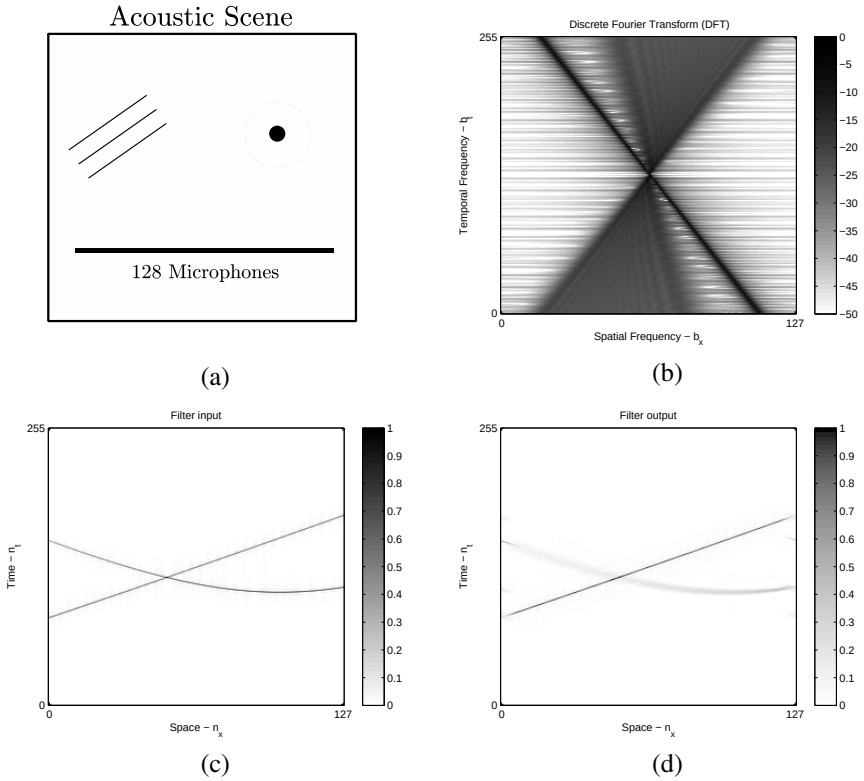


Figure 8.12: Example of filtering *directly* in the spatio-temporal Fourier domain (with no lapped transform). (a) The acoustic scene consists of two Dirac sources, where one is in the intermediate-field and the other in the far-field. The goal is to suppress the dashed source. (b) DFT along the entire spatial axis. (c) Filter input. (d) Filter output.

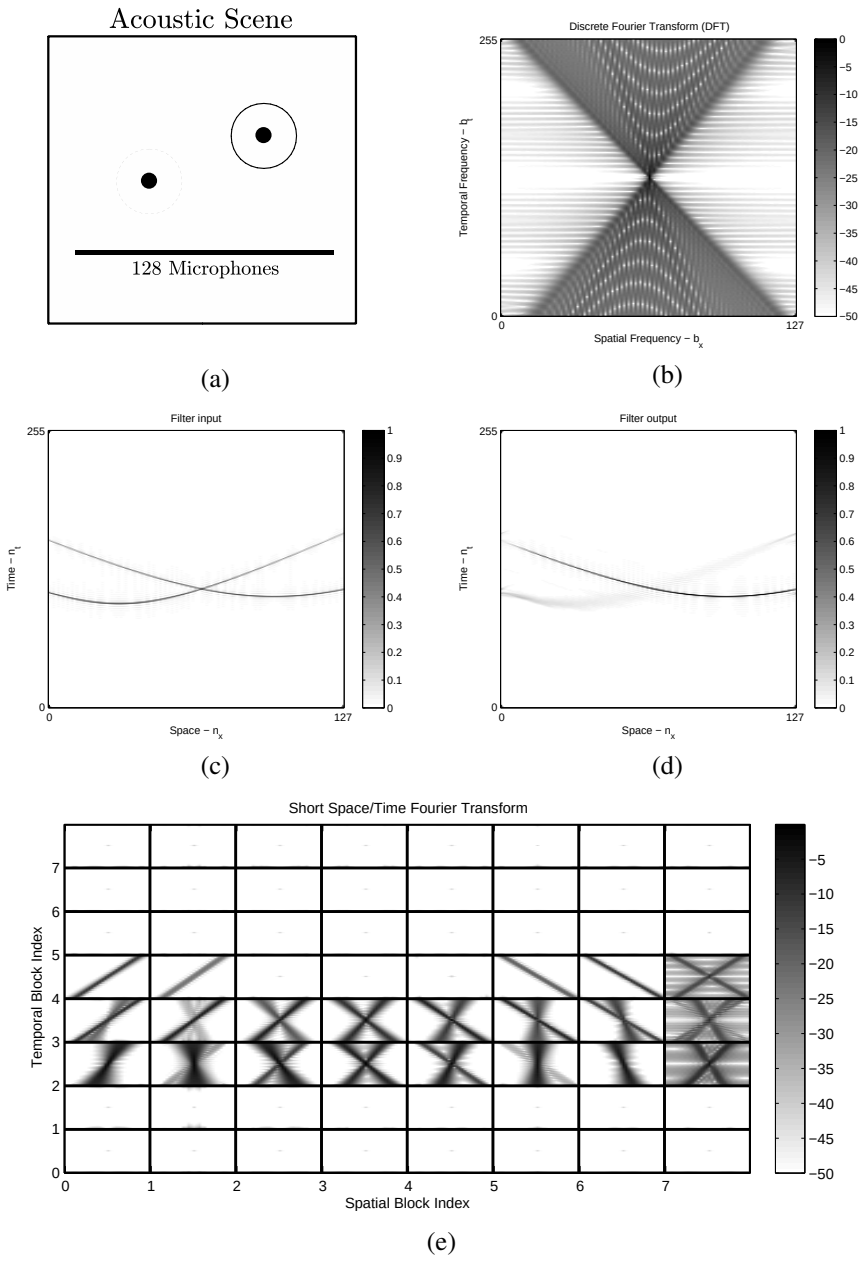


Figure 8.13: Example of filtering in the short space/time Fourier domain. (a) The acoustic scene consists of two Dirac sources in the intermediate-field. The goal is to suppress the dashed source. (b) DFT along the entire spatial axis. (c) Filter input. (d) Filter output. (e) Short space/time Fourier transform.

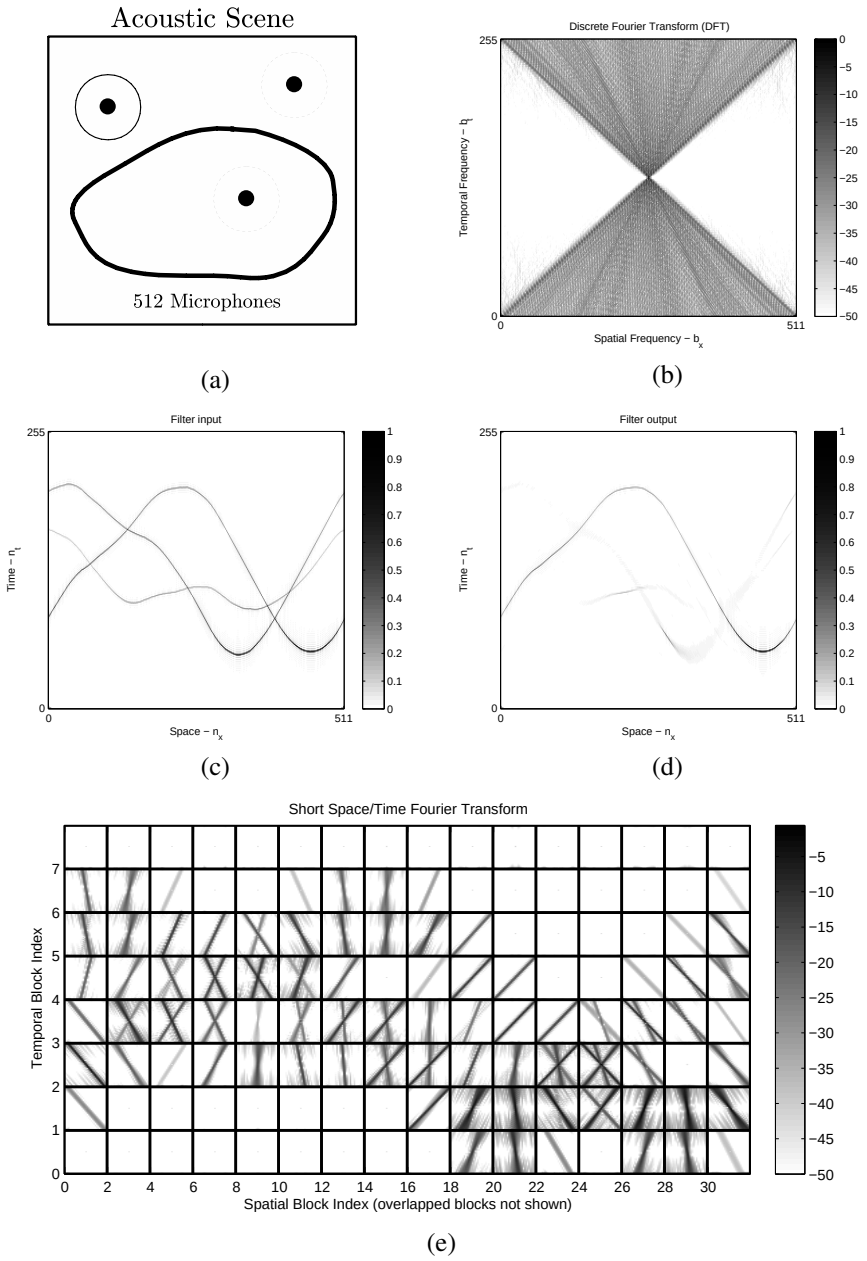


Figure 8.14: Example of filtering in the short space/time Fourier domain. (a) The acoustic scene consists of three Dirac sources in the intermediate-field, and a curved microphone array. The goal is to suppress the dashed sources. (b) DFT along the entire spatial axis. (c) Filter input. (d) Filter output. (e) Short space/time Fourier transform.

Chapter 9

Acoustic Wave Field Coding

9.1 Introduction

Since the early days of digital signal processing (DSP), the question of how to represent signals efficiently in a suitable mathematical framework has been paired with the question of how to efficiently store them in a digital medium. The storage of digital audio, in particular, has been marked by two major breakthroughs: (i) the development of pulse-code modulation (PCM) [67], and (ii) the development of perceptual audio coding [55]. These techniques were popularized, respectively, by their use in Compact Disc technology and MP3 compression; both had a deep impact on the entire industry of audio storage.

The development of PCM is a by-product of the invention of the A/D converter. When the analog signal is converted into a digital signal, the discretization occurs not only in the temporal axis but also in the amplitude of each sample. The amplitude discretization of the analog signal is known as *quantization*, and is performed by a component of the A/D converter called the *quantizer*. The quantizer is a non-linear system that approximates an analog function by a "staircase function", where each "stair" is an N-bit approximation of the analog amplitude. If N_{bit} is the maximum number bits supported by the quantizer, than the output signal has a maximum of $2^{N_{\mathrm{bit}}}$ approximation levels.

In Compact Disc technology, the audio data is stored directly in PCM format, with some additional error correcting information, whereas in MP3 the audio data is stored in the transform domain [15; 51]. More precisely, MP3 operates by transforming the PCM signal to a Fourier-based domain through the use of a uniform filter bank, where the amplitude of the frequency coefficients is again quantized. The key breakthrough is that the number of bits used for quantizing each coefficient is variable, and, most importantly, dependent on their perceptual significance. Psychoacoustic studies show that a great portion of the signal is actually redundant on a perceptual level. This is related to the way the inner ear processes mechanical waves: the wave is decomposed into frequencies by the cochlea, where each frequency stimulates a local group of sensory cells. If a given frequency is close to another frequency with higher amplitude, it will not be strong enough to overcome the stimulation caused by the stronger frequency, and therefore will not be perceived. For this reason, the use of perceptual criteria in the quantization process yields an average compression ratio of $1/10$, over the use of PCM.

In the spatio-temporal analysis of acoustic wave fields, the question arises of how much

relevant information is contained in the wave field, and what is the best way of storing it. When the sound pressure is captured by the multiple microphones to be processed by a computer, there is an implicit amplitude quantization of the pressure values in $p[\mathbf{n}]$. The spatio-temporal signal obtained by a linear array, for instance, is in fact a 2-D PCM signal. With modern optical media such as Double Layer DVD (approximately 8.55GB of storage capacity), we can store about 24 audio channels with 80 minutes of raw (uncompressed) PCM data. However, if the goal is to store in the order of 100 channels, it is imperative that the data be compressed as efficiently as possible.

The most relevant work on joint compression of audio channels dates back to the development of perceptual audio coding in the early 90's. When it was realized that mono PCM audio could be efficiently compressed using filter banks theory and perceptual models, the techniques were immediately extended to stereo PCM audio (see, *e.g.*, Johnston *et al.* [56] and Herre *et al.* [47]), and later to an unlimited number of PCM audio channels (see, *e.g.*, Faller *et al.* [32] and Herre *et al.* [48]). The basic premise of these techniques is that the audio channels are highly correlated and therefore can be jointly encoded with high efficiency, using a parametric approach. The correlation criteria can be both mathematically based—for example, using the theory of dimensionality reduction of data sets [89]; or perceptually motivated—for example, based on the ability of humans to localize sound sources in space [13]. However, what all these techniques have in common is that they treat the multichannel audio data (*i.e.*, the acoustic wave field) as multiple functions of time, and not as a function of space and time as the wave equation demands.

The theory of spatio-temporal signal processing presented in this thesis allows us to follow a radically different approach to compressing the acoustic wave field. Instead of treating the multichannel audio data as multiple functions of time, as is the current paradigm, we can treat the entire wave field as a single multidimensional function of space and time, and perform the actual coding in the multidimensional Fourier domain. When the spatio-temporal signal $p[\mathbf{n}]$ is transformed into the spatio-temporal Fourier domain, there is an implicit decorrelation of the multichannel audio data. This decorrelation is optimal for harmonic sources in the far-field, as these are the basic elements of the spatio-temporal Fourier transform. As a consequence, by quantizing the transform coefficients in $P[\mathbf{b}]$ instead of jointly coding the multichannel signals $p[0, n_t]$, $p[1, n_t]$, ..., $p[N_x - 1, n_t]$, we are directly coding the elementary components of the wave field, which are the plane wave coefficients.

In this chapter, we show how the acoustic wave field can be encoded (and decoded) non-parametrically in the spatio-temporal Fourier domain by using an intuitively simple extension of the traditional mono coder. We show what are the practical bounds for encoding plane waves, and how $P[\mathbf{b}]$ can be quantized using spatio-temporal perceptual criteria. In particular, we propose a method for combining the perceptual models used in MP3 with models of spatial perception. Finally, we briefly discuss a possible bit-packing format for streaming the compressed wave field data.

9.2 Encoding plane waves

In the previous chapters, we saw that a plane wave is characterized by a frequency of oscillation Ω_0 and a propagation vector $\mathbf{u}_0$ indicating its direction. If the wave field is observed on a straight line—typically the x-axis—and the propagation vector forms an angle α_0 with the line, then the sound pressure generated by the plane wave is given by

$p(x,t) = A_0 e^{j\Omega_0(t+\cos\alpha_0 \frac{x}{c})}$, where A_0 is the plane wave amplitude. The respective spatio-temporal Fourier transform is given by

$$P(\Phi_x, \Omega) = A_0(2\pi)^2 \delta(\Omega - \Omega_0)\delta\Big(\Phi_x - \cos\alpha_0 \frac{\Omega}{c}\Big).$$

In the 2-D spatio-temporal Fourier domain, this represents a single point located at $(\Phi_x, \Omega) = \big(\cos\alpha_0 \frac{\Omega_0}{c}, \Omega_0\big)$ with amplitude $A_0(2\pi)^2$. Since the plane wave is completely defined by its amplitude, frequency, and angle, the entire wave front can be encoded by storing the parameters A_0, Ω_0, and α_0. If two plane waves were being generated at the scene, the coding overhead would be an additional set of the same parameters (recall that, in continuous space and time, the spectral representation of a given number of plane waves consists of an equal number of Diracs points). In the general case, a given number of plane waves requires three times as many parameters in order to be encoded. In practice, however, we will show that there is an upper-bound to the number of required bits.

The spatio-temporal Fourier domain reduces plane waves to their defining parameters, A_0, Ω_0, and α_0, and is therefore the optimal domain for encoding plane waves. As discussed in previous chapters, plane waves are the elementary components in the spatio-temporal analysis of the wave field, the same way frequencies are the elementary components in the temporal analysis of signals. The same is true in the directional representation of the wave field, as $P(\alpha, \Omega)$ is simply a scaled version of $P(\Phi_x, \Omega)$ in polar coordinates. Thus, intuitively, we can state that the spatio-temporal Fourier domain is the optimal domain for encoding acoustic wave fields that are essentially composed of plane waves. If the sources are in the near-field, however, or the space is reverberant, the statement is debatable.

It is important to notice that, in continuous space and time, having a spatial dimension in the signal does not increase the amount of information other than by the position of the Diracs across the Φ_x or α-axis, compared to not having a spatial dimension where only A_0 and Ω_0 are required. In discrete space and time, however, if the number of spatial samples is finite, the plane waves are not based on ideal Diracs but on smooth support functions. As a consequence, to fully encode a single plane wave the entire spectral support must be encoded. This is not a big issue if we consider that typical window functions have most of their spectral energy concentrated in the main lobe. If a fair amount of distortion is allowed, the main lobe of the spectral support can be encoded with high efficiency.

The trade-off between the bit rate used for encoding the plane waves and the resulting distortion is discussed in the next section.

9.3 Rate-distortion analysis

Rate-distortion analysis is a branch of information theory that addresses the problem of encoding a random sequence of samples with a limited number of bits, and measuring the resulting distortion [6; 22]. The goal is to obtain a curve that expresses the output distortion as a function of the coding rate, typically using the mean squared error (MSE) distortion metric. The rate-distortion curve can be obtained analytically as a closed form expression or as a pair of parametric equations [6].

Analytical expressions are known to be difficult to derive, and only a few particular cases are manageable in purely analytical terms. This is particularly troublesome in a multidimensional space such as the spatio-temporal domain. It is much easier, in this case, to estimate the rate-distortion function directly from numerical simulations. For this purpose, we use a

simple Matlab implementation of a spectral coding algorithm that operates with a given rate, and directly calculate the distortion on the decoded output. This way, we obtain a sufficient number of points to construct a rate-distortion curve.

Suppose we want to compress the wave field observed on a straight line with N_x spatial points, by encoding the coefficients of $P[\mathbf{b}]$ in the transform domain. The first step is to quantize the amplitude of $P[\mathbf{b}]$, so that a limited number of bits is needed to encode the amplitudes of each coefficient. One way to quantize $P[\mathbf{b}]$ is by defining $P_{\mathrm{Q}}[\mathbf{b}]$ such that

$$P_{\mathrm{Q}}[\mathbf{b}] = \operatorname{sign}\{P[\mathbf{b}]\} \left\lfloor \mathrm{SF}[\mathbf{b}]\, |P[\mathbf{b}]| \right\rfloor, \tag{9.1}$$

where $\mathrm{SF}[\mathbf{b}]$ contains the scale factors of each coefficient, and $\lfloor \cdot \rfloor$ denotes rounding to the closest lower integer. The purpose of the scale factors is to scale the coefficients of $P[\mathbf{b}]$ such that the rounding operation yields the desired quantization noise.

Conversely, the noisy reconstruction of $P[\mathbf{b}]$ can be obtained as

$$\hat{P}[\mathbf{b}] = \operatorname{sign}\{P_{\mathrm{Q}}[\mathbf{b}]\} \left(\frac{1}{\mathrm{SF}[\mathbf{b}]} |P_{\mathrm{Q}}[\mathbf{b}]| \right). \tag{9.2}$$

To determine the number of bits required to encode the quantized coefficients, we need to associate the amplitude values to a binary code book—preferably one that achieves the entropy. In this experiment, we use a Huffman code book similar to the one used in the MPEG standard [51], where code words are organized such that less bits are used to describe lower amplitude values.

Defining Huffman $\{A\}$ as an operator that maps the amplitude value A to the corresponding set of bits (or code word) in the Huffman code book, the number of bits $R[\mathbf{b}]$ required for each coefficient is given by

$$R[\mathbf{b}] = |\text{Huffman}\{P_{\mathrm{Q}}[\mathbf{b}]\}| \tag{9.3}$$

where $|\cdot|$ denotes the size of the set. Note that (9.3) is a deterministic expression, representing the number of bits used in one iteration of the experiment.

Using the MSE between the input signal $p[\mathbf{n}]$ and the output signal $\hat{p}[\mathbf{n}]$ as a measure of distortion, the rate-distortion function $D(R)$ is given by the parametric pair

$$\boxed{\begin{array}{l} R = \sum_{\mathbf{b}=\mathbf{0}}^{\mathbf{N1}-\mathbf{1}} R[\mathbf{b}] \\ D = \frac{1}{\det \mathbf{N}} \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N1}-\mathbf{1}} (p[\mathbf{n}] - \hat{p}[\mathbf{n}])^2 \end{array}} \tag{9.4}$$

where R and D are functions of $\mathrm{SF}[\mathbf{b}]$, with $0 \leq \mathrm{SF}[\mathbf{b}] < \infty$. In the limiting cases, where $\mathrm{SF}[\mathbf{b}] \to 0$ and $\mathrm{SF}[\mathbf{b}] \to \infty$ for all $\mathbf{b}$, (9.4) yields respectively

$$\lim_{\mathrm{SF}[\mathbf{b}]\to 0} R = 0$$

$$\lim_{\mathrm{SF}[\mathbf{b}]\to 0} D = \operatorname{var}\{p[\mathbf{n}]\}$$

and

$$\lim_{\mathrm{SF}[\mathbf{b}]\to\infty} R = \infty$$

$$\lim_{\mathrm{SF}[\mathbf{b}]\to\infty} D = 0$$

where $\mathrm{var}\{\cdot\}$ denotes the signal variance (or power), and the last equality comes from the limiting case $\lim_{\mathrm{SF}[\mathbf{b}]\to\infty} \hat{P}[\mathbf{b}] = P[\mathbf{b}]$.

Figure 9.1 shows examples of rate-distortion curves that result of encoding the acoustic wave field observed on a straight line, using some of the spatio-temporal transforms discussed in this thesis—namely, the short space/time Fourier transform, the MDCT, and the directional Fourier transform. In these examples, the acoustic scene is composed of white-noise sources, in order to reduce the influence of the temporal behavior of the wave field in the bit rate R. Also, since we are evaluating the influence of the number of spatial points N_x in the final number of bits required, the bit rate is expressed in units of "bits per time sample".

The rate-distortion curves obtained for the three spatio-temporal transforms do not differ much from each other. The most striking result (but somehow expected) is that, in all cases, the increase in the number of spatial points N_x does not increase the bit rate proportionally, but it actually converges to an upper-bound. The reason is that, even though doubling N_x also duplicates the number of transform coefficients, the support functions are narrowed to half the width, and the trade-off tends to balance itself out. Thus, increasing N_x past a certain limit does not increase the spectral information, since all it adds are zero values (*i.e.*, amplitude values that are quantized to zero). The directional Fourier domain is characterized by a slightly higher upper-bound, mostly due to the smoothing effect of the half-band filters in the directional filter bank.

It can also be observed in Figure 9.1 that for lower bit-rates—in the order of those used by perceptual audio coders [15]—the difference between one channel and a large number of channels is low in terms of MSE. For example, in Figure 9.1-(a), the number of bits required to encode 256 channels is 11.3 bits/time-sample, as opposed to the 2.6 bits/time-sample required for encoding 1 channel. To have a fair comparison, we can consider that a practical codec would require about 20% of bit-rate overhead with decoding information [51], and thus increase the average rate to 13.6 bits/time-sample. Still, compared to encoding 1 channel, the total bit rate required to support the additional 255 channels is only 5 times higher.

Another interesting result is that, similarly to what happens when N_x is increased, the increase in the number of sources does not increase the bit rate proportionally; again, it converges to an upper-bound. Equally interesting is that the upper-bound is the same for all domains. This is because the bit rate only increases until the entire domain is filled up with information. Once this happens, the spectral support generated by additional sources will simply overlap with the existing ones.

9.4 Perceptual wave field coding

The coding paradigm introduced in this chapter is conceptually similar to that of a traditional transform domain coder, except that the spectral coefficients represent plane waves instead of regular sinusoids. The plane waves can be encoded with a limited number of bits and, yet, be reconstructed with an arbitrarily low mean squared error (MSE). It is a known fact, however, that the MSE is not a suitable measure of distortion for audio signals—particularly music

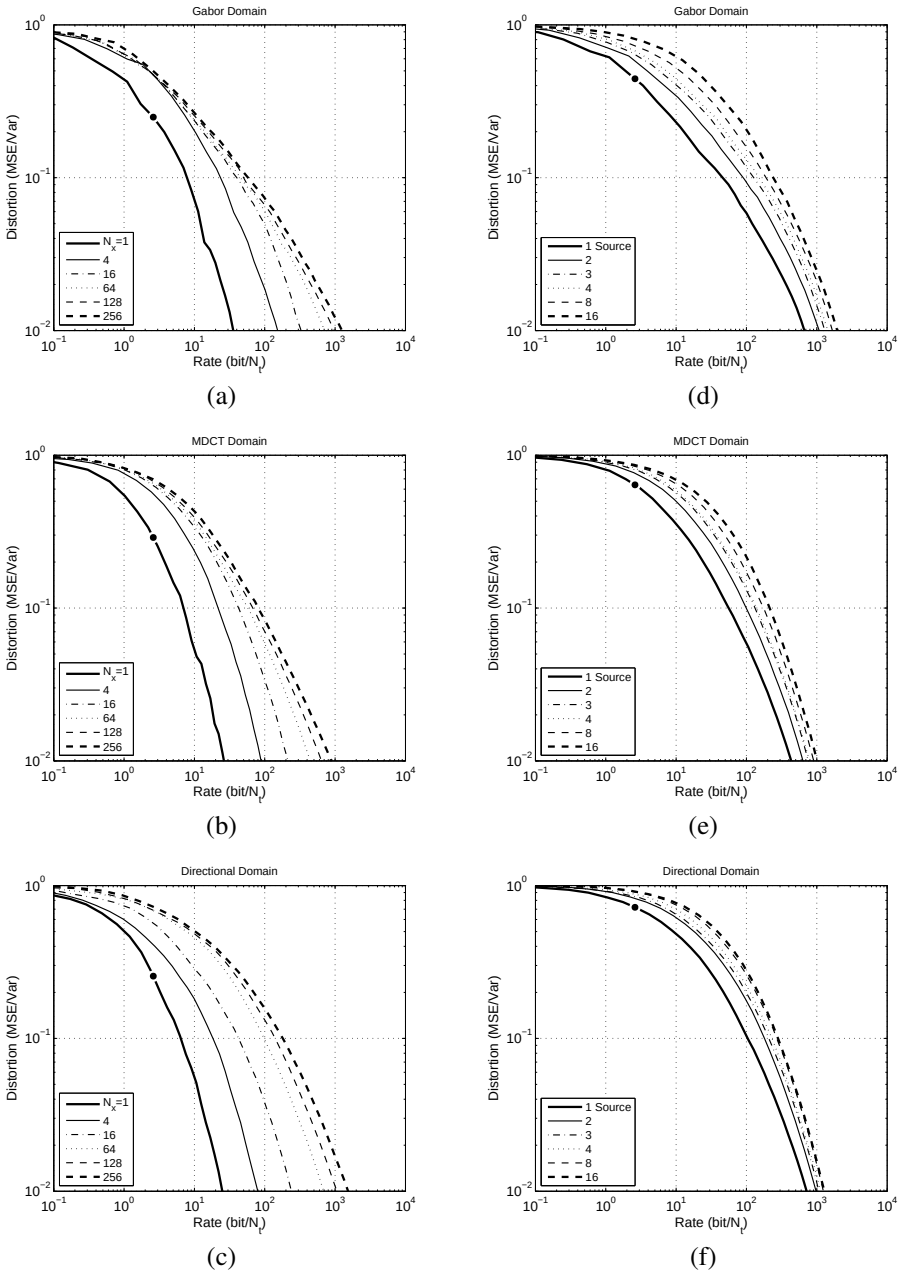


Figure 9.1: Experimental rate-distortion curves for white-noise sources in the far-field observed on a straight line. On the left, the $D(R)$ curves are shown for one source encoded in: (a) the short space/time Fourier domain (Gabor domain) with no overlapping, (b) the MDCT domain with 50% overlapping, and (c) the directional domain with no overlapping. In (a), (b), and (c), the source is fixed at $\alpha = \frac{\pi}{3}$ and the number of spatial points N_x is variable. On the right, the $D(R)$ curves are shown for multiple sources encoded in: (d) the short space/time Fourier domain, (e) the MDCT domain, and (f) the directional domain. In (d), (e), and (f), the number of spatial points is fixed, $N_x = 64$, and the sources are placed at random angles. The black circle shown in each plot indicates $R = 2.6$ bits/time-sample, which is the average rate of state-of-art perceptual coders.

signals—since it is tangentially related to the way humans perceive sound. For example, delaying a signal by a fraction of a sample results in a high MSE, but it is unperceivable to the human ear.

Perceptual audio coders such as MP3 use a distortion measure based on actual listening tests, where the listener classifies the quality of the decoded sound compared to the original one. The amount of perceivable noise as a result of quantization is predicted by an empirical model known as the *psychoacoustic model*, which provides an estimation of the quantization noise power for each frequency based on the amplitude of the neighboring frequencies. The model is based on an anatomical phenomenon known as *frequency masking*, which characterizes the frequency selectivity of the cochlea in the sensory analysis of sound.

In this section, we show how a perceptual wave field coder (WFC) can be implemented non-parametrically using the same concepts that represent the foundation of state-of-art perceptual audio coders. We discuss in particular: (i) the basic structure of the WFC, and how it is inspired by traditional mono coders; (ii) what is right spatio-temporal filter bank for a WFC application; (iii) how to derive a spatio-temporal psychoacoustic model from existing models; (iv) how to construct a bit-stream format that includes the quantized data and side information necessary for decoding.

It is important to mention, however, that the proposed method represents an entirely new paradigm of spatial audio coding, published for the first time in 2008 [72]. A considerable amount future research is still needed to fully determine the potential of this method.

9.4.1 From MP3 to WFC: A simple step

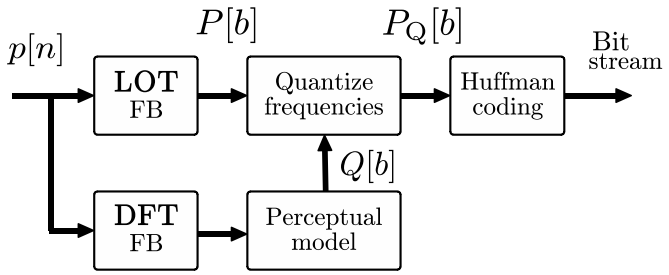
The basic structure of a perceptual audio coder consists of a filter bank that translates the input signal into the frequency domain, an amplitude quantizer regulated by a psychoacoustic model, and a lossless entropy coder. This is illustrated in Figure 9.2-(a).

A perceptual wave field coder is simply a multidimensional generalization of this structure—for simplicity, we consider only one spatial dimension. The 2-D filter bank translates the input wave field into the plane wave domain, and the amplitude of each plane wave is quantized according to its perceptual significance. Then, an entropy coder maps the amplitude values into a sequence of bits that can be packed into a bit-stream. This is illustrated in Figure 9.2-(b).

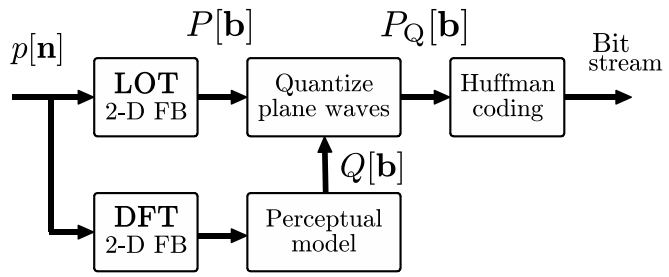
9.4.2 Choosing the right filter bank

The use of different spatio-temporal transforms in the context of a coding application has different advantages. On the one hand, the MDCT is a critically sampled lapped orthogonal transform (LOT) that allows overlapping between blocks in both space and time, as opposed to the directional transform which does not preserve critical sampling (in the strict mathematical sense) if overlapping is used. This gives the MDCT an advantage in terms of coding gain. On the other hand, in a perceptual coding application, the directional transform provides a more suitable representation of the wave field in terms of directional sound, which, as we will see, simplifies the use of psychoacoustic models. A cascade of the two filter banks is also an interesting case to consider [73; 30].

When the directional transform is used with overlapping, it is possible in practice to obtain an approximation of a critically sampled output—particularly when the number of spatial points is high and the Nyquist sampling conditions are satisfied. Under these conditions, we have seen that at least half of the entire spectrum has very low energy values (see Figure 4.4),



(a)



(b)

Figure 9.2: Wave field encoder as a generalization of a perceptual mono encoder. (a) In a typical mono encoder, the input signal is transformed into the frequency domain by a filter bank (FB) implementing a lapped orthogonal transform (LOT). The resulting transform coefficients (frequencies) are quantized with the aid of a perceptual model, which regulates the amount of quantization noise. The quantized coefficients are then encoded into a bit-stream by a Huffman encoder. The perceptual model typically requires a DFT input [15]. (b) In a wave field encoder, the input spatio-temporal signal is transformed into the plane wave domain by a 2-D lapped orthogonal transform. The resulting transform coefficients (plane waves) are quantized with the aid of a spatio-temporal perceptual model, obtained by combining a frequency masking model with a directional masking model. Finally, the quantized coefficients are encoded into a bit-stream by a Huffman encoder.

which results in an effective low-MSE coding gain of 2 when encoding the wave field in the Fourier domain. This inherent gain can be exchanged for a 50% overlapping in one of the domains, resulting in a critically sampled directional transform with overlapping.

9.4.3 Spatio-temporal perception

The psychoacoustic model is the component that defines a perceptual audio coder. The knowledge of how the inner ear operates, and the model's ability to emulate the frequency selectivity of the cochlea, is the major reason why a compression ratio of $1/10$ is achieved by MP3 with a near perceptual transparency.

The mathematical description of the frequency masking phenomenon was obtained throughout the years by means of perceptual studies conducted on human volunteers (see, *e.g.*, Katz *et al.* [58] and Plack *et al.* [75]). Similar studies were performed on the directional perception of sound (see, *e.g.*, Blauert *et al.* [13]). The directional masking effect (or spatial masking) occurs when two sources are close to each other, but one has a high enough amplitude to "silence" the other. Such models are very important in a WFC application, since localized sound tends to be dominant over diffuse sound.

The two types of models can be combined in order to construct a spatio-temporal psychoacoustic model that describes the perception of plane waves by human listeners, as illustrated in Figure 9.3. The goal of a spatio-temporal psychoacoustic model is to estimate, for each transform block $P[\mathbf{n}]$ of size $N_x \times N_t$, a block $Q[\mathbf{b}]$ of equal size that contains the maximum quantization noise power that each plane wave can sustain without causing perceivable artifacts. If a directional transform is used, the estimation of $Q[\mathbf{b}]$ as a combination of frequency and directional masking effects is simplified, since the two dimensions are orthogonal: frequency masking occurs along the Ω-axis, while directional masking occurs along the α-axis.

As our previous experiments have shown [72; 71], the use of a spatio-temporal psychoacoustic model to regulate the quantization of plane waves allows the wave field to be encoded and decoded without causing perceivable artifacts.

9.4.4 Quantization using perceptual criteria

The main purpose of $Q[\mathbf{b}]$ is to provide a means of deriving the scale factors $\mathrm{SF}[\mathbf{b}]$, necessary for the quantization formulas defined in (9.1) and (9.2). In uniform quantization, the quantization noise power is given by $\Delta_{\mathrm{Q}}^2/12$, where Δ_{Q} is the quantization step. In (9.1), $\mathrm{SF}[\mathbf{b}]$ must be defined such that $\Delta_{\mathrm{Q}} = 1$. Therefore, to have a quantization free of perceivable artifacts, the scale factors must satisfy the condition

$$\mathrm{SF}[\mathbf{b}] \geq \sqrt{12Q[\mathbf{b}]}. \tag{9.5}$$

A typical difficulty with this method is that, in order for the quantized coefficients to be re-scaled using (9.2), the scale factors have to be transmitted to the decoder. In a practical implementation, it is not possible to have one scale factor per coefficient, otherwise no compression would be achieved. Instead, a scale factor is assigned to one critical band [15], such that all coefficients within the same critical band are quantized with the same scale factor. In WFC, the critical bands are 2-D and the scale factors $\mathrm{SF}[\mathbf{b}]$ are approximated by a piecewise constant surface. The critical bands can be organized, for example, as shown in Figure 9.4.

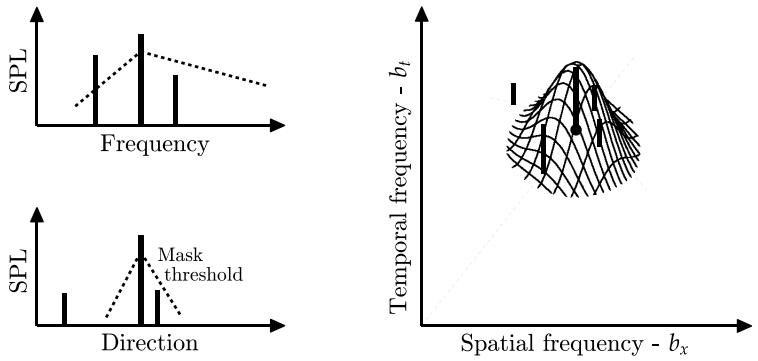


Figure 9.3: Derivation of a spatio-temporal psychoacoustic model from two separate models. The masking threshold generated by the frequency of a given plane wave is combined with the masking threshold generated by its direction. In the frequency axis, the masking effect caused by the center frequency is enough to mask the lower SPL (sound pressure level) frequency on the right, but not enough to mask the frequency of the left. In the directional axis, the center source masks the source on the right but not the one on the left. In the spatio-temporal Fourier domain, the combined masking effect forms a dome-shaped masking threshold.

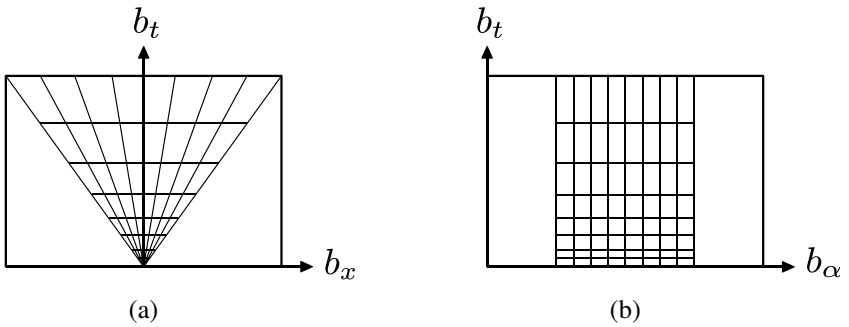


Figure 9.4: Spatio-temporal distribution of critical bands. (a) In the spatio-temporal Fourier domain, the critical bands can have a non-uniform partitioning along the temporal frequency axis, such that it emulates the frequency selectivity of the cochlea (higher resolution at lower frequencies). In the spatial axis, the boundary of the critical bands can be defined as diagonal lines with uniformly distributed slopes, such that different directions are separated with the same resolution. The entire evanescent wave region can be characterized by a single coding band. If the MDCT is used, only the $b_x \geq 0$ plane is needed. (b) In the directional Fourier domain, the critical bands can be more easily defined as rectangular tiles, with a non-uniform partitioning along the temporal frequency axis and a uniform partitioning along the directional axis. Similarly, the evanescent wave region can be characterized by a single coding band.

9.4.5 Huffman coding

After quantization, the spectral coefficients can be converted into the binary base through the use of entropy coding. A Huffman code book with a certain range can be assigned to each critical band, and all the coefficients belonging to a given band are encoded with the same code book.

The use of entropy coding is possible because the values generated by the MDCT for various types of music has a fairly stable probability distribution. An MDCT occurrence histogram, for different signal samples, clearly shows that small absolute values are more likely than large absolute values, and that most of the values fall within the range of -20 to 20. For this reason, state-of-art audio coders use a predefined set of Huffman code books that cover all the ranges up to a certain value r. If any coefficient is bigger than r or smaller than $-r$, it is encoded with a fixed number of bits using PCM. In addition, adjacent values $(P[\mathbf{b}], P[\mathbf{b}+\mathbf{1}])$ are coded in pairs, instead of individually. Each Huffman code book covers all combinations of values from $(P[\mathbf{b}], P[\mathbf{b}+\mathbf{1}]) = (-r, -r)$ up to $(P[\mathbf{b}], P[\mathbf{b}+\mathbf{1}]) = (r, r)$.

A Huffman code book of this type can be derived by using a probabilistic model for the spatio-temporal MDCT coefficients, defined for example by [71]

$$\text{probability}\left\{A_0, A_1\right\} = \frac{\left(\text{mean}\left\{A_0, A_1\right\} + \text{var}\left\{A_0, A_1\right\} + 1\right)^{-1}}{\sum_{A_0=-r}^{r} \sum_{A_1=-r}^{r} \left(\text{mean}\left\{A_0, A_1\right\} + \text{var}\left\{A_0, A_1\right\} + 1\right)^{-1}}.$$

This comes from the assumption that (A_0, A_1) is more likely to have both values A_0 and A_1 within a small amplitude range, and that (A_0, A_1) has no sharp variations between A_0 and A_1.

9.4.6 Bit-stream format

The final step in WFC is to organize all binary data into a time series of bits called the *bit-stream*, in such a way that the decoder can parse the data and use it reconstruct the spatio-temporal signal $p[\mathbf{n}]$. The basic components of the bit-stream are the main header and the frames that contain the encoded transform coefficients from each block, as illustrated in Figure 9.5. The frames themselves have a small header with side information necessary to decode the spectral data.

The main header is located at the beginning of the bit-stream, and contains information about the sampling frequencies in space and time, the window type, the MDCT block size, and any additional parameters that remain fixed for the whole duration of the spatio-temporal signal.

The frame format is repeated for each spectral block $\mathbf{Y}_{g,l}$, and organized in the following order:

$$\mathbf{Y}_{0,0} \ldots \mathbf{Y}_{0,N_l-1} \mathbf{Y}_{N_g-1,0} \ldots \mathbf{Y}_{N_g-1,N_l-1}$$

such that, for each time instance, all spatial blocks are consecutive. Each block $\mathbf{Y}_{g,l}$ is encapsulated in a frame, with a header that contains the scale factors used by $\mathbf{Y}_{g,l}$ and the Huffman code book identifiers.

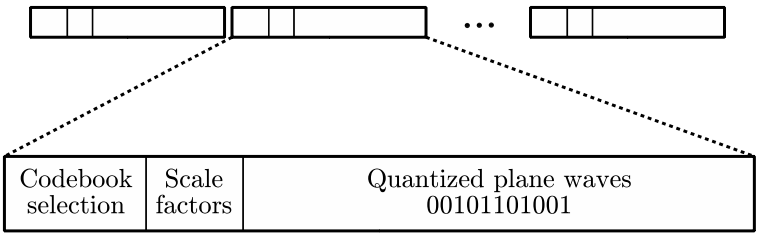


Figure 9.5: Bit-stream format of a perceptual wave field coder. Each frame contains the quantized and encoded plane waves, plus a header with side information including the Huffman codebook selection and the scale factors.

9.4.7 Decoding

The decoding stage of WFC, illustrated in Figure 9.6 consists of three steps: Huffman decoding, re-scaling, and the inverse filter bank. The decoding is controlled by a state machine representing the Huffman code book assigned to each critical band [71]. Since Huffman coding generates prefix-free binary sequences, the decoder knows immediately how to parse the encoded transform coefficients. Once the coefficients are decoded, the amplitudes are re-scaled using (9.2) and the scale factor is associated to each critical band. Finally, the inverse LOT is applied to the transform blocks, and the output signal is obtained through spatio-temporal overlap-and-add. The decoded multichannel signal $p[\mathbf{n}]$ can be interpolated into $p(x,t)$ without loss of information as long as the anti-aliasing conditions are satisfied.

9.4.8 Experimental method

A possible way of determining the compression ratio achieved by the proposed method, using a Matlab implementation, is to perform the following 6 steps:

1. Generate the acoustic wave field from a music sound file;
2. Generate the input signal $p[\mathbf{n}]$ by computing the sound pressure on the N_x microphones;
3. Run the WFC algorithm using the input $p[\mathbf{n}]$, in order to obtain the decoded signal $\hat{p}[\mathbf{n}]$;
4. Listen to each channel $\hat{p}[0,n_t]$, $\hat{p}[1,n_t]$, $\ldots$, $\hat{p}[N_x,n_t]$ individually, in order to check for artifacts;
5. Repeat 3 and 4 for lower bit rates until a threshold is obtained, under which the artifacts start being audible;
6. Plot the input and output signals, and visually verify that the curvature of the wave front is preserved.

To this moment, we have solely verified that the coding process is, in fact, transparent—*i.e.*, it does not generate audible artifacts, and the curvature of the wave front is preserved—for $N_x = 4$ and a constant bit rate of $R = 2.6$ bits per time-sample and per channel, which is equivalent to encoding each channel independently using MP3. This means that we have not

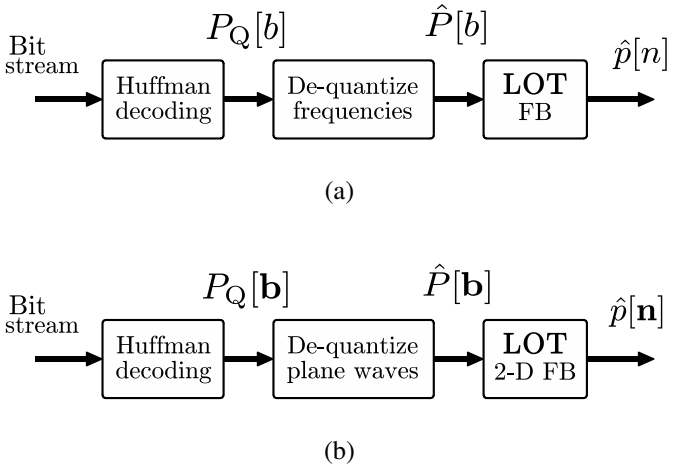


Figure 9.6: Wave field decoder as a generalization of a perceptual mono decoder. (a) In a typical mono decoder, the input bit-stream is decoded by the corresponding Huffman codebook, and the quantized frequencies are re-scaled with the aid of the scale factors transmitted as side information. Then, the recovered transform blocks are transformed back into a time-domain signal. (b) In a wave field decoder, the input bit-stream is decoded by the corresponding Huffman codebook, and, similarly, the quantized plane waves are re-scaled with the aid of the scale factors transmitted as side information. Finally, the recovered transform blocks are transformed back into a spatio-temporal signal.

yet tested the algorithm for compression, but only for transparency. A systematic study of the compression ratio achieved by this method for a different number of channels is part of our future work.

9.5 Summary notes

- Perceptual audio coders operate by quantizing and encoding the frequency coefficients based on a model of human auditory perception;
- The acoustic wave field can be compressed by quantizing and encoding the plane waves in the spatio-temporal Fourier domain;
- The rate-distortion curves generated by encoding plane waves in the spatio-temporal Fourier domain converge to an upper-bound, either by increasing the number of channels or the number of sources;
- A perceptual wave field coder can be easily constructed as a generalization of a perceptual mono coder;
- A spatio-temporal psychoacoustic model that characterizes the human perception of plane waves can be derived through a combination of a frequency masking model and a directional masking model;
- The quantized and encoded plane waves can be packed into a bit-stream with side information necessary for the decoding process.

Conclusions and Future Work

We presented a signal processing framework for manipulating acoustic wave fields that are sampled with arrays of microphones, based on a multidimensional Fourier representation of the wave equation. The thesis builds on the idea that arrays of uniformly spaced sensors organized on a grid are, from a signal processing perspective, equivalent to sampling the sound pressure along the three dimensions of space. The output of this sampling operation is a 4-D discrete spatio-temporal signal, representing the entire acoustic wave field in space and time. Given the dynamics of the wave equation, it can be verified that the wave field generated by a harmonic source has necessarily the same harmonic behavior along the three dimensions of space. This gives a clear indication that the oscillation patterns of the wave field across space are potentially sparse in the spatial Fourier domain. Using existing tools of multidimensional Fourier analysis, the 4-D signal can be represented and analyzed efficiently in both continuous and discrete space and time. In addition, using the Huygens principle and a technique known as wave field synthesis, the sampled spatio-temporal signal can be reconstructed back into an "analog" wave front after being manipulated by a digital computer, or stored in a digital medium. Based on this premise, we dedicated five chapters of this thesis to extending some of the most important concepts of 1-D signal processing to the 4-D spatio-temporal domain. In particular, we extended the theory of linear time-invariant (LTI) systems to a theory of linear space- and time-invariant (LSTI) systems based on the Huygens principle, by proving the properties of linearity and shift-invariance of spatio-temporal systems that take wave fronts as input and generate wave fronts as output. We revisited the main results of spatio-temporal Fourier analysis of the wave field, gradually obtained by other authors in the study of field equations, and we generalized these results to account for a local windowed analysis in space and time, resulting in what we call space/time–frequency representation of the acoustic wave field. We also proposed an alternative representation of the spatio-temporal spectra as a function of the direction of wave propagation, rather than the non-intuitive notion of spatial frequency or wave number. The last three chapters of the thesis were dedicated to processing the acoustic wave field in discrete space and time. We showed how various types of spatio-temporal transforms, discussed in the context of continuous space and time, can be implemented by a digital computer in the discrete domain through the use of multidimensional filter bank theory. Using these discrete representations, we again extended to the spatio-temporal domain two important concepts of 1-D signal processing: filtering and coding. We showed that spatio-temporal filters can be easily designed in the spatio-temporal Fourier domain in order to enhance or suppress sources in the acoustic scene, both by direct application of a linear filter or through spatio-temporal adaptive filtering. Finally, we proposed an entirely new paradigm of spatial audio coding, based on the encoding of the wave field elements in the spatio-temporal Fourier domain (known as plane waves), both

using a mean squared error distortion measure and using criteria based on the human auditory perception.

Future work

This thesis started as study on transparent coding of acoustic wave fields, based on a decomposition of the wave field into elementary components—a considerable shift from the traditional parametric multichannel coding methods, such as MPEG surround. Eventually, however, we realized that there were no satisfying tools for processing the wave field in discrete space and time. Tools that are critical for the efficient implementation of state-of-art mono and stereo audio coders, such as time–frequency analysis, were poorly understood in the context of acoustic wave fields. Once we developed the concept of space/time–frequency analysis, only then we were able to determine the bit rate requirements and trade-offs of a practical wave field coder. As mentioned in the introduction, the theory of signal processing was mainly developed and optimized for 1-D time-varying signals. The use of microphone arrays for recording spatio-temporal audio data is a new trend, and therefore many of these 1-D signal processing concepts will have to be extended to the spatial domain. Some examples include spatio-temporal linear prediction for encoding speech wave fronts, spatio-temporal sparse sampling for the localization of shot events, and spatio-temporal wavelet analysis for handling sharply curved microphone arrays. With these additional tools, as well as the ones presented in this thesis, many other applications can be envisioned. Microphone arrays could potentially be used, for example, to turn any flat surface into a computer touchpad, to track moving sources across rooms by analyzing the wall vibrations, and to perform acoustical "auto-focus" on target sources when recording sound. In summary, the more sensor array technology grows in popularity and decreases in cost, the more will grow the need for a solid theoretical framework and for efficient computational tools that are capable of translating creative ideas into practical implementations.

Bibliography

[1] T. Ajdler. *The plenacoustic function and its applications*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2006.

[2] T. Ajdler, L. Sbaiz, and M. Vetterli. The plenacoustic function and its sampling. *IEEE Trans. Sig. Proc.*, 54:3790–3804, 2006.

[3] J. Allen and D. Berkley. Image method for efficiently simulating small-room acoustics. *J. Acoustic Soc. America*, 65:943–950, 1979.

[4] R. Bamberger and M. Smith. A filter bank for the directional decomposition of images: Theory and design. *IEEE Trans. Signal Process.*, 40:882–893, 1992.

[5] K. Becker, M. Becker, and J. Schwarz. *String Theory and M-Theory: A Modern Introduction*. Cambridge University Press, 2007.

[6] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, 1971.

[7] A. Berkhout. *Applied seismic wave theory*. Elsevier Publishing Company, 1987.

[8] A. Berkhout. A holographic approach to acoustic control. *J. Audio Eng. Soc.*, 36(12):977–995, 1988.

[9] A. Berkhout. Wave-front synthesis: A new direction in electroacoustics. In *Audio Eng. Soc. 93th Conv.*, 1992.

[10] A. Berkhout, D. de Vries, J. Baan, and B. van den Oetelaar. A wave field extrapolation approach to acoustical modeling in enclosed spaces. *J. Acoustic Soc. America*, 105(3):1725–1733, 1999.

[11] A. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *J. Acoustic Soc. America*, 93:2764–2778, 1993.

[12] E. Berliner. Gramophone, 1896.

[13] J. Blauert. *Spatial hearing: The psychophysics of human sound localization*. The MIT Press, 1996.

[14] M. Boone. Acoustic rendering with wave field synthesis. In *ACM SigGraph Campfire: Acoustic Render.*, 2001.

[15] M. Bosi and R. Goldberg. *Introduction to digital audio coding and standards*. Springer, 2002.

[16] P. Bremaud. *Mathematical Principles of Signal Processing*. Springer, 2002.

[17] L. Brown. *A radar history of World War II: Technical and military imperatives*. Taylor & Francis, 1999.

[18] D. Champeney. *A Handbook of Fourier Theorems*. Cambridge University Press, 1989.

[19] V. Chappelier, C. Guillemot, and S. Marinkovic. Image coding with iterated contourlet and wavelet transforms. In *IEEE Inter. Conf. Image Process.*, 2004.

[20] I. Ciufolini and V. Gorini. *Gravitational Waves*. Taylor & Francis, 2001.

[21] TDK Corporation. Acoustic components: World's smallest mems microphones with a digital interface. *EPCOS Press Release*, March 4, 2010.

[22] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[23] J. d'Alembert. *Essai d'une nouvelle theorie de la resistance des fluides [Essay on a new theory of the resistance of fluids]*. 1752.

[24] P. Dirichlet. Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données [on the convergence of trigonometric series which serve to represent an arbitrary function between given limits]. *Journal für die reine und angewandte Mathematik*, 4:157–169, 1829.

[25] M. Do. *Directional Multiresolution Image Representations*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2001.

[26] M. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans. Image Process.*, 14:2091–2106, 2005.

[27] D. Dudgeon. *Multidimensional Digital Signal Processing*. Prentice Hall, 1984.

[28] T. Edison. Improvement in phonograph or speaking machines, 1878.

[29] F. Engel. Documents on the invention of magnetic recording in 1878: An appreciation on the 150th anniversary of the inventor's birth. In *Audio Engineering Society 88th Convention*, 3 1990.

[30] R. Eslami and H. Radha. A new family of nonredundant transforms using hybrid wavelets and directional filter banks. *IEEE Trans. Image Process.*, 16:1152–1167, 2007.

[31] L. Euler. De la propagation du son [on the propagartion of sound]. *Mémoires de l'académie des sciences de Berlin*, pages 185–209, 1766.

[32] C. Faller and F. Baumgarte. Binaural cue coding: a novel and efficient representation of spatial audio. In *IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, volume 2, pages 1841–1844, 2002.

[33] J. Fourier. Mémoire sur la propagation de la chaleur dans les corps solides [memoir on the propagation of heat in solid bodies]. *Nouveau Bulletin des sciences par la Société philomatique de Paris*, 6:215–221, 1807.

[34] O. Frost. An algorithm for linearly constrained adaptive array processing. *Proc. IEEE*, 60:926–935, 1972.

[35] D. Gabor. Theory of communication. *Journal I.E.E.*, 93:429–457, 1946.

[36] B. Gardner and K. Martin. Hrtf measurements of a kemar dummy-head microphone. Technical report, Massachusetts Institute of Technology, 1994.

[37] M. Gerzon. Practical periphony: The reproduction of full-sphere sound. In *Audio Eng. Soc. 65th Conv.*, 1980.

[38] I. Gradshteyn and I. Ryzhik. *Table of Integrals, Series and Products*. Academic Press, 2000.

[39] G. Green. *An essay on the application of mathematical analysis to the theories of electricity and magnetism*. Self-published, 1828.

[40] L. Griffiths and C. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas, Propagation*, 30:27–34, 1982.

[41] B. Gross and H. Pelzer. Relations between delta functions. In *Proceedings of the Royal Society of London*, volume 210, pages 434–437, 1951.

[42] A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15:723–736, 1984.

[43] M. Hadhoud and D. Thomas. The two-dimensional adaptive lms (tdlms) algorithm. *IEEE Trans. on Circuits and Systems*, 35(5):485 –494, may 1988.

[44] T. Hankins. *Jean d'Alembert: Science and the Enlightenmen*. Informa Healthcare, 1990.

[45] D. Harris and R. Mersereau. A comparison of algorithms for minimax design of two-dimensional linear phase fir digital filters. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 25(6):492 – 500, dec 1977.

[46] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 2001.

[47] J. Herre, K. Brandenburg, and D. Lederer. Intensity stereo coding. In *Audio Eng. Soc. 96th Conv.*, 1994.

[48] J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hoelzer, and C. Spenger. Mp3 surround: Efficient and compatible coding of multi-channel audio. In *Audio Eng. Soc. 116th Conv.*, 2004.

[49] http://rr.epfl.ch/31/. Multimedia files associated with the journal paper entitled "space/time-frequency processing of acoustic wave fields: Theory, algorithms, and applications", by f. pinto and m. vetterli, 2009.

[50] C. Huygens. *Traité de la lumière [Treatise on light]*. Pieter van der Aa, 1690.

[51] ISO/IEC. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s - part 3: Audio, 1993. JTC1/SC29/WG11.

[52] H. Jahnke. *A History of Analysis*. American/London Mathematical Society, 1999.

[53] A. Johnson, J. Princen, and H.Chan. Frequency scalable video coding using the mdct. In *IEEE Inter. Conf. Acoustics, Speech, Signal Process.*, 1994.

[54] D. Johnson and D. Dudgeon. *Array Signal Processing*. Prentice Hall, 1993.

[55] J. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE J. Selected Areas Commun.*, 5(2):314–323, 1988.

[56] J. Johnston and A. Ferreira. Sum-difference stereo transform coding. In *IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, volume 2, pages 569–572, 1992.

[57] Y. Kamp and J. Thiran. Chebyshev approximation for two-dimensional nonrecursive digital filters. *IEEE Trans. Circuits and Systems*, 22(3):208 – 218, mar 1975.

[58] J. Katz. *Handbook of Clinical Audiology*. Lippincott Williams & Wilkins, 2001.

[59] W. Kühnel. *Differential Geometry: Curves - Surfaces - Manifolds*. American Mathematical Society, 2005.

[60] J. Kovacevic, D. Le Gall, and M. Vetterli. Image coding with windowed modulated filter banks. In *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 1989.

[61] J. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer, 2007.

[62] H. Malvar. *Signal processing with lapped transforms*. Artech House Publishers, 1992.

[63] V. Martin and T. Guignard. Graft of the boundary integral method onto the image-source method for vehicle acoustics. In *14th Inter. Cong. Sound Vibration*, 2007.

[64] J. McClellan. The design of two-dimensional digital filters by transformation. In *Proc. 7th Annual Princeton Conference on Information Sciences and Systems*, pages 247–251, 1973.

[65] P. Morse and K. Ingard. *Theoretical acoustics*. Princeton University Press, 1987.

[66] D. Morton. *Sound Recording: The Life Story of a Technology*. Johns Hopkins University Press, 2006.

[67] B. Oliver and C. Shannon. Communication system employing pulse code modulation, 1957.

[68] A. Oppenheim and R. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, 1998.

[69] A. Oppenheim and A. Willsky. *Signals and Systems*. Prentice Hall, 1997.

[70] S. Phoong, C. Kim, P. Vaidyanathan, and R. Ansari. A new class of two-channel biorthogonal filter banks and wavelet bases. *IEEE Trans. Signal Process.*, 43:649–665, 1995.

[71] F. Pinto and M. Vetterli. Bitstream format for spatio-temporal wave field coder. In *Audio Eng. Soc. 124th Conv.*, 2008.

[72] F. Pinto and M. Vetterli. Wave field coding in the spacetime frequency domain. In *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 2008.

[73] F. Pinto and M. Vetterli. Coding of spatio-temporal audio spectra using tree-structured directional filterbanks. In *IEEE Work. Appl. Signal Process. Audio and Acoustic*, 2009.

[74] F. Pinto and M. Vetterli. Near-field adaptive beamforming and source localization in the spacetime frequency domain. In *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 2010.

[75] C. Plack. *The Sense of Hearing*. Psychology Press, 2005.

[76] J. Princen and A. Bradley. Analysis/synthesis filter bank design based on time domain aliasing cancellation. *IEEE Trans. Acoustic, Speech, Signal Process.*, 34:1153–1161, 1986.

[77] J. Princen, A. Johnson, and A. Bradley. Subband/transform coding using filter bank designs based on time domain aliasing cancellation. In *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 1987.

[78] L. Rabiner, J. McClellan, and T. Parks. Fir digital filter design techniques using weighted chebyshev approximation. In *Proc. IEEE*, volume 63, 1975.

[79] J. Rosen. Researchers play tune recorded before edison. *The New York Times*, March 27, 2008.

[80] R. Sadek and C. Kyriakakis. A novel multichannel panning method for standard and arbitrary loudspeaker configurations. In *Audio Eng. Soc. 117th Convention*, 10 2004.

[81] T. Sarkar, R. Mailloux, A. Oliner, M. Salazar-Palma, and D. Sengupta. *History of wireless*. Wiley-IEEE Press, 2006.

[82] A. Sommerfeld. *Partielle Differentialgleichungen der Physik [Partial Differential Equations in Physics]*. Dieterich'sche Verlagsbuchhandlung, 1947.

[83] E. Torick. Highlights in the history of multichannel sound. *J. Audio Eng. Soc.*, 46(1/2):27–31, 1998.

[84] P. Vaidyanathan. *Multirate Systems And Filter Banks*. Prentice Hall, 1992.

[85] P. Vandewalle, J. Kovacevic, and M. Vetterli. Reproducible research in signal processing - what, why, and how. *IEEE Signal Process. Mag.*, 2009.

[86] M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice Hall, 1995.

[87] B. Widrow, K. Duvall, R. Gooch, and W. Newman. Signal cancellation phenomena in adaptive antennas: Causes and cures. *IEEE Trans. Antennas, Propagation*, 30:469–478, 1982.

[88] E. Williams. *Fourier Acoustics*. Academic Press, 1999.

[89] D. Yang, A. Hongmei, C. Kyriakakis, and C.-C. Kuo. High-fidelity multichannel audio coding with karhunen-loeve transform. *IEEE Trans. Speech, Audio Process.*, 11(4):365–380, 2003.

FRANCISCO M. PINTO

November 2010

HOME ADDRESS:
Av. Tir-Fédéral 92, Apt 423
1024 Ecublens
Switzerland

CONTACTS:
OFFICE: (+41) 216937594
MOBILE: (+41) 792808619
E-MAIL: francisco.pinto@epfl.ch

PERSONAL DETAILS:
DATE OF BIRTH: December 5, 1981
NATIONALITY: Portuguese

EDUCATION

2006 – 2010 — **SWISS FEDERAL INSTITUTE OF TECHNOLOGY** — Lausanne, Switzerland

PhD in Computer, Communication and Information Sciences;

THESIS: Spatio-temporal processing of acoustic wave fields;
SUPERVISOR: Prof. Martin Vetterli.

1999 – 2004 — **UNIVERSITY OF PORTO** — Porto, Portugal

Licenciatura (5-year degree) in Electrical and Computer Engineering;

CLASS RANK: 10th out of 160 (top 6%);
MAJOR: Telecommunication systems, electronics and computers;
THESIS: Wireless acoustic communications using spread spectrum techniques;
SUPERVISOR: Prof. Aníbal Ferreira.

PROFESSIONAL EXPERIENCE

Summer 2009 — **PHONAK AG / SONOVA** — Zürich, Switzerland

POSITION: Summer intern;
AREA OF ACTIVITY: Hearing aids technology.

ACADEMIC EXPERIENCE

Since 2006 — **SWISS FEDERAL INSTITUTE OF TECHNOLOGY** — Lausanne, Switzerland

POSITION: Teaching Assistant / Lecturer;
COURSES: Statistical Signal Processing and Applications (with Dr. Andrea Ridolfi);
Signal Processing for Speech, Audio and Acoustics (with Dr. Christof Faller);
Mathematical Principles of Signal Processing (with Dr. Andrea Ridolfi).

2004 - 2006 — **INSTITUTE FOR SYSTEMS AND COMPUTER ENGINEERING** — Porto, Portugal

POSITION: Research assistant;
AREAS OF ACTIVITY: Wireless acoustic communications; Digital equalization of room acoustics.

2003 — **INSTITUTE FOR BIOMEDICAL ENGINEERING** — Porto, Portugal

POSITION: Research assistant;
AREA OF ACTIVITY: Biomedical signal processing and computer vision.

AWARDS

2008 *EPFL Award for Outstanding Achievements*, awarded by the Swiss Federal Institute of Technology;

2008 *ICASSP 2008 Best Student Paper Award*, awarded by IEEE (among 1300 papers);

2006 *Fundação para a Ciência e Tecnologia Fellowship*, awarded by the Portuguese Government;

2006 *Calouste Gulbenkian Fellowship*, awarded by the Calouste Gulbenkian Foundation;

2005 *Top 10 Engineering Students Honour*, awarded by the University of Porto.

PUBLICATIONS

Journal / Long papers (more than 10 pages)

2010 F. Pinto, M. Vetterli, *"Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications"*, IEEE Transactions on Signal Processing; (published)

2008 F. Pinto, M. Vetterli, *"Bitstream format for spatio-temporal wave field coder"*, Audio Engineering Society 124th Convention; (published)

2006 A. Rocha, F. Pinto, A. Leite, A. Ferreira, *"Adaptive audio equalization of rooms based on a technique of transparent insertion of acoustic probe signals"*, Audio Engineering Society 120th Convention; (published)

Conference papers

2010 F. Pinto, M. Vetterli, *"Near-field adaptive beamforming and source localization in the spacetime frequency domain"*, IEEE International Conference on Acoustics, Speech, and Signal Processing; (published)

2009 F. Pinto, M. Vetterli, *"Coding of spatio-temporal audio spectra using tree-structured directional filterbanks"*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics; (published)

2008 F. Pinto, M. Vetterli, *"Wave field coding in the spacetime frequency domain"*, IEEE International Conference on Acoustics, Speech, and Signal Processing; **(published, with best student paper award)**

PATENTS

2009 F. Pinto, M. Vetterli, *"Audio wave field encoding"*, EPO Application EP09156817, filed March 31, 2009;

2008 F. Pinto, M. Vetterli, *"Audio wave field encoding"*, USPO Application 12058988, filed March 31, 2008.

TECHNICAL SKILLS

AREAS OF PROFICIENCY: Speech, signal, and image processing (theory and practice); filter banks theory; array signal processing; acoustic wave theory; artificial neural networks; perceptual audio coding;
OPERATING SYSTEMS: Linux and Windows;
COMPUTER LANGUAGES: Assembly, C/C++, Matlab/Simulink, Java, JavaScript, PHP, LaTeX, HTML/CSS.

SPOKEN LANGUAGES

PORTUGUESE (native language); **ENGLISH** (fluent); **FRENCH** (good); **SPANISH** (basics); **FINNISH** (basics).