T-Labs Series in Telecommunication Services

Jens Ahrens

Analytic Methods of Sound Field Synthesis

Springer

T-Labs Series in Telecommunication Services

Series Editors

For further volumes:
http://www.springer.com/series/10013

Jens Ahrens

Analytic Methods of Sound Field Synthesis

Springer

Jens Ahrens
Deutsche Telekom Laboratories
Technische Universität Berlin
Ernst-Reuter-Platz 7
10587 Berlin
Germany

ISSN 2192-2810 e-ISSN 2192-2829
ISBN 978-3-642-25742-1 e-ISBN 978-3-642-25743-8
DOI 10.1007/978-3-642-25743-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2011945029

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The present book summarizes the work that I have performed in the context of my doctoral dissertation and my subsequent activities at the Quality and Usability Lab, which is jointly run by the University of Technology Berlin and Deutsche Telekom Laboratories. The initial motivation for this work has been the question of how the two best-known methods of sound field synthesis, namely Wave Field Synthesis and Near-field Compensated Higher Order Ambisonics, relate. The answer to this question had been discussed in the research communities for years but a convincing conclusion had not been found. I present in this book a general formulation for the problem of sound field synthesis that allows for identifying above methods as particular solutions so that a juxtaposition is straightforward. Practical applications and synthesis of sound fields with diverse properties are then treated based on the general framework, which further facilitates the interpretation. The website http://www.soundfieldsynthesis.org accompanying this book makes available for download MATLAB/Octave scripts for all included simulations so that the reader can perform further investigations without having to start from scratch.

As with any book, the people who deserve acknowledgements are too numerous to list. I therefore mention only those who receive my very special acknowledgements. All others who have contributed to my research work and who are not mentioned here shall be aware of my appreciation.

Special thanks go to Sebastian Möller for putting immeasurable efforts in providing perfect working conditions and for giving me the freedom to work on the topic of sound field synthesis. And, of course, I thank him for reviewing my doctoral dissertation. Irene Hube-Achter's efforts have also contributed to a considerable extent to the pleasantness of my working conditions which I am also very thankful for.

Jens Blauert deserves general acknowledgements for exciting and inspiring discussions over the years; and he deserves special acknowledgements for reviewing my dissertation and for giving valuable comments and suggestions.

Frank Schultz has also given valuable comments on my dissertation.

I wish to thank all of my colleagues at Quality and Usability Lab, most notably Matthias Geier, Karim Helwani and Hagen Wierstorf of the audio technology group, Warcel Wältermann and Alexander Raake, and I wish to thank the management of Deutsche Telekom Laboratories for their support and enthusiasm for spatial audio.

The last and thus most important paragraph is dedicated to Sascha Spors who deserves most pronounced acknowledgments for various efforts including introducing me to the topic of sound field synthesis, guiding me through all these years that I have spent at Quality and Usability Lab and Deutsche Telekom Laboratories, and also for organizing my employment after a single phone call. And finally, I am especially thankful for the fact that we have shared and do still share so many of our interests and for the coincidence that brought us together.

Berlin, August 2011 Jens Ahrens

Contents

Symbols

c	speed of sound in air ($c = 343\text{m/s}$ is assumed)
i	imaginary unit, $i^2 = -1$
$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$	position vector in Cartesian coordinates (Appendix A)
$\mathbf{x}^T = [x\, y\, z]$	transposition of vector $\mathbf{x}$
$\|x\|$	absolute value (Weisstein 2002)
$\|\mathbf{x}\|$	vector norm (Weisstein 2002)
$\mathbf{e}_x$	unit vector pointing in x-direction
(α, β)	direction specified by azimuth α and colatitude β
arccos	inverse cosine (Weisstein 2002)
$\delta(\cdot)$	Dirac delta function
$\delta_{nn'}$	Kronecker Delta, defined in (2.26)
$\partial\Omega$	boundary enclosing volume Ω_i
Ω_i	volume enclosed by boundary $\partial\Omega$
Ω_e	domain exterior to boundary $\partial\Omega$
∇	gradient defined in (2.5) and (2.12)
$\Re\{\cdot\}$	real part
$\Im\{\cdot\}$	imaginary part
$G(\mathbf{x}, \mathbf{x}_0, \omega)$	Green's function for excitation at $\mathbf{x}_0$
$G_0(\mathbf{x} - \mathbf{x}_0, \omega)$	free-field Green's function for excitation at $\mathbf{x}_0$ given by (2.66)
$j_n(\cdot)$	n-th order spherical Bessel function (Arfken and Weber 2005)
$y_n(\cdot)$	n-th order spherical Neumann function (Arfken and Weber 2005)
$h_n^{(1,2)}(\cdot)$	n-th order spherical Hankel function of first and second kind defined in (2.15)
$Y_n^m(\beta, \alpha)$	spherical harmonic of n-th degree and m-th order, defined in (2.23)

$\mathring{S}_n^m(r,\omega)$	spherical harmonics expansion coefficients of sound field $S(\mathbf{x},\omega)$, defined in (2.31)
$\breve{S}_{n,i}^m(\omega)$ or $\breve{S}_n^m(\omega)$	interior expansion coefficients of sound field $S(\mathbf{x},\omega)$, defined in (2.32a)
$\breve{S}_{n,e}^m(\omega)$	exterior expansion coefficients of sound field $S(\mathbf{x},\omega)$, defined in (2.32b)
$\tilde{S}(k_x,k_y,z,\omega)$	sound field $S(\mathbf{x},\omega)$ considered in wavenumber domain with respect to k_x and k_y
$\check{S}(\cdot)$	angular spectrum representation of $S(\cdot)$, defined in (2.55a)
$\bar{S}(\cdot)$	signature function of sound field $S(\cdot)$, defined in (2.45)
$P_n^m(\cdot)$	associated Legendre function of n-th degree and m-th order (Gumerov and Duraiswami 2004)
S_R^2	2-sphere (i.e. a spherical surface) of radius R (Weisstein 2002)
S_R^1	1-sphere (i.e. a circle) of radius R (Weisstein 2002)
$(I\|I)_{n\,n'}^{m\,m'}(\cdot)$	translation coefficient for interior-to-interior translation
$(E\|I)_{n\,n'}^{m\,m'}(\cdot)$	translation coefficient for exterior-to-interior translation
sinc x	sinus cardinalis, sinc x $= \sin(\pi x)/(\pi x)$
$\langle\cdot\rangle$	inner product (Weisstein 2002)
$\gamma_{n_1,n_2,n}^{m_1,m_2,m}$	Gaunt coefficient, defined in (D.6)
$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$	Wigner 3j-Symbol as defined in (Weisstein 2002)
$\mathcal{E}\begin{pmatrix} m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{pmatrix}$	E-symbol, defined in (D.7)
${}_3F_2(\cdot)$	generalized hypergeometric function (Arfken and Weber 2005)
$(\cdot)!$	factorial (Weisstein 2002)
$(\cdot)!!$	double factorial (Weisstein 2002)
$\frac{\partial}{\partial\mathbf{n}}$	gradient in direction $\mathbf{n}$, refer to (2.61)
$dA(\mathbf{x}_0)$	infinitesimal surface element
$S(\omega)^*$	complex conjugate of $S(\omega)$
$\circ\!\!-\!\!\!-\!\!\bullet$	Fourier transform
$\bullet\!\!-\!\!\!-\!\!\circ$	inverse Fourier transform
$\lfloor x \rfloor$	floor function, gives the largest integer not greater than x (Weisstein 2002)
$\lceil x \rceil$	ceiling function, gives the smallest integer not smaller than x (Weisstein 2002)
$\mathrm{sign}(x)$	signum function, defined in (5.73)
$J_m(\cdot)$	Bessel function of m-th order (Weisstein 2002)
$H_m^{(2)}(\cdot)$	m-th order Hankel function of second kind (Weisstein 2002)

Acronyms

ASW	Apparent Source Width
ASDF	Audio Scene Description Format
BRIR	Binaural Room Impulse Response
BRTF	Binaural Room Transfer Function
FIR	Finite Impulse Response
FOS	First-Order Section
HOA	Higher Order Ambisonics
HRIR	Head-related Impulse Response
HRTF	Head-related Transfer Function
IIR	Infinite Impulse Response
LEV	Listener Envelopment
NFC-HOA	Near-field Compensated Higher Order Ambisonics
SDM	Spectral Division Method
SOS	Second-Order Section
SpatDIF	Spatial Sound Description Interchange Format
WFS	Wave Field Synthesis

Chapter 1
Introduction

1.1 Nomenclature

The notational conventions employed in this book are outlined in the following.

For scalar variables, lower case denotes time domain and upper case denotes time-frequency domain, e.g., $s(\mathbf{x}, t)$ vs. $S(\mathbf{x}, \omega)$. f denotes the time frequency, which is related to the radian frequency ω via $\omega = 2\pi f$.

Vectors are denoted by lower case boldface, e.g., $\mathbf{k}$. The three-dimensional position vector in Cartesian coordinates is given as $\mathbf{x} = [x\ y\ z]^T$; the coordinate systems employed are presented in Appendix A. The definition of the Fourier transform is outlined in Appendix B.

When it is referred to a *sound field* $s(\mathbf{x}, t)$ in this book, it is referred to the *sound pressure*, i.e., the *local pressure deviation* from the ambient pressure (in the present case the atmospheric pressure) caused by a sound wave. The SI (*Système international d'unités*) unit of sound pressure is the *pascal* ($1\,\text{Pa} = 1\,\text{N/m}^2$) (Bureau International des Poids et Mesures 2006). The time-frequency spectrum of a sound field $S(\mathbf{x}, \omega)$, i.e., the spectral amplitude density of $s(\mathbf{x}, t)$, is thus given in $\text{Pa} \cdot \text{s}$ or Pa/Hz respectively (Girod et al. 2001). For convenience, $S(\mathbf{x}, \omega)$ is referred to as a "sound field represented in time-frequency domain" or "sound pressure in time-frequency domain" in this book.

Angles are given in radians if not indicated as different. When a quantity is given in a logarithmic scale, the reference value is always the underlying unit.

The following two examples of a plane wave and a spherical wave sound field illustrate further notational conventions. The sound pressure deviation $S_{\text{pw}}(\mathbf{x}, \omega)$ in time-frequency domain caused by a plane wave sound field propagating in direction $\mathbf{k}_{\text{pw}}$ is given by (Williams 1999, Eq. (2.24), p. 21)

$$S_{\text{pw}}(\mathbf{x}, \omega) = \hat{S}_{\text{pw}}(\omega) e^{-i\mathbf{k}_{\text{pw}}^T \mathbf{x}}, \tag{1.1}$$

J. Ahrens, *Analytic Methods of Sound Field Synthesis*,
T-Labs Series in Telecommunication Services, DOI: 10.1007/978-3-642-25743-8_1,

with

$$\mathbf{k}_{\mathrm{pw}}^T = [k_{\mathrm{pw},x}\ k_{\mathrm{pw},y}\ k_{\mathrm{pw},z}] \tag{1.2}$$

$$= k_{\mathrm{pw}} \cdot [\cos\theta_{\mathrm{pw}} \sin\phi_{\mathrm{pw}} \quad \sin\theta_{\mathrm{pw}} \sin\phi_{\mathrm{pw}} \quad \cos\phi_{\mathrm{pw}}] \tag{1.3}$$

and $(\theta_{\mathrm{pw}}, \phi_{\mathrm{pw}})$ being the propagation direction of the plane wave in spherical coordinates. i denotes the imaginary unit $\left(i^2 = -1\right)$.

The right hand side of (1.1) is composed of two components:

1. A time-frequency component $\hat{S}_{\mathrm{pw}}(\omega)$, which represents the information with respect to time such as a sine wave or a music signal.
2. A spatial transfer function $e^{-i\mathbf{k}_{\mathrm{pw}}^T \mathbf{x}}$ representing the spatial structure of the sound field.

The spatial transfer function $e^{-i\mathbf{k}_{\mathrm{pw}}^T \mathbf{x}}$ as used in (1.1) is of dimension 1 so that $\hat{S}_{\mathrm{pw}}(\omega)$ has to be of the unit Pa/Hz in order that (1.1) is correct.

Now consider an outgoing spherical wave sound field $S_{\mathrm{sw}}(\mathbf{x}, \omega)$ originating from the coordinate origin given by

$$S_{\mathrm{sw}}(\mathbf{x}, \omega) = \hat{S}_{\mathrm{sw}}(\omega) \frac{e^{-i\frac{\omega}{c}r}}{r}. \tag{1.4}$$

The spatial transfer function in (1.4) is of unit 1/m so that $\hat{S}_{\mathrm{sw}}(\omega)$ has to be of unit Ns/m.

This inconsistency is a consequence of the simplifying notational conventions. In order to explicitly account for the physical meaning of the involved functions, quantities like the density of the medium in which the sound wave propagates and alike have to be considered explicitly (Williams 1999). For notational simplicity, this book employs the convention applied widely in the scientific literature of exclusively considering the spatial transfer function of a given sound field or similar quantity under consideration neglecting the time information as well as constant factors. The explicit composition of the involved time components such as $\hat{S}_{\mathrm{pw}}(\omega)$ and $\hat{S}_{\mathrm{sw}}(\omega)$ is not relevant in the presented investigation and is therefore not treated. The reader is referred to (Williams 1999).

As is common in electrical engineering, complex notation is used for purely real harmonic time-domain signals (Girod et al. 2001). I.e., a unit amplitude cosine wave $\hat{s}_{\cos}(t)$ of radian frequency ω_0 is notated as

$$\hat{s}_{\cos}(t) := e^{i\omega_0 t}. \tag{1.5}$$

The actual time-domain signal is then obtained by considering exclusively the real part of $\hat{s}_{\cos}(t)$ as

$$\Re\{\hat{s}_{\cos}(t)\} = \cos(\omega_0 t). \tag{1.6}$$

1.2 A Brief Overview of Audio Presentation Methods

The concept of *sound field synthesis* as presented in this book constitutes one of a number of rather recent advancements in the field of the spatial audio presentation. The term *spatial audio* has become popular in this context, though a commonly accepted definition does not exist. The usage covers all ranges from the reference to an audio signal that contains spatial information to the concept of *Gestalt* perception (Bregman 1990).

Note that some authors prefer the term *audio reproduction* or *sound reproduction*, e.g., (Toole 2008, pp. 3–4). As discussed ibidem, "reproduction" implies recreation of some sort. The methods presented in this book go beyond recreation or imitation of a given perceptual quantity and allow for the generation of—primarily spatial—information to a considerable extent. Therefore, the term *audio presentation* is used in this book, which is considered a more general term that also covers reproduction.

Since the invention of the telephone, the first electro-acoustic communication device patented by Alexander Graham Bell in 1876 (Bell 1876), a great variety of audio presentation methods both headphone-based and loudspeaker-based have evolved. A brief yet incomplete overview of these methods will be given with a focus on those branches of the evolution that lead to the proposition of sound field synthesis.

Due to the single loudspeaker that is employed in the telephone, only *monaural* auditory cues such as timbre and perceived distance can be controlled, a circumstance that limits the presentable spatial information (Blauert 1997). As early as in 1881, two parallel telephone channels were used in order to transmit performances from the Paris Opera House (du Moncel 1881; Torick 1998). The service was commercialized a few years later and termed *Théâtrophone*. The enabled provision of controllable *binaural* auditory cues essentially extended the transmittable spatial information.

Occasionally, some of the physical theory presented in later chapters is anticipated though without detailed outline in order not to distract the reading flow. It is recommended that the interested reader revisits this chapter after familiarizing with the details of the theory outlined later on.

1.2.1 Audio Presentation Based on Head-Related Transfer Functions

The acoustical properties of the human body, most notably the head and the outer ears, are imitated by the presentation system in order to create auditory events with specific spatial attributes. These acoustical properties are described by *head-related transfer functions* (HRTFs) and are individual (Blauert 1997). Typically, a given scene is recorded with a mannequin or a person with ear-mounted microphones, or HRTFs obtained from measurements are imposed on the signals (Hammershøi and Møller 2002). The approach is also referred to as *binaural reproduction* or