LOCAL SOUND FIELD SYNTHESIS

Dissertation
zur
Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)
der Fakultät für Informatik und Elektrotechnik
der Universität Rostock

vorgelegt von
M.Sc. Fiete Winter,
geb. am 29.11.1988 in Rostock
aus Rostock

Gutachter:
Prof. Dr.-Ing. Sascha Spors, Universität Rostock
Prof. Dr.-Ing. Alexander Raake, Technische Universität Ilmenau
Prof. Dr.-Ing. Jens Ahrens, Chalmers University of Technology

Datum der Einreichung:	25.06.2019
Datum der Annahme:	23.09.2019
Datum der Verteidigung:	17.10.2019
Datum der Endversion:	18.11.2019

Digital Object Identifier (DOI): 10.18453/rosdok_id00002568

Contents

1 Introduction

1.1 Overview

One of the outstanding capabilities of the human auditory system is to recover information on acoustic scenarios out of a mixture of sounds.[1] With only the two ears as their acoustic sensors, humans are able to analyse complex scenes and form auditory objects[2] from the ear signals. This internal representation allows to interpret and potentially focus on particular objects. The cocktail-party effect[3] is a well-known example, where the listener focuses the attention to a target speaker and actively segregates it from its interfering milieu[4]. The remarkable degree of complexity which humans are able to resolve also implies challenges for the presentation of spatial audio content via electronic devices. Electro-acoustic transducers such as headphones or loudspeaker are used to provide the listener with selected aspects of a desired scenario. The presentation has to be of sufficient level of detail in order to invoke the intended perceptual impression.

Presenting spatial audio content via loudspeakers has a long history and is well-established in everyday life for decades. According to the reviews of Rumsey,[5] and Ahrens,[6] the traditional two-channel stereophony was patented in the 1930s by Blumlein.[7] A typical stereo setup is depicted in Fig. 1.1a. The method was extended towards Ambisonics[8] and Quadraphony[9] to augment the spatial impression with a larger variety of sound directions. Five-channel Surround Sound (5.0) with an optional low-frequency channel (5.1) constitutes another extension and is standardised by the International Telecommunication Union (ITU).[10] In other reproduction techniques additional loudspeakers are placed at layers above and below the listeners to present three-dimensional (3D) sound. The number of suggested loudspeaker layouts is vast.[11] A well-established method for arbitrary 3D geometries is Vector Based Amplitude Panning (VBAP).[12]

The mentioned techniques heavily rely on certain properties of the human auditory system and the perception of sound connected to it. For example, stereo panning exploits the effect of summing localisation:[13] By adjusting the amplitude or the delay between the two loudspeaker signals, the perceived direction of the so-called phantom source is shifted in between the loudspeakers. However, the exploited mechanism does only lead to the desired spatial perception

[1] Cherry (1953). "Some Experiments on the Recognition of Speech, with One and with Two Ears". In: *J. Acoust. Soc. Am.* 25.5, pp. 975–979.

[2] Bregman (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT press, Ch. 1.

[3] Cherry, op. cit., Sec. 2.

[4] Hawley et al. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer". In: *J. Acoust. Soc. Am.* 115.2, pp. 833–843, Sec. I.

[5] Rumsey (2001). *Spatial Audio*. Focal Press, Sec. 1.4.

[6] Ahrens (2010). "The Single-layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources". PhD thesis. Berlin, Germany: Technische Universität Berlin, Sec. 1.1.

[7] Blumlein (1931). *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems*. British Patent Specification 394,325.

[8] Gerzon (1973). "Periphony: With-Height Sound Reproduction". In: *J. Aud. Eng. Soc.* 21.1, pp. 2–10.

[9] Torick (1998). "Highlights in the History of Multichannel Sound". In: *J. Aud. Eng. Soc.* 46.1/2, pp. 27–31.

[10] *Multichannel stereophonic sound system with and without accompanying picture* (Aug. 2012). Standard ITU-R BS.775-3. International Telecommunication Union Radiocommunication Assembly.

[11] *Advanced sound system for programme production* (July 2018). Standard ITU-R BS.2051-2. International Telecommunication Union Radiocommunication Assembly.

[12] Pulkki (1997). "Virtual Sound Source Positioning Using Vector Base Amplitude Panning". In: *J. Aud. Eng. Soc.* 45.6, pp. 456–466.

[13] Blauert (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press, Sec. 3.3.1.

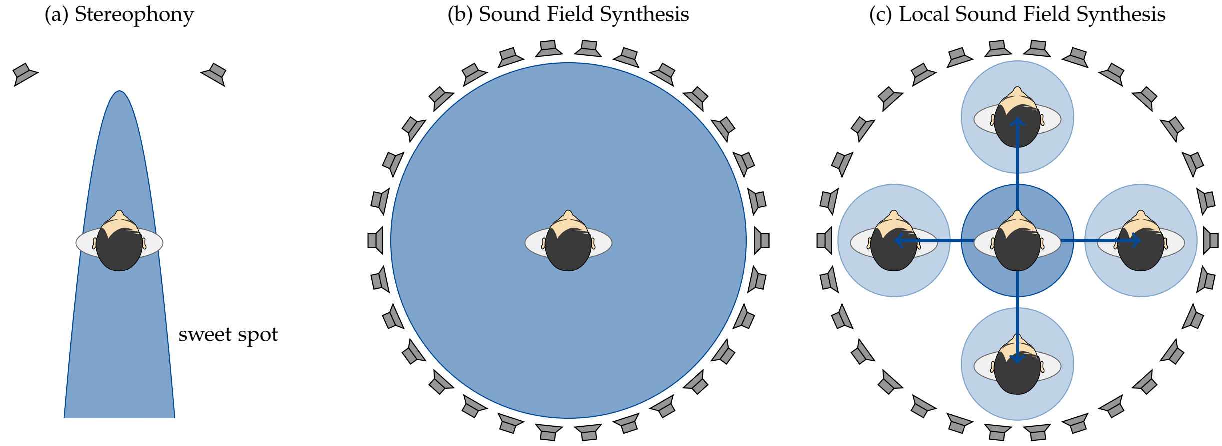


Figure 1.1: Comparison of different loudspeaker reproduction scenarios. The blue area marks the target region, where the desired perceptual impression is pursued to be invoked correctly. For Local Sound Field Synthesis, this region can be shifted according to the position of the listener. The figure is adapted from Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Fig. 1.4 and Ahrens (2012). *Analytic Methods of Sound Field Synthesis*. T-Labs Series in Telecommunication Services. Berlin Heidelberg, Germany: Springer-Verlag, Fig. 1.1.

for a very limited area, where the listener is supposed to be situated in front of the loudspeakers. In the context of stereophony and surround sound, it is known as the sweet spot. With the growing number of employed loudspeakers it becomes feasible to pursue a physically accurate synthesis of a desired sound field within an extended target region. It is hereby assumed that humans cannot distinguish between the desired sound field and its perfect physical reconstruction. Presentation methods following this paradigm are commonly subsumed under the term Sound Field Synthesis (SFS). Two well-known representatives are Near-Field-Compensated Higher-Order Ambisonics (NFCHOA)[14] and Wave Field Synthesis (WFS).[15] An exemplary scenario is depicted in Fig. 1.1b. For an accurate synthesis, a distribution of densely spaced loudspeakers surrounding the target region is necessary. The required distance between two adjacent loudspeakers is inversely proportional to the highest frequency for which correct synthesis is supposed to be achieved. The audible range up to 20 kHz demands a distance of approximately 1 cm.[16] Violating this condition potentially leads to spatial aliasing impairing the accuracy. Such small distances cannot be realised with today's loudspeaker technology as the minimal distance is constrained by the cabinet size. Moreover, an extended target region with a boundary of several metres length leads to a high logistical and financial effort including a vast number of transducers, amplifiers, and digital-to-analog (D/A) converters. Thus, most SFS systems employ up to hundreds of loudspeakers and typically achieve an artefact-free synthesis with an upper frequency limit of 1 kHz to 2 kHz.

A compromise between the available listening area and the synthesis accuracy is made in Local Sound Field Synthesis (LSFS):[17] A more accurate reproduction inside an area which is smaller than the area surrounded by the loudspeakers is pursued. Stronger artefacts outside the prioritised area are permitted. It is shown in Fig. 1.1c, that LSFS is useful for applications, where the listener's position is restricted to a small region of interest or is tracked using a suitable technology. As various approaches to LSFS exist, an in-depth discussion on them will be given in Ch. 3.

[14] Daniel (May 2003). "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format". In: *Proc. of 23rd Intl. Aud. Eng. Soc. Conf. on Signal Processing in Audio Recording and Reproduction*. Copenhagen, Denmark.

[15] Berkhout (1988). "A Holographic Approach to Acoustic Control". In: *J. Aud. Eng. Soc.* 36.12, pp. 977–995.

[16] based upon the half-wavelength criterion, exempli gratia (e.g.) Van Trees (2004). *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons, Eq. (2.130).

[17] Local Sound Field Synthesis was initially introduced as a term in Spors and Ahrens (Oct. 2010b). "Local Sound Field Synthesis by Virtual Secondary Sources". In: *Proc. of 40th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Tokyo, Japan, Sec. 1.

Although the mentioned techniques for spatial audio presentation are differently motivated by physics and/or perception, their performance has to be finally assessed and compared based on the quality judgement given by the listener. Typical terms used in conjunction with audio technology are sound quality and Quality of Experience (QoE).[18] Whereas sound quality restricts investigations to the perceivable influence of the technical system, QoE takes the whole listening experience into account. According to Raake and Wierstorf, it is challenging to directly assess QoE due to its holistic character and most studies focus on sound quality. Blauert and Jekosch[19] identified the concepts of plausibility and authenticity as closely related to sound quality. Plausibility is related to the internally build-up expectation of the listener and how well the presentation agrees with her or his implicit reference. Asking for the authenticity of a sound presentation method implies the provision of an explicit reference. If the listener is not able to distinguish between the presentation and the reference, the method can be regarded as authentic or transparent. Authenticity may be investigated using the paradigm of fidelity,[20] where different aspects or attributes of an audio reproduction are assessed separately. In the context of surround sound, Rumsey et al.[21] identified spatial and timbral fidelity as significant aspects of the overall sound quality. The perceived quality of (L)SFS is also subject to recent research: Wittek et al.[22,23] compared stereophony and WFS with respect to (w.r.t.) spatial and timbral fidelity. Wierstorf[24] investigated the horizontal localisation in NFCHOA and WFS as an aspect of spatial fidelity. Further, he conducted experiments to assess the timbral fidelity of WFS.[25,26]

[18] Raake and Wierstorf (Sept. 2016). "Assessment of audio quality and experience using binaural-hearing models". In: *Proc. of 22nd Int. Congress on Acoustics (ICA)*. Buenos Aires, Argentina, Sec. 1.

[19] Blauert and Jekosch (Aug. 2003). "Concepts Behind Sound Quality: Some Basic Considerations". In: *Proc. of International Congress and Exposition on Noise Control Engineering (INTERNOISE)*. Seogwipo, Korea.

[20] Raake and Wierstorf, loc. cit.

[21] Rumsey et al. (2005). "On the relative importance of spatial and timbral fiedelities in judgements of degraded multichannel audio quality". In: *J. Acoust. Soc. Am.* 118.2, pp. 968–976.

[22] Wittek (2007). "Perceptual differences between wavefield synthesis and stereophony". PhD thesis. University of Surrey.

[23] Wittek et al. (Oct. 2007). "On the Sound Color Properties of Wavefield Synthesis and Stereo". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA.

[24] Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Sec. 5.1.

[25] Wierstorf et al. (Aug. 2014). "Coloration in Wave Field Synthesis". In: *Proc. of 55th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Helsinki, Finland.

[26] Wierstorf, op. cit., Sec. 5.2.

1.2 Goals and Structure

The goal of this thesis is to investigate the physical and perceptual properties of selected methods for LSFS and compare them to conventional SFS. As LSFS potentially enhances the synthesis accuracy around the listener's position, the question arises, whether this has an positive effect on aspects related to sound quality as well. Moreover, this work focusses on how technical parameters such as the position of the listener or the size of target region in LSFS influences physics and perception. Although panning techniques such as Ambisonics[27] and VBAP[28] are not explicitly considered within the present work, results from the literature are incorporated into the discussions whenever it is suitable.

[27] Gerzon, op. cit.

[28] Pulkki, op. cit.

In Ch. 2, the mathematical foundations of sound propagation in linear acoustics are introduced. The integral notation of the linearised wave equation will be revisited in particular since it builds the basis for (L)SFS. As a key concept for the later discussions on the physical properties, the local wavenumber vector[29] will be introduced.

[29] Firtha et al. (2017). "Improved Referencing Schemes for 2.5D Wave Field Synthesis Driving Functions". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 25.5, pp. 1117–1127, Eq. (15).

The fundamental problem of (L)SFS is verbally and mathematically formulated in Ch. 3. WFS and NFCHOA are revisited as representatives of the conventional solution to it. They will later be used as the baseline for the comparison. Based upon prior discussions on the benefits and drawbacks of different LSFS strategies, Local Wave Field Synthesis using Spatial Bandwidth Limitation (LWFS-SBL)[30] and Local Wave Field Synthesis using Virtual Secondary Sources (LWFS-VSS)[31] are selected for further investigations. For all methods, special attention is drawn to the implementation of the methods in the discrete-time domain as a consequence of digital signal processing. The influence of the parametrisation of the mentioned techniques on the properties of the synthesised sound fields is examined on a qualitative level.

[30] Hahn et al. (June 2016). "Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain". In: *Proc. of 140th Aud. Eng. Soc. Conv.* Paris, France.

[31] Spors and Ahrens, op. cit.

Several theoretical treatises covered the trade-off between the spatial extent of the listening area, the number of actuators, and the frequency up to which artefact-free synthesis is possible in (L)SFS. For example, Kennedy et al.[32] derived lower bounds for the mentioned frequency assuming arbitrary sound fields. If and how the synthesis accuracy can be enhanced by incorporating additional knowledge about the desired scenario, remains an open research question. Ch. 4 introduces a geometric model, which predicts the artefact-free frequency bound as a function of the mentioned dependencies. The model is further used to compare the selected (L)SFS methods w.r.t. their physical properties.

[32] Kennedy et al. (2007). "Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields". In: *IEEE Trans. Signal Process.* 55.6, pp. 2542–2556.

The perceptual evaluation of the approaches is presented in Ch. 5 and Ch. 6. For spatial fidelity, the azimuthal localisation of a point source as the desired sound field is investigated. Timbral fidelity is assessed by measuring the perceived colouration of the synthesis with the point source as the reference. A summary of the thesis is given in Ch. 7.

1.3 Open Science, Reproducibility, and External Resources

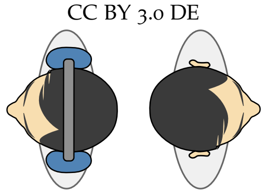


Figure 1.2: Pictograms of listener with and without headphones.

The success of scientific research in general heavily relies on its credibility and acceptance in the broader society. In the past years, striking terms as *fake news*[33] and *fake science*[34] had emerged as synonyms for the intentional misinformation eroding the trust in reliable research. Making scientific results publicly accessible and reproducible is one building block to counteract this process. As a publicly financed researcher, the author also sees a moral obligation to the public. The provision of means to reproduce the published results should moreover be common practice. It allows other researchers to comprehend and validate the outcome of the research. Most of the present work relies on the implementation of signal processing chains as computer programs, which is by definition prone to errors. Thus, *with some exceptions, anything less than the release of source programs is intolerable for results that depend on computation.*[35]

[33] Lazer et al. (2018). "The science of fake news". In: *Science* 359.6380, pp. 1094–1096.

[34] Eckert and Hornung (July 2018). *Fake Science – Die Lügenmacher*. Das Erste. television documentary.

[35] Ince et al. (Feb. 2012). "The case for open computer programs". In: *Nature* 482.7386, pp. 485–488.

The approach towards the mentioned aspects is very much inspired by the work of Wierstorf:[36] Several features of the present document rely on hyperlinks allowing to navigate between different

[36] Wierstorf, op. cit.

parts of thesis and providing access to external resources. Thus, the author highly recommends the usage of the electronic version of this document for maximum comfort. For all figures, the symbol provides a link to the according folder of the code repository[37] containing the necessary scripts and data to reproduce the figure. All (L)SFS algorithms presented in Ch. 3 are implemented – partly by the author – in the Sound Field Synthesis Toolbox for MATLAB.[38] It is publicly available either in the code repository or as a citable snapshot of the used version.[39] Whenever possible, a link to the distinct function of the toolbox implementing a currently discussed feature is given as a side note. The perceptive model used to extract auditory features from binaural signals in Ch. 5 and Ch. 6 is part of the Two!Ears Auditory Model.[40] The layout of the thesis is based on a LaTeX class which mimics the style of the books of Tufte,[41] whereas the style for chapter headings stems from Firtha.[42] The pictograms for a listener with and without headphones depicted in Fig. 1.2 are part of the thesis of Wierstorf. All colormaps used in the plots are published by Brewer and are implemented by Schneider and Wierstorf.[43,44]

[37] snapshot available at Winter (2019). *Local Sound Field Synthesis – Supplements 1.0.1*. DOI: 10.5281/zenodo.3543537.

[38] Wierstorf and Spors (Apr. 2012). "Sound Field Synthesis Toolbox". In: *Proc. of 132nd Aud. Eng. Soc. Conv.* Budapest, Hungary.

[39] Sound Field Synthesis Toolbox
Wierstorf et al. (Mar. 2019). *Sound Field Synthesis Toolbox 2.5.0*. DOI: 10.5281/zenodo.2597212

[40] Two!Ears Auditory Model
Two!Ears Team (Oct. 2018). *Two!Ears Auditory Model 1.5*. DOI: 10.5281/zenodo.1458420

[41] tufte-latex
e.g. Tufte (May 2006). *Beautiful Evidence*. 1st ed. Cheshire, Connecticut: Graphics Press, LLC

[42] Firtha (2019). "A Generalized Wave Field Synthesis Framework with Application for Moving Virtual Sources". PhD thesis. Budapest University of Technology and Economics.

[43] colorbrewer2.org
Brewer (2005). *Designing Better Maps: A Guide for GIS Users*. New York, USA: ESRI Press

[44] Schneider and Wierstorf (June 2014). *Gnuplot-colorbrewer: ColorBrewer color schemes for gnuplot*. DOI: 10.5281/zenodo.10282.

1.4 Mathematical Preliminaries

A position vector $\mathbf{x}$ in the three-dimensional, right-hand coordinate system is defined by its Cartesian (x, y, z), its cylindrical (ρ, ϕ, z) or its spherical representation (r, ϑ, ϕ), see Fig. 1.3. The orthogonal distance of the vector from the z-axis is denoted by ρ, while ϕ describes the azimuth angle between the x-axis and the projection of $\mathbf{x}$ onto the xy-plane. The 2-norm of $\mathbf{x}$, id est (i.e.) its distance to the coordinates' origin, is given as $r = |\mathbf{x}|$. Its polar angle is denoted as ϑ. These representations are connected via

$$x = \rho \cos\phi = r \sin\vartheta \cos\phi\,, \tag{1.1a}$$
$$y = \rho \sin\phi = r \sin\vartheta \sin\phi\,, \tag{1.1b}$$
$$z = z \qquad = r \cos\vartheta\,. \tag{1.1c}$$

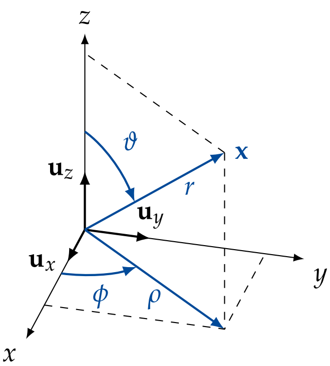


Figure 1.3: Coordinate representations

In general, column vectors are assumed in Cartesian coordinates, i.e. $\mathbf{x} = [x, y, z]^{\mathrm{T}}$, with T denoting the transposition operator. Elements of a vector are denoted using the same subscripted indices of their corresponding vector. For example, x_i, y_i, and z_i belong to $\mathbf{x}_i$. The scalar product of two vectors is given by the notation[45]

$$\langle \mathbf{x}_1 | \mathbf{x}_2 \rangle := x_1 x_2 + y_1 y_2 + z_1 z_2\,. \tag{1.2}$$

A special case of the scalar product is the squared 2-norm $|\mathbf{x}|^2 = \langle \mathbf{x} | \mathbf{x} \rangle$. In Cartesian coordinates, the gradient of a scalar function f w.r.t. $\mathbf{x}$ is defined as[46,47]

$$\nabla_{\mathbf{x}} f(\mathbf{x}) := \frac{\partial f(\mathbf{x})}{\partial x}\mathbf{u}_x + \frac{\partial f(\mathbf{x})}{\partial y}\mathbf{u}_y + \frac{\partial f(\mathbf{x})}{\partial z}\mathbf{u}_z\,. \tag{1.3}$$

The unit vectors along the coordinate axes are denoted as $\mathbf{u}_x$, $\mathbf{u}_y$, and $\mathbf{u}_z$. The directional derivative of f along an unit vector $\mathbf{n}$

$$\nabla_{\mathbf{x},\mathbf{n}} f(\mathbf{x}) := \langle \nabla_{\mathbf{x}} f(\mathbf{x}) | \mathbf{n} \rangle \tag{1.4}$$

[45] Arfken and Weber (2005). *Mathematical Methods for Physicists*. 6th ed. Elsevier, Eq. (1.24).

[46] Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, Eq. (2.4).

[47] Arfken and Weber, op. cit., Eq. (1.61).

is related to the gradient via the scalar product. The Laplace operator w.r.t. $\mathbf{x}$ is defined as[48,49]

$$\nabla_{\mathbf{x}}^2 f(\mathbf{x}) := \frac{\partial^2 f(\mathbf{x})}{\partial x^2} + \frac{\partial^2 f(\mathbf{x})}{\partial y^2} + \frac{\partial^2 f(\mathbf{x})}{\partial z^2}\,. \tag{1.5}$$

48 Williams, op. cit., p. 15.

49 Arfken and Weber, op. cit., Eq. (1.81a).

2 Fundamentals of Sound Fields

This chapter introduces the reader to sound propagation in linear acoustics which builds the basis for SFS. As a well-established model for sound propagation in air, the linearised wave equation is presented in Sec. 2.1. Different alternatives to represent the sound fields fulfilling the wave equation are discussed in Sec. 2.2. They will later be used to derive selected methods for SFS. Finally, Sec. 2.3 focusses on the definition of the local wavenumber vector as one of the key concepts for the calculi in the subsequent chapters.

2.1 The Linearised Wave Equation

The theoretical foundation of wave propagation in linear acoustics is given by the linearised wave equation, which will be simply termed wave equation in the course of this thesis. As the term linearised already suggests, it derives from more general, non-linear principles of fluid dynamics, namely the equations of mass, momentum, energy, and state.[1] Several assumptions especially about the medium (air) in which the waves propagate and the amplitude of the waves have to be made in order to achieve linearisation. For a more detailed discussion, the reader is referred to Blackstock[2] or Pierce.[3] The three upcoming sections discuss the wave equation in its differential and integral representation together with selected solutions.

[1] Blackstock (2000). *Fundamentals of physical acoustics*. John Wiley & Sons, p. 83.

[2] Ibid., Cha. 2.

[3] Pierce (2014). "Basic Linear Acoustics". In: *Springer Handbook of Acoustics*. Ed. by Thomas D Rossing. 2nd ed. Springer, pp. 29–115, Sec. 3.3.

2.1.1 Differential Formulation

The inhomogeneous wave equation is defined in its differential form as[4]

$$\nabla^2_{\mathbf{x}}\, p(\mathbf{x},t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(\mathbf{x},t) = -q_p(\mathbf{x},t)\,, \tag{2.1}$$

where the Laplace operator $\nabla^2_{\mathbf{x}}$ is defined by (1.5). The speed of sound is denoted as c and is fixed to 343 m/s for all simulations within this thesis. The position- and time-dependent sound pressure field $p(\mathbf{x},t)$ has to fulfil this equation in order to be a valid model in terms of linear acoustics. The source density $q_p(\mathbf{x},t)$ describes possibly existing sound sources also known as (a.k.a.) inhomogenities. If the density is zero everywhere in the 3D space $\mathbb{R}^3$, Eq. (2.1) is generally referred to as the homogeneous wave equation.[5]

[4] Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, Eq. (8.39).

[5] Ibid., Eq. (2.1).

The temporal Fourier Transform for a time-signal $f(t)$ is defined as[6]

[6] Girod et al. (2001). *Signal and Systems*. Wiley, Eq. (9.1).

$$F(\omega) := \int_{-\infty}^{\infty} f(t)\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t\,, \tag{2.2}$$

whereas $F(\omega)$ denotes the according Fourier spectrum. It should be noted, that other treatises[7,8] use a different convention with a plus-sign in the exponential. The angular frequency ω is related to the temporal frequency f via $\omega = 2\pi f$. Along with its differentiation theorem,[9] the transform is applied to Eq. (2.1) to obtain an expression for the wave equation in the temporal frequency domain:[10]

$$\nabla^2_{\mathbf{x}} P(\mathbf{x},\omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{x},\omega) = -Q_P(\mathbf{x},\omega)\,. \tag{2.3}$$

The Fourier spectra of the sound pressure field and the source density are denoted as $P(\mathbf{x},\omega)$ and $Q_P(\mathbf{x},\omega)$, respectively. Eq. (2.3) is generally referred to as the inhomogeneous Helmholtz equation.[11] Again, $Q_P(\mathbf{x},\omega)$ equals zero in the homogeneous case. Although (2.1) and (2.3) are equivalent, most of the upcoming calculi are carried out in the temporal frequency domain as the absence of the derivative w.r.t. time allows for more convenient derivations.

The general solution to (2.3) is given by the sum of the particular solution to (2.3) and the general solution to the respective homogeneous equation.[12] A versatile tool to express arbitrary solutions of the inhomogeneous wave equation in the absence of boundary conditions is the 3D free-field Green's function. With the used convention for the Fourier Transform in (2.2), it is given as[13]

$$G(\mathbf{x}|\mathbf{x}_\mathrm{s},\omega) = \frac{\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_\mathrm{s}|}}{4\pi|\mathbf{x}-\mathbf{x}_\mathrm{s}|}\,. \tag{2.4}$$

Its according density reads

$$Q_G(\mathbf{x}|\mathbf{x}_\mathrm{s},\omega) = \delta(\mathbf{x}-\mathbf{x}_\mathrm{s}) := \delta(x-x_\mathrm{s})\,\delta(y-y_\mathrm{s})\,\delta(z-z_\mathrm{s}) \tag{2.5}$$

with $\delta(\cdot)$ denoting the Dirac delta distribution.[14] It can be interpreted as an impulse-like excitation located at $\mathbf{x}_\mathrm{s}$.[15] The resulting specialised inhomogeneous Helmholtz equation reads[16]

$$\nabla^2_{\mathbf{x}} G(\mathbf{x}|\mathbf{x}_\mathrm{s},\omega) + \left(\frac{\omega}{c}\right)^2 G(\mathbf{x}|\mathbf{x}_\mathrm{s},\omega) = -\delta(\mathbf{x}-\mathbf{x}_\mathrm{s})\,. \tag{2.6}$$

The free-field Green's function in (2.4) is only one particular solution to (2.6) because solutions to the homogeneous Helmholtz equation may be added in order to form the general solution. According to Williams,[17] the sifting property[18,19]

$$\int_{-\infty}^{\infty} f(x_\mathrm{s})\delta(x_\mathrm{s}-x)\mathrm{d}x_\mathrm{s} = f(x) \tag{2.7}$$

of the Dirac delta distribution can be used to express the source density of an arbitrary sound field in (2.3) as[20]

$$Q_P(\mathbf{x},\omega) = \iiint_{\mathbb{R}^3} Q_P(\mathbf{x}_\mathrm{s},\omega)\,\delta(\mathbf{x}-\mathbf{x}_\mathrm{s})\,\mathrm{d}V_\mathrm{s}\,. \tag{2.8}$$

A suitably chosen differential volume element for integration is denoted as $\mathrm{d}V_\mathrm{s} = \mathrm{d}V_\mathrm{s}(\mathbf{x}_\mathrm{s})$. The equation postulates a 3D convolution

[7] Gumerov and Duraiswami (2004). *Fast multipole methods for the Helmholtz equation in three dimensions*. Oxford, UK: Elsevier Science, Eq. (1.1.14).

[8] Williams, op. cit., Eq. (1.6).

[9] Girod et al., op. cit., p. 564.

[10] Williams, op. cit., Eq. (8.39).

[11] Jensen et al. (2011). *Computational Ocean Acoustics*. 2nd ed. Modern Acoustics and Signal Processing. New York, USA: Springer, Eq. (2.29).

[12] Ibid., p. 78.

[13] Williams, op. cit., Eq. (8.41).

[14] Girod et al., op. cit., Sec. 8.3.

[15] Williams, op. cit.

[16] Jensen et al., op. cit., Eq. (2.60).

[17] Williams, op. cit., Sec. 8.6.

[18] Girod et al., op. cit., Eq. (8.15).

[19] Gel'fand and Shilov (1964). *Generalized functions: Vol. 1. Properties and operations*. New York, London: Academic Press, p. 4.

[20] Williams, op. cit., Eq. (8.42).

of the source density of the free-field Green's function and the source density of the sound field over the whole $\mathbb{R}^3$ space. After substituting the Dirac delta distribution in (2.8) with the left-hand side of (2.6), the order of differentiation and integration in the resulting equation is interchanged. A comparison with (2.3) then reveals that[21]

[21] Ibid., Eq. (8.43).

$$P(\mathbf{x},\omega) = \iiint_{\mathbb{R}^3} Q_P(\mathbf{x}_\mathrm{s},\omega) G(\mathbf{x}|\mathbf{x}_\mathrm{s},\omega)\,\mathrm{d}V_\mathrm{s} \tag{2.9}$$

holds. A particular solution $P(\mathbf{x},\omega)$ of (2.3) may be expressed by its corresponding source density $Q_P(\mathbf{x},\omega)$ together with the free-field Green's function. The integral postulates $P(\mathbf{x},\omega)$ as the result of a 3D convolution. In free-field, the wave equation can be interpreted as a linear time and space invariant (LTSI) system, with $Q_P(\mathbf{x},\omega)$ and $P(\mathbf{x},\omega)$ as its input and output, respectively. The free-field Green's function in the temporal Fourier domain can be interpreted as the spatial impulse response and temporal transfer function of the wave equation. It allows to compute any sound field for a given source density. In the following section, Eq. (2.9) is used to derive solutions to (2.3) for selected source densities.

2.1.2 Selected Solutions

This section presents a selection of analytically expressible sound fields which are the solution to the inhomogeneous Helmholtz equation (2.3) for relatively simple source densities $Q_P(\mathbf{x},\omega)$.

`greens_function_mono.m`
`greens_function_imp.m`

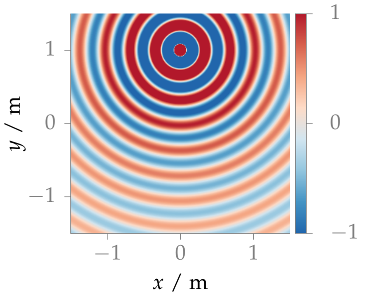


Figure 2.1: Real part of the sound pressure $P_\mathrm{ps}(\mathbf{x}|\mathbf{x}_\mathrm{ps},\omega)$ of a point source (2.11) located at $\mathbf{x}_\mathrm{ps} = [0,1,0]^\mathrm{T}$ m emitting a monochromatic ($f = 1$ kHz) source signal. The plot is normalised to the pressure magnitude at the coordinates' origin.

Monopole Point Source: The sound field $P_\mathrm{ps}(\mathbf{x}|\mathbf{x}_\mathrm{ps},\omega)$ of a monopole point source located at $\mathbf{x}_\mathrm{ps}$ is a generalisation of the free-field Green's function (2.4) with its density

$$Q_\mathrm{ps}(\mathbf{x}|\mathbf{x}_\mathrm{ps},\omega) = \tilde{S}(\omega)\delta(\mathbf{x}-\mathbf{x}_\mathrm{ps})\,, \tag{2.10}$$

where $\tilde{S}(\omega)$ describes the Fourier spectrum of the source signal $\tilde{s}(t)$, which is emitted by the point source. The sound pressure field is given as[22]

$$P_\mathrm{ps}(\mathbf{x}|\mathbf{x}_\mathrm{ps},\omega) = \tilde{S}(\omega)\frac{\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_\mathrm{ps}|}}{4\pi|\mathbf{x}-\mathbf{x}_\mathrm{ps}|} \tag{2.11}$$

which is exemplarily shown in Fig. 2.1. The position of the point source will either be parametrised by its Cartesian $(x_\mathrm{ps}, y_\mathrm{ps}, z_\mathrm{ps})$ or its spherical representation $(r_\mathrm{ps}, \vartheta_\mathrm{ps}, \phi_\mathrm{ps})$.

[22] Spors (2006). "Active listening room compensation for spatial sound reproduction systems". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg., Eq. (2.27)

Monopole Line Source: The sound field $P_\mathrm{ls}(\mathbf{x}|\mathbf{x}_\mathrm{ls},\omega)$ of an infinite-length line source in z-direction located at $\mathbf{x}_\mathrm{ls} = [x_\mathrm{ls}, y_\mathrm{ls}, 0]^\mathrm{T}$ satisfies (2.3) with the density[23,24]

[23] Williams, op. cit., p. 265.

[24] Spors, op. cit., Eq. (2.38).

$$Q_\mathrm{ls}(\mathbf{x},\omega) = \tilde{S}(\omega)\delta(x-x_\mathrm{ls})\delta(y-y_\mathrm{ls})\,. \tag{2.12}$$

It is alternatively parametrised in cylindrical coordinates by $\mathbf{x}_\mathrm{ls} = \rho_\mathrm{ls}[\cos\phi_\mathrm{ls}, \sin\phi_\mathrm{ls}, 0]^\mathrm{T}$. A comparison between (2.10) and (2.12) shows, that the line source can be interpreted as a continuous, infinite, linear distribution of monopole point sources, here along the z-direction.[25]

[25] Williams, op. cit., Sec. 8.6.1.

The corresponding sound pressure field is given as[26]

$$P_{\mathrm{ls}}(\mathbf{x}|\mathbf{x}_{\mathrm{ls}},\omega) = -\frac{\mathrm{j}}{4}\,\tilde{S}(\omega)\,H_0^{(2)}\left(\frac{\omega}{\mathrm{c}}\sqrt{(x-x_{\mathrm{ls}})^2+(y-y_{\mathrm{ls}})^2}\right) \tag{2.13}$$

and is exemplarily shown in Fig. 2.2. The amplitude decay of a line source is approximately 3 dB per distance doubling, which is half decay of a monopole point source.[27] $H_0^{(2)}(\cdot)$ denotes the cylindrical Hankel function of second kind and zeroth order.[28]

Dipole Point Source: A dipole point source $P_{\mathrm{dps}}(\mathbf{x}|\mathbf{x}_{\mathrm{dps}},\mathbf{n}_{\mathrm{dps}},\omega)$ is parametrised by its position $\mathbf{x}_{\mathrm{dps}}$ and the unit vector $\mathbf{n}_{\mathrm{dps}}$, which describes the orientation of the dipole radiation pattern. The corresponding source density reads

$$Q_{\mathrm{dps}}(\mathbf{x}|\mathbf{x}_{\mathrm{dps}},\mathbf{n}_{\mathrm{dps}},\omega) = \tilde{S}(\omega)\nabla_{\mathbf{x},\mathbf{n}_{\mathrm{dps}}}\delta(\mathbf{x}-\mathbf{x}_{\mathrm{dps}})\,, \tag{2.14}$$

where the directional derivative $\nabla_{\mathbf{x},\mathbf{n}_{\mathrm{dps}}}$ along $\mathbf{n}_{\mathrm{dps}}$ is defined by (1.4). Inserting the density into (2.9) yields

$$P_{\mathrm{dps}}(\mathbf{x}|\mathbf{x}_{\mathrm{dps}},\mathbf{n}_{\mathrm{dps}},\omega) = \tilde{S}(\omega)\iiint_{\mathbb{R}^3}\nabla_{\mathbf{x}_{\mathrm{s}},\mathbf{n}_{\mathrm{dps}}}\delta(\mathbf{x}_{\mathrm{s}}-\mathbf{x}_{\mathrm{dps}})G(\mathbf{x}|\mathbf{x}_{\mathrm{s}},\omega)\,\mathrm{d}V_{\mathrm{s}}\,. \tag{2.15}$$

For the derivation, $\mathbf{n}_{\mathrm{dps}}$ may chosen to $\mathbf{u}_x$[29] without loss of generality since the underlying coordinate system may be rotated such that the $\mathbf{n}_{\mathrm{dps}}$ and the x-axis are aligned. The directional derivative in (2.15) simplifies to the derivative with respect to x_{s}. While the integrals w.r.t. y_{s} and z_{s} are solved via the sifting theorem in (2.7), the differentiation theorem[30]

$$\int_{-\infty}^{\infty} f(x_{\mathrm{s}})\frac{\mathrm{d}\delta(x_{\mathrm{s}}-x)}{\mathrm{d}x}\mathrm{d}x_{\mathrm{s}} = -\left.\frac{\mathrm{d}f(x_{\mathrm{s}})}{\mathrm{d}x_{\mathrm{s}}}\right|_{x_{\mathrm{s}}=x} \tag{2.16}$$

of the Dirac delta distribution is applied to solve the remaining one. After the $\mathbf{n}_{\mathrm{dps}}$ has been re-generalised to arbitrary directions, the sound pressure field reads[31]

$$P_{\mathrm{dps}}(\mathbf{x}|\mathbf{x}_{\mathrm{dps}},\mathbf{n}_{\mathrm{dps}},\omega) = -\tilde{S}(\omega)\nabla_{\mathbf{x},\mathbf{n}_{\mathrm{dps}}}G(\mathbf{x}|\mathbf{x}_{\mathrm{dps}},\omega) \tag{2.17}$$

describing the dipole point source as the directional derivative of a monopole point source along $\mathbf{n}_{\mathrm{dps}}$. The sound pressure is finally given as[32]

$$P_{\mathrm{dps}}(\mathbf{x}|\mathbf{x}_{\mathrm{dps}},\mathbf{n}_{\mathrm{dps}},\omega) = \tilde{S}(\omega)\left(\mathrm{j}\frac{\omega}{\mathrm{c}}+\frac{1}{|\mathbf{x}-\mathbf{x}_{\mathrm{dps}}|}\right)\underbrace{\frac{\langle\mathbf{x}-\mathbf{x}_{\mathrm{dps}}|\mathbf{n}_{\mathrm{dps}}\rangle}{|\mathbf{x}-\mathbf{x}_{\mathrm{dps}}|}}_{\text{directivity}}\underbrace{\frac{\mathrm{e}^{-\mathrm{j}\frac{\omega}{\mathrm{c}}|\mathbf{x}-\mathbf{x}_{\mathrm{dps}}|}}{4\pi|\mathbf{x}-\mathbf{x}_{\mathrm{dps}}|}}_{\text{monopole}} \tag{2.18}$$

and is exemplary shown in Fig. 2.3. As shown in Eq. (2.18), the dipole source can be described as a monopole point source weighted by a directivity pattern and the term given in the brackets. For close distances $|\mathbf{x}-\mathbf{x}_{\mathrm{dps}}|$ and low frequencies ω, the second addend in the bracket becomes dominant leading to a total amplitude decay of

[26] Ibid., Eq. (8.47).

[27] Möser (2009). *Engineering Acoustics*. Springer, Sec. 3.1 and 3.2.

[28] Abramowitz and Stegun (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. 55. Courier Corporation, Sec. 9.1.4.

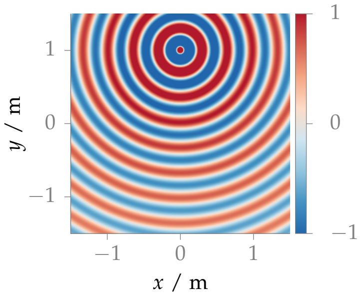


Figure 2.2: Real part of the sound pressure $P_{\mathrm{ls}}(\mathbf{x}|\mathbf{x}_{\mathrm{ls}},\omega)$ of a line source (2.13) located at $\mathbf{x}_{\mathrm{ls}} = [0,1,0]^{\mathrm{T}}$ m emitting a monochromatic ($f = 1\,\mathrm{kHz}$) source signal. The plot is normalised to the pressure magnitude at the coordinates' origin.

[29] unit vector pointing in positive x-direction, see Sec. 1.4

[30] Gel'fand and Shilov, op. cit., p. 26.

[31] Schultz (2016). "Sound Field Synthesis for Line Source Array Applications in Large-Scale Sound Reinforcement". PhD thesis. University of Rostock, Eq. (2.17).

[32] Ibid., Eq. (2.18).

12 dB per distance doubling. In the far-field, the first addend leads to a high-pass characteristic with the distance decay of a monopole point source. The directivity pattern describes a so-called figure-of-eight, which is the cosine of the angle between $\mathbf{x} - \mathbf{x}_{\mathrm{dps}}$ and $\mathbf{n}_{\mathrm{dps}}$.

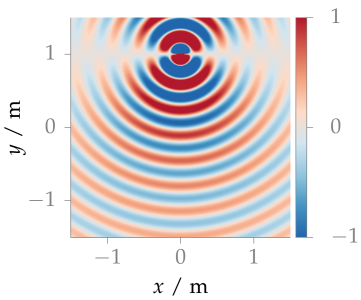


Figure 2.3: Real part of the sound pressure $P_{\mathrm{dps}}(\mathbf{x}|\mathbf{x}_{\mathrm{dps}}, \mathbf{n}_{\mathrm{dps}}, \omega)$ of a dipole source (2.18) located at $\mathbf{x}_{\mathrm{dps}} = [0, 1, 0]^{\mathrm{T}}$ m and oriented along $\mathbf{n}_{\mathrm{dps}} = [0, -1, 0]^{\mathrm{T}}$ emitting a monochromatic ($f = 1$ kHz) source signal. The plot is normalised to the pressure magnitude at the coordinates' origin.

Plane Wave: A plane wave $P_{\mathrm{pw}}(\mathbf{x}|\mathbf{n}_{\mathrm{pw}}, \omega)$ with its propagation direction $\mathbf{n}_{\mathrm{pw}} = [\cos\phi_{\mathrm{pw}} \sin\vartheta_{\mathrm{pw}}, \sin\phi_{\mathrm{pw}} \sin\vartheta_{\mathrm{pw}}, \cos\vartheta_{\mathrm{pw}}]^{\mathrm{T}}$ is often regarded as a solution to the homogeneous Helmholtz equation.[33] Although the author does not raise doubts against its mathematical validity, stating the corresponding source density $Q_{\mathrm{pw}}(\mathbf{x}, \omega)$ to be zero, raises the question what kind of sound source creates a plane wave. As described by Junger and Feit,[34] a plane wave is excited by a uniformly vibrating plate of infinite size and infinitesimal thickness. It is modelled by Spors[35] as a continuous, infinite, planar distribution of monopole point sources which is shown in Fig. 2.4. The location of the plane is parametrised by its normal vector $\mathbf{n}_{\mathrm{pw}}$ and its signed distance d to the coordinates' origin $\mathbf{0}$. For the derivation, Spors chose the plane to be located at the coordinates origin, i.e. $d = 0$. However, the propagation direction of the plane wave is only correct in the halfspace "in front" of the plate. Behind the plate, the propagation direction is $-\mathbf{n}_{\mathrm{pw}}$. Thus, it is necessary to consider the limiting case where $d \to \infty$. The corresponding source density reads[36]

[33] Williams, op. cit., Sec. 2.6.2.

[34] Junger and Feit (1986). *Sound, Structures, and Their Interaction*. Cambridge, USA: MIT Press.

[35] Spors, op. cit., Sec. 2.4.4.

[36] Ibid., modification of Eq. 2.45.

$$Q_{\mathrm{pw}}(\mathbf{x}|\mathbf{n}_{\mathrm{pw}}, \omega) = 2\mathrm{j}\frac{\omega}{\mathrm{c}}\,\tilde{S}(\omega) \lim_{d\to\infty} \delta\big(\langle \mathbf{n}_{\mathrm{pw}}|\mathbf{x}\rangle + d\big)\mathrm{e}^{+\mathrm{j}\frac{\omega}{\mathrm{c}}d}\,, \tag{2.19}$$

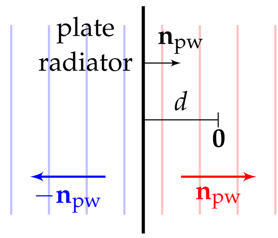


Figure 2.4: Infinite vibrating plate and its created sound field.

whereas the exponential term compensates for the propagation delay. Similar to the calculus for the dipole point source, $\mathbf{n}_{\mathrm{pw}}$ is chosen to $\mathbf{u}_x$ without loss of generality. After inserting (2.19) into (2.3) the sifting theorem (2.7) is used to solve the integral w.r.t. x_{s}. The resulting integral[37]

[37] Ibid., modification of Eq. 2.46.

$$P_{\mathrm{pw}}(\mathbf{x}|\mathbf{u}_x, \omega) = \tag{2.20}$$
$$\frac{2\mathrm{j}\frac{\omega}{\mathrm{c}}}{4\pi} \lim_{d\to\infty} \mathrm{e}^{+\mathrm{j}\frac{\omega}{\mathrm{c}}d} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{j}\frac{\omega}{\mathrm{c}}\sqrt{(x+d)^2+(y-y_{\mathrm{s}})^2+(z-z_{\mathrm{s}})^2}}}{\sqrt{(x+d)^2+(y-y_{\mathrm{s}})^2+(z-z_{\mathrm{s}})^2}}\,\mathrm{d}y_{\mathrm{s}}\,\mathrm{d}z_{\mathrm{s}}$$

is solved using the identities from Gradshteyn and Ryzhik.[38] It can be re-generalised to arbitrary propagation directions $\mathbf{n}_{\mathrm{pw}}$ as

[38] Gradshteyn and Ryzhik (2007). *Tables of Integrals, Series, and Products*. Ed. by Alan Jeffrey and Daniel Zwillinger. 7th ed. Academic Press, Eq. 3.876–1/2, 6.677–3/4.

$$P_{\mathrm{pw}}(\mathbf{x}|\mathbf{n}_{\mathrm{pw}}, \omega) = \tilde{S}(\omega)\mathrm{e}^{-\mathrm{j}\frac{\omega}{\mathrm{c}}\langle \mathbf{n}_{\mathrm{pw}}|\mathbf{x}\rangle}\,, \tag{2.21}$$

which is the sound field pressure of a plane wave. As shown in Fig. 2.5, the amplitude of the sound pressure is constant. The wave fronts are planar and oriented perpendicular to the propagation direction.

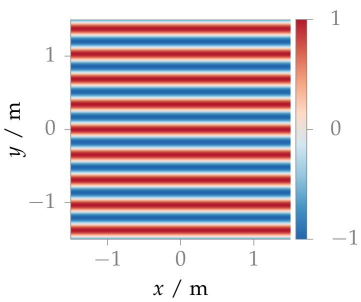


Figure 2.5: Real part of the sound pressure $P_{\mathrm{pw}}(\mathbf{x}|\mathbf{n}_{\mathrm{pw}}, \omega)$ of a monochromatic ($f = 1$ kHz) plane wave (2.21) with a propagation direction of $\mathbf{n}_{\mathrm{pw}} = [0, -1, 0]^{\mathrm{T}}$.

2.1.3 Integral Formulation

The integral notation of the wave equation or equivalently of the Helmholtz equation builds the theoretical basis of all techniques for SFS presented later in this thesis. The geometry used for the explanations is shown in Fig. 2.6. Given a bounded volume Ω in $\mathbb{R}^3$ with its smooth boundary $\partial\Omega$, the position vector $\mathbf{x} \in \Omega$, and an arbitrary sound pressure field $P(\mathbf{x}, \omega)$ with its corresponding source density $Q_P(\mathbf{x}, \omega)$, the integral notation of the wave equation is given as[39]

[39] Schultz, op. cit., Eq. (2.13).

$$
\begin{aligned}
P(\mathbf{x}, \omega) = & \iiint_{\Omega} G(\mathbf{x}|\mathbf{x}_\mathrm{s}, \omega) Q_P(\mathbf{x}_\mathrm{s}, \omega) \,\mathrm{d}V_\mathrm{s} \\
& + \iint_{\partial\Omega} P(\mathbf{x}_0, \omega) \nabla_{\mathbf{x}_0, \mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0, \omega) - G(\mathbf{x}|\mathbf{x}_0, \omega) \nabla_{\mathbf{x}_0, \mathbf{n}_0} P(\mathbf{x}_0, \omega) \,\mathrm{d}A_0 \,.
\end{aligned} \tag{2.22}
$$

$\mathrm{d}V_\mathrm{s} = \mathrm{d}V_\mathrm{s}(\mathbf{x}_\mathrm{s})$ and $\mathrm{d}A_0 = \mathrm{d}A_0(\mathbf{x}_0)$ describe suitably chosen differential volume and boundary elements, respectively. The boundary integral represents a distribution of monopole and dipole point sources[40] located at $\mathbf{x}_0 \in \partial\Omega$. The dipoles are oriented along the inward pointing surface normal $\mathbf{n}_0 = \mathbf{n}_0(\mathbf{x}_0)$ of $\partial\Omega$. The source signals $\tilde{S}(\omega)$ of the dipoles and the monopoles are the boundary sound pressure $P(\mathbf{x}_0, \omega)$ and its directional derivative $\nabla_{\mathbf{x}_0, \mathbf{n}_0} P(\mathbf{x}_0, \omega)$, respectively. Contrary to (2.9), $P(\mathbf{x}, \omega)$ is not defined by a single volume integral over the whole source density $Q_P(\mathbf{x}, \omega)$ in $\mathbb{R}^3$, but rather by a volume integral including all sources inside Ω and a boundary integral along $\partial\Omega$. The boundary integral describes the contributions of all sound sources defined by $Q_P(\mathbf{x}, \omega)$, which are located outside Ω. Thus,

[40] see Sec. 2.1.2

$$
\begin{aligned}
& \iiint_{\bar{\Omega}} G(\mathbf{x}|\mathbf{x}_\mathrm{s}, \omega) Q_P(\mathbf{x}_\mathrm{s}, \omega) \,\mathrm{d}V_\mathrm{s} = \\
& \qquad \iint_{\partial\Omega} P(\mathbf{x}_0, \omega) \nabla_{\mathbf{x}_0, \mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0, \omega) - G(\mathbf{x}|\mathbf{x}_0, \omega) \nabla_{\mathbf{x}_0, \mathbf{n}_0} P(\mathbf{x}_0) \,\mathrm{d}A_0
\end{aligned} \tag{2.23}
$$

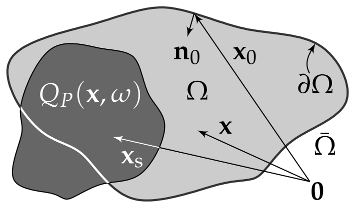


Figure 2.6: Geometry for the Integral Formulation of the Helmholtz Equation. The area shaded in dark grey illustrates a non-zero source density $Q_P(\mathbf{x}, \omega)$.

with $\bar{\Omega} := \mathbb{R}^3 / \Omega$ can be easily derived by subtracting (2.9) from (2.22). The sound field $P(\mathbf{x}, \omega)$ may be split into two parts: its homogeneous component

$$
\begin{aligned}
& \mathcal{H}_P^{\Omega}(\mathbf{x}, \omega) = \\
& \iint_{\partial\Omega} P(\mathbf{x}_0, \omega) \nabla_{\mathbf{x}_0, \mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0, \omega) - G(\mathbf{x}|\mathbf{x}_0, \omega) \nabla_{\mathbf{x}_0, \mathbf{n}_0} P(\mathbf{x}_0, \omega) \,\mathrm{d}A_0 \,,
\end{aligned} \tag{2.24}
$$

is source-free within Ω and describes all contributions from $\bar{\Omega}$. Its inhomogeneous counterpart

$$
\mathcal{I}_P^{\Omega}(\mathbf{x}, \omega) := \iiint_{\Omega} G(\mathbf{x}|\mathbf{x}_\mathrm{s}, \omega) Q_P(\mathbf{x}_\mathrm{s}, \omega) \,\mathrm{d}V_\mathrm{s} \tag{2.25}
$$

defines contributions from all sources inside Ω. According to (2.22), it is not possible to define sources inside Ω using the distribution of monopole and dipole source along $\partial\Omega$ as this would require monopole point sources inside Ω. $\mathcal{I}_P(\mathbf{x}, \omega)$ can therefore be regarded as the systemic error introduced when a sound field with non-zero source density $Q_P(\mathbf{x}, \omega)$ inside Ω is described by the boundary integral alone.

The Helmholtz-Integral-Equation: In the special case of $P(\mathbf{x}, \omega)$ being source-free within Ω, i.e. $Q_P(\mathbf{x}, \omega) = 0 \,\forall\, \mathbf{x} \in \Omega$, the volume integral in (2.22) vanishes which leads to the interior Helmholtz-Integral-Equation (HIE)[41,42]

[41] Williams, op. cit., Eq. (8.15).

[42] Colton and Kress (2013). *Inverse acoustic and electromagnetic scattering theory*. 3rd ed. Vol. 93. Applied Mathematical Sciences. Springer Science & Business Media, Eq. (2.5).

$$P(\mathbf{x},\omega) = \tag{2.26}$$
$$\iint_{\partial\Omega} P(\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0,\omega) - G(\mathbf{x}|\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} P(\mathbf{x}_0,\omega)\ \mathrm{d}A_0\,.$$

In this case, $\mathcal{I}_P^{\Omega}(\mathbf{x},\omega)$ vanishes and $P(\mathbf{x},\omega) = \mathcal{H}_P^{\Omega}(\mathbf{x},\omega)$ holds. Remarkably, independent of whether or not $P(\mathbf{x},\omega)$ exhibits an inhomogeneous component $\mathcal{I}_P^{\Omega}(\mathbf{x},\omega)$, the boundary integral does only describe the homogeneous part given by (2.24). In other words, driving the monopole and dipole distribution with the overall field $P(\mathbf{x},\omega)$ is identical to driving it with only the homogeneous component $\mathcal{H}_P^{\Omega}(\mathbf{x},\omega)$. Mathematically speaking,

$$\iint_{\partial\Omega} \mathcal{H}_P(\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0,\omega) - G(\mathbf{x}|\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} \mathcal{H}_P(\mathbf{x}_0,\omega)\ \mathrm{d}A_0$$
$$= \iint_{\partial\Omega} P(\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0,\omega) - G(\mathbf{x}|\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} P(\mathbf{x}_0,\omega)\ \mathrm{d}A_0\,. \tag{2.27}$$

Boundary Conditions: The necessity of the surface pressure and the surface pressure gradient to define the sound field within Ω is one potential drawback of the HIE.[43] It is evident from the right-hand side of (2.26) and has been studied extensively in the literature, that this can be resolved by applying boundary conditions either to the free-field Green's function[44] or to the sound field itself.[45,46,47] The second option will be discussed here: A sound field $P(\mathbf{x},\omega)$ being source-free inside Ω may be modified to

$$P_T(\mathbf{x},\omega) := P(\mathbf{x},\omega) + P_D(\mathbf{x},\omega)\,, \tag{2.28}$$

where the $P_D(\mathbf{x},\omega)$ is purely inhomogeneous within Ω, i.e. its homogeneous component $\mathcal{H}_{P_D}^{\Omega}(\mathbf{x},\omega)$ is zero. The addends $P(\mathbf{x},\omega)$ and $P_D(\mathbf{x},\omega)$ can hence be regarded as the homogeneous and inhomogeneous part of the total sound field $P_T(\mathbf{x},\omega)$, respectively. Following the same argumentation as for (2.27) the original sound field is defined as

$$P(\mathbf{x},\omega) = \tag{2.29}$$
$$\iint_{\partial\Omega} P_T(\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} G(\mathbf{x}|\mathbf{x}_0,\omega) - G(\mathbf{x}|\mathbf{x}_0,\omega)\nabla_{\mathbf{x}_0,\mathbf{n}_0} P_T(\mathbf{x}_0,\omega)\ \mathrm{d}A_0\,.$$

The addend $P_D(\mathbf{x},\omega)$ is used to let the total sound field fulfil a chosen boundary condition w.r.t. to the surface pressure or its directional derivative. Within this treatise, the homogenous Dirichlet (sound-soft) boundary condition[48]

$$P_T(\mathbf{x}_0,\omega) = 0\,\forall\,\mathbf{x}_0 \in \partial\Omega \tag{2.30}$$

is of special interest, which further simplifies Eq. (2.29) to

$$P(\mathbf{x},\omega) = -\iint_{\partial\Omega} \nabla_{\mathbf{x}_0,\mathbf{n}_0} P_T(\mathbf{x}_0,\omega) G(\mathbf{x}|\mathbf{x}_0,\omega)\ \mathrm{d}A_0\,. \tag{2.31}$$

As the remaining task, $\nabla_{\mathbf{x}_0,\mathbf{n}_0} P_T(\mathbf{x}_0,\omega)$ has to be determined based on the sound field $P(\mathbf{x},\omega)$ and on the shape of the boundary $\partial\Omega$. It

[43] Williams, op. cit., Sec. 8.7.

[44] Ibid., Sec. 8.8.

[45] Ibid., Sec. 8.7.

[46] Schultz and Spors (2014). "Comparing Approaches to the Spherical and Planar Single Layer Potentials for Interior Sound Field Synthesis". In: *Acta Acustica united with Acustica* 100.5, pp. 900–911, Sec. 4.2.

[47] Fazi and Nelson (2013). "Sound field reproduction as an equivalent acoustical scattering problem". In: *J. Acoust. Soc. Am.* 134.5, pp. 3721–3729, Sec. IV.

[48] Ibid., Eq. (14).

has been shown by Zotter and Spors,[49] and Fazi and Nelson,[50] that $\nabla_{\mathbf{x}_0,\mathbf{n}_0} P_T(\mathbf{x}_0,\omega)$ is equivalent to the directional pressure gradient on the surface of a sound-soft scatterer, whose shape coincides with the shape of Ω. A solution of the scattering problem with $P(\mathbf{x},\omega)$ as the sound field impinging on the scatterer is also a solution to the original task. In general, it is not straightforward to solve the scattering problem for arbitrary shapes. However, analytic solutions for simple boundaries, like e.g. planes, cylinders and spheres, exist. They build the basis for the SFS methods presented in Ch. 3.

[49] Zotter and Spors (Sept. 2013). "Is sound field control determined at all frequencies? How is it related to numerical acoustics?" In: *Proc. of 52nd Intl. Aud. Eng. Soc. Conf. on Sound Field Control - Engineering and Perception*. Guildford, UK, Sec. 3.

[50] Fazi and Nelson, op. cit., Sec. IV.

2.2 Sound Field Representations

This section presents possibilities to expand arbitrary sound fields into a superposition of basis functions weighted by their respective expansion coefficients. In general, all presented expansions depend on the coordinate $\mathbf{x}_\mathrm{c}$ around which they are expanded. However, without loss of generality the coordinates' origin $\mathbf{0}$ is chosen as the expansion centre for most of the explanations. Whenever the dependency on $\mathbf{x}_\mathrm{c}$ is omitted for the respective expansion coefficients, $\mathbf{x}_\mathrm{c} = \mathbf{0}$ holds.

2.2.1 Expansion into Spherical Basis Functions

The definition of the spherical coordinates r, ϑ, and ϕ used in this treatise is given in Sec. 1.4. The general solution of the Helmholtz Equation (2.3) for this coordinate system is the Spherical Harmonics expansion a.k.a. the inverse Spherical Harmonics transform

$$P(\mathbf{x},\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{P}_n^m(r,\omega) \Psi_n^m(\vartheta,\phi) \tag{2.32}$$

where $\mathring{P}_n^m(r,\omega)$ denotes the Spherical Harmonics expansion coefficients. The according Spherical Harmonics transform is given as

$$\mathring{P}_n^m(r,\omega) = \int_0^{2\pi} \int_0^{\pi} P(\mathbf{x},\omega)\, \Psi_n^{-m}(\vartheta,\phi) \sin\vartheta \,\mathrm{d}\vartheta \,\mathrm{d}\phi \,. \tag{2.33}$$

The angular basis functions a.k.a. Spherical Harmonics are defined as[51]

$$\Psi_n^m(\vartheta,\phi) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}}\, L_n^{|m|}(\cos\vartheta) \mathrm{e}^{+\mathrm{j}m\phi} \,. \tag{2.34}$$

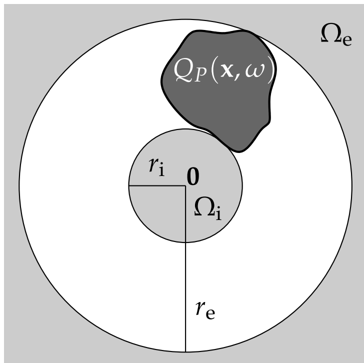


Figure 2.7: Example for interior and exterior domain (grey shaded areas) in the spherical coordinate system.

$L_n^m(\cos\vartheta)$ are the associated Legendre functions.[52] For the structure of $\mathring{P}_n^m(r,\omega)$, two cases have to be considered: As shown in Fig. 2.7, the coefficients fulfil the Helmholtz equation (2.3) in the interior domain Ω_i or in the exterior domain Ω_e. The former region describes the source-free area, where the radius r is smaller than the minimal radius r_i of the sound source distribution. Hence, all sources are located outside this domain. The exterior domain $r > r_\mathrm{i}$ surrounds the sound sources. Due to the linearity of the Helmholtz equation the shown source distribution may be split up into sub-distributions.

[51] Gumerov and Duraiswami, op. cit., Eq. (2.1.59).

[52] Ibid., Eq. (2.1.47).

Source type	$\breve{P}_n^m$	Conditions
Point Source	$-\mathrm{j}\frac{\omega}{c}h_n^{(2)}(\frac{\omega}{c}r_{\mathrm{ps}})\Psi_n^{-m}(\vartheta_{\mathrm{ps}},\phi_{\mathrm{ps}})$	$r < r_{\mathrm{ps}}$
Line Source	$\pi\,\mathrm{j}^{m-n-1}\,H_m^{(2)}(\frac{\omega}{c}\rho_{\mathrm{ls}})\Psi_n^{-m}\left(\frac{\pi}{2},\phi_{\mathrm{ls}}\right)$	$r < \rho_{\mathrm{ls}},\ \vartheta = \frac{\pi}{2}$
Plane Wave	$4\pi\mathrm{j}^{-n}\Psi_n^{-m}(\vartheta_{\mathrm{pw}},\phi_{\mathrm{pw}})$	

Table 2.1: Interior spherical expansion coefficients of various source types. The temporal spectrum of the source signal $\tilde{S}(\omega)$ is omitted, for brevity.

The interior and exterior domain can then be defined for each sub-distribution, individually. The Spherical Harmonics expansion specialises to[53]

$$P(\mathbf{x},\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\begin{Bmatrix}\breve{P}_n^m(\omega)j_n(\frac{\omega}{c}r) & \forall\,\mathbf{x}\in\Omega_\mathrm{i}\\ \mathring{P}_n^m(\omega)h_n^{(2)}(\frac{\omega}{c}r) & \forall\,\mathbf{x}\in\Omega_\mathrm{e}\end{Bmatrix}\Psi_n^m(\vartheta,\phi)\,, \quad (2.35)$$

with the nth-order spherical Bessel function $j_n(\cdot)$.[54] For the used definition of the Fourier Transform[55], the nth-order spherical Hankel function of second kind $h_n^{(2)}(\cdot)$ has to be used for the exterior domain. $\breve{P}_n^m(\omega)$ and $\mathring{P}_n^m(\omega)$ denote the respective interior and exterior spherical expansion coefficients of the sound field. For the former, Tab. 2.1 shows the coefficients for some of the sound sources introduced in Sec. 2.1.2. The radial and the angular basis functions are subsumed under[56]

$$I_n^m(\mathbf{x},\omega) = j_n\left(\frac{\omega}{c}r\right)\Psi_n^m(\vartheta,\phi)\,, \quad (2.36a)$$

$$E_n^m(\mathbf{x},\omega) = h_n^{(2)}\left(\frac{\omega}{c}r\right)\Psi_n^m(\vartheta,\phi) \quad (2.36b)$$

where $I_n^m(\mathbf{x},\omega)$ and $E_n^m(\mathbf{x},\omega)$ denote the interior and the exterior spherical basis functions, respectively.

2.2.2 Expansion into Circular Basis Function[57]

Any two-dimensional (2D) sound field whose pressure is independent of the z coordinate may be expanded into[58]

$$P(\mathbf{x},\omega) = \sum_{m=-\infty}^{\infty}\mathring{P}_m(\rho,\omega)\mathrm{e}^{+\mathrm{j}m\phi}\,, \quad (2.37)$$

with the Circular Harmonics expansion coefficients given via the Circular Harmonics Transform (CHT)

$$\mathring{P}_m(\rho,\omega) = \frac{1}{2\pi}\int_0^{2\pi}P(\mathbf{x},\omega)\,\mathrm{e}^{-\mathrm{j}m\phi}\,\mathrm{d}\phi\,. \quad (2.38)$$

The polar coordinates ρ and ϕ are defined within the cylindrical coordinate system in Sec. 1.4. Analogue to the Spherical Harmonics expansion, the interior and exterior domains have to be handled separately. The domains are defined via the minimum and maximum radii $\rho_{\{\mathrm{i,e}\}}$ of the non-zero source density. The expansion reads

$$P(\mathbf{x},\omega) = \sum_{m=-\infty}^{\infty}\begin{Bmatrix}\breve{P}_m(\omega)J_m(\frac{\omega}{c}\rho) & \forall\,\mathbf{x}\in\Omega_\mathrm{i}\\ \mathring{P}_m(\omega)H_m^{(2)}(\frac{\omega}{c}\rho) & \forall\,\mathbf{x}\in\Omega_\mathrm{e}\end{Bmatrix}\mathrm{e}^{+\mathrm{j}m\phi}\,, \quad (2.39)$$

[53] Ahrens (2012). *Analytic Methods of Sound Field Synthesis*. T-Labs Series in Telecommunication Services. Berlin Heidelberg, Germany: Springer-Verlag, Eq. (2.32).

[54] Abramowitz and Stegun, op. cit., Sec. 10.1.

[55] see (2.2)

[56] Gumerov and Duraiswami, op. cit., modification of Eq. (2.1.101) and (2.1.102), since different definition of Fourier Transform w.r.t. time is used.

[57] Parts of this section are published in Hahn et al. (June 2016). "Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain". In: *Proc. of 140th Aud. Eng. Soc. Conv.* Paris, France.

[58] Williams, op. cit., Eq. (4.49).

where $J_m(\cdot)$ denotes the mth-order cylindrical Bessel function. The cylindrical Hankel function of second kind and mth order is given as $H_m^{(2)}(\cdot)$. $\breve{P}_m(\omega)$ and $\hat{P}_m(\omega)$ denote the interior and exterior circular expansion coefficients, respectively. Tab. 2.2 lists some examples for interior circular expansion coefficients. Analogous to (2.36), the subsumptions

circexp_mono_ps.m
circexp_mono_pw.m

$$I_m(\mathbf{x},\omega) = J_m\left(\frac{\omega}{c}\rho\right) \mathrm{e}^{+\mathrm{j}m\phi} \tag{2.40a}$$

$$E_m(\mathbf{x},\omega) = H_m^{(2)}\left(\frac{\omega}{c}\rho\right) \mathrm{e}^{+\mathrm{j}m\phi} \tag{2.40b}$$

with the interior and exterior circular basis functions $I_m(\mathbf{x},\omega)$ and $E_m(\mathbf{x},\omega)$ are used.

The interior circular and spherical expansion coefficients can be converted into each other. For the interior spherical expansion (2.35), the order of summation may be reordered to[59,60]

[59] Hahn et al., op. cit., Eq. (28).

[60] Ahrens, op. cit., Eq. (2.34).

$$P(\mathbf{x},\omega) = \sum_{m=-\infty}^{\infty} \left[\sum_{n=|m|}^{\infty} \breve{P}_n^m(\omega) j_n\left(\frac{\omega}{c}r\right) \Psi_n^m(\vartheta,0) \right] \mathrm{e}^{+\mathrm{j}m\phi}, \tag{2.41}$$

where $\Psi_n^m(\vartheta,\phi) = \Psi_n^m(\vartheta,0)\mathrm{e}^{+\mathrm{j}m\phi}$ from (2.34) has been utilised. A comparison of (2.41) and the interior case of (2.39) yields

$$\breve{P}_m(\omega) J_m\left(\frac{\omega}{c}\rho\right) = \sum_{n=|m|}^{\infty} \breve{P}_n^m(\omega) j_n\left(\frac{\omega}{c}r\right) \Psi_n^m(\vartheta,0)\,. \tag{2.42}$$

Using the circular and spherical expansion of a plane wave (see Tab. 2.1 and 2.2) allows to express the cylindrical Bessel function in terms of spherical bessel functions

$$J_m\left(\frac{\omega}{c}\rho\right) = \sum_{n=|m|}^{\infty} 4\pi \mathrm{j}^{m-n} \Psi_n^{-m}\left(\frac{\pi}{2},0\right) j_n\left(\frac{\omega}{c}r\right) \Psi_n^m(\vartheta,0)\,. \tag{2.43}$$

After the substitution of $J_m(\frac{\omega}{c}\rho)$ in (2.42), the conversion relation[61]

[61] Hahn and Spors (May 2015b). "Sound Field Synthesis of Virtual Cylindrical Waves Using Circular and Spherical Loudspeaker Arrays". In: *Proc. of 138th Aud. Eng. Soc. Conv.* Warsaw, Poland, Eq. (14).

$$\breve{P}_n^m(\omega) = 4\pi \mathrm{j}^{m-n} \breve{P}_m(\omega) \Psi_n^{-m}\left(\frac{\pi}{2},0\right) \tag{2.44}$$

from the circular to the spherical expansion coefficients is obtained. This relation is, for example, used to derive the regular spherical expansion coefficients of a line source shown in Tab. 2.2. The inverse conversion is given as[62]

[62] Hahn et al., op. cit., Eq. (32).

$$\breve{P}_m(\omega) = \frac{\mathrm{j}^{|m|-m}}{4\pi \Psi_{|m|}^{-m}\left(\frac{\pi}{2},0\right)} \breve{P}_{|m|}^m(\omega)\,. \tag{2.45}$$

It is exact only for 2D, i.e. z-independent, sound fields. For 3D sound fields expressed via $\breve{P}_n^m(\omega)$, the conversion to a circular expansion states an approximation, which coincides with the original sound field only at the expansion centre.[63] Combining (2.45) and (2.44) allows for an approximation

[63] Ibid., Sec. 2.7.

$$\breve{P}_n^m(\omega) \approx \mathrm{j}^{|m|-n} \frac{\Psi_n^{-m}\left(\frac{\pi}{2},0\right)}{\Psi_{|m|}^{-m}\left(\frac{\pi}{2},0\right)} \breve{P}_{|m|}^m(\omega) \tag{2.46}$$

of the spherical expansion via its sectorial[64] ($n = |m|$) subset. Again, this sectorial approximation becomes exact for 2D sound fields.

[64] Gumerov and Duraiswami, op. cit., Sec. 2.1.2.2.

Source type	$\check{P}_m$	Conditions
Point Source	$\frac{\mathrm{j}^{\|m\|-m}}{4\pi}(-\mathrm{j}\frac{\omega}{\mathrm{c}})h^{(2)}_{\|m\|}(\frac{\omega}{\mathrm{c}}r_{\mathrm{ps}})\mathrm{e}^{-\mathrm{j}m\phi_{\mathrm{ps}}}$	$\rho < r_{\mathrm{ps}}$, $\vartheta_{\mathrm{ps}} = \frac{\pi}{2}$
Line Source	$-\frac{\mathrm{j}}{4}H^{(2)}_m(\frac{\omega}{\mathrm{c}}\rho_{\mathrm{ls}})\mathrm{e}^{-\mathrm{j}m\phi_{\mathrm{ls}}}$	$\rho < \rho_{\mathrm{ls}}$
Plane Wave	$\mathrm{j}^{-m}\mathrm{e}^{-\mathrm{j}m\phi_{\mathrm{pw}}}$	$\vartheta_{\mathrm{pw}} = \frac{\pi}{2}$

Table 2.2: Interior circular expansion coefficients for various source types. The temporal spectrum of the source signal $\breve{S}(\omega)$ is omitted, for brevity. Since the point source is a 3D sound field, the coefficients are an approximation based on Eq. (2.45).

2.2.3 Expansion into Plane Waves

Another representation form which does only hold for the interior domain Ω_i is the plane wave expansion or Plane Wave Decomposition (PWD)[65]

$$P(\mathbf{x},\omega) = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} \bar{P}(\phi_{\mathrm{pw}},\vartheta_{\mathrm{pw}},\omega)\mathrm{e}^{-\mathrm{j}\frac{\omega}{\mathrm{c}}\langle\mathbf{n}_{\mathrm{pw}}|\mathbf{x}\rangle}\sin\vartheta_{\mathrm{pw}}\,\mathrm{d}\vartheta_{\mathrm{pw}}\,\mathrm{d}\phi_{\mathrm{pw}}\,. \tag{2.47}$$

[65] Ahrens, op. cit., Eq. (2.45).

It constitutes a superposition of plane waves with their propagation directions continuously distributed on the unit sphere. Each individual plane wave is hereby weighted by its corresponding plane wave coefficient $\bar{P}(\phi_{\mathrm{pw}},\vartheta_{\mathrm{pw}},\omega)$ sometimes referred to as the signature function.[66] A conversion between the 3D PWD and regular spherical expansion coefficients (2.35) of same sound field is achieved via[67]

[66] Ibid., Sec. 2.2.4.

[67] Ibid., Eq. (E-16) and (E-12).

$$\bar{P}(\phi_{\mathrm{pw}},\vartheta_{\mathrm{pw}},\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\mathrm{j}^n\check{P}_n^m(\omega)\Psi_n^m(\vartheta_{\mathrm{pw}},\phi_{\mathrm{pw}})\,,\text{ and} \tag{2.48a}$$

$$\check{P}_n^m(\omega) = \mathrm{j}^{-n}\int_0^{2\pi}\int_0^{\pi}\bar{P}(\phi_{\mathrm{pw}},\vartheta_{\mathrm{pw}},\omega) \cdot\Psi_n^{-m}(\vartheta_{\mathrm{pw}},\phi_{\mathrm{pw}})\sin\vartheta_{\mathrm{pw}}\,\mathrm{d}\vartheta_{\mathrm{pw}}\,\mathrm{d}\phi_{\mathrm{pw}}\,, \tag{2.48b}$$

respectively. For a 2D sound field, (2.47) can be further simplified to a 2D plane wave expansion[68]

[68] Kuntz (2009). "Wave Field Analysis Using Virtual Circular Microphone Arrays". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, Eq. (2.246).

$$P(\mathbf{x},\omega) = \frac{1}{2\pi}\int_0^{2\pi}\bar{P}(\phi_{\mathrm{pw}},\omega)\mathrm{e}^{-\mathrm{j}\frac{\omega}{\mathrm{c}}\langle\mathbf{n}_{\mathrm{pw}}|\mathbf{x}\rangle}\,\mathrm{d}\phi_{\mathrm{pw}} \tag{2.49}$$

with the corresponding coefficients $\bar{P}(\phi_{\mathrm{pw}},\omega)$. A similar conversion relation is given for the interior circular expansion coefficients (2.39):[69]

[69] Ibid., Eq. (4.91).

$$\bar{P}(\phi_{\mathrm{pw}},\omega) = \sum_{m=-\infty}^{\infty}\mathrm{j}^m\check{P}_m(\omega)\mathrm{e}^{+\mathrm{j}m\phi_{\mathrm{pw}}} \tag{2.50a}$$

$$\check{P}_m(\omega) = \mathrm{j}^{-m}\int_0^{2\pi}\bar{P}(\phi_{\mathrm{pw}},\omega)\mathrm{e}^{-\mathrm{j}m\phi_{\mathrm{pw}}}\,\mathrm{d}\phi_{\mathrm{pw}} \tag{2.50b}$$

2.3 The Local Wavenumber Vector

The concept of the local wavenumber vector was introduced in the context of SFS by Firtha et al..[70] The temporal frequency spectrum of a sound field $P(\mathbf{x},\omega)$ may be expressed via its real-valued amplitude $A_P(\mathbf{x},\omega)$ and phase $\Phi_P(\mathbf{x},\omega)$ as

[70] Firtha et al. (2017). "Improved Referencing Schemes for 2.5D Wave Field Synthesis Driving Functions". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 25.5, pp. 1117–1127, Eq. (15).

$$P(\mathbf{x},\omega) = A_P(\mathbf{x},\omega)\,\mathrm{e}^{+\mathrm{j}\Phi_P(\mathbf{x},\omega)}\,. \tag{2.51}$$

One may use this notation together with the product rule of differentiation to express the gradient of a sound field as

$$\nabla_{\mathbf{x}} P(\mathbf{x},\omega) = \left(\frac{\nabla_{\mathbf{x}} A_P(\mathbf{x},\omega)}{A_P(\mathbf{x},\omega)} + \mathrm{j}\nabla_{\mathbf{x}}\Phi_P(\mathbf{x},\omega) \right) P(\mathbf{x},\omega)\,. \tag{2.52}$$

For the used Fourier Transform convention, the local wavenumber vector is defined as

$$\mathbf{k}_P(\mathbf{x},\omega) := -\nabla_{\mathbf{x}}\Phi_P(\mathbf{x},\omega) = -\mathrm{Im}\left(\frac{\nabla_{\mathbf{x}} P(\mathbf{x},\omega)}{P(\mathbf{x},\omega)}\right) \tag{2.53}$$

as the gradient of the phase $\Phi_P(\mathbf{x},\omega)$ or as the imaginary part of the normalised gradient $\frac{\nabla_{\mathbf{x}} P(\mathbf{x},\omega)}{P(\mathbf{x},\omega)}$. The vector points towards the local propagation direction of the sound field at a given coordinate $\mathbf{x}$. Its normalised vector $\hat{\mathbf{k}}_P(\mathbf{x},\omega)$ is occasionally used to describe this direction. An example is shown for a point source in Fig. 2.8: The vector is oriented perpendicular to the wave fronts, i.e. to the surfaces of equal phase $\Phi_P(\mathbf{x},\omega)$. This is a general property of vector fields resulting from a gradient operator, where the direction is always perpendicular to surfaces of constant value of the underlying scalar field.[71] For the fundamental sound fields presented in Sec. 2.1.2, $\mathbf{k}_P(\mathbf{x},\omega)$ fulfils the local dispersion relation, i.e. its length is fixed to $\frac{\omega}{c}$.[72] For arbitrary sound fields, this statement is true for asymptotically high frequencies.[73] Generally

$$\mathbf{k}_P(\mathbf{x},\omega) \overset{\omega\to\infty}{\approx} \frac{\omega}{c}\,\hat{\mathbf{k}}_P(\mathbf{x},\omega) \tag{2.54}$$

holds. The approximation is based upon assumption, that the relative change of the amplitude expressed via $\frac{\nabla_{\mathbf{x}} A_P(\mathbf{x},\omega)}{A_P(\mathbf{x},\omega)}$ in (2.52) is much smaller than the change of the phase $\nabla_{\mathbf{x}}\Phi_P(\mathbf{x},\omega)$ for high frequencies. This assumption may be further used to omit the amplitude term in (2.52). The sound field gradient is then approximated via[74]

$$\nabla_{\mathbf{x}} P(\mathbf{x},\omega) \overset{\omega\to\infty}{\approx} -\mathrm{j}\frac{\omega}{c}\hat{\mathbf{k}}_P(\mathbf{x},\omega) P(\mathbf{x},\omega)\,. \tag{2.55}$$

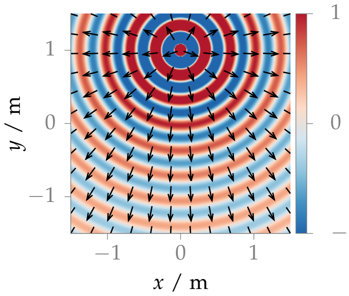


Figure 2.8: Real part of the sound pressure $P_{\mathrm{ps}}(\mathbf{x}|\mathbf{x}_{\mathrm{ps}},\omega)$ of a monochromatic ($f = 1$ kHz) point source (2.11) located at $\mathbf{x}_{\mathrm{ps}} = [0,1,0]^{\mathrm{T}}$ m. The plot is normalised to the pressure magnitude the coordinates' origin. The black arrows illustrate the according local wavenumber vector (2.53).

[71] Arfken and Weber (2005). *Mathematical Methods for Physicists*. 6th ed. Elsevier, p. 35.

[72] Firtha (2019). "A Generalized Wave Field Synthesis Framework with Application for Moving Virtual Sources". PhD thesis. Budapest University of Technology and Economics, Tab. 3.1.

[73] Kinsler et al. (Dec. 1999). *Fundamentals of Acoustics*. 4th ed. John Wiley & Sons, Inc., Sec. 5.14.

[74] Firtha, op. cit., Sec. 3.12.

3 Selected Methods for (Local) Sound Field Synthesis

As already stated in Sec. 1.1, methods for SFS pursue a physically accurate reconstruction of a desired sound field inside a defined target region. The representation of the sound field can be categorised into two fundamental principles:[1,2] In model-based rendering, the virtual sound field is a composition of elementary sources whose pressure is described via simple parametric models such as point sources or plane waves. Additional examples are given in Sec. 2.1.2. For each elementary source a (dry) source signal has to be provided. Data-based rendering synthesises a sound field whose spatio-temporal structure is acquired via Sound Field Analysis (SFA) techniques using e.g. microphone arrays.[3,4,5] The theory of SFS requires a continuous distribution of so-called secondary sources to be positioned around the target region in order to achieve correct synthesis. Each secondary source has to be fed with its driving signal such that the superposition of the sound fields emitted by all sources coincide with the virtual sound field. In practice, a limited number (up to hundreds) of individually driven loudspeakers approximate this continuous distribution. The synthesis accuracy is mainly limited by spatial sampling artefacts, which are introduced to the synthesised sound field due to the finite resolution of this discretisation. These artefacts can be avoided as long as a critical number of actuators are deployed,[6] which grows linearly with the spatial scale of the target region and the temporal frequency. The relation serves as a motivation for LSFS: A more accurate reproduction inside a downsized area which is smaller than the area surrounded by the Secondary Source Distribution (SSD) is pursued. To achieve this, stronger artefacts outside the prioritised area are permitted. This is sensible for applications, where the listener's position is restricted to a small region of interest or is tracked using a suitable technology.

Approaches to derive the driving signals of the SSD may be classified into numerical and analytical methods: Many numerical techniques to SFS[7,8,9,10] spatially sample the target region in order to establish a set of control points. They relate the desired sound pressure at these points to the loudspeaker driving signals via a linear equation system. It is solved using suitable methods for (regularised) matrix inversion. Hereby, no distinction between model and data-

This introduction was partly published in Winter et al. (May 2016). "On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.5, pp. 914–926

[1] Spors et al. (Oct. 2011). "Efficient realization of model-based rendering for 2.5-dimensional near-field compensated higher order Ambisonics". In: *Proc. of 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, USA, Sec. 1.

[2] Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Sec. 2.4.

[3] Teutsch (2006). "Wavefield Decomposition Using Microphone Arrays and Its Application to Acoustic Scene Analysis". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg.

[4] Kuntz (2009). "Wave Field Analysis Using Virtual Circular Microphone Arrays". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg.

[5] Rafaely (2015). *Fundamentals of Spherical Array Processing*. Heidelberg, Germany: Springer.

[6] Kennedy et al. (2007). "Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields". In: *IEEE Trans. Signal Process.* 55.6, pp. 2542–2556.

[7] Miyoshi and Kaneda (Feb. 1988). "Inverse filtering of room acoustics". In: *IEEE Trans. Acoust., Speech, Signal Process.* 36.2, pp. 145–152

[8] Kirkeby and Nelson (1993). "Reproduction of plane wave sound fields". In: *J. Acoust. Soc. Am.* 94.5

[9] Kirkeby et al. (1996). "Local sound field reproduction using digital signal processing". In: *J. Acoust. Soc. Am.* 100.3.

[10] Kolundžija et al. (2011). "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion". In: *J. Aud. Eng. Soc.* 59.10, pp. 721–734.

based rendering is made. Alternatively, the control points are established on the boundary of the target region in order to achieve accurate synthesis inside.[11] The inversion problem may be transferred to the Spherical Harmonics domain which is also known as mode matching.[12,13] Although most approaches perform calculations in the frequency domain, numerical methods for the time domain do exist.[14] A related yet different field of research that received significant attention in the recent years is multizone sound field reproduction. It aims at independently controlling the synthesised sound field in two or more portions of the target area.[15,16,17] In order to do so, these methods create bright zones with significant sound pressure and dark zones with minimum sound pressure. The dark zones can then be exposed to another bright zone that is ideally independent of the initial bright zone. The drawback of numerical approaches is the large computational effort which has to be spent to solve the system of linear equations. If the computation is carried out in the frequency domain the calculation of the driving signals typically has to be repeated for different temporal frequencies. Moreover, the required number of control points increases with increasing frequency. For time-variant sound fields e.g. with moving sound sources, the equation system has to be solved for each time step.

The driving signals of analytic SFS methods are derived from the parametric description of the virtual sound field. While this seems to restrict these techniques to the model-based rendering paradigm, data-based approaches are possible. As an example, the spatio-temporal structure of a sound field measured by a microphone array, can be converted into a PWD. Each plane wave is then synthesised individually using the parametric approach. Due to the mathematical expression for the driving signals, analytic SFS methods are simple to realise compared to the numerical methods. Moreover, the expressions allow to change parameters over time in order to enable time-variant synthesis scenarios. A better analysis of the connection between the parametrisation and the artefacts occurring in the synthesised sound field is also possible. Ahrens argues, that specific scenarios in numerical SFS have to considered individually and the fundamental properties are difficult to deduce.[18] Because of their benefits, this thesis focuses on analytic SFS methods.

This chapter introduces the reader to selected SFS methods and discusses potential challenges regarding their implementation. In Sec. 3.1, the fundamental synthesis problem is formulated. Solutions to it including their driving signals in the frequency domain and their realisation in the discrete-time domain are presented in the subsequent sections. They include two well-established representatives for conventional SFS, namely WFS[19] and NFCHOA.[20] The two LSFS methods termed LWFS-SBL[21] and LWFS-VSS[22] are discussed afterwards.

[11] Ise (1999). “A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems”. In: *Acta Acustica united with Acustica* 85.1, pp. 78–87

[12] Poletti (2005). “Three-dimensional surround sound systems based on spherical harmonics”. In: *J. Aud. Eng. Soc.* 11.53, pp. 1004–1025.

[13] Hannemann and Donohue (2008). “Virtual Sound Source Rendering Using a Multipole-Expansion and Method-of-Moments Approach”. In: *J. Aud. Eng. Soc.* 56.6, pp. 473–481.

[14] Feng et al. (2018). “Time-domain sound field reproduction using the group Lasso”. In: *J. Acoust. Soc. Am.* 143.2, EL55–EL60.

[15] Wu and Abhayapala (Aug. 2011). “Spatial Multizone Soundfield Reproduction: Theory and Design”. In: *IEEE Trans. Audio, Speech, Language Process.* 19.6, pp. 1711–1720.

[16] Coleman et al. (2014). “Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array”. In: *J. Acoust. Soc. Am.* 135.4.

[17] Jin and Kleijn (Dec. 2015). “Theory and Design of Multizone Soundfield Reproduction Using Sparse Methods”. In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 23.12, pp. 2343–2355.

[18] Ahrens (2010). “The Single-layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources”. PhD thesis. Berlin, Germany: Technische Universität Berlin

[19] Berkhout (1988). “A Holographic Approach to Acoustic Control”. In: *J. Aud. Eng. Soc.* 36.12, pp. 977–995

[20] Daniel (May 2003). “Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format”. In: *Proc. of 23rd Intl. Aud. Eng. Soc. Conf. on Signal Processing in Audio Recording and Reproduction*. Copenhagen, Denmark

[21] Hahn et al. (June 2016). “Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain”. In: *Proc. of 140th Aud. Eng. Soc. Conv.* Paris, France

[22] Spors and Ahrens (Oct. 2010b). “Local Sound Field Synthesis by Virtual Secondary Sources”. In: *Proc. of 40th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Tokyo, Japan

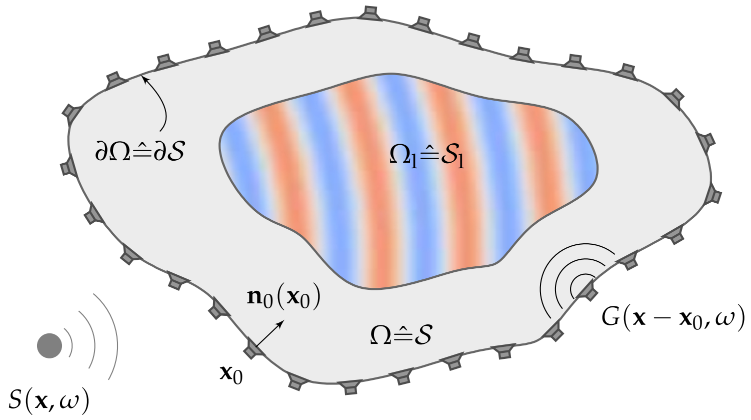


Figure 3.1: The illustration shows the geometry for the synthesis a virtual sound field $S(\mathbf{x}, \omega)$. For the 3D scenario, synthesis is supposed to be achieved inside Ω_l using the secondary sources positioned around Ω, i.e. along its boundary $\partial\Omega$. The counterparts for the 2D problem are denoted as $\mathcal{S}_l$, $\mathcal{S}$, and $\partial\mathcal{S}$, respectively.

3.1 Problem Formulation

The fundamental task is to synthesise a desired a.k.a. virtual sound field $s(\mathbf{x}, t)$ with its frequency spectrum $S(\mathbf{x}, \omega)$ within a defined target region $\Omega_l \subseteq \Omega$, see Fig. 3.1. For the case where $\Omega_l = \Omega$, approaches are usually referred to as conventional SFS. The remaining case, i.e. $\Omega_l \subset \Omega$, is usually termed LSFS. In order to achieve correct synthesis a distribution of secondary sources is positioned along the regions' boundary $\partial\Omega$. Each secondary source is oriented along the inward pointing boundary normal $\mathbf{n}_0$. Although approaches for SFS exist, where the loudspeakers' directivity is incorporated,[23] the sound field emitted by an individual secondary source is commonly modelled by a monopole point source, cf. Eq. (2.11). As mentioned by Schultz,[24] this model is in reasonable agreement with today's loudspeakers, especially at low frequencies. The sound field is thus given by the free-field Green's function $G(\mathbf{x}|\mathbf{x}_0, \omega) = G(\mathbf{x} - \mathbf{x}_0, \omega)$. Each individual secondary source is driven by its respective driving signal $D(\mathbf{x}_0, \omega)$, whereas $\mathbf{x}_0 \in \partial\Omega$ denotes the position of the secondary source on the boundary. The resulting superposition of all secondary sources constitutes the reproduced sound field $P(\mathbf{x}, \omega)$. The driving signals have to be chosen such that the reproduced and the desired sound field coincide within Ω_l. Mathematically, this is subsumed by the 3D Single Layer Potential (SLP)

$$S(\mathbf{x}, \omega) \stackrel{!}{=} P(\mathbf{x}, \omega) = \iint_{\partial\Omega} D(\mathbf{x}_0, \omega) G(\mathbf{x}|\mathbf{x}_0, \omega) \, \mathrm{d}A_0 \quad \forall \mathbf{x} \in \Omega_l \,. \tag{3.1}$$

Based on the discussions on the HIE and the usefulness of boundary conditions in Sec. 2.1.3, finding the correct driving signal is equivalent to solving the scattering problem stated in (2.31). Hence,

$$D(\mathbf{x}_0, \omega) = -\nabla_{\mathbf{x}_0, \mathbf{n}_0} S_T(\mathbf{x}_0, \omega) \tag{3.2}$$

where $\nabla_{\mathbf{x}_0, \mathbf{n}_0} S_T(\mathbf{x}_0, \omega)$ describes the directional derivative of the total sound field on the surface of a sound-soft scatterer with the same shape as Ω and $S(\mathbf{x}, \omega)$ as the impinging sound field.

For many practical applications, sound field synthesis is restricted to the horizontal plane, i.e. $z = 0$, with the secondary sources positioned on a contour $\partial\mathcal{S}$ enclosing the target area $\mathcal{S}_l \subseteq \mathcal{S}$, i.e. $z_0 = 0$.

[23] Ahrens and Spors (Aug. 2009). "An Analytical Approach to 2.5D Sound Field Reproduction Employing Circular Distributions of Non-Omnidirectional Loudspeakers". In: *Proc. of 17th European Signal Processing Conference (EUSIPCO)*. Glasgow, UK.

[24] Schultz (2016). "Sound Field Synthesis for Line Source Array Applications in Large-Scale Sound Reinforcement". PhD thesis. University of Rostock, Sec. 2.2.1.

Within the described scenario, it is assumed that the virtual sound field does only propagate in horizontal directions. The z-component of its local wavenumber vector $\mathbf{k}_S(\mathbf{x}, \omega)$ defined by (2.53) is zero. The 3D SLP specialises to the line integral

$$S(\mathbf{x}, \omega) \stackrel{!}{=} P(\mathbf{x}, \omega) = \int_{\partial\mathcal{S}} D(\mathbf{x}_0, \omega) G(\mathbf{x}|\mathbf{x}_0, \omega)\, \mathrm{d}l_0 \quad \forall \mathbf{x} \in \mathcal{S}_l\,, \qquad (3.3)$$

which will be referred to as the 2D SLP. A suitably chosen differential line segment is denoted as $\mathrm{d}l_0 = \mathrm{d}l(x_0, y_0)$. According to Williams,[25] the required sound field emitted by an individual secondary source to achieve perfect synthesis would have to coincide with the one of a line source.[26] As already mentioned, the usually employed loudspeakers with closed cabinets exhibit point-source-like radiation characteristics. This dimensionality mismatch is usually subsumed under the term 2½-dimensional (2.5D) synthesis.[27] As a consequence, the amplitude decay of the synthesized sound field differs from that of the desired one. These deviations of the amplitude decay are systemic so that there is no general cure. The synthesized sound field can be referenced to a given contour or location inside $\mathcal{S}$ on which/at which the complex amplitude is correct. This was exemplarily shown for WFS[28] and NFCHOA.[29]

[25] Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, Sec. 8.6.2.

[26] see Sec. 2.1.2

[27] Verheijen (1997). "Sound Reproduction by Wave Field Synthesis". PhD thesis. Delft University of Technology, Sec. 2.3.

[28] Firtha et al. (2017). "Improved Referencing Schemes for 2.5D Wave Field Synthesis Driving Functions". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 25.5, pp. 1117–1127.

[29] Winter et al. (May 2016). "On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.5, pp. 914–926, Sec. II.B.

3.2 Wave Field Synthesis (WFS)

According to the extensive overview by Schultz,[30] the history of WFS dates back to the late 1980s. Berkhout proposed an approach for acoustic holography using loudspeaker arrays[31] which was later termed Wave Field Synthesis.[32] Research and development of WFS was dominated by the Delft University of Technology until the beginning of the next millennium including important dissertations by Vogel,[33] Start,[34] and Verheijen.[35] The theory of WFS was revisited by Spors et al.,[36] where it was put in the greater context by considering it to be an implicit solution to the SLP. A modern framework for WFS using the concept of the local wavenumber vector and linking the above approaches was published by Firtha.[37] It had, moreover, a special focus on moving virtual sound sources.

[30] Schultz, op. cit., pp. 14–16.

[31] Berkhout, op. cit.

[32] Berkhout et al. (1993). "Acoustic control by wave field synthesis". In: *J. Acoust. Soc. Am.* 93.5, pp. 2764–2778.

[33] Vogel (1993). "Application of wave field synthesis in room acoustics". PhD thesis. Delft University of Technology.

[34] Start (1997). "Direct Sound Enhancement by Wave Field Synthesis". PhD thesis. Delft University of Technology.

[35] Verheijen, op. cit.

[36] Spors et al. (May 2008). "The theory of Wave Field Synthesis revisited". In: *Proc. of 124th Aud. Eng. Soc. Conv.* Amsterdam, The Netherlands.

[37] Firtha (2019). "A Generalized Wave Field Synthesis Framework with Application for Moving Virtual Sources". PhD thesis. Budapest University of Technology and Economics.

3.2.1 Driving Signals in the Frequency Domain

Driving Signal for 3D Scenarios: WFS is based upon the Kirchhoff approximation[38] of the equivalent scattering problem: For high frequencies, i.e. short wavelengths, a convex scatterer can be assumed to be locally planar with $\mathbf{n}_0$ being the normal vector of the plane at $\mathbf{x}_0$. The directional gradient of the total sound field then reads[39]

$$\nabla_{\mathbf{x}_0,\mathbf{n}_0} S_T(\mathbf{x}_0, \omega) \approx a_S(\mathbf{x}_0, \omega) 2 \nabla_{\mathbf{x}_0,\mathbf{n}_0} S(\mathbf{x}_0, \omega)\,, \qquad (3.4)$$

where $a_S(\mathbf{x}_0, \omega)$ divides the $\partial\Omega$ into a region "illuminated" by the impinging sound field and a shadowed area. With the help of the local wavenumber vector it is generally expressed as[40]

[38] Colton and Kress (2013). *Inverse acoustic and electromagnetic scattering theory*. 3rd ed. Vol. 93. Applied Mathematical Sciences. Springer Science & Business Media, p. 57.

[39] Ibid., Eq. (3.35).

[40] Firtha et al., op. cit., Eq. (46).

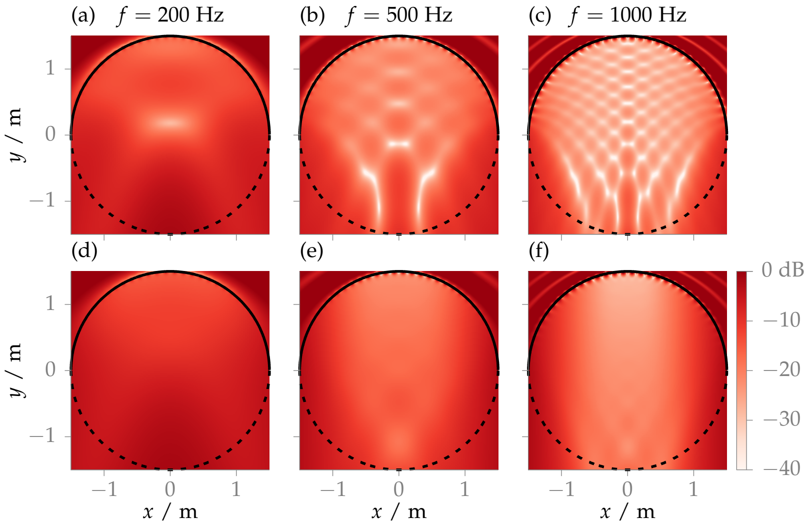


Figure 3.2: The plots show the normalised absolute error

$$20\log_{10}\left|\frac{P(\mathbf{x},\omega)-P_{\mathrm{pw}}(\mathbf{x},\omega)}{P_{\mathrm{pw}}(\mathbf{x},\omega)}\right|$$

between synthesised sound field $P(\mathbf{x},\omega)$ (3.1) and a virtual monochromatic plane wave $P_{\mathrm{pw}}(\mathbf{x},\omega)$ with the propagation direction $\mathbf{n}_{\mathrm{pw}} = [0,-1,0]^{\mathrm{T}}$, see Eq. (2.21) and Fig. 2.5. A quasi-continuous (2450 secondary sources), spherical SSD centred around the coordinates' origin with a radius of 1.5 m (black line) is driven by 3D WFS, see (3.6b). The dashed part of the SSD is inactive due to the secondary source selection criterion given by (3.5). For the bottom row an additional tapering is applied, see Eq. (3.7).

$$a_S(\mathbf{x}_0,\omega) = \begin{cases} 1 & \text{, if } \langle \mathbf{k}_S(\mathbf{x}_0,\omega) | \mathbf{n}_0 \rangle \geq 0 \text{ and} \\ 0 & \text{, otherwise.} \end{cases} \tag{3.5}$$

In the context of WFS, $a_S(\mathbf{x}_0,\omega)$ is often referred to as the secondary source selection criterion.[41] The 3D WFS driving signal is given as[42]

$$D_{3\mathrm{D}}^{\mathrm{WFS}}(\mathbf{x}_0,\omega) = -2\, a_S(\mathbf{x}_0,\omega)\, \nabla_{\mathbf{x}_0,\mathbf{n}_0} S(\mathbf{x}_0,\omega) \tag{3.6a}$$

$$\approx 2\mathrm{j}\frac{\omega}{c} a_S(\mathbf{x}_0,\omega)\, \langle \hat{\mathbf{k}}_S(\mathbf{x}_0,\omega) | \mathbf{n}_0 \rangle S(\mathbf{x}_0,\omega)\,. \tag{3.6b}$$

The high-frequency approximation of the sound field gradient given by (2.55) was used for the second expression.[43] The secondary source selection can be interpreted as a rectangular truncation window of the driving signal. For simple source models, e.g. plane waves or point sources, $a_S(\mathbf{x}_0,\omega)$ can be derived by geometrical considerations and is independent of the frequency. A list of the relevant criteria is later given in Tab. 3.1.

In order to demonstrate the effect of the truncation, synthesis examples are shown in Fig. 3.2. The plots show the error between the synthesised sound field and a plane wave serving as the virtual sound field. For the plots in top row, the secondary source selection defined by (3.5) was used. The error generally decreases as the frequency increases. Here, the Kirchhoff approximation becomes more accurate for high frequencies. The observed fluctuations in the sound field are due to the secondary source selection or, more specifically, due to its discontinuity at the transition between the illuminated and the shadowed SSD (solid and dashed part). It was discussed by Verheijen[44] and Spors[45] that the employment of window functions other than the rectangular one, usually referred to as tapering, has a homogenising effect on the synthesis error. This can be observed in bottom row of Fig. 3.2, where the tapered secondary selection criterion

$$a_S(\mathbf{x}_0,\omega) = \begin{cases} \langle \mathbf{k}_S(\mathbf{x}_0,\omega) | \mathbf{n}_0 \rangle & \text{, if } \langle \mathbf{k}_S(\mathbf{x}_0,\omega) | \mathbf{n}_0 \rangle \geq 0 \text{ and} \\ 0 & \text{, otherwise.} \end{cases} \tag{3.7}$$

[41] Spors (Oct. 2007). "Extension of an Analytic Secondary Source Selection Criterion for Wave Field Synthesis". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA.

[42] Ahrens, op. cit., Eq. 2.54.

[43] Firtha, op. cit., Eq. (4.4).

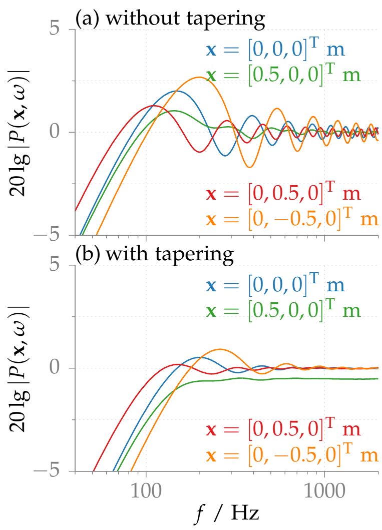


Figure 3.3: The plots show the magnitude spectra of the sound field synthesised using WFS. The same synthesis setup as for Fig. 3.2 is used. In (b), the additional tapering defined in (3.7) is employed.

[44] Verheijen, op. cit., Sec. 2.4.1.

[45] Spors, op. cit., Sec. 5.

	Point Source $D^{\text{WFS}}_{\text{2.5D,ps}}(\mathbf{x}_0\vert\mathbf{x}_{\text{ps}},\omega)$	Plane Wave $D^{\text{WFS}}_{\text{2.5D,pw}}(\mathbf{x}_0\vert\mathbf{n}_{\text{pw}},\omega)$	Focused Point Source $D^{\text{WFS}}_{\text{2.5D,fs}}(\mathbf{x}_0\vert\mathbf{x}_{\text{fs}},\mathbf{n}_{\text{fs}},\omega)$
$a_S(\mathbf{x}_0,\omega)=1$, if ...	$\langle\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert\mathbf{n}_0\rangle\geq 0$	$\langle\mathbf{n}_{\text{pw}}\vert\mathbf{n}_0\rangle\geq 0$	$\langle\mathbf{x}_{\text{fs}}-\mathbf{x}_0\vert\mathbf{n}_{\text{fs}}\rangle\geq 0$ and $\langle\mathbf{x}_{\text{fs}}-\mathbf{x}_0\vert\mathbf{n}_0\rangle\geq 0$
$\Delta_S(\mathbf{x}_0)$	$\frac{\vert\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert\vert\mathbf{x}_0-\mathbf{x}_{\text{ref}}\vert}{\vert\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert+\vert\mathbf{x}_0-\mathbf{x}_{\text{ref}}\vert}$	$\vert\mathbf{x}_0-\mathbf{x}_{\text{ref}}\vert$	$-\vert\mathbf{x}_0-\mathbf{x}_{\text{fs}}\vert\left[1+\frac{\vert\mathbf{x}_0-\mathbf{x}_{\text{fs}}\vert}{\vert\mathbf{x}_{\text{ref}}-\mathbf{x}_{\text{fs}}\vert}\right]$
$\langle\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)\vert\mathbf{n}_0\rangle$	$\frac{\langle\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert\mathbf{n}_0\rangle}{\vert\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert}$	$\langle\mathbf{n}_{\text{pw}}\vert\mathbf{n}_0\rangle$	$\frac{\langle\mathbf{x}_{\text{fs}}-\mathbf{x}_0\vert\mathbf{n}_0\rangle}{\vert\mathbf{x}_0-\mathbf{x}_{\text{fs}}\vert}$
$S(\mathbf{x}_0,\omega)$	$\tilde{S}(\omega)\frac{\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\vert\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert}}{4\pi\vert\mathbf{x}_0-\mathbf{x}_{\text{ps}}\vert}$	$\tilde{S}(\omega)\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle\mathbf{x}_0\vert\mathbf{n}_{\text{pw}}\rangle}$	$\tilde{S}(\omega)\frac{\mathrm{e}^{+\mathrm{j}\frac{\omega}{c}\vert\mathbf{x}_0-\mathbf{x}_{\text{fs}}\vert}}{4\pi\vert\mathbf{x}_0-\mathbf{x}_{\text{fs}}\vert}$

Table 3.1: Quantities for the 2.5D Wave Field Synthesis (WFS) driving signal in (3.9a).

is applied. Although the error for the tapered criterion is generally higher than for the original one, fewer fluctuations are visible. The findings are confirmed by the magnitude spectra of the synthesised sound field shown in Fig. 3.3. The tapering results in fewer fluctuations in the spectra. As a drawback, the pressure loss at low frequencies is more pronounced for the tapered driving signals. Also, an incorrect magnitude different from the desired 0 dB is observed at $\mathbf{x}=[0.5,0,0]^{\mathrm{T}}$ m.

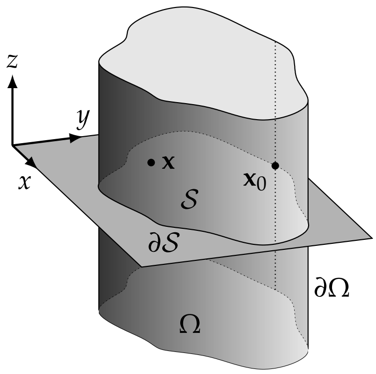


Figure 3.4: The geometry for the derivation of the 2.5D WFS approach considers a 2D area $\mathcal{S}$ which is extruded along the z-direction to form the 3D region Ω.

Driving Signal for 2.5D Scenarios: For the derivation of the 2.5D WFS approach, the volume Ω shown in Fig. 3.4 is considered: In the xy-plane, a 2D convex area $\mathcal{S}$ is defined. The 3D region Ω is constructed by extruding $\mathcal{S}$ in z-direction towards $\pm\infty$. Hence, the shape of Ω and its surface normal $\mathbf{n}_0$ are independent from the z-coordinate. The SLP (3.1) together with the 3D WFS driving signal (3.6a) migrates to

$$S(\mathbf{x},\omega)\approx\int_{\partial S}\int_{-\infty}^{\infty}D^{\text{WFS}}_{\text{3D}}(\mathbf{x}_0,\omega)G(\mathbf{x}\vert\mathbf{x}_0,\omega)\,\mathrm{d}z_0\,\mathrm{d}l_0\,. \tag{3.8}$$

The integral w.r.t. z_0 is solved using the Stationary Phase Approximation (SPA). It is introduced in App. A. Since the approximation involves the whole integrand of (3.8), the calculus heavily depends on the specific virtual sound field. For detailed derivation, the reader is referred to Firtha et al.[46,47] The remaining integral over $\partial\mathcal{S}$ matches the 2D SLP given by (3.3). The generic 2.5D WFS driving signal and approximate solution of the 2D SLP is given by[48]

[46] Firtha et al., op. cit.

[47] Firtha, op. cit., Sec. 4.1

[48] Firtha et al., op. cit., Eq. (47)

$$D^{\text{WFS}}_{\text{2.5D}}(\mathbf{x}_0,\omega)=-a_S(\mathbf{x}_0,\omega)\sqrt{\frac{8\pi\Delta_S(\mathbf{x}_0)}{\mathrm{j}\frac{\omega}{c}}}\nabla_{\mathbf{x}_0,\mathbf{n}_0}S(\mathbf{x}_0,\omega) \tag{3.9a}$$

$$\approx a_S(\mathbf{x}_0,\omega)\sqrt{\mathrm{j}\frac{\omega}{c}8\pi\Delta_S(\mathbf{x}_0)}\langle\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)\vert\mathbf{n}_0\rangle S(\mathbf{x}_0,\omega)\,. \tag{3.9b}$$

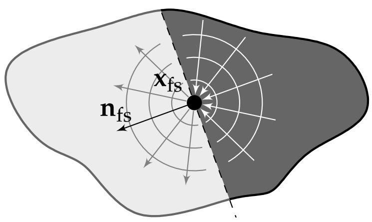


Figure 3.5: For a focused point source located at $\mathbf{x}_{\text{fs}}$ and oriented along $\mathbf{n}_{\text{fs}}$. A plane along the dashed black line separates the converging (white arrows) from the diverging (grey arrows) half spaces from each other.

Again, the high-frequency approximation of the sound field gradient given by (2.55) was used for the second expression.[49] The distance factor $\Delta_S(\mathbf{x}_0)$ can be used to reference the synthesised sound field to a given contour or position on which/at which asymptotically

[49] Firtha, op. cit., Eq. (4.14).

`driving_function_mono_wfs`

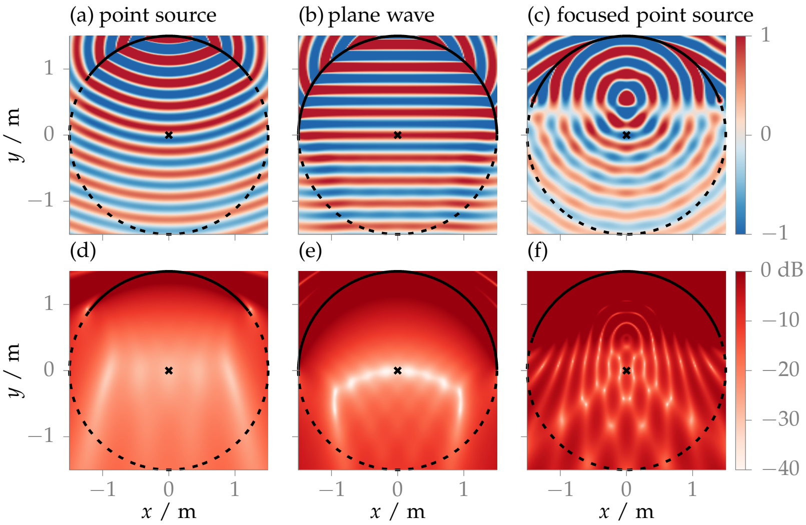


Figure 3.6: The top plots show the synthesis sound field $P(\mathbf{x}, \omega)$ (3.3) using a quasi-continuous (1024 secondary sources), circular SSD centred around the coordinates' origin with a radius of 1.5 m (black line). It is driven by 2.5D WFS (3.9b) with the quantities given in Tab. 3.1 according to the titles above each plot. The dashed part of the SSD is inactive due to the secondary source selection criterion. The bottom plots show the normalised absolute error

$$20 \log_{10} \left| \frac{P(\mathbf{x}, \omega) - S(\mathbf{x}, \omega)}{S(\mathbf{x}, \omega)} \right|$$

to the desired sound fields $S(\mathbf{x}, \omega)$, which are a point source at $[0, 2.5, 0]^{\mathrm{T}}$ m, a plane wave with $\mathbf{n}_{\mathrm{pw}} = [0, -1, 0]^{\mathrm{T}}$, and a point source at $[0, 0.5, 0]^{\mathrm{T}}$ m with a monochromatic ($f = 1$ kHz) source signal. The reference coordinate $\mathbf{x}_{\mathrm{ref}}$ is set to $[0, 0, 0]^{\mathrm{T}}$ m (black cross).

($\omega \to \infty$) correct synthesis is achieved. For more details, the reader is referred to Firtha et al.[50] The quantities of the 2.5D WFS driving signal for a virtual point source and a plane wave[51] are listed in Tab. 3.1. The involved coordinate $\mathbf{x}_{\mathrm{ref}}$ defines the reference position where asymptotically correct synthesis is achieved.

[50] Firtha et al., op. cit.

[51] Schultz, op. cit., Eq. (2.137) and (2.177).

It was explained in Sec. 2.1.3 that the synthesis of inhomogeneities such as point sources inside $\mathcal{S}$ is not possible with an SSD located at $\partial\mathcal{S}$. However, amongst other SFS methods, WFS allows for the synthesis of focused sound sources. They act as approximations of the virtual inhomogeneities inside a part of $\mathcal{S}$. For a point source, this can be achieved by emitting a sound field that converges towards a focus point $\mathbf{x}_{\mathrm{fs}}$ and diverges afterwards. As shown in Fig. 3.5, converging and diverging half space are separated via a plane with the normal vector $\mathbf{n}_{\mathrm{fs}}$ located at the focus point $\mathbf{x}_{\mathrm{fs}}$. The underlying principle is termed acoustic focusing by time reversal/phase conjugation.[52,53,54] The according quantities for the driving signal[55] are also given in Tab. 3.1.

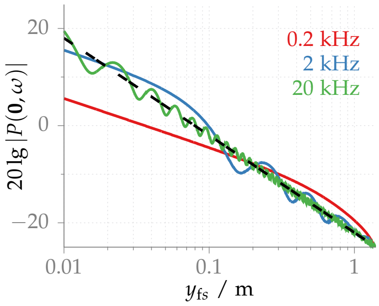


Figure 3.7: Magnitude of the synthesised monochromatic focused point source located at $\mathbf{x}_{\mathrm{fs}} = [0, y_{\mathrm{fs}}, 0]$ evaluated at the coordinates' origin $\mathbf{0}$. The simulation was performed for different frequencies (coloured lines). For the synthesis, the same setup as for Fig. 3.6 was used. For comparison, the magnitude of a point source at $\mathbf{x}_{\mathrm{fs}} = [0, y_{\mathrm{fs}}, 0]$ is plotted in dashed black.

An example for the synthesis resulting from the three different driving functions is shown in Fig. 3.6. For the point source and the plane wave, the wave fronts match the expectations. The corresponding plots at bottom indicate, that the error near the reference position $\mathbf{x}_{\mathrm{ref}}$ (black cross) is reduced but not negligible. Perfect synthesis at this location can be only achieved for asymptotically high frequencies and distances, as WFS constitutes a high-frequency/far-field approximation. For the focused point source, ripples in the wave fronts are visible resulting in an comparatively high synthesis error. In Fig. 3.7, the magnitude of the focused point source at the reference position is compared to the desired point source for different frequencies and source positions. It can be deduced, that the focused source approximates the point source best for high frequencies and large distances, i.e. high y_{fs}.

[52] Yon et al. (2003). "Sound focusing in rooms: The time-reversal approach". In: *J. Acoust. Soc. Am.* 113.3.

[53] de Vries and Berkhout (1981). "Wave theoretical approach to acoustic focusing". In: *J. Acoust. Soc. Am.* 70.3.

[54] Fink (Sept. 1992). "Time reversal of ultrasonic fields. I. Basic principles". In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 39.5, pp. 555–566.

[55] Winter et al. (June 2019a). "A Geometric Model for Prediction of Spatial Aliasing in 2.5D Sound Field Synthesis". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 27.6, pp. 1031–1046, Eq. (64).

3.2.2 Practical Realisation of Model-Based Rendering[56]

The 2D SLP in (3.3) constitutes the integration of the driving signal over a continuous SSD, which cannot be implemented with today's loudspeaker technology. For the implementation of WFS, it is hence necessary to spatially discretise the SSD. A finite number N_0 of loudspeakers has to be placed at discrete points on the boundary $\partial\mathcal{S}$. The SLP transforms to a sum over finite set of secondary source positions $\mathbf{x}_0^{(\nu)}$. It is known, that this discretisation may lead to spatial aliasing, which will be discussed extensively in Ch. 4. As a consequence of digital audio processing, temporal sampling of the driving signals is also considered. It is conveniently modelled by multiplying a time-continuous signal $\tilde{s}(t)$ with a Dirac comb.[57] A discrete-time signal $\tilde{s}[n]$ corresponds to the samples $\tilde{s}(nT_\mathrm{s})$ where T_s defines the sampling period as the reciprocal of the sample rate f_s. The discrete-time signal is connected to the desired continuous frequency spectrum $\tilde{S}(\omega)$ via the Discrete Time Fourier Transform (DTFT)[58], which should not be confused with the Discrete Fourier Transform (DFT)[59]. For the virtual source types listed in Tab. 3.1, the 2.5D WFS driving signal exhibit the general structure of

$$D_{2.5\mathrm{D}}^{\mathrm{WFS}}(\mathbf{x}_0, \omega) = \tilde{S}(\omega) \underbrace{\sqrt{\pm \mathrm{j}\frac{\omega}{\mathrm{c}}}}_{\text{pre-filter}} \underbrace{w(\mathbf{x}_0)}_{\text{weight}} \underbrace{\mathrm{e}^{-\mathrm{j}\omega\tau(\mathbf{x}_0)}}_{\text{delay}} \tag{3.10}$$

stating a geometry-independent pre-filtering of the source signal followed by geometry-dependent real-valued weight $w(\mathbf{x}_0)$ and delay $\tau(\mathbf{x}_0)$. The option $\pm$ for the pre-filter depends on the sign of the distance factor $\Delta_\mathrm{S}(\mathbf{x}_0)$ in the 2.5D WFS driving signal. The equation covers the driving signals for a single virtual source. In practice, the driving signals for different virtual sources are added in order to render the whole scene. This is illustrated by the addition operators in Fig. 3.8. The results are then D/A converted, amplified and played back over the loudspeakers. The rendering for a single virtual source is implemented in two essential steps described in the following.

Pre-Filtering: As the pre-filter $h_\mathrm{pre}[n]$ is geometry independent it can be either directly applied to the source signal $\tilde{s}[n]$ (option I) or to the driving signals of each loudspeaker (option II). The first strategy is more efficient, if the number of loudspeakers is larger than the number of virtual sources. If all source signals are known a-priori, an offline pre-filtering may also be considered. As the number of loudspeakers is subject to tighter practical constraints than the number of virtual sources, option I is a sensible standard strategy. However, option II was chosen in some implementations of WFS.[60,61]

An analytic formulation for $h_\mathrm{pre}[n]$ with the desired transfer function was derived by Schultz.[62] Since it is of infinite length, Schultz designed an Finite Impulse Response (FIR) approximation by applying the windowing method.[63] An alternative approach uses an Infinite Impulse Response (IIR) approximation.[64] However, the men-

[56] Parts of this section are published in Winter and Spors (Aug. 2016). "On Fractional Delay Interpolation for Local Wave Field Synthesis". In: *Proc. of 24th European Signal Processing Conference (EUSIPCO)*. Budapest, Hungary, pp. 2415–2419; and in Winter and Spors (Mar. 2017). "Time-Domain Realisations of 2.5-Dimensional Local Sound Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Kiel, Germany, pp. 1134–1137.

`driving_function_imp_wfs`

[57] Girod et al. (2001). *Signal and Systems*. Wiley, Sec. 11.3.1.

[58] Manolakis and Ingle (June 2011). *Applied Digital Signal Processing: Theory and Practice*. Cambridge, USA: Cambridge University Press, Sec. 4.3.2.

[59] Ibid., Sec. 7.2.

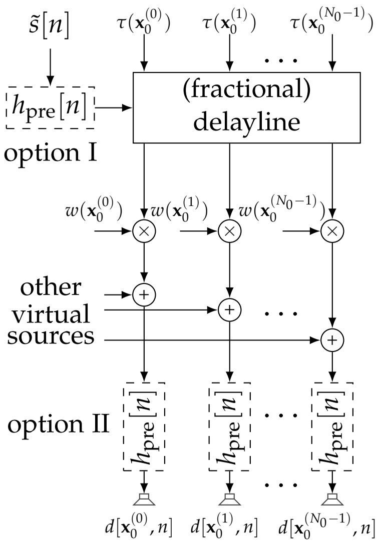


Figure 3.8: Block diagram showing the time domain realisation of WFS for one virtual source. Contributions from other virtual sources are incorporated via the addition-operators.

[60] Geier and Spors (Nov. 2012). "Spatial Audio with the SoundScape Renderer". In: *Proc. of 27th Tonmeistertagung – VDT Internation Convention*. Cologne, Germany.

[61] Wierstorf and Spors (Apr. 2012). "Sound Field Synthesis Toolbox". In: *Proc. of 132nd Aud. Eng. Soc. Conv.* Budapest, Hungary.

[62] Schultz, op. cit., Sec. 2.5.

[63] Manolakis and Ingle, op. cit., Sec. 10.3.

[64] Salvador (May 2010). "Wave Field Synthesis Using Fractional Order Filters and Fractional Delays". In: *Proc. of 120th Aud. Eng. Soc. Conv.* London, UK.

tioned solutions do not allow to incorporate additional practical aspects for WFS:[65] Due to the diffraction artefacts, a pressure loss of approximately -3 dB per Octave in the low frequency region can be observed in the synthesised sound field.[66] Further, spatial aliasing leads to an energy boost at high frequencies with an average increase of approximately 3 dB per Octave.[67] This is exemplarily shown for a virtual plane wave in Fig. 3.10b. A shelving of the pre-filter depicted in Fig. 3.10a flattens out the mentioned effects. Within this thesis, a least-squares FIR approximation of the ideal shelved spectrum is used. It allows to define the two corner frequencies for the low and high-frequency shelve in a flexible manner. Alternatively, shelved IIR pre-filters were presented by Schultz et al.[68]

[65] Schultz et al. (Mar. 2013). "Derivation of IIR-pre-filters for soundfield synthesis using linear secondary source distributions". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Meran, Italy, Sec. "IIR Filter Design".

[66] Spors and Ahrens (May 2010a). "Analysis and Improvement of Pre-equalization in 2.5-dimensional Wave Field Synthesis". In: *Proc. of 128th Aud. Eng. Soc. Conv.* London, UK, Sec. 7.2.

[67] Ibid., Sec. 6.3.

`wfs_fir_prefilter`

[68] Schultz et al., op. cit.

`delayline`

(Fractional) Delay and Weighting: Afterwards, the potentially pre-filtered source signal $\tilde{s}[n]$ is stored in a delayline. A delayline is essentially a signal buffer, from which delayed versions of the source signal can be requested. The geometry-dependent weighting $w(\mathbf{x}_0)$ may also be included in each request. It is worth noting, that the source signal is only written once into the delayline while the number of requests scales linearly with the number of loudspeakers N_0. In terms of computational effort, the request process is most critical. The required delays $\tau(\mathbf{x}_0)$ are generally not an integer multiple of the sample period T_s. Furthermore, the change of the scene geometry over time, e.g. moving sound sources, results in time-variant delays. Interpolation has to be applied to the sampled source signal in order to retrieve signal values for inter-sample positions. An extensive overview about the realisation of fractional delay interpolation and arbitrary sample rate conversion for WFS is given by Franck.[69] It is sufficient here to list some of the most practical interpolation strategies.

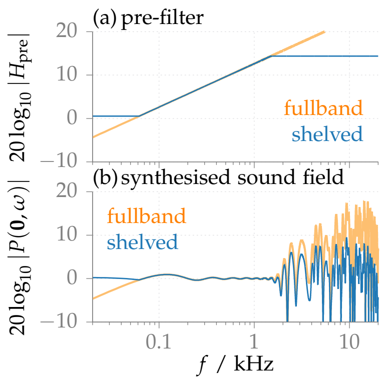


Figure 3.9: Plot (a) shows the idealised magnitude spectra of the fullband and shelved pre-filters. The resulting spectra of the sound field synthesised are plotted for the coordinates' origin in (b). A circular SSD with a radius of 1.5 m and 56 equi-angularly positioned secondary sources was employed to synthesise a virtual plane wave. The WFS driving signal is given by (3.9b) together with Tab. 3.1.

The nearest neighbour interpolation w.r.t. the sample position is regarded as the simplest and computationally cheapest method: It has been shown by Ahrens et al.[70] that this approach is perceptually sufficient for stationary scenarios reproduced with a circular loudspeaker array (1.5 m radius, 56 loudspeakers) at a sampling frequency of 44.1 kHz. Own research[71] suggests, that this transparency is due to significant spatial aliasing artefacts introduced by the reproduction setup. They dominate the artefacts caused by the delay interpolation. As a rule of thumb, a better delay interpolation is required, if the spatial aliasing artefacts are less prominent. According to Franck et al.,[72] audible artefacts occur in dynamic scenarios even for slow source movements with this interpolation strategy.

The second option interpolates the source signal upon request e.g. by applying a suitable fractional delay filter depending on the requested delay. Two well-known filter types are the Lagrange FIR filter and the Thiran IIR Allpass filter.[73] The accuracy of the interpolation generally increases with the number of filter coefficients. Due to the filtering and the (re-)calculation of the filter coefficients, the computational cost per request are comparatively high.

[69] Franck (2011). "Efficient Algorithms for Arbitrary Sample Rate Conversion with Application to Wave Field Synthesis". PhD thesis. Technische Universität Ilmenau, p. 269.

[70] Ahrens et al. (May 2010). "Perceptual Assessment of Delay Accuracy and Loudspeaker Misplacement in Wave Field Synthesis". In: *Proc. of 128th Aud. Eng. Soc. Conv.* London, UK, Sec. 7.

[71] Winter and Spors, "On Fractional Delay Interpolation for Local Wave Field Synthesis".

[72] Franck et al. (Sept. 2007). "Reproduction of Moving Sound Sources by Wave Field Synthesis: An Analysis of Artifacts". In: *Proc. of 32nd Intl. Aud. Eng. Soc. Conf. on DSP For Loudspeakers*. Hilleroed, Danmark, Sec. 3.2.

[73] Laakso et al. (Jan. 1996). "Splitting the unit delay [FIR/all pass filters design]". In: *IEEE Signal Process. Mag.* 13.1, pp. 30–60.

A delay independent preprocessing is an efficient extension to the prior category. A prominent approach is the oversampling of the signal about an integer factor. After the upsampling, missing samples are calculated via a delay-independent interpolation filter, which is efficiently implemented using polyphase structures.[74] The oversampled signal is stored in the delayline. Upon request, the delay is rounded to the next integer or is interpolated by a low-order Fractional Delay (FD) filter in the oversampled domain. The delayed signal is then downsampled, again. As a drawback, the memory required to store the oversampled signal scales with the upsampling factor. Within own work,[75] an oversampling of factor 8 with a Parks-McClellan[76] linear-phase FIR interpolation filter of 512 taps and a 9th-order Lagrange interpolator for the individual delays showed close to no artefacts for a sampling frequency of 44.1 kHz. It will be used as the reference, high accuracy FD method in this thesis. A comparison between the reference method and the nearest-neighbour interpolation is conducted in Fig. 3.10: For the discrete SSD, the pressure of the synthesised sound field is dominated by spatial aliasing artefacts at high frequencies. The chosen delay interpolation has close to no influence. Without spatial aliasing in the continuous, the impact becomes observable.

[74] Manolakis and Ingle, op. cit., Sec. 12.2

[75] Winter and Spors, op. cit.

[76] McClellan and Parks (2005). "A personal history of the Parks-McClellan algorithm". In: *IEEE Signal Process. Mag.* 22.2, pp. 82–86

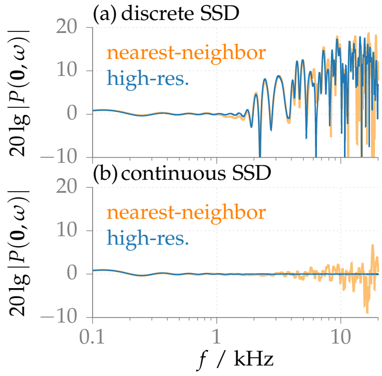


Figure 3.10: Plots shows the magnitude spectra of the synthesised sound field at the coordinates' origin. A circular SSD with a radius of 1.5 m is driven by WFS to synthesise a virtual plane wave. The driving signal is given (3.9b) together with the quantities in Tab. 3.1. A discrete SSD consist of 56 equi-angularly positioned secondary sources. The coloured lines indicate the used delay interpolation method.

3.3 Near-Field-Compensated Higher-Order Ambisonics

The complicated terminology of Near-Field-Compensated Higher-Order Ambisonics (NFCHOA) already suggest, that this SFS method is an extension to Ambisonics which was originally published by Gerzon.[77] According to Ahrens,[78] Ambisonics assumes the virtual sound field and the sound field emitted by each secondary source to be plane waves. The driving signals are then calculated in the Spherical Harmonics domain. Zotter et al.[79] give an alternative interpretation for Ambisonics, where the a virtual point source is co-located on a unit sphere together with all secondary sources. Both interpretations yield the same driving signals. Fazi[80] mentioned in his review, that early approaches were restricted to the zeroth and the first order of Spherical Harmonics although the theory of Gerzon did not involve any restrictions in that regard. Several extensions towards higher orders subsumed under Higher Order Ambisonics (HOA) were presented, from which the works of Bamford[81] and Daniel[82] are mentioned exemplary. Finally, the near-field compensation[83] models the secondary sources as point sources and constitutes NFCHOA. The approach was further generalised to incorporate directive secondary sources.[84]

[77] Gerzon (1973). "Periphony: With-Height Sound Reproduction". In: *J. Aud. Eng. Soc.* 21.1, pp. 2–10.

[78] Ahrens (2012). *Analytic Methods of Sound Field Synthesis*. T-Labs Series in Telecommunication Services. Berlin Heidelberg, Germany: Springer-Verlag, Sec. 1.2.4 and Sec. 3.3.5.

[79] Zotter et al. (May 2010). "Ambisonic Decoding with and without Mode-Matching: A Case Study Using the Hemisphere". In: *Proc. of 2nd Int. Symp. on Ambisonics and Spherical Acoustics*. Paris, France, Sec. 2.

[80] Fazi (2010). "Sound field reproduction". PhD thesis. University of Southampton, Sec. 1.1.2.

[81] Bamford (1995). "An Analysis of Ambisonics Sound Systems of First and Second Order". MA thesis. Waterloo, Canada: University of Waterloo.

[82] Daniel (2001). "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia". PhD thesis. Université Paris 6.

[83] Idem, "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format".

[84] Ahrens and Spors, op. cit.

Source type	Denotations		$\mathring{D}_m(\omega)$	Condition
	Driving Function	Coefficients		
Point Source	$D^{\text{NFCHOA}}_{\text{2.5D,ps}}(\mathbf{x}_0\|\mathbf{x}_{\text{ps}},\omega)$	$\mathring{D}^{\text{NFCHOA}}_{m,\text{ps}}(\omega)$	$\tilde{S}(\omega)\dfrac{h^{(2)}_{\|m\|}(\frac{\omega}{c}r_{\text{ps}})}{2\pi R\, h^{(2)}_{\|m\|}(\frac{\omega}{c}R)}\mathrm{e}^{-\mathrm{j}m\phi_{\text{ps}}}$	$R < r_{\text{ps}},\ \vartheta_{\text{ps}} = \frac{\pi}{2}$
Line Source	$D^{\text{NFCHOA}}_{\text{2.5D,ls}}(\mathbf{x}_0\|\mathbf{x}_{\text{ls}},\omega)$	$\mathring{D}^{\text{NFCHOA}}_{m,\text{ls}}(\omega)$	$\tilde{S}(\omega)\dfrac{\mathrm{j}^{m-\|m\|}\, H^{(2)}_{m}(\frac{\omega}{c}\rho_{\text{ls}})}{2\frac{\omega}{c}R\, h^{(2)}_{\|m\|}(\frac{\omega}{c}R)}\mathrm{e}^{-\mathrm{j}m\phi_{\text{ls}}}$	$R < \rho_{\text{ls}}$
Plane Wave	$D^{\text{NFCHOA}}_{\text{2.5D,pw}}(\mathbf{x}_0\|\mathbf{n}_{\text{pw}},\omega)$	$\mathring{D}^{\text{NFCHOA}}_{m,\text{pw}}(\omega)$	$\tilde{S}(\omega)\dfrac{2\,\mathrm{j}^{1-\|m\|}}{\frac{\omega}{c}R\, h^{(2)}_{\|m\|}(\frac{\omega}{c}R)}\mathrm{e}^{-\mathrm{j}m\phi_{\text{pw}}}$	$\vartheta_{\text{pw}} = \frac{\pi}{2}$

Table 3.2: Circular Harmonics coefficients $\mathring{D}_m$ of the 2.5D NFCHOA driving signal (3.17) for selected virtual sound fields.

3.3.1 Driving Signals in the Frequency Domain

Driving Signal for 3D Scenarios: NFCHOA states the solution to the specialised SLP[85]

$$S(\mathbf{x},\omega) \stackrel{!}{=} \int_0^{2\pi}\int_0^{\pi} D(\mathbf{x}_0,\omega)G(\mathbf{x}|\mathbf{x}_0,\omega)R^2\sin\vartheta_0\,\mathrm{d}\vartheta_0\,\mathrm{d}\phi_0 \quad \forall \mathbf{x}|r<R \tag{3.11}$$

for a spherical SSD of constant radius R centred at the coordinates' origin. The equivalent scattering problem is solved in the Spherical Harmonics domain and the driving signal is given as[86,87,88]

$$D^{\text{NFCHOA}}_{\text{3D}}(\mathbf{x}_0,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{\breve{S}^m_n(\omega)}{-\mathrm{j}\frac{\omega}{c}R^2 h^{(2)}_n(\frac{\omega}{c}R)}\Psi^m_n(\vartheta_0,\phi_0)\,, \tag{3.12}$$

where $\breve{S}^m_n(\omega)$ denotes the interior spherical expansion coefficients[89] of the virtual sound field. Since, the 3D NFCHOA driving signal is the exact analytic solution to the SLP in (3.11), the corresponding synthesis matches the virtual sound field everywhere inside the spherical SSD. As this was already confirmed by Ahrens[90,91] via numerical simulations, examples for the synthesised sound field are omitted, here.

Driving Signal for 2.5D Scenarios: In the 2.5D synthesis scenario, the spherical secondary source distribution reduces to a circle, here in the xy-plane. The synthesis problem specialises to

$$S(\mathbf{x},\omega) \stackrel{!}{=} \int_0^{2\pi} D(\mathbf{x}_0,\omega)G(\mathbf{x}|\mathbf{x}_0,\omega)R\,\mathrm{d}\phi_0 \quad \forall \mathbf{x}|\rho<R, z=0\,, \tag{3.13}$$

which states a circular convolution. Ahrens and Spors[92] derived the driving signal directly from the synthesis integral using the convolution theorem of the Circular Harmonics.[93] However, this derivation does only provide little insight into the connection between the 3D and the 2.5D synthesis problem. An alternative derivation based on the 3D driving signal (3.12) is presented here: expanding the fraction in (3.12) about $\Psi^{-m}_n(\frac{\pi}{2},0)$ yields

$$D^{\text{NFCHOA}}_{\text{3D}}(\mathbf{x}_0,\omega) = \frac{1}{R^2}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{\breve{S}^m_n(\omega)\Psi^{-m}_n(\frac{\pi}{2},0)}{(-\mathrm{j}\frac{\omega}{c})h^{(2)}_n(\frac{\omega}{c}R)\Psi^{-m}_n(\frac{\pi}{2},0)}\Psi^m_n(\vartheta_0,\phi_0) \tag{3.14}$$

[85] Schultz and Spors (2014). "Comparing Approaches to the Spherical and Planar Single Layer Potentials for Interior Sound Field Synthesis". In: *Acta Acustica united with Acustica* 100.5, pp. 900–911.
`driving_function_mono_nfchoa`

[86] Ahrens, op. cit., Sec. 3.4.

[87] Fazi and Nelson (2013). "Sound field reproduction as an equivalent acoustical scattering problem". In: *J. Acoust. Soc. Am.* 134.5, pp. 3721–3729, Eq. (22).

[88] Schultz and Spors, op. cit., Sec. 5.2.

[89] see Sec. 2.2.1

[90] Ahrens, "The Single-layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources", Sec. 3.2.2.

[91] Idem, *Analytic Methods of Sound Field Synthesis*, Sec. 3.3.2.

[92] Ahrens and Spors (2008). "An Analytical Approach to Sound Field Reproduction using Circular and Spherical Loudspeaker Distributions". In: *Acta Acustica united with Acustica* 94.6, pp. 988–999.

[93] Williams, op. cit., Sec. 1.2.

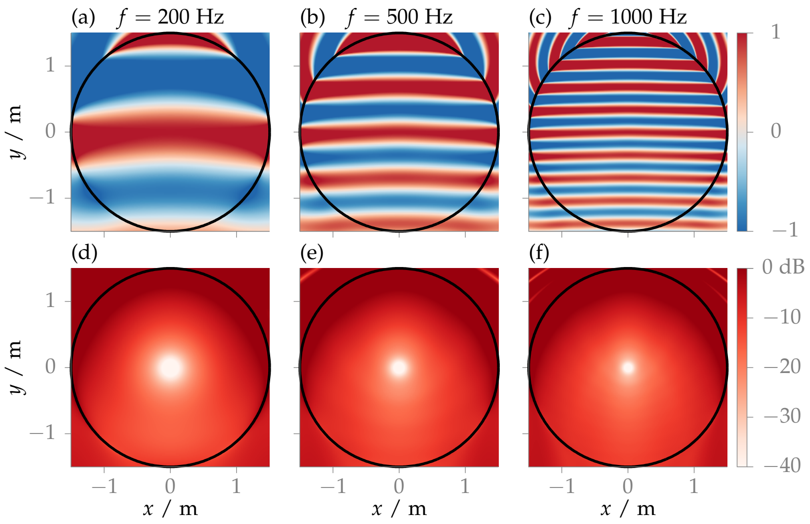


Figure 3.11: The top plots show the reproduced sound field (3.3) using a quasi-continuous, circular secondary source distribution centred around the coordinates' origin with a radius of 1.5 m (black line). It is driven by 2.5D NFCHOA (3.17) to synthesise a monochromatic plane wave $P_{\text{pw}}(\mathbf{x}, \omega)$ with the propagation direction $\mathbf{n}_{\text{pw}} = [0, -1, 0]^{\text{T}}$ and varying frequency f. The bottom plots show corresponding normalised absolute error defined as

$$20 \lg \left| \frac{P(\mathbf{x}, \omega) - P_{\text{pw}}(\mathbf{x}, \omega)}{P_{\text{pw}}(\mathbf{x}, \omega)} \right| .$$

Taking Tab. 2.1 into account, the denominator constitutes the regular spherical expansion coefficients of a point source or more precisely, the free-field Green's function with $\mathbf{x}_{\text{s}} = [R, 0, 0]^{\text{T}}$. Both, the numerator and the denominator are replaced by their sectorial approximations (2.46) yielding

$$D_{3\text{D}}^{\text{NFCHOA}}(\mathbf{x}_0, \omega) \approx \frac{1}{R^2} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{\breve{S}_{|m|}^{m}(\omega)}{(-\text{j}\frac{\omega}{c}) h_{|m|}^{(2)}(\frac{\omega}{c}R) \Psi_{|m|}^{m}(\frac{\pi}{2}, 0)} \Psi_n^{-m}\left(\frac{\pi}{2}, 0\right) \Psi_n^{m}(\vartheta_0, \phi_0) \tag{3.15}$$

The summation can be written as $\sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty}$, whereas the summation w.r.t. n

$$\sum_{n=|m|}^{\infty} \Psi_n^{-m}\left(\frac{\pi}{2}, 0\right) \Psi_n^{m}(\vartheta_0, \phi_0) = \frac{1}{2\pi} \delta(\cos(\vartheta_0)) \text{e}^{+\text{j}m\phi_0} \tag{3.16}$$

is derived in Sec. B.1. Eq. (3.15) is inserted together with (3.16) into (3.11). The integral w.r.t. ϑ_0 is solved with the sifting property[94,95] of the Dirac delta distribution. The resulting line integral coincides with the original synthesis problem in (3.13). The driving signal is given as

$$D_{2.5\text{D}}^{\text{NFCHOA}}(\mathbf{x}_0, \omega) = \sum_{m=-\infty}^{\infty} \underbrace{\frac{\breve{S}_{|m|}^{m}(\omega)}{-\text{j}2\pi \frac{\omega}{c} R h_{|m|}^{(2)}(\frac{\omega}{c}R) \Psi_{|m|}^{m}(\frac{\pi}{2}, 0)}}_{\mathring{D}_m(\omega)} \text{e}^{+\text{j}m\phi_0} . \tag{3.17}$$

The Circular Harmonics coefficients of the driving signal are denoted as $\mathring{D}_m(\omega)$. Tab. 3.2, shows the coefficients for selected virtual sound fields. The derivation reveals, that the 2.5D NFCHOA driving signal is based on the conversion of the desired sound field and the free-field Green's function into their 2D approximations, which are only exact at the coordinates' origin. The plots in Fig. 3.11 confirm that the synthesised and virtual sound field only coincide at this

window type	$\mathring{w}_m^M$ for $\lvert m \rvert \leq M$
rectangular	1
max-$\mathbf{r}_E$	$\cos\left(\frac{\pi}{2} \frac{m}{M+1}\right)$

Table 3.3: Selected modal weighting functions a.k.a. modal windows $\mathring{w}_m^M$. The windows always yield zero for $|m| > M$.

[94] Girod et al., op. cit., Eq. (8.15).

[95] Gel'fand and Shilov (1964). *Generalized functions: Vol. 1. Properties and operations*. New York, London: Academic Press, p. 4.

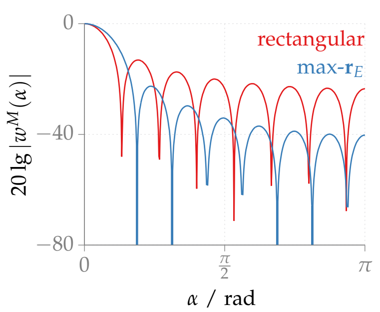


Figure 3.12: The plots show the angular spectra $w^M(\phi)$ for the rectangular and max-$\mathbf{r}_E$ window with $M = 7$. The negative abscissa is omitted due to symmetry.

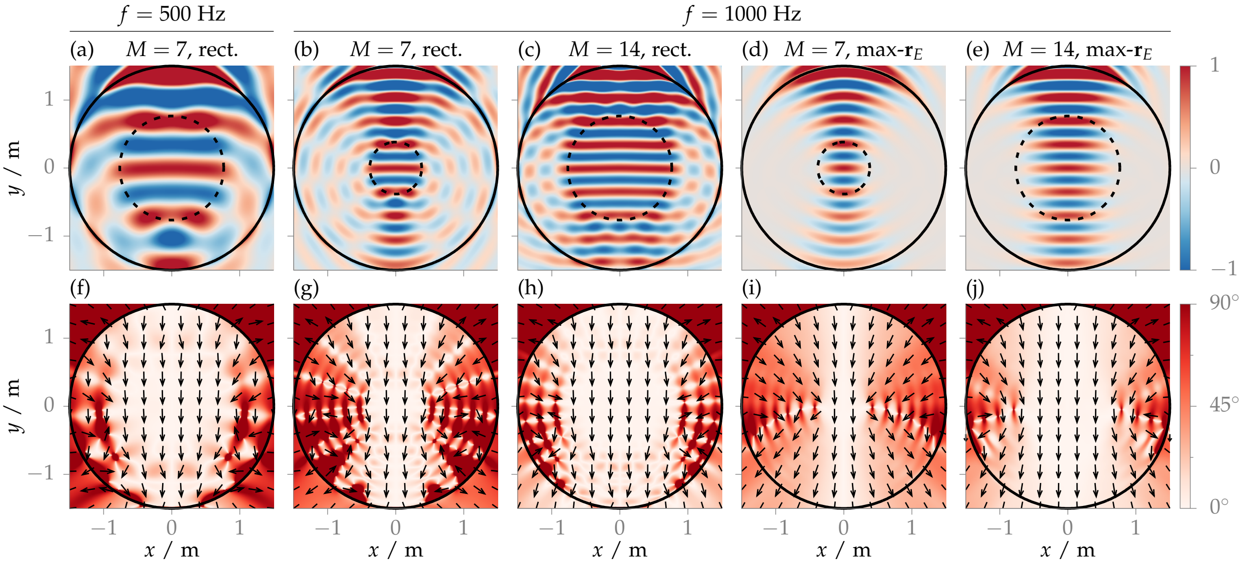


Figure 3.13: The top plots show the reproduced sound field (3.3) using a quasi-continuous, circular secondary source distribution centred around the coordinates' origin with a radius of 1.5 m (solid black line). It is driven by 2.5D bandwidth-limited NFCHOA (3.18) to synthesise a monochromatic plane wave $P_{\mathrm{pw}}(\mathbf{x}, \omega)$ with the propagation direction $\mathbf{n}_{\mathrm{pw}} = [0, -1, 0]^{\mathrm{T}}$. The dashed black circles mark the area of high synthesis accuracy with a radius of $^{Mc}/_{2\pi f}$. The temporal frequency, the modal bandwidth M, and used modal window type are indicated above the plots. The black arrows in the bottom plots show the according normalised local wavenumber vector $\hat{\mathbf{k}}_P(\mathbf{x}, \omega)$ describing the local propagation direction of the synthesised sound field. It is numerically computed using Eq. (2.53). The colour encodes the angle between the vector and the desired propagation direction $\mathbf{n}_{\mathrm{pw}}$ in degrees. The angle is defined as

$$\arccos\left(\langle \hat{\mathbf{k}}_P(\mathbf{x}, \omega) | \mathbf{n}_{\mathrm{pw}} \rangle\right) .$$

position. Here, the error between the two quantities is negligible compared to other positions. Opposite to WFS, no diffraction artefacts are visible since no discontinuous truncation by a secondary source selection criterion is applied.

Spatial Bandwidth Limitation: In order to be computationally feasible, the summation in (3.17) has to be truncated to sensible value. This is usually termed Spatial Bandwidth Limitation (SBL) and modelled by multiplying $\mathring{D}_m(\omega)$ with a finite-length weighting function a.k.a. modal window $\mathring{w}_m^M$. Assuming symmetric truncation, the window is only non-zero between $\pm M$, which usually referred to as the modal bandwidth. The truncated driving signal then reads

$$D_{2.5\mathrm{D}}^{\mathrm{NFCHOA}}(\mathbf{x}_0, \omega) = \sum_{m=-M}^{M} \mathring{w}_m^M \mathring{D}_m(\omega) \mathrm{e}^{+\mathrm{j}m\phi_0} . \tag{3.18}$$

The window types of interest are listed in Tab. 3.3: A simple truncation of the sum at $\pm M$ is equivalent to the rectangular weighting. It has been shown in the context of Ambisonics that the max-$\mathbf{r}_E$ weighting[96] has a positive effect on spatial perception.[97] To compare the different window types, the angular spectrum

$$w^M(\phi) = \sum_{m=-M}^{M} \mathring{w}_m^M \mathrm{e}^{+\mathrm{j}m\phi} \tag{3.19}$$

constituting the Inverse Circular Harmonics Transform (ICHT)[98] of $\mathring{w}_m^M$ is used. It can be deduced from Fig. 3.12, that the max-$\mathbf{r}_E$ window suppresses the side-lobes in the angular spectrum at the cost of a broader main lobe compared to the rectangular window. As a consequence, the overall emitted energy is concentrated to fewer loudspeakers in direction of the virtual source, if the max-$\mathbf{r}_E$ window is used. Formally, the max-$\mathbf{r}_E$ window maximises the length of the so-called energy-vector $\mathbf{r}_E$ which is defined as[99,100]

[96] Daniel et al. (Sept. 1998). "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions". In: *Proc. of 105th Aud. Eng. Soc. Conv.* San Francisco, USA, Eq. (19).

[97] Frank (June 2013a). "Phantom Sources using Multiple Loudspeakers in the Horizontal Plane". PhD thesis. Graz, Austria: Institute of Electronic Music and Acoustics, p. 31ff.

[98] see Eq. (2.37)

[99] Daniel et al., op. cit., space-continuous version of Eq. (14).

[100] Zotter et al. (2012). "Energy-Preserving Ambisonic Decoding". In: *Acta Acustica united with Acustica* 98.1, space-continuous version of Eq. (19).

$$|\mathbf{r}_E| = \frac{\int_0^{2\pi} |w^M(\phi)|^2 \cos(\phi)\, \mathrm{d}\phi}{\int_0^{2\pi} |w^M(\phi)|^2\, \mathrm{d}\phi}\,. \tag{3.20}$$

In Fig. 3.13, the effect of the SBL on the synthesised sound field can be observed. Due to the limitation, the sound field is only synthesised accurately in a circular area around the array centre, see dashed circle in the top plots. Its radius can be approximated by $|\mathbf{x}| \approx {}^{Mc}/_{2\pi f}$.[101] The linear dependency between radius and modal bandwidth motivates to set M as high as possible. However, in order to include the whole area surrounded by the circular loudspeaker array of radius R up to a maximum frequency f_{max}, a modal bandwidth of $M = {}^{2\pi R f_{\mathrm{max}}}/_{c}$ is necessary. As an example, $f_{\mathrm{max}} = 20$ kHz and $R = 1.5$ m would lead to $M \approx 550$ resulting in a considerable computational effort. Furthermore, it will be later shown for discrete SSDs in Ch. 4, that spatial aliasing can be reduced by decreasing M. The top plots in Fig. 3.13 further show, that the max-$\mathbf{r}_E$ window leads to a smoother synthesis of the desired sound field compared to its rectangular counterpart. Here, ripples in the wave front are observable especially near the boundary of the dashed circles. The according normalised local wavenumber vectors $\hat{\mathbf{k}}_P(\mathbf{x}, \omega)$ of the synthesised sound fields are shown in the remaining plots. Here, the max-$\mathbf{r}_E$ window leads to less fluctuations of the propagation direction.

[101] Ahrens, op. cit., Eq. (2.41).

In order to further investigate the impact of the SBL on the spectral properties of the synthesised sound field, some examples are plotted in Fig. 3.14. As already shown for the sound fields in Fig. 3.11, the approximation $f = {}^{Mc}/_{2\pi|\mathbf{x}|}$ by Ahrens[102] yields a reasonable threshold for an artefact-free synthesis. In Fig. 3.14, the dashed line indicates these frequencies for $|\mathbf{x}| = 0.5$ m. Above this frequencies, the arising artefacts heavily depend on the direction along which the evaluation position is shifted away from the centre relative to the propagation direction of the plane wave. If the shift is perpendicular, e.g. $\mathbf{x} = [0.5, 0, 0]^{\mathrm{T}}$ m (weakly saturated lines), a significant loss in sound pressure is observable. The effect is stronger for the max-$\mathbf{r}_E$ window (blue) in comparison to the rectangular window (red). Shifting the evaluation position along or antiparallel to the propagation causes only minor distortions of the spectrum. This can be recognised by the spectra for $\mathbf{x} = [0, 0.5, 0]^{\mathrm{T}}$ m (strongly saturated lines). Here, the fluctuations are more pronounced for the rectangular window. The findings agree with the work of Hahn and Spors,[103] where the error induced by the SBL of a plane wave also showed a strong dependence on the azimuth of $\mathbf{x}$ relative to its propagation direction $\mathbf{n}_{\mathrm{pw}}$.

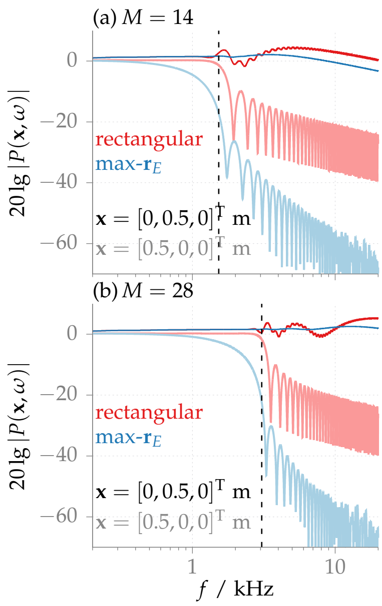


Figure 3.14: Magnitude spectra of the synthesised sound field for the same scenario as in Fig. 3.13. Two different modal bandwidth M and windows $\mathring{w}_m^M$ are used. The two evaluation positions $\mathbf{x}$ are encoded by the saturation of the line colour. The black dashed line indicates $f = {}^{Mc}/_{2\pi|\mathbf{x}|}$.

[102] Ibid., Eq. (2.41).

[103] Hahn and Spors (Mar. 2015a). "Modal Bandwidth Reduction in Data-based Binaural Synthesis including Translatory Head-movements". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 1122–1125.

3.3.2 Practical Realisation of Model-Based Rendering

`driving_function_imp_nfchoa`

As for WFS, spatial and temporal sampling has to be applied to the driving signals, which results in

$$d_{2.5\mathrm{D}}^{\mathrm{NFCHOA}}[\mathbf{x}_0^{(\nu)}, n] = \sum_{m=-M}^{M} \mathring{w}_m^M \mathring{d}_m[n] \mathrm{e}^{+\mathrm{j}m\phi_0^{(\nu)}}\,. \tag{3.21}$$

If the loudspeakers are equi-angularly distributed, i.e. $\phi_0^{(\nu)} = 2\pi\nu/N_0$, the summation states an Inverse Discrete Fourier Transform (IDFT)[104] and is efficiently implemented via an Inverse Fast Fourier Transform (IFFT), see bottom of Fig. 3.15.
Since the final driving signals and $\mathring{w}_m^M$ are real-valued, $\mathring{d}_m[n]$ has to be conjugate symmetric, i.e. the coefficients for $\pm m$ have to form conjugate complex pairs. This simplifies the computation, as only positive m have to be considered. As the remaining task, the time-discrete Circular Harmonics coefficients $\mathring{d}_m[n]$ have to be realised. While no stable realisation for the virtual line source is known to the author, the coefficients for a plane wave and a point source are implemented via IIR filters. A short outline of the major design steps is presented here. For details, the reader is referred to the original publication[105] and revisiting discussions.[106,107] With the Laplace-domain ($\mathrm{j}\omega \to s$) representation of the spherical Hankel function,[108] the coefficients in Tab. 3.2 for the point source and the plane wave can be expressed in the frequency domain as

$$\mathring{D}^{\text{NFCHOA}}_{\text{pw},m}(\omega) = \tilde{S}(\omega)\underbrace{\mathrm{e}^{-\mathrm{j}\frac{\omega}{\mathrm{c}}(r_{\text{ps}}-R)}}_{\text{delay }\tau}\underbrace{\frac{\mathrm{e}^{-\mathrm{j}m\phi_{\text{ps}}}}{2\pi r_{\text{ps}}}}_{\text{weight }g_m}\underbrace{\left[\frac{\theta_{|m|}\left(\frac{r_{\text{ps}}}{\mathrm{c}}s\right)}{\theta_{|m|}\left(\frac{R}{\mathrm{c}}s\right)}\right]_{s=\mathrm{j}\omega}}_{\mathring{H}_{m,\text{pw}}(\omega)}, \text{ and} \tag{3.22a}$$

$$\mathring{D}^{\text{NFCHOA}}_{\text{ps},m}(\omega) = \tilde{S}(\omega)\underbrace{\mathrm{e}^{+\mathrm{j}\frac{\omega}{\mathrm{c}}R}}_{\text{delay }\tau}\underbrace{2(-1)^m\mathrm{e}^{-\mathrm{j}m\phi_{\text{pw}}}}_{\text{weight }g_m}\underbrace{\left[\frac{s^{|m|}}{\theta_{|m|}\left(\frac{R}{\mathrm{c}}s\right)}\right]_{s=\mathrm{j}\omega}}_{\mathring{H}_{m,\text{ps}}(\omega)}. \tag{3.22b}$$

The expressions state, that the source signal is first delayed about τ. The delay depends on the radius of the array R and on the distance of the point source r_{ps}, if applicable. As for WFS, it is applied to the source signal via a (fractional) delay line, see top of Fig. 3.15. The computational effort for the delay interpolation is however less critical as for WFS, since the number of delay operations is independent of the number of loudspeakers.

The delayed signal is filtered by a system whose transfer function (brackets) is given in the Laplace domain as a rational function, where $\theta_{|m|}$ denotes the $|m|$th-order reverse Bessel polynomial.[109] As all coefficients of the polynomial are real-valued, the zeros and poles of the ratio are either real-valued or conjugate complex pairs. The ratio can be factorised into first- and second-order real-valued rational functions a.k.a. first- and second-order-sections. In order to achieve factorisation, the non-trivial zeros and poles have to be calculated by finding the roots of the reverse Bessel polynomials. As pointed out by Hahn and Spors,[110] root-finding-algorithms which are specialised to the reserve Bessel polynomials outperform standard methods, especially for high orders. It is worth noting, that the roots can be computed offline w.r.t. s and are then scaled on-the-fly by c/R or c/r_{ps}. In order to derive digital filters, the bilinear transform[111] is applied to

[104] Manolakis and Ingle, op. cit., Sec. 7.2.

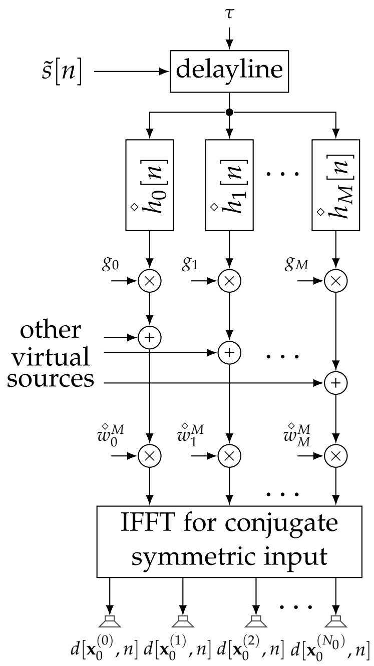


Figure 3.15: Block-Diagram showing the time-domain realisation of NF-CHOA for one virtual point source or plane wave emitting the source signal $s[n]$. Contributions from other virtual sources are incorporated via the addition-operators.

[105] Daniel, op. cit., Sec. 3.

[106] Spors et al., "Efficient realization of model-based rendering for 2.5-dimensional near-field compensated higher order Ambisonics".

[107] Hahn and Spors (May 2017). "Further Investigations on the Design of Radial Filters for the Driving Functions of Near-Field Compensated Higher-Order Ambisonics". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany.

[108] Pomberger (2008). "Angular and Radial Directivity Control for Spherical Loudspeaker Arrays". Diploma Thesis. University of Music and Dramatic Arts, Graz, Austria, Eq. (3.20).

[109] Grosswald (1978). *Bessel Polynomials*. Vol. 698. Lecture Notes in Mathematics. Berlin Heidelberg, Germany: Springer-Verlag, Cha. 2.

[110] Hahn and Spors, op. cit.

[111] Manolakis and Ingle, op. cit., Sec. 11.3.2.

the first- and second-order-sections. In the literature, alternative filter design techniques such as the matched z-transform method[112] or the Corrected Impulse Invariance Method (CIIM)[113] were investigated for the current implementation. Compared to the bilinear transform, their performance was reported to be only subtly better[114] or much worse,[115] respectively.

After filtering, the complex conjugate symmetric gain factor denoted as g_m is applied. As the contributions from other virtual sources are added afterwards, the computational effort for the modal windowing and the IFFT is independent of the number of virtual sources.

[112] Ifeachor and Jervis (2002). *Digital Signal Processing: A Practical Approach*. 2nd ed. Pearson Education, Sec. 8.7.

[113] Jackson (Oct. 2000). "A correction to impulse invariance". In: *IEEE Signal Process. Lett.* 7.10, pp. 273–275.

[114] Hahn and Spors, op. cit., Sec. 5.

[115] Spors et al., op. cit., Sec. 4.2.

3.4 Local Wave Field Synthesis using Spatial Bandwidth Limitation[116]

It was shown for NFCHOA, that the SBL establishes a region of high synthesis accuracy around the centre of a circular loudspeaker array. This motivates an extended approach, where the location of this region can be shifted away from the centre. Analytic extensions to NFCHOA published in[117] achieve this using multipole re-expansion[118] in the circular/spherical harmonics domain. A time-domain implementation of the required re-expansion coefficients using IIR filter banks was presented by Baumgartner.[119] The complexity of this approach increases drastically with the spatial bandwidth M. As its major drawback, NFCHOA restricts the array geometry to circles or spheres. As WFS is more flexible in that regard its extension called LWFS-SBL[120] is presented, here. It uses the same SBL mechanism.

[116] Parts of this section are published in Winter et al. (Aug. 2017c). "Time-Domain Realisation of Model-Based Rendering for 2.5D Local Wave Field Synthesis Using Spatial Bandwidth-Limitation". In: *Proc. of 25th European Signal Processing Conference (EUSIPCO)*. Kos Island, Greece, pp. 688–692.

[117] Winter et al., "On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources", Sec. IV.A.

[118] Gumerov and Duraiswami (2004). *Fast multipole methods for the Helmholtz equation in three dimensions*. Oxford, UK: Elsevier Science, Sec. 3.2.

[119] Baumgartner (2011). "Time Domain Fast-Multipole Translation for Ambisonics". MA thesis. Graz, Austria: Institute of Electronic Music and Acoustics.

[120] Hahn et al., op. cit.

3.4.1 Driving Signals in the Frequency Domain

`driving_function_mono_localwfs_sbl`

The SBL is applied to the interior circular expansion (2.39) of the virtual sound field around the expansion centre $\mathbf{x}_\mathrm{c}$. The resulting approximation of the sound field reads

$$S(\mathbf{x},\omega) \approx S_M^\mathrm{B}(\mathbf{x},\mathbf{x}_\mathrm{c},\omega) = \sum_{m=-M}^{M} \mathring{w}_m^M \breve{S}_m(\mathbf{x}_\mathrm{c},\omega) J_m(\mathbf{x}^\dagger,\omega)\,. \qquad (3.23)$$

The bandwidth limited sound field is denoted as $S_M^\mathrm{B}(\mathbf{x},\mathbf{x}_\mathrm{c},\omega)$. The vector $\mathbf{x}^\dagger = \mathbf{x} - \mathbf{x}_\mathrm{c}$ describes a position in a shifted coordinate frame with the expansion centre $\mathbf{x}_\mathrm{c}$ as its origin. A similar approach can be taken to apply SBL to the regular spherical expansion of the virtual sound field.[121] However, for 2.5D synthesis scenarios, the circular expansion serves as a sufficient example. While in the original publication[122] the WFS driving signal is directly computed from the truncated interior circular expansion in (3.23), an intermediate representation is chosen here: The bandwidth-limited virtual sound field is converted to a 2D PWD (2.49) with its coefficients

[121] Winter et al., op. cit.

[122] Hahn et al., op. cit.

$$\bar{S}_M^\mathrm{B}(\phi_\mathrm{pw},\mathbf{x}_\mathrm{c},\omega) = \sum_{m=-M}^{M} \mathring{w}_m^M \,\mathrm{j}^m\, \breve{S}_m(\mathbf{x}_\mathrm{c},\omega) \mathrm{e}^{+\mathrm{j}m\phi_\mathrm{pw}} \qquad (3.24)$$

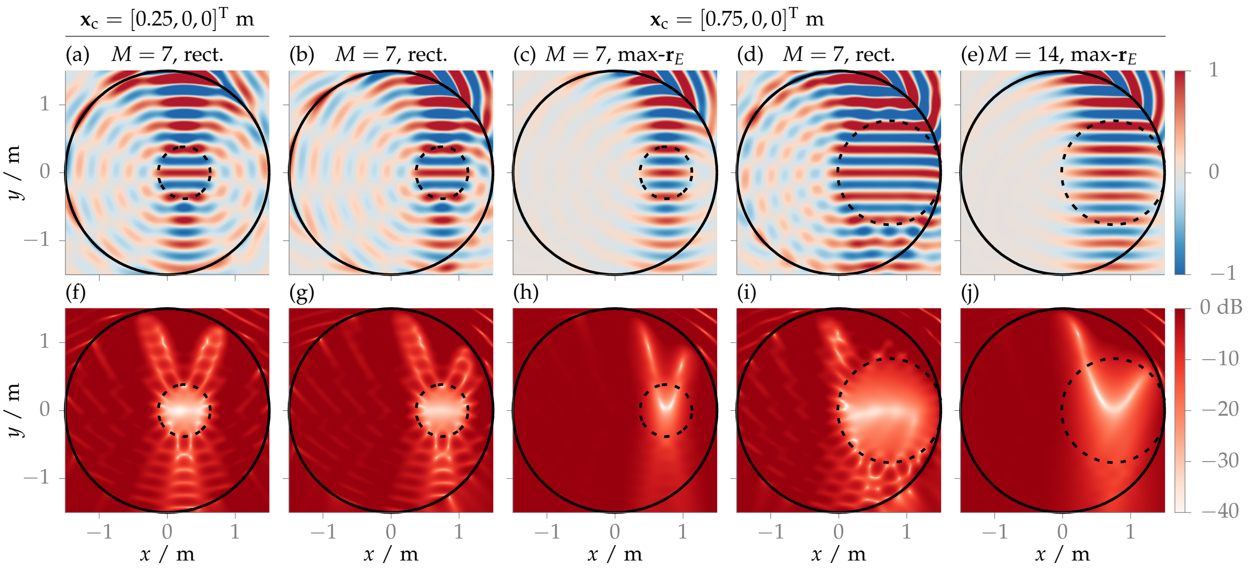


Figure 3.16: The top plots show the reproduced sound field (3.3) using a quasi-continuous (512), circular SSD centred around the coordinates' origin with a radius of 1.5 m (black line). It is driven by 2.5D LWFS-SBL (3.25) to synthesise a monochromatic ($f = 1000$ Hz) plane wave $P_{\mathrm{pw}}(\mathbf{x}, \omega)$ with the propagation direction $\mathbf{n}_{\mathrm{pw}} = [0, -1, 0]^T$. The expansion centre $\mathbf{x}_c$, the modal bandwidth M, and used modal window type are indicated above the plots. The dashed black circles mark the area of high synthesis accuracy around $\mathbf{x}_c$ with a radius of $M c/\omega$. The bottom plots show corresponding normalised absolute error defined as

$$20 \lg \left| \frac{P(\mathbf{x}, \omega) - P_{\mathrm{pw}}(\mathbf{x}, \omega)}{P_{\mathrm{pw}}(\mathbf{x}, \omega)} \right| .$$

computed from the circular expansion coefficients. Because the integral in the PWD (2.49) is a linear operation, the WFS driving signal can be applied to each individual plane wave. Finally, the LWFS-SBL driving signal is given

$$D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL}}(\mathbf{x}_0, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \bar{S}_M^{\mathrm{B}}(\phi_{\mathrm{pw}}, \mathbf{x}_c, \omega) D_{2.5\mathrm{D,pw}}^{\mathrm{WFS}}(\mathbf{x}_0^{\dagger} | \mathbf{n}_{\mathrm{pw}}, \omega) \, \mathrm{d}\phi_{\mathrm{pw}} \tag{3.25}$$

It states the superposition of conventional 2.5D WFS driving functions $D_{2.5\mathrm{D,pw}}^{\mathrm{WFS}}(\mathbf{x}_0^{\dagger} | \mathbf{n}_{\mathrm{pw}}, \omega)$ for an ensemble of plane waves with their propagation direction $\mathbf{n}_{\mathrm{pw}} = [\cos\phi_{\mathrm{pw}}, \sin\phi_{\mathrm{pw}}, 0]^T$ continuously distributed over the unit circle. The position of the secondary source $\mathbf{x}_0^{\dagger}$ is also given in the shifted coordinate frame. The driving signal for each plane wave is weighted by the according expansion coefficient $\bar{S}_M^{\mathrm{B}}(\phi_{\mathrm{pw}}, \mathbf{x}_c, \omega)$.

An example of the synthesised sound fields is shown in Fig. 3.16. The effect of the modal bandwidth M and the weighting function is similar to the effect observed for NFCHOA. A region of high accuracy with the radius $M c/2\pi f$ evolves around the expansion centre $\mathbf{x}_c$. The error plots in the bottom row confirm an increased synthesis accuracy. As a major benefit compared to NFCHOA, the location of the area can now be defined.

3.4.2 Practical Realisation of Model-Based Rendering

`driving_function_imp_localwfs_sbl`

In addition to the obligatory spatial sampling of the SSD and the temporal sampling of the involved signals, discretisation has to be applied to the PWD coefficients. The continuous expansion in (3.25) is approximated by a sum over N_{pw} equidistant samples on the unit circle, i.e. $\phi_{\mathrm{pw}}^{(l)} = 2\pi l/N_{\mathrm{pw}}$. The discrete-time driving signal and the

PWD coefficients of the bandwidth-limited sound field read

$$d^{\text{LWFS-SBL}}_{2.5\text{D}}[\mathbf{x}_0^{(\nu)}, n] \approx \frac{1}{N_{\text{pw}}} \sum_{l=0}^{N_{\text{pw}}-1} \bar{s}^{\text{B}}_M[\phi^{(l)}_{\text{pw}}, \mathbf{x}_{\text{c}}, n] *_n d^{\text{WFS}}_{2.5\text{D,pw}}[\mathbf{x}_0^{\dagger,(\nu)} | \mathbf{n}^{(l)}_{\text{pw}}, n]\,, \tag{3.26a}$$

$$\bar{s}^{\text{B}}_M[\phi^{(l)}_{\text{pw}}, \mathbf{x}_{\text{c}}, n] = \sum_{m=-M}^{M} \mathring{w}^M_m \, \mathrm{j}^m \breve{s}_m[\mathbf{x}_{\text{c}}, n] \, \mathrm{e}^{+\mathrm{j}\frac{2\pi}{N_{\text{pw}}} lm}\,. \tag{3.26b}$$

The discrete-time convolution is denoted by $*_n$. The realisation of LWFS-SBL is illustrated in Fig. 3.17. A conventional WFS renderer (bottom) presented in Sec. 3.2.2 is used to implement (3.26a). For each $\phi^{(l)}_{\text{pw}}$, a virtual plane wave with the according propagation direction $\mathbf{n}^{(l)}_{\text{pw}}$ and source signal $\bar{s}[\phi^{(l)}_{\text{pw}}, \mathbf{x}_{\text{c}}, n]$ is rendered via WFS. The required signals are defined in (3.26b) as an IDFT w.r.t. m. Since the plane wave signals are expected to be real-valued, an efficient implementation is, again, given by an IFFT for conjugate symmetric input. Hence, the remaining structure shown in Fig. 3.17 is very similar to the NFCHOA implementation (red frame). The coefficients for an interior circular expansion around $\mathbf{x}_{\text{c}}$ are given in the frequency domain as

$$\breve{S}_{\text{pw},m}(\mathbf{x}_{\text{c}}, \omega) = \tilde{S}(\omega) \underbrace{\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle \mathbf{x}_{\text{c}} | \mathbf{n}_{\text{pw}} \rangle}}_{\text{delay } \tau} \underbrace{\mathrm{j}^{-m} \, \mathrm{e}^{-\mathrm{j}m\phi_{\text{pw}}}}_{\text{weight } g_m}, \text{ and} \tag{3.27a}$$

$$\breve{S}_{\text{ps},m}(\mathbf{x}_{\text{c}}, \omega) = \tilde{S}(\omega) \underbrace{\mathrm{e}^{-\mathrm{j}\frac{\omega}{c} r^\dagger_{\text{ps}}}}_{\text{delay } \tau} \underbrace{\frac{\mathrm{j}^m \mathrm{e}^{-\mathrm{j}m\phi^\dagger_{\text{ps}}}}{4\pi r^\dagger_{\text{ps}}}}_{\text{weight } g_m} \underbrace{\left[\frac{\theta_{|m|}\left(\frac{r^\dagger_{\text{ps}}}{c} s\right)}{s^{|m|}} \right]_{s=\mathrm{j}\omega}}_{\mathring{H}_{m,\text{ps}}(\omega)}. \tag{3.27b}$$

The azimuth $\phi^\dagger_{\text{ps}}$ and the radius $r^\dagger_{\text{ps}}$ define the position $\mathbf{x}^\dagger_{\text{ps}} = \mathbf{x}_{\text{ps}} - \mathbf{x}_{\text{c}}$ of the point source in the shifted coordinate frame. The implementation for the virtual plane wave is straight-forward, as the signal is delayed and weighted. Contrary to the realisation of the radial filters in NFCHOA, an IIR implementation of $\mathring{H}_{m,\text{ps}}(\omega)$ is not possible without further treatment. Due to its $|m|$-order pole, the filter is not stable. Also an FIR implementation, e.g. via the frequency-sampling method,[123] is challenging as the pole in $\mathring{H}_{m,\text{ps}}(\omega)$ induces very high amplitudes for low frequencies. In the context of SFA using spherical microphone arrays, Lösler and Zotter[124] used Linkwitz-Riley (LR)[125] filters to stabilise the radial filters. This approach is adapted, here. LR filters are squared Butterworth filters and do only exist for even orders 2η. A LR pair consists of a lowpass $H^{\text{LP}}_{2\eta}(\omega)$ and the highpass $H^{\text{HP}}_{2\eta}(\omega)$ with the same cut-off frequency. They have a joined allpass characteristic of

$$H^{\text{AP}}_{2\eta}(\omega) = H^{\text{LP}}_{2\eta}(\omega) + H^{\text{HP}}_{2\eta}(\omega)\,. \tag{3.28}$$

with $|H^{\text{AP}}_{2\eta}(\omega)| = 1$. All three filter-types have the same phase response. In the original approach,[126] an individual LR highpass filter of order $2\eta = 2\lceil |m|/2 \rceil$ was applied to $\mathring{H}_{m,\text{ps}}(\omega)$ to compensate the pole at $s = 0$. Since different filters were used for each m, different

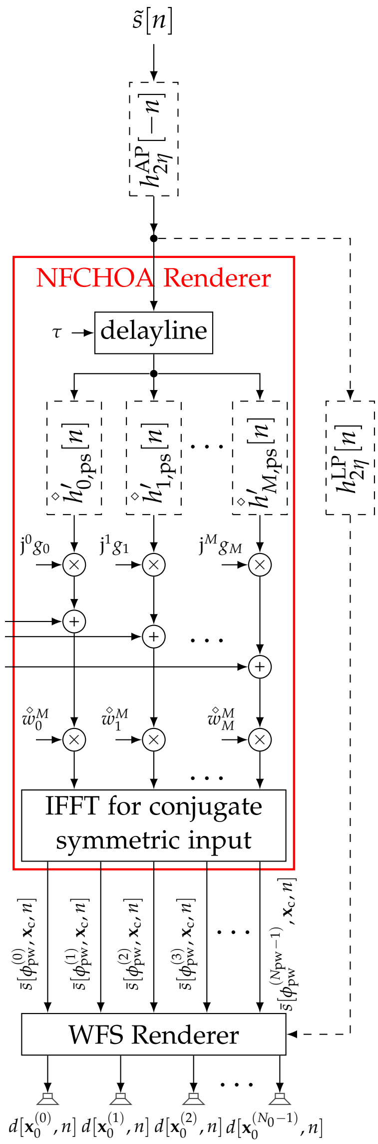


Figure 3.17: Block-Diagram showing the time-domain realisation of LWFS-SBL for one virtual source. Contributions from other virtual sources are incorporated via the addition-operators. The dashed elements are only necessary for the synthesis of a virtual point source.

[123] Manolakis and Ingle, op. cit., Sec. 10.4.

[124] Lösler and Zotter (Mar. 2015). "Comprehensive radial filter design for practical higher-order Ambisonic recording". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 452–455.

[125] Linkwitz (1976). "Active Crossover Networks for Noncoincident Drivers". In: *J. Aud. Eng. Soc.* 24.1, pp. 2–8.

[126] Lösler and Zotter, op. cit.

kind of phase distortions are implied for each mode. When recombining all modes via the IDFT these phase mismatches would lead to undesired destructive interferences. Phase match between the modes is preserved by a cascade of $(M-1)$ LR allpass filters with the same phase response as for the LR highpass filter applied to the other modes. This results in a cascade of M filters for each mode. While this is feasible for a low M (≈ 4) which is common for spherical microphone arrays, model-based SFS may require much higher M. In the current approach, computational complexity is reduced by applying the same highpass filter to all modes. The necessary order of the LR highpass filter to compensate all poles is $2\eta = 2\lceil M/2 \rceil$. The resulting stabilised radial filters read

$$\mathring{H}'_{m,\mathrm{ps}}(\omega) = \mathring{H}_{m,\mathrm{ps}}(\omega) H^{\mathrm{HP}}_{2\eta}(\omega) = \left[\frac{\theta_{|m|}\left(\frac{r'_{\mathrm{ps}}}{c} s\right) s^{2\eta-|m|}}{[B_\eta(\frac{s}{\omega_c})]^2} \right]_{s=\mathrm{j}\omega} . \tag{3.29}$$

The Butterworth polynomials[127] are denoted as $B_\eta(\cdot)$ and the crossover or cut-off frequency is defined via $\omega_c = 2\pi f_c$. As for NF-CHOA the digital IIR implementation $\mathring{h}'_{m,\mathrm{ps}}[n]$ is achieved via first- and second-order sectioning and the bilinear transform. Since the same highpass filter is applied to all modes, the entire driving function is highpass filtered. The effective driving signal in frequency domain becomes

$$D_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) = D^{\mathrm{LWFS-SBL}}_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) H^{\mathrm{HP}}_{2\eta}(\omega) . \tag{3.30}$$

This results in a lack of energy at lower frequencies for the reproduced sound field. Since, conventional WFS is expected to be accurate at low frequencies, the LWFS-SBL is augmented by the WFS driving signal within this range. The source signal is filtered by the LR lowpass filter $H^{\mathrm{LP}}_{2\eta}(\omega)$ corresponding to the highpass filter used for the stabilisation of the radial filters. The filtered signal is directly fed into the WFS renderer as the source signal for a virtual point source with the same parameters as for the LWFS-SBL driving signal. This is illustrated by the dashed signal path in Fig. 3.17. At the output of the renderer low- and high-frequency component combine to

$$D_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) = D^{\mathrm{WFS}}_{2.5\mathrm{D,ps}}(\mathbf{x}_0, \omega) H^{\mathrm{LP}}_{2\eta}(\omega) + D^{\mathrm{LWFS-SBL}}_{2.5\mathrm{D,ps}}(\mathbf{x}_0, \omega) H^{\mathrm{HP}}_{2\eta}(\omega) . \tag{3.31}$$

as the effective driving signal. ω_c may be chosen according to the aliasing frequency, i.e. the frequency up to which WFS does not introduce significant aliasing to the reproduced sound field. As an optional processing step, the allpass characteristic of the crossover pair is compensated by applying its inverse $H^{\mathrm{AP}}_{2\eta}(\omega)^{-1}$ directly to the source signal. Since the inverse of an allpass generally results in an unstable filter, the so-called backward filtering approach is utilised: The inverse of an allpass filter is equivalent to its conjugate complex, which corresponds to a time inversion of the allpass' impulse response $h^{\mathrm{AP}}_{2\eta}[n]$. Latter behaviour can also be achieved by

[127] Lipshitz and Vanderkooy (1983). "A Family of Linear-Phase Crossover Networks of High Slope Derived by Time Delay". In: *J. Aud. Eng. Soc.* 31.1/2, pp. 2–20, Tab. 1.

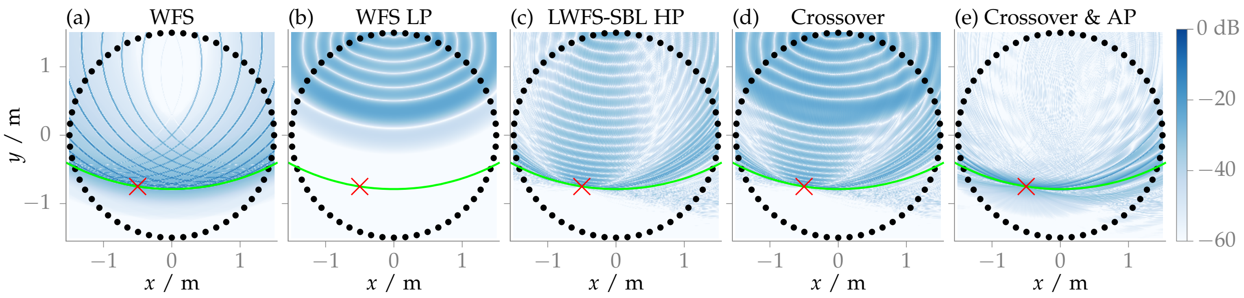


Figure 3.18: The plots show the time-snapshots of the reproduced sound fields using different (filtered) parts of the LWFS-SBL driving signal. The colour encodes the logarithmically scaled pressure magnitude. A circular SSD with a radius of 1.5 m and 56 equi-angularly positioned secondary sources was employed to synthesise a point source at $[0, 2.5, 0]^{\mathrm{T}}$ m emitting a broadband impulse. For LWFS-SBL, the expansion centre $\mathbf{x}_c = [-0.5, -0.75, 0]^{\mathrm{T}}$ and modal the bandwidth $M = 28$. The time in each snapshot is set to $t = |\mathbf{x}_{\mathrm{ps}} - \mathbf{x}_c|/c$. The green line indicates of the wave front of the virtual sound field as the ground truth. For the Linkwitz-Riley LR Filter pair, the crossover frequency $f_c = 1$ kHz and the order $2\eta = 28$.

time-inverting the input signal and time-inverting the resulting output signal, again. An online, block-based approach for backward filtering was published by Powell and Chau.[128] The effective driving signal reads

$$D_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) = D^{\mathrm{WFS}}_{2.5\mathrm{D,ps}}(\mathbf{x}_0, \omega) |H^{\mathrm{LP}}_{2\eta}(\omega)| + D^{\mathrm{LWFS-SBL}}_{2.5\mathrm{D,ps}}(\mathbf{x}_0, \omega) |H^{\mathrm{HP}}_{2\eta}(\omega)| \,. \tag{3.32}$$

The plots in Fig. 3.18 illustrate the effect of the processing steps on the synthesised sound field: In Fig. 3.18a, the conventional WFS driving signals result in the characteristic first wave front followed by the additional aliasing wave fronts. Applying the LR lowpass filter $H^{\mathrm{LP}}_{2\eta}(\omega)$ as in (3.31), suppresses the aliasing contributions in Fig. 3.18b . Due to the non-linear phase of the filter, the wave front is smeared and exhibits a visible delay compared to the ground truth (green line). The sound field for the stabilised LWFS-SBL driving signal (3.30) is plotted Fig. 3.18c, where a decent reconstruction of the first wave front at the expansion centre $\mathbf{x}_c$ (red cross) is observable. Again, subsequent contributions are caused by the phase distortions of the LR highpass $H^{\mathrm{HP}}_{2\eta}(\omega)$. As shown Fig. 3.18d, the crossover (3.31) combines the filtered driving signals, resulting in a fullband sound field with significant phase distortions. The additional allpass compensation (3.32) leads to the desired aliasing-free wave front at the expansion centre, see Fig. 3.18e.

[128] Powell and Chau (Nov. 1991). "A technique for realizing linear phase IIR filters". In: *IEEE Trans. Signal Process.* 39.11, pp. 2425–2435.

3.5 Local Wave Field Synthesis using Virtual Secondary Sources[129]

Focused point sources as introduced in Sec. 3.2.1 approximate the sound field of a point source located inside the area surrounded by the SSD. Since a point source is equivalent to the free-field Green's function, a set of focused point source may be utilised to approximately synthesise a second, virtual SSD. The basic concept of LWFS-VSS[130] is to distribute this virtual secondary sources around the target region. They are driven by WFS in order to reproduce the desired sound field within the target region. The focused sources are then synthesised by the real loudspeakers. As for LWFS-SBL, the basic principle can also be implemented via NFCHOA. However, due to its limitations w.r.t. to the geometry and the simplicity of WFS, the latter is preferred and presented here.

[129] Parts of this section are published in Winter and Spors (May 2015). "Physical Properties of Local Wave Field Synthesis using Linear Loudspeaker Arrays". In: *Proc. of 138th Aud. Eng. Soc. Conv.* Warsaw, Poland.

[130] Spors and Ahrens, "Local Sound Field Synthesis by Virtual Secondary Sources".

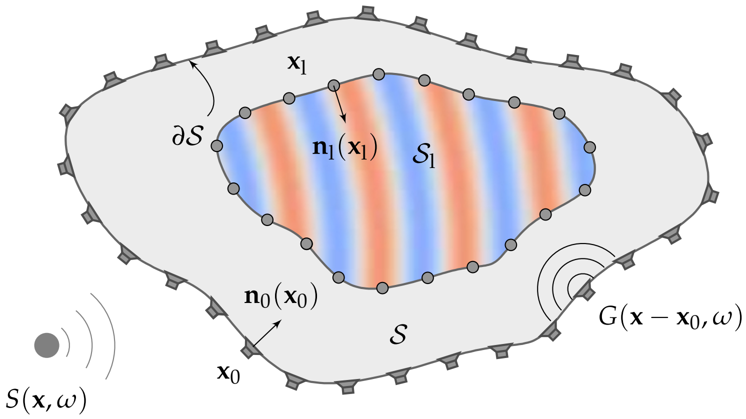


Figure 3.19: The geometry for 2.5D LWFS-VSS shows the focused sources (grey dots) distributed along the boundary of the target region $\mathcal{S}_l$ as a virtual SSD. The "real" SSD along $\partial\mathcal{S}$ is used to synthesis the focused sources.

3.5.1 Driving Signals in the Frequency Domain

driving_function_mono_localwfs_vss

The underlying geometry for the 2.5D synthesis is shown in Fig. 3.19. The concept can, however, be easily generalised towards 3D synthesis scenarios. Assuming an SSD (grey dots) on the boundary $\partial\mathcal{S}_l$ of the target region, the 2D SLP is modified to

$$S(\mathbf{x},\omega) \stackrel{!}{=} P(\mathbf{x},\omega) = \oint_{\partial\mathcal{S}_l} D(\mathbf{x}_0,\omega) G(\mathbf{x}|\mathbf{x}_l,\omega)\,\mathrm{d}l_l \quad \forall \mathbf{x} \in \mathcal{S}_l\,. \tag{3.33}$$

The free-field Green's function $G(\mathbf{x}|\mathbf{x}_l,\omega)$ states an inhomogeneity inside $\mathcal{S}$. According to the discussion on the integral formulation of the wave equation Sec. 2.1.3, an inhomogeneity cannot be synthesised correctly using the SSD along $\partial\mathcal{S}$. It is approximated via a focused source whose sound field is given as

$$G(\mathbf{x}|\mathbf{x}_l,\omega) \approx \oint_{\partial\mathcal{S}} D^{\text{WFS}}_{\text{2.5D,fs}}(\mathbf{x}_0|\mathbf{x}_l,\mathbf{n}_l,\omega) G(\mathbf{x}|\mathbf{x}_0,\omega)\,\mathrm{d}l_0\,. \tag{3.34}$$

The 2.5D WFS driving function $D^{\text{WFS}}_{\text{2.5D,fs}}(\mathbf{x}_0|\mathbf{x}_l,\mathbf{n}_l,\omega)$ for a focused source at $\mathbf{x}_l$ and oriented along $\mathbf{n}_l$ is given by (3.9b) and Tab. 3.1. Since the sound field of the focused source is divided into a diverging and converging part, it only yields a sensible approximation of the free-field Green's function in one half-space defined by $\mathbf{n}_l$. If the target region $\mathcal{S}_l$ is non-convex, there potentially exist $\mathbf{x} \in \partial\mathcal{S}_l$, which are part of the converging half-space. It is assumed, that $\mathcal{S}_l$ is convex in order to avoid this problem. After inserting (3.34) into (3.33) and rearranging the order of the integrals, the reproduced sound field is expressed by

$$P(\mathbf{x},\omega) = \oint_{\partial\mathcal{S}} \underbrace{\oint_{\partial\mathcal{S}_l} D(\mathbf{x}_l,\omega) D^{\text{WFS}}_{\text{2.5D,fs}}(\mathbf{x}_0|\mathbf{x}_l,\mathbf{n}_l,\omega)\,\mathrm{d}l_l}_{D^{\text{LWFS}-\text{VSS}}_{\text{2.5D}}(\mathbf{x}_0,\omega)} G(\mathbf{x}|\mathbf{x}_0,\omega)\,\mathrm{d}l_0 \tag{3.35}$$

with $D^{\text{LWFS}-\text{VSS}}_{\text{2.5D}}(\mathbf{x}_0,\omega)$ as the 2.5D LWFS-VSS driving signal. In order to support arbitrary convex target regions, 2.5D WFS is used to drive the virtual SSD. The resulting LWFS-VSS driving signal reads

$$D^{\text{LWFS}-\text{VSS}}_{\text{2.5D}}(\mathbf{x}_0,\omega) = \oint_{\partial\mathcal{S}_l} D^{\text{WFS}}_{\text{2.5D}}(\mathbf{x}_l,\omega) D^{\text{WFS}}_{\text{2.5D,fs}}(\mathbf{x}_0|\mathbf{x}_l,\mathbf{n}_l,\omega)\,\mathrm{d}l_l\,. \tag{3.36}$$

Fig. 3.20 shows examples for a circular virtual SSD (dashed line). Contrary to NFCHOA and LWFS-SBL, the size of the target region

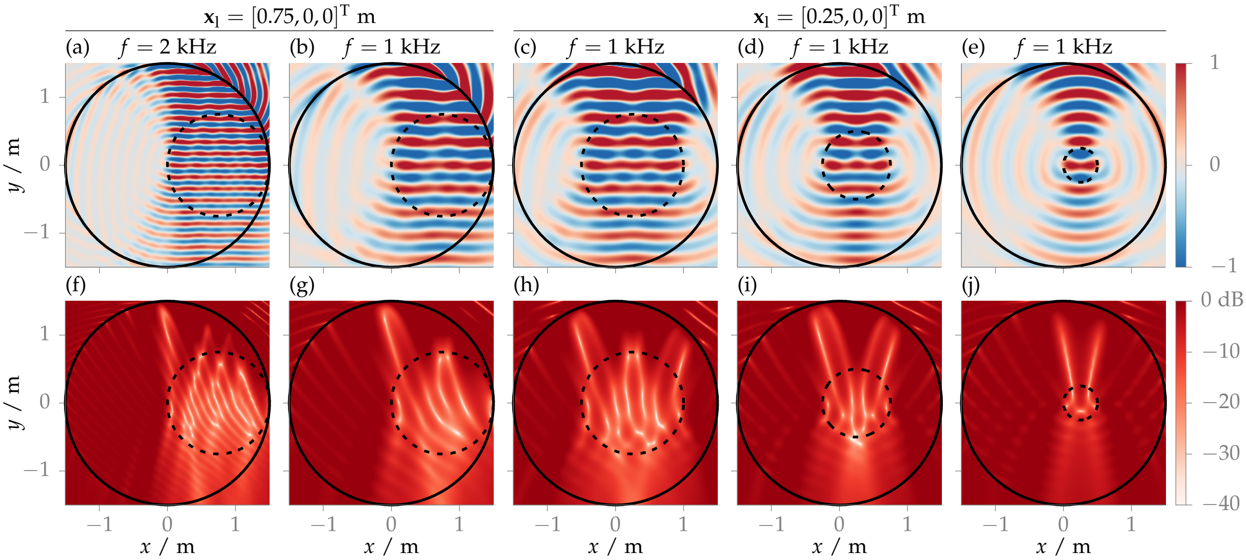


Figure 3.20: The top plots show the reproduced sound field (3.3) using a quasi-continuous, circular secondary source distribution centred around the coordinates' origin with a radius of 1.5 m (black line). It is driven by 2.5D LWFS-VSS (3.36) to synthesise a monochromatic plane wave $P_\mathrm{pw}(\mathbf{x}, \omega)$ with the propagation direction $\mathbf{n}_\mathrm{pw} = [0, -1, 0]^\mathrm{T}$. The dashed black circles marks the quasi-continuous circular virtual SSDs of different radii R_l centred around $\mathbf{x}_\mathrm{l}$. The bottom plots show corresponding normalised absolute error defined as

$$20 \lg \left| \frac{P(\mathbf{x}, \omega) - P_\mathrm{pw}(\mathbf{x}, \omega)}{P_\mathrm{pw}(\mathbf{x}, \omega)} \right| .$$

is not a function of frequency without further modifications. Although the characteristics of the planar wave front are generally reconstructed correctly (top row), ripples in the synthesised sound field are observable. This is further substantiated by the plots in the bottom row, where the synthesis error is reduced but is also heavily fluctuating inside the target region. These artefacts are partially caused by diffraction artefacts stemming from the secondary source selection criterion. In LWFS-VSS, the criterion is applied to the SSD for each focused source and to the virtual SSD. It was furthermore discussed in Sec. 3.2.1, that the focused source approximates a monopole point source most accurate for high frequencies and large distances. Fig. 3.21a confirms, that the size of the target region and, consequentially, the distance to the focused sources have a significant influence on the magnitude fluctuations. The artefacts are reduced for larger radii R_l. Adding a cosine tapering to the secondary source selection in Fig. 3.21b results in a smoothing of the magnitude spectra. As a drawback, this leads to a deviation from the desired magnitude of 0 dB and a stronger loss of power at low frequencies. The shelving of the pre-filter[131] has to be adjusted accordingly.

[131] see Sec. 3.2.2

3.5.2 Practical Realisation of Model-Based Rendering

`driving_function_imp_localwfs_vss`

Similar to the realisation of LWFS-SBL, additional discretisation has to be applied to the virtual SSD. The discrete time driving signals read

$$d_\mathrm{2.5D}^\mathrm{LWFS-VSS}[\mathbf{x}_0^{(\nu)}, n] = \sum_{m=0}^{N_\mathrm{l}-1} d_\mathrm{2.5D}^\mathrm{WFS}[\mathbf{x}_\mathrm{l}^{(l)}, n] \; d_\mathrm{2.5D,fs}^\mathrm{WFS}[\mathbf{x}_0^{(\nu)} | \mathbf{x}_\mathrm{l}^{(l)}, \mathbf{n}_\mathrm{l}^{(l)}, n] \,. \tag{3.37}$$

The position of the N_l virtual secondary sources are denoted as $\mathbf{x}_\mathrm{l}^{(l)}$. For the implementation of this formula, Spors and Ahrens[132] proposed to use two concatenated WFS renderers. A sketch is shown in Fig. 3.22. The source signal is fed into the first renderer to cal-

[132] Ibid., Sec. 2.2.

culate the driving signals $d_{2.5\mathrm{D}}^{\mathrm{WFS}}[\mathbf{x}_\mathrm{l}^{(l)}, n]$ for the virtual SSD. For each $\mathbf{x}_\mathrm{l}^{(l)}$, a focused source with the according position and orientation is rendered in the second WFS instance.
The driving signals $d_{2.5\mathrm{D}}^{\mathrm{WFS}}[\mathbf{x}_\mathrm{l}^{(l)}, n]$ are fed into in the renderer as the source signals for the focused sources. A demonstrator combining two instances of an existing WFS real-time rendering software[133] was successfully implemented by Immohr.[134]

3.6 Summary

This chapter revisited the problem of (L)SFS and presented a selection of analytic methods for its solution. In particular, the impact of various parameters, possible design choices, and practical implications have been discussed. For all four methods, spatial discretisation of the SSD potentially leads to spatial aliasing artefacts. For the two LWFS methods, the sampling of the involved PWD or the virtual SSD states another source of aliasing. This will be discussed extensively in Ch. 4.

For 3D WFS, the driving signals were derived using the Kirchhoff approximation of the equivalent scattering problem. Due to the involved secondary source criterion, the SSD is discontinuously truncated and diffraction artefacts are observable by magnitude fluctuations in the synthesised sound field. They lead to a pressure loss at low frequencies. The usage of a more smooth spatial window a.k.a. tapering leads to smaller fluctuations at the cost of a more pronounced loss in magnitude. The 2.5D driving signal is derived by applying the SPA to the 3D SLP in order to reduce its dimensionality. In addition to the diffraction artefacts, characteristic 2.5D magnitude mismatches are now present in the synthesised sound field. In the discrete-time domain, model-based rendering is realised by a geometry-independent pre-filter and a geometry-dependent delay and weighting. A sensible design of the pre-filter can be used to partly compensate the pressure loss from the diffraction and a potential pressure boost due to spatial aliasing artefacts caused by the discretised SSD. For the geometry-dependent part, the effects of delay interpolation have been discussed.

3D NFCHOA was revisited as a solution of the equivalent scattering problem for spherical SSDs. Its 2.5D counterpart for a circular SSD was derived by using the approximative Circular Harmonics representation for the virtual sound field and free-field Green's function. It provided a closer connection between the two driving signals than alternative derivations using the convolution theorem of the Circular Harmonics.[135] Similar to WFS, 2.5D magnitude mismatches occur in the synthesised sound field. For the chosen driving signals, the sound field is only exact at the centre of the circular SSD. As a practical consideration, the inverse Circular Harmonics transform of the driving signal has to be truncated. This SBL restricts the area of approximately correct synthesis to a circle around the centre, whose radius is frequency-dependent. Here, the smoother shape

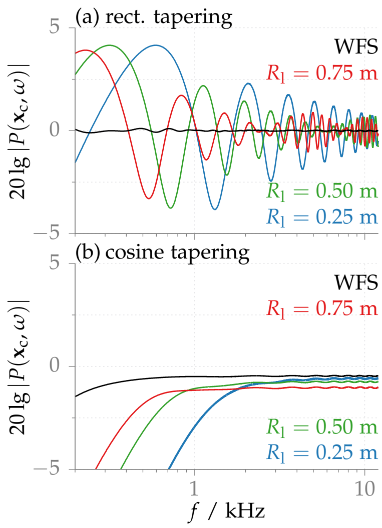


Figure 3.21: The plots show the magnitude spectra of the sound field synthesised using Local Wave Field Synthesis (LWFS) (colored) and WFS (black). The same synthesis setup as for Fig. 3.20 is used. For LWFS, $\mathbf{x}_c = [0.25, 0, 0]^\mathrm{T}$ and different radii R_l of the target region were used. The reference position $\mathbf{x}_\mathrm{ref}$ was set to $\mathbf{x}_c$. The spectra are evaluated at $\mathbf{x}_c$. In (b), an additional cosine-shaped tapering was employed. For the simulations, a fullband pre-filter was used.

[133] Geier and Spors, op. cit.

[134] Implementation was part of Immohr (Mar. 2017). "Zuhörerverfolgung für lokale Schallfeldsynthese". Bachelor's Thesis. University of Rostock, which was supervised by the author.

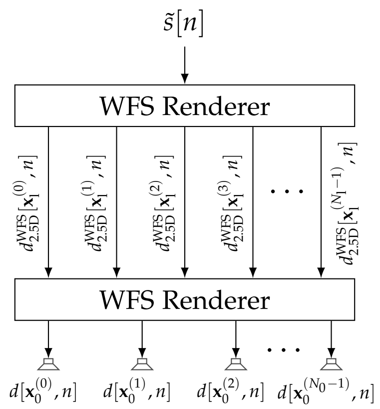


Figure 3.22: Block-Diagram showing the time-domain realisation of LWFS-VSS for one virtual source.

[135] Ahrens and Spors, op. cit.

of the truncation window, e.g. max-$\mathbf{r}_E$, leads to fluctuations in the synthesised sound field. For the discrete-time domain, an IIR filter bank and an IFFT are the essential parts of the implementation.

LWFS-SBL extends NFCHOA about an adjustable position of the prioritised synthesis area. It uses a spatially bandwidth-limited Circular Harmonics representation of the virtual sound field, converts it into a PWD, and synthesises each individual plane wave with conventional WFS. It shares similar properties regarding the SBL with NFCHOA. LWFS-SBL is implemented as a combination of a modified NFCHOA renderer and a WFS renderer. For a virtual point source, LR highpass filters are necessary to stabilise the involved IIR filters. The magnitude loss at low frequencies in the driving signals is compensated by adding a lowpass filtered driving signals of conventional WFS. The crossover frequency between the two contributions has to be sensibly chosen.

LWFS-VSS utilises focused sources as a virtual SSD surrounding the target region. The focused sources are driven by conventional WFS to synthesise the virtual sound field. They are synthesised by the real SSD using the according WFS driving signals. Thus, LWFS-VSS inherits some of the properties from WFS: The secondary source selection of the virtual SSD leads to additional diffraction artefacts. As for WFS, a smooth tapering suppresses these fluctuations with the drawback of larger magnitude losses as low frequencies. LWFS-VSS is implemented as two concatenated WFS renderers.

4 Spatial Discretisation and Aliasing

The probably most prominent implication for the practical realisation of SFS is the employment of a finite set of loudspeakers as opposed to the continuous secondary source distributions required by the synthesis integrals of the previous chapters. Selected reasons for this limitation have been presented in Sec. 1.1. The spatial discretisation of the SSD leads to spatial aliasing artefacts impairing the synthesis accuracy. Several theoretical treatises investigated spatial aliasing in SFS with a dedicated focus on the aliasing frequency. This frequency describes the largest temporal frequency up to which aliasing artefacts are negligible for a given synthesis scenario. Exceeding this frequency can be regarded as a violation of the anti-aliasing criterion. For linear and circular SSDs driven by WFS, criteria were derived for fundamental virtual sound fields such as plane and spherical waves.[1,2] A comparison to NFCHOA with respect to the aliasing properties was presented by Spors and Ahrens.[3] The found criteria are listening position independent. However, numerical simulation of the synthesised sound fields suggest a spatial heterogeneity of the aliasing frequency. Corteel et al.[4] used a time-domain model based on path-lengths to predict the position-dependent aliasing frequency for virtual point sources. It was further utilised by Oldfield[5] for his investigations on focused point sources in WFS. Within own work,[6] a model was published that predicts the occurrence of spatial aliasing for virtual plane waves synthesised by an NFCHOA approach for LSFS.[7] It was further used to predict the method's optimal parametrisation to avoid aliasing. The model was extended and applied to multizone SFS by Donley et al.[8] There, the impact of the spatial aliasing caused by the synthesis of the bright zone on the sound pressure inside the quiet zone was modelled. Analytic solutions for the aliasing frequency were derived for virtual plane waves synthesised by linear and circular SSDs. Firtha discussed spatial aliasing within his unified WFS framework.[9] He utilised the concept of the local wavenumber vector to derive anti-aliasing criteria.

This chapter introduces a geometric model to predict spatial aliasing in SFS. It heavily relies on the local wavenumber vector as a description of the local propagation direction of the virtual sound field. Other than Firtha,[10] the model does not explicitly utilise a spatial frequency domain to describe the aliasing. The approaches

This introduction was published in major parts as Winter et al. (June 2019a). "A Geometric Model for Prediction of Spatial Aliasing in 2.5D Sound Field Synthesis". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 27.6, pp. 1031–1046, Sec. I

[1] Spors and Rabenstein (May 2006). "Spatial Aliasing Artifacts Produced by Linear and Circular Loudspeaker Arrays used for Wave Field Synthesis". In: *Proc. of 120th Aud. Eng. Soc. Conv.* Paris, France.

[2] Spors and Ahrens (Oct. 2009). "Spatial Sampling Artifacts of Wave Field Synthesis for the Reproduction of Virtual Point Sources". In: *Proc. of 126th Aud. Eng. Soc. Conv.* Munich, Germany.

[3] Spors and Ahrens (Oct. 2008). "A comparison of wave field synthesis and higher-order Ambisonics with respect to physical properties and spatial sampling". In: *Proc. of 125th Aud. Eng. Soc. Conv.* San Fransisco, USA.

[4] Corteel et al. (May 2008). "Wave Field Synthesis Rendering with Increased Aliasing Frequency". In: *Proc. of 124th Aud. Eng. Soc. Conv.* Amsterdam, The Netherlands.

[5] Oldfield (Apr. 2013). "The analysis and improvement of focused source reproduction with wave field synthesis". PhD thesis. Salford, UK: University of Salford.

[6] Winter et al. (May 2016). "On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.5, pp. 914–926.

[7] Ahrens (2012). *Analytic Methods of Sound Field Synthesis.* T-Labs Series in Telecommunication Services. Berlin Heidelberg, Germany: Springer-Verlag, Sec. 4.4.5.

[8] Donley et al. (June 2018). "Multizone Soundfield Reproduction With Privacy- and Quality-Based Speech Masking Filters". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 26.6, pp. 1041–1055.

[9] Firtha (2019). "A Generalized Wave Field Synthesis Framework with Application for Moving Virtual Sources". PhD thesis. Budapest University of Technology and Economics, Sec. 4.4.

[10] Ibid., Sec. 4.4.

are however closely related, which will be shown in the upcoming sections. The model predicts the spatial occurrence of aliasing artefacts as a function of the listening position, the geometry of the SSD, and the virtual sound field. It generalises the approaches from the literature towards the mentioned dependencies. The framework is based on a high-frequency, i.e. ray-based, approximation of the underlying SFS problem.

As a baseline for discussion, a traditional approach to model spatial aliasing in SFS for circular SSDs is presented in Sec. 4.1. An anti-aliasing criterion independent of the listening position and the virtual sound field is derived, which easily generalises to other SSD geometries. In Sec. 4.2, the geometric model for WFS is developed for a linear SSD and then further generalised towards arbitrary convex SSDs. The model is applied to different synthesis scenarios including the effect of non-uniform discretisation of the SSD. Also, optimal sampling schemes w.r.t. the aliasing frequency are discussed. The model is extended towards NFCHOA in Sec. 4.3 incorporating effects of the SBL. The two LWFS approaches are covered in the subsequent sections. Here, the additional discretisation of the PWD and the virtual SSD involved in the practical realisation of the two methods is discussed.

4.1 Traditional Model[11]

[11] Parts of this section are published as Winter et al., op. cit., Sec. III-C.

In the following, a conventional approach to describe spatial aliasing in SFS is revisited. Although a circular SSD is used as an example, the theory can be easily generalised towards other geometries such as lines,[12,13] planes,[14] or spheres.[15,16] The sound field reproduced by the circular discretised SSD consisting of N_0 equiangularly spaced secondary sources reads

[12] Ahrens, op. cit., Sec. 4.6.

[13] Spors and Rabenstein, op. cit., Sec. 3.

[14] Ahrens, op. cit., Sec. 4.5.

[15] Fazi (2010). "Sound field reproduction". PhD thesis. University of Southampton, Sec. 7.1.

[16] Ahrens, op. cit., Sec. 4.3.

$$P^{\mathrm{S}}(\mathbf{x},\omega) = \sum_{\nu=0}^{N_0-1} D(\mathbf{x}_0^{(\nu)},\omega) G(\mathbf{x}-\mathbf{x}_0^{(\nu)},\omega)\Delta_\phi R\,, \tag{4.1}$$

with $\mathbf{x}_0^{(\nu)} = R[\cos(\nu\Delta_\phi), \sin(\nu\Delta_\phi), 0]^{\mathrm{T}}$ and the angular spacing of $\Delta_\phi = 2\pi/N_0$. This results in a difference in arc length between the secondary source of $\Delta_{\mathbf{x}_0} = \Delta_\phi R$. A commonly used model to describe the sampling process is the multiplication of the continuous driving signal by a Dirac impulse comb $\text{III}(\cdot)$.[17] The sampled driving signal reads

[17] Girod et al. (2001). *Signal and Systems*. Wiley, Sec. 11.3.1.

$$D^{\mathrm{S}}(\mathbf{x}_0,\omega) = D(\mathbf{x}_0,\omega)\, \underbrace{\Delta_\phi \sum_{\nu=0}^{N_0-1} \delta(\phi_0 - \nu\Delta_\phi)}_{:=\text{III}\left(\frac{\phi_0}{\Delta_\phi}\right)} \tag{4.2}$$

with $\delta(\phi_0 - \nu\Delta_\phi)$ being the Dirac delta distribution[18] imposed at $\nu\Delta_\phi$. Note, that $D^{\mathrm{S}}(\mathbf{x}_0,\omega)$ is still a continuous function, only non-zero at integer multiples of Δ_ϕ. It is treated as the original driving signal $D(\mathbf{x}_0,\omega)$ and is inserted into the SLP for circular SSD given by (3.13).

[18] Ibid., Sec. 8.3.

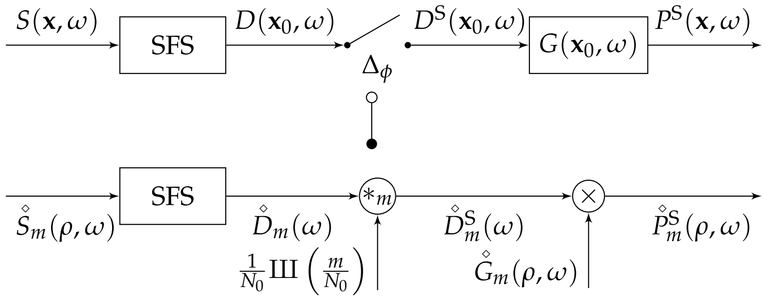


Figure 4.1: The block diagrams illustrate the sampling of the driving signal and reconstruction of the samples using the free-field Green's function. The spatial domain is shown at the top, the Circular Harmonics domain at the bottom.

The resulting sound field reproduced by the discrete SSD is given by

$$P^{\mathrm{S}}(\mathbf{x},\omega) = \int_0^{2\pi} D^{\mathrm{S}}(\mathbf{x}_0,\omega) G(\mathbf{x}-\mathbf{x}_0,\omega) R \,\mathrm{d}\phi_0\,, \tag{4.3}$$

which states the circular convolution of the sampled driving signal with the free-field Green's function. It may be interpreted as a continuous spatial filter used to reconstruct the reproduced sound field $P^{\mathrm{S}}(\mathbf{x},\omega)$ given the samples of the driving signal. The sampling-and-reconstruction process is illustrated in Fig. 4.1 (top): The continuous driving signal is computed from the virtual sound field depending on the employed SFS method. After sampling with the angular distance Δ_ϕ, the filtering with the free-field Green's function constitutes the reproduced sound field. As for any sampling-and-reconstruction process, two types of artefacts can occur: First, an insufficient sampling rate (or distance) violating the Nyquist-Theorem[19,20] introduces aliasing in the sampled quantity. Second, the lowpass characteristic of the reconstruction filter is not ideal, such that undesired parts of signal spectrum caused by the prior sampling are not sufficiently suppressed.[21] These two phenomenons are usually referred to as pre- and postaliasing, respectively.[22,23]

For the discussion of both artefacts, the spectral a.k.a. Circular Harmonics domain is more suitable. The according block diagram in Fig. 4.1 (bottom) results from the CHT of (4.3). Following the multiplication theorem of the Fourier series,[24] a multiplication with the Dirac comb in the spatial domain results in a convolution with its coefficients in the Circular Harmonics domain. They are given by a Dirac comb $\frac{1}{N_0}Ш\left(\frac{m}{N_0}\right)$, again. Using the sifting property[25,26] of the Dirac delta distribution, the Circular Harmonics coefficients of the sampled driving signal are given by[27]

$$\mathring{D}^{\mathrm{S}}_m(\omega) = \sum_{\eta=-\infty}^{\infty} \mathring{D}_{m-\eta N_0}(\omega)\,. \tag{4.4}$$

The equation constitutes the superposition of shifted versions of $\mathring{D}_m$ occuring at integer multiples of N_0. As an example, the 2.5D NF-CHOA driving signal for a virtual plane wave is considered.[28] The according coefficients $\mathring{D}_m$ plotted in Fig. 4.2a exhibit a lowpass characteristic w.r.t. m. The spatial bandwidth, i.e. the range of coefficients with a magnitude considerably greater than zero, increases with the frequency f. The range is approximated by $m \approx \pm\frac{\omega}{c}R$,[29] see yellow

[19] Ibid., Sec. 11.3.2.

[20] Manolakis and Ingle (June 2011). *Applied Digital Signal Processing: Theory and Practice*. Cambridge, USA: Cambridge University Press, Sec. 6.1.

[21] Ibid., Sec. 6.2.

[22] Fazi and Nelson (2013). "Sound field reproduction as an equivalent acoustical scattering problem". In: *J. Acoust. Soc. Am.* 134.5, pp. 3721–3729, Sec. 7.1.5.

[23] Schultz (2016). "Sound Field Synthesis for Line Source Array Applications in Large-Scale Sound Reinforcement". PhD thesis. University of Rostock.

[24] Ahrens, op. cit., Sec. D.1.

[25] Girod et al., op. cit., Eq. (8.15).

[26] Gel'fand and Shilov (1964). *Generalized functions: Vol. 1. Properties and operations*. New York, London: Academic Press, p. 4.

[27] Spors and Rabenstein, op. cit., Eq. (15).

[28] see Eq. (3.17) and Tab. 3.2

[29] Ahrens, op. cit., Eq. (2.40).

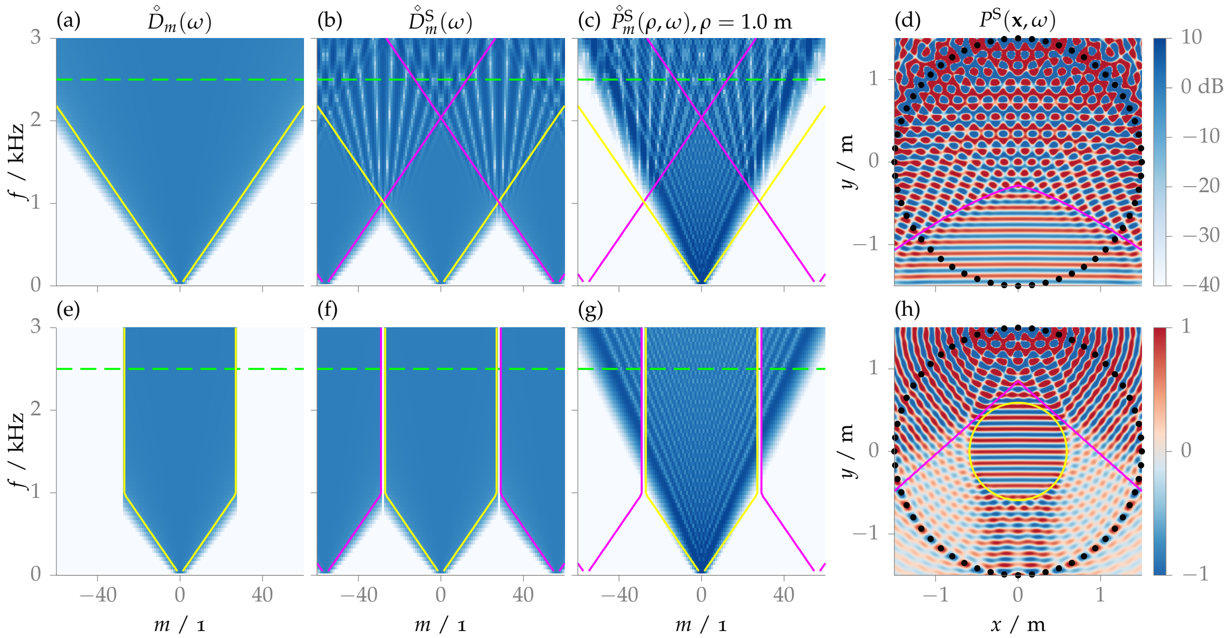


Figure 4.2: (a) shows the Circular Harmonics coefficients of the continuous 2.5D NFCHOA driving signal for a virtual plane wave with the propagation direction $\mathbf{n}_{\mathrm{pw}} = [0, -1, 0]^{\mathrm{T}}$ and $R = 1.5$ m, see Eq. (3.17). Its sampled counterpart given by (4.4) with $N_0 = 56$ is plotted in (b). The coefficients of the synthesised sound field are given in (c). All coefficients are normalised to their respective value at $m = 0, f = 1$ kHz and are logarithmically scaled. The yellow and magenta lines correspond to the approximated bandwidth of $\mathring{D}_m(\omega)$ and its spectral repetitions, respectively. The green line indicates the frequency $f = 2.5$ kHz, for which the synthesised sound field is plotted in (d). The area above the magenta line is corrupted by spatial aliasing. The remaining figures in the bottom row show the analogous quantities for the driving function weighted by a rectangular modal window $\mathring{w}_m^M$ with $M = 27$, see Eq. (3.18). The yellow circle in (h) indicates the area $|\mathbf{x}| \leq {}^{Mc}/_{2\pi f}$.

lines. The coefficients for the sampled driving signal are depicted in Fig. 4.2b. A considerable overlap of the spectral repetitions ($\eta \neq 0$, yellow) and the original spectrum ($\eta = 0$, magenta) above a given temporal frequency can be observed. Using the mentioned approximations for the spatial bandwidth of the spectra, the limit is given by

$$f^{\mathrm{S}} \approx \frac{N_0 c}{4\pi R} = \frac{c}{2\Delta_\phi R} = \frac{c}{2\Delta_{\mathbf{x}_0}} \tag{4.5}$$

which constitutes the aliasing frequency f^{S}. It is related to the well-known half-wavelength criterion:[30] The distance between two actuators, here given by arc length $\Delta_{\mathbf{x}_0} = \Delta_\phi R$, has to be smaller than half of the wavelength in order to avoid spatial aliasing. For frequencies above this limit, aliasing is present in the coefficients of the synthesised sound field, see interference patterns in Fig. 4.2c. This leads to spatial aliasing artefacts in the synthesised sound field as plotted in Fig. 4.2d. However, the artefacts are not homogeneously distributed in space. For positions below the magenta line no considerable spatial aliasing is observable. The presented criterion only detects, if and not where spatial aliasing occurs. It is, thus, unable to explain this phenomenon.

A common approach to avoid the pre-aliasing is to apply a lowpass filter a.k.a. antialiasing pre-filter to the continuous signal before the sampling.[31,32,33] In the chosen example, this is equivalent to the multiplication of the coefficients $\mathring{D}_m$ with the modal window $\mathring{w}_m^M$. This is also known as SBL and was already introduced by (3.18). The coefficients filtered by a rectangular window are shown in Fig. 4.2e. After sampling no overlap of the original spectrum and the repetitions is present, see Fig. 4.2f. The largest aliasing-free modal bandwidth is given as $M = \lfloor {}^{(N_0-1)}/_2 \rfloor$.[34] Although the overlap of the spectral repetitions is prevented, the coefficients of the synthe-

[30] Van Trees (2004). *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons, Eq. (2.130).

[31] Manolakis and Ingle, op. cit., Sec. 6.5.1.

[32] Girod et al., op. cit., p. 282.

[33] Schultz, op. cit., Sec. 3.1.

[34] Ahrens, op. cit., Sec. 4.26.

sised sound field in Fig. 4.2g still contain post-aliasing (outside the yellow lines). This is due to the imperfect lowpass characteristics of the free-field Green's function as a spatial reconstruction filter. It was discussed by various authors[35,36,37,38] that the deployment of directive secondary sources results in a stronger suppression of the spectral repetitions. As already outlined in Sec. 3.3.1, the bandwidth limitation leads to a restriction of the area of correct synthesis to a circular area of radius $Mc/2\pi f$,[39] see yellow circle in Fig. 4.2h. Due to the imperfect reconstruction undesired contributions to the sound field are still present above the magenta line. Their spatial occurrence cannot be predicted by the presented traditional model.

[35] Start (1997). "Direct Sound Enhancement by Wave Field Synthesis". PhD thesis. Delft University of Technology, Sec. 5.3 and 5.4.

[36] Ahrens, op. cit., Sec. 4.4.6 and 4.6.6.

[37] Schultz, loc. cit.

[38] Firtha, op. cit., Sec. 4.4.3.

[39] Ahrens, op. cit., Eq. (2.41).

4.2 Geometric Model for Wave Field Synthesis[40]

[40] Major parts of this section are published as Winter et al. (June 2019a). "A Geometric Model for Prediction of Spatial Aliasing in 2.5D Sound Field Synthesis". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 27.6, pp. 1031–1046, Sec. III and IV.

In the previous section, SFS with a discrete SSD was modelled as the concatenation of spatial sampling and reconstruction using the free-field Green's function. Spatial aliasing resulted from an overlap of spectral repetitions in the spatial frequency domain. The approach was not able to describe the spatial structure of the aliasing artefacts occurring in the synthesised sound field. In the following, a derivation of a geometric model for spatial aliasing in 2.5D WFS will be presented. In order to establish this, a high frequency approximation of the underlying synthesis problem, which treats sound waves as rays is reasonable. It was discussed in Sec. 3.2, that 2.5D WFS with the involved Kirchhoff approximation states the solution to the 2D SLP for asymptotically high frequencies. Hence, its 2.5D driving signal given by Eq. (3.9b) fulfils the requirements for ray based modelling. First, the derivations will be carried out for a linear SSD and then generalised to convex geometries.

4.2.1 Continuous Linear Secondary Source Distribution

In the following, a continuous linear SSD along the x-axis is assumed, see grey line in Fig. 4.3. The secondary source positions are denoted by $\mathbf{x}_0 = [x_0, 0, 0]^\mathrm{T}$ and the boundary normal vector $\mathbf{n}_0$ points into the positive y-direction. Correct synthesis is supposed to be achieved inside the positive y-half plane. The SLP given by (3.3) specialises to an 1D convolution integral

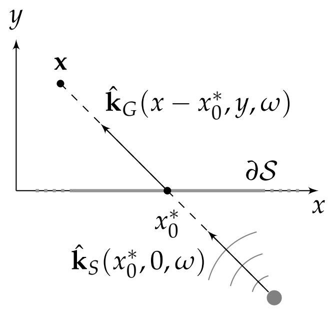


Figure 4.3: The image shows an exemplary synthesis scenario for a continuous linear Secondary Source Distribution (SSD) along the x-axis. A virtual point source (grey dot) is to be synthesised in the upper half plane ($y > 0$). The stationary phase point x_0^* is defined as the coordinate, where $\hat{\mathbf{k}}_G$ and $\hat{\mathbf{k}}_S$ point into the same direction.

$$S(x,y,\omega) \stackrel{!}{=} P(x,y,\omega) = \int_{-\infty}^{\infty} D(x_0,\omega) G(x - x_0, y, \omega) \,\mathrm{d}x_0\,, \quad \forall y > 0\,, \tag{4.6}$$

where the dependencies on the spatial variables are split, for clarity. The 3D free-field Green's function

$$G(x - x_0, y, \omega) = \underbrace{\frac{1}{4\pi\sqrt{(x-x_0)^2 + y^2}}}_{A_G(x-x_0,y,\omega)} \underbrace{\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\sqrt{(x-x_0)^2+y^2}}}_{\mathrm{e}^{+\mathrm{j}\Phi_G(x-x_0,y,\omega)}} \tag{4.7}$$

is expressed via the amplitude-phase notation which was introduced in Sec. 2.3. The 2.5D WFS driving signal (3.9b) for this geometry

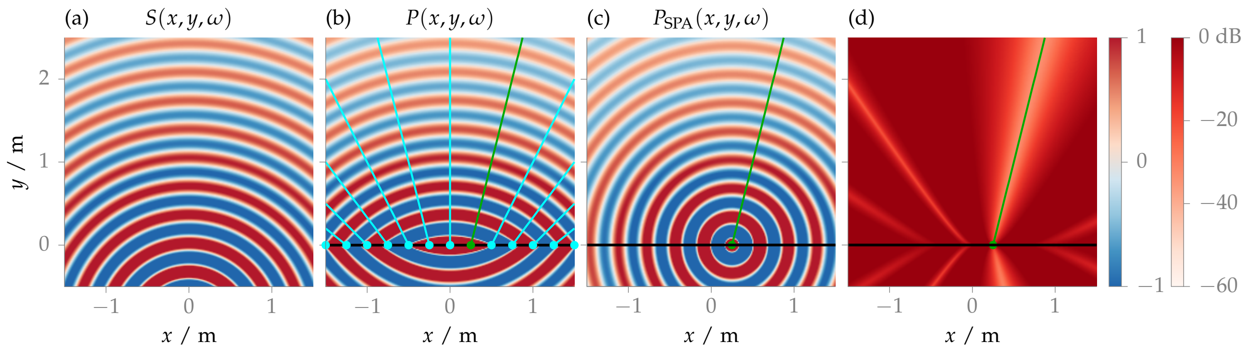


Figure 4.4: (a) shows the real part of a monochromatic ($f = 1$ kHz) point source located at $\mathbf{x}_{\mathrm{ps}} = [0, -1, 0]^{\mathrm{T}}$ m which serves as the virtual sound field. The sound field (4.9) synthesised by a quasi-continuous linear SSD along the x-axis is plotted in (b). Each coloured line indicates the positions $\mathbf{x}$ for which the secondary source at the start of the according line (circles) is the stationary secondary source $\mathbf{x}_0^*$. For the secondary source at the green circle, (c) shows the SPA of the synthesised sound field (4.10) and (d) plots the normalised error $20\lg\left|\frac{P_{\mathrm{SPA}}(x,y,\omega)-P(x,y,\omega)}{P(x,y,\omega)}\right|$.

specialises to

$$D_{2.5\mathrm{D}}^{\mathrm{WFS}}(x_0,\omega) = \sqrt{\mathrm{j}\frac{\omega}{\mathrm{c}}}\sqrt{8\pi\Delta_S(x_0)}\,\hat{k}_{S,y}(x_0,0,\omega)A_S(x_0,0,\omega)\mathrm{e}^{+\mathrm{j}\Phi_S(x_0,0,\omega)}, \tag{4.8}$$

where the phase-amplitude notation of the virtual sound field was used. The virtual sound field is evaluated at $y = 0$ as the secondary sources are distributed along the x-axis. The y component of the normalised local wavenumber vector is denoted as $\hat{k}_{S,y}$. Inserting the driving signal and the free-field Green's function into the synthesis integral (4.6) yields

$$P(x,y,\omega) = \sqrt{\mathrm{j}\frac{\omega}{\mathrm{c}}}\int_{-\infty}^{\infty}\sqrt{8\pi\Delta_S(x_0)}\,\hat{k}_{S,y}(x_0,0,\omega) \cdot A_S(x_0,\omega)\,A_G(x-x_0,y,\omega)\mathrm{e}^{+\mathrm{j}(\Phi_S(x_0,0,\omega)+\Phi_G(x-x_0,y,\omega))}\,\mathrm{d}x_0\,. \tag{4.9}$$

It is approximated by using the SPA defined in (A.1). It reads

$$P_{\mathrm{SPA}}(x,y,\omega) = 4\pi\sqrt{\mathrm{j}\frac{\omega}{\mathrm{c}}}\,S(x_0^*,0,\omega)\,G(x-x_0^*,y,\omega)\,\hat{k}_{S,y}(x_0^*,0,\omega) \cdot\sqrt{\frac{\Delta_S(x_0^*)}{|\Phi_S''(x_0^*,0,\omega)+\Phi_G''(x-x_0^*,y,\omega)|}}\cdot\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\operatorname{sgn}(\Phi_S''(x_0^*,0,\omega)+\Phi_G''(x-x_0^*,y,\omega))} \tag{4.10}$$

and states that the major part of the reproduced sound field at $\mathbf{x}$ is contributed by an individual secondary source located at $\mathbf{x}_0^* = [x_0^*, 0, 0]$. The stationary phase point x_0^* has to fulfil the condition

$$0 \stackrel{!}{=} \Phi_S'(x_0^*,y,\omega) + \Phi_G'(x-x_0^*,y,\omega)\,. \tag{4.11}$$

The terms $\Phi_\cdot'(\cdot)$ and $\Phi_\cdot''(\cdot)$ denote the first- and second-order derivative of the phase w.r.t. x_0 evaluated at the according arguments. With the definition of the normalised local wavenumber vector in (2.54), the condition is equivalent to

$$\hat{k}_{S,x}(x_0^*,0,\omega) \stackrel{!}{=} \hat{k}_{G,x}(x-x_0^*,y,\omega) \tag{4.12}$$

where $\hat{k}_{\cdot,x}$ denotes the x-component of the respective vector. For the 2.5D synthesis scenarios defined in Sec. 3.1, the z-component of the involved vectors is fixed to zero. With their unit length, the normalised local wavenumber vectors are determined by one of their remaining components (x and y) despite an unknown sign of the other component. For the virtual sound field, latter ambiguity can be resolved taking the synthesis scenario in Fig. 4.3 into account: As all virtual sources are supposed to be located in the negative-y halfspace, the local wavenumber vector has a positive y-component. The normalised local wavenumber vector of the 3D free-field Green's function is given by

$$\hat{\mathbf{k}}_G(x - x_0^*, y, \omega) = \frac{1}{\sqrt{(x - x_0^*)^2 + y^2}} \begin{bmatrix} x - x_0^* \\ y \\ 0 \end{bmatrix}, \tag{4.13}$$

which also has a positive y-component for the target region $y > 0$. Further, $\hat{\mathbf{k}}_G$ is independent of the angular frequency ω. Hence,

$$\hat{\mathbf{k}}_S(x_0^*, 0, \omega) \stackrel{!}{=} \hat{\mathbf{k}}_G(x - x_0^*, y) \quad \forall y > 0, \tag{4.14}$$

is the equivalent condition for the normalised local wavenumber vectors. This relation is illustrated in Fig. 4.3: At the stationary phase point x_0^*, the vector $\hat{\mathbf{k}}_S$ and $\hat{\mathbf{k}}_G$ are aligned. Eqs. (4.13) and (4.14) are solved for $\mathbf{x}$ yielding

$$\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} x_0^* \\ 0 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} \hat{k}_{S,x}(x_0^*, 0, \omega) \\ \hat{k}_{S,y}(x_0^*, 0, \omega) \\ 0 \end{bmatrix}, 0 \leq \gamma \leq \infty \tag{4.15}$$

which is the parametric definition of a ray starting at $\mathbf{x}_0^*$ with the direction $\hat{\mathbf{k}}_S(x_0^*, 0, \omega)$. For asymptotically high frequencies, the reproduced sound field along the given ray is mainly determined by the secondary source located at $\mathbf{x}_0^* = [x_0^*, 0, 0]^{\mathrm{T}}$. Fig. 4.4 compares the sound field of a virtual point source with the corresponding synthesis and its SPA. The rays (coloured lines in Fig. 4.4b) are perpendicular to the wave fronts of the synthesised sound field. Fig. 4.4d shows the error between synthesised sound field and its SPA for a distinct $\mathbf{x}_0^*$ (green circle): Along the corresponding ray (green line), the error is significantly lower. The derived approximation can be regarded as reasonable.

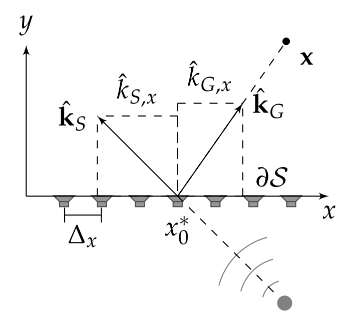


Figure 4.5: The image shows an exemplary synthesis scenario for a discrete linear Secondary Source Distribution (SSD) along the x-axis with the sampling distance Δ_x. A virtual point source (grey dot) is to be synthesised in the upper half plane ($y > 0$).

4.2.2 Uniformly Discretised Linear Secondary Source Distribution

For the uniform discretisation of a linear SSD depicted in Fig. 4.5, the reproduced sound field $P(x, y, \omega)$ is given by

$$P^{\mathrm{S}}(x, y, \omega) = \sum_{\nu=-\infty}^{\infty} D(\nu\Delta_x, \omega)\, G(x - \nu\Delta_x, y, \omega)\, \Delta_x \tag{4.16}$$

with the sampling distance denoted by Δ_x. As already discussed in Sec. 4.1, a commonly used model to describe this sampling process is the multiplication of the continuous quantity by a Dirac impulse comb $\mathrm{III}(\cdot)$.[41] The sampled driving signal reads

[41] Girod et al., op. cit., Sec. 11.3.1.

$$D^{\mathrm{S}}(x_0,\omega) = D(x_0,\omega)\,\underbrace{\Delta_x \sum_{\nu=-\infty}^{\infty} \delta(x_0 - \nu\Delta_x)}_{:=\text{III}\left(\frac{x_0}{\Delta_x}\right)} = \sum_{\eta=-\infty}^{\infty} \underbrace{D(x_0,\omega)\mathrm{e}^{-\mathrm{j}2\pi\eta\frac{x_0}{\Delta_x}}}_{:=D^{\mathrm{S}}_{\eta}(x_0,\omega)}\,. \tag{4.17}$$

The second equality follows from the Fourier series of the Dirac comb.[42] The η-th aliasing component of the discrete driving signal is denoted by $D^{\mathrm{S}}_{\eta}(x_0,\omega)$, whereas the zeroth component is the original continuous driving signal. The η-th aliasing component of the sound field $P^{\mathrm{S}}(x,y,\omega)$ synthesised by the discrete SSD is given by

[42] Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, Sec. 1.7.

$$P^{\mathrm{S}}_{\eta}(x,y,\omega) = \int_{-\infty}^{\infty} D^{\mathrm{S}}_{\eta}(x_0,\omega)G(x-x_0,y,\omega)\,\mathrm{d}x_0\,. \tag{4.18}$$

Superimposing $P^{\mathrm{S}}_{\eta}(x,y,\omega)$ for all η will result in $P^{\mathrm{S}}(x,y,\omega)$ given by (4.16). Since the aliasing components are individually accessible, they can be approximated separately via the SPA given by (A.1). The resulting approximation reads

$$\begin{aligned}P^{\mathrm{S}}_{\eta,\mathrm{SPA}}(x,y,\omega) =& 4\pi\sqrt{\mathrm{j}\frac{\omega}{c}}\,S(x_0^*,0,\omega)\,G(x-x_0^*,y,\omega)\mathrm{e}^{-\mathrm{j}2\pi\eta\frac{x_0^*}{\Delta_x}}\\ &\cdot \hat{k}_{S,y}(x_0^*,0,\omega)\sqrt{\frac{\Delta_S(x_0^*)}{|\Phi''_S(x_0^*,0,\omega)+\Phi''_G(x-x_0^*,y,\omega)|}}\\ &\cdot \mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\operatorname{sgn}(\Phi''_S(x_0^*,0,\omega)+\Phi''_G(x-x_0^*,y,\omega))}\,.\end{aligned} \tag{4.19}$$

Compared to (4.11), the condition for the stationary phase point x_0^* is extended with an additional phase term belonging to the aliasing component of the discrete driving signal. It reads

$$0 \overset{!}{=} \Phi'_S(x_0^*,y,\omega) + \Phi'_G(x-x_0^*,y,\omega) - \eta\frac{2\pi}{\Delta_x}\,. \tag{4.20}$$

The equivalent condition for the x-components of the normalised local wavenumber vectors is given by

$$\hat{k}_{S,x}(x_0^*,0,\omega) + \frac{\eta c}{\Delta_x f} \overset{!}{=} \hat{k}_{G,x}(x-x_0^*,y)\,. \tag{4.21}$$

At the stationary phase point x_0^*, the difference between the x-components of $\hat{\mathbf{k}}_G$ and $\hat{\mathbf{k}}_S$ is an integer multiple of the wavelength $\lambda = c/f$ normalised by the sampling distance Δ_x. The same approach as of Sec. 4.2.1 is taken to solve the equation for $\mathbf{x}$. The corresponding ray equation for the aliasing components reads

$$\begin{bmatrix}x\\y\\0\end{bmatrix} = \begin{bmatrix}x_0^*\\0\\0\end{bmatrix} + \gamma\begin{bmatrix}\hat{k}_{S,x}(x_0^*,0,\omega)+\frac{\eta c}{\Delta_x f}\\ \sqrt{1-\left(\hat{k}_{S,x}(x_0^*,0,\omega)+\frac{\eta c}{\Delta_x f}\right)^2}\\ 0\end{bmatrix}, 0 \le \gamma \le \infty\,. \tag{4.22}$$

It is evident from the square-root-term defining the y-component of the ray's direction vector, that the condition

$$\left|\hat{k}_{S,x}(x_0^*,0,\omega) + \frac{\eta c}{\Delta_x f}\right| \le 1 \tag{4.23}$$

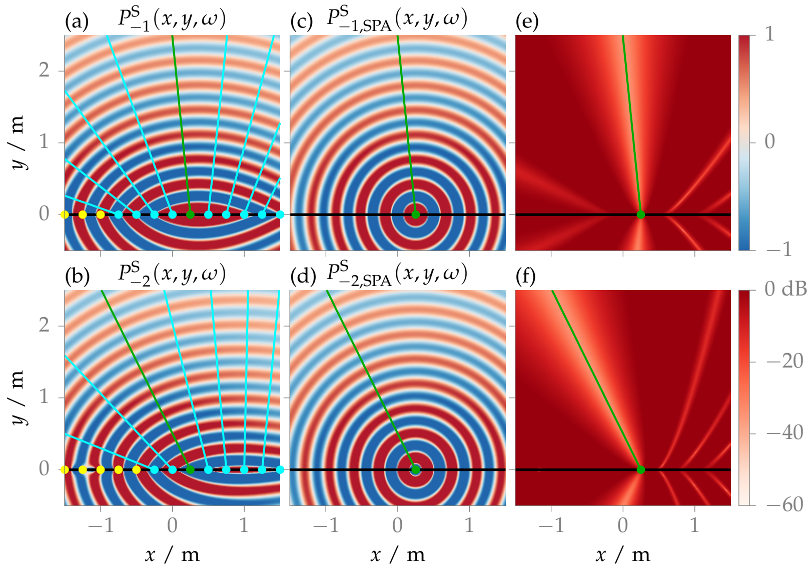


Figure 4.6: (a) and (b) show the real part of aliasing components $P^{S}_{\eta}(x,y,\omega)$ for $\eta = -1$ and -2, respectively. The same virtual sound field as in Fig. 4.4 was synthesised using using a discrete linear SSD along the x-axis with $\Delta_x = 1\,\mathrm{m}$. Each coloured line indicates the positions $\mathbf{x}$ for which the secondary source at the start of the respective line (circles) is the stationary secondary source $\mathbf{x}_0^*$. The yellow circles mark secondary sources, for which the condition in (4.23) is not fulfilled. For the secondary source at the green circle, (c) and (d) plot the according SPA of the aliasing component (4.19). (e) and (f) show the normalised error

$$20\lg\left|\frac{P^{S}_{\eta,\mathrm{SPA}}(x,y,\omega) - P^{S}_{\eta}(x,y,\omega)}{P^{S}_{\eta}(x,y,\omega)}\right| .$$

Due to the axial symmetry the plots for positive η can be generated by negating the x-coordinate.

has to be fulfilled in order to have a real-valued solution. Otherwise the η-th aliasing component is not excited by the secondary source located at x_0^*. It is a generalisation of the condition derived by Spors[43] for virtual plane waves towards arbitrary sound fields. Fig. 4.6 shows an example for the aliasing components and their corresponding ray approximations. As for the continuous SSD, the rays are locally perpendicular to the wavefront curvature of the sound fields. The normalised error between the aliasing component and its SPA is significantly lower along the given ray indicating that the SPA is a meaningful tool.

[43] Spors (Mar. 2008). "Investigation of spatial aliasing artifacts of wave field synthesis in the temporal domain". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Dresden, Germany, Eq. (3).

4.2.3 Discrete Convex Secondary Source Distribution

The presented model for the linear SSD will now be extended towards general convex boundaries, including non-uniform sampling of the SSD. The boundary $\partial\mathcal{S}$ is described as a curve $\mathbf{x}_0(u)$ depending on the parameter $u \in [u_{\min}, u_{\max}]$, see Fig. 4.7. The component-wise derivative of $\mathbf{x}_0$ w.r.t. u is denoted as $\mathbf{x}_0' = \mathbf{x}_0'(u)$. It is oriented along the unit tangent vector $\mathbf{t}_0$. The inward pointing boundary normal vector $\mathbf{n}_0$ is perpendicular to $\mathbf{x}_0'$ and $\mathbf{t}_0$. The 2D SLP in (3.3) is rewritten as the line integral

$$P(\mathbf{x},\omega) = \int_{u_{\min}}^{u_{\max}} D(\mathbf{x}_0(u),\omega) G(\mathbf{x}-\mathbf{x}_0(u),\omega) |\mathbf{x}_0'(u)| \,\mathrm{d}u. \tag{4.24}$$

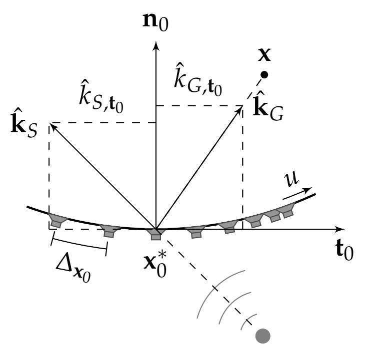


Figure 4.7: The sketch shows a synthesis scenario for a discrete convex SSD (black arc).

There exist an infinite number of parametrisations describing the same boundary. For example, $\mathbf{x}_0 = [u, 0, 0]^{\mathrm{T}}$ and $\mathbf{x}_0 = [u^3, 0, 0]^{\mathrm{T}}$ describe the same linear SSD for $u \in [-\infty, \infty]$. However, an equidistant sampling w.r.t. u would lead to different sampling schemes w.r.t. $\mathbf{x}_0$. Limiting the upcoming discussion to equidistant sampling for u is sufficient since any deterministic non-uniform scheme can be realised with a suitable parametrisation. Analogous to the linear SSD, the sampling results in aliasing components for the driving signal and synthesised sound field. For the SPA of the sound field components, the stationary phase point u^* has to fulfil

$$\eta \frac{2\pi}{\Delta_u} \stackrel{!}{=} \left. \frac{\partial \Phi_S(\mathbf{x}_0(u),\omega) + \Phi_G(\mathbf{x}-\mathbf{x}_0(u),\omega)}{\partial u} \right|_{u=u^*}, \tag{4.25}$$

with Δ_u being the sampling distance in the u-domain. The chain rule for differentiation is used together with the definition of the local wavenumber vector in (2.53) to formulate the equivalent condition

$$\langle\, \mathbf{x}_0'^* \,|\, \mathbf{k}_S(\mathbf{x}_0^*,\omega) \,\rangle + \eta \frac{2\pi}{\Delta_u} \stackrel{!}{=} \langle\, \mathbf{x}_0'^* \,|\, \mathbf{k}_G(\mathbf{x}-\mathbf{x}_0^*,\omega) \,\rangle \tag{4.26}$$

for the local wavenumber vectors. The asterisk denotes the according entities evaluated at u^*. Normalising all involved vectors while preserving equality yields

$$\frac{\omega}{c} \underbrace{\langle\, \mathbf{t}_0^* \,|\, \hat{\mathbf{k}}_S(\mathbf{x}_0^*,\omega) \,\rangle}_{:=\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0^*,\omega)} + \eta \frac{2\pi}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0^*)} \stackrel{!}{=} \frac{\omega}{c} \underbrace{\langle\, \mathbf{t}_0^* \,|\, \hat{\mathbf{k}}_G(\mathbf{x}-\mathbf{x}_0^*) \,\rangle}_{:=\hat{k}_{G,\mathbf{t}_0}(\mathbf{x}-\mathbf{x}_0^*)}, \tag{4.27}$$

where $\hat{k}_{\cdot,\mathbf{t}_0}$ denotes the tangential component of the respective vector.[44] The length of $\mathbf{x}_0'$ and the sampling distance Δ_u are combined to $\Delta_{\mathbf{x}_0}(\mathbf{x}_0) := |\mathbf{x}_0'|\Delta_u$, which can be interpreted as the local sampling distance in Cartesian space. The equation establishes a connection between the tangential components of the normalised local wavenumber vectors. Analogous to their x and y components, the tangential and normal components of the unit vectors cannot be chosen independently. Thus, (4.27) uniquely defines $\hat{\mathbf{k}}_G(\mathbf{x}-\mathbf{x}_0^*)$ for $\mathbf{x} \in \mathcal{S}$ under the assumption that $\partial\mathcal{S}$ is convex. Solving it for $\mathbf{x}$ yields the desired ray equation

[44] see Fig. 4.7

$$\mathbf{x} = \mathbf{x}_0^* + \gamma\, \mathbf{R}_0^* \begin{bmatrix} \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0^*,\omega) + \frac{\eta c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0^*,\omega))f} \\ \sqrt{1 - \left(\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0^*,\omega) + \frac{\eta c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0^*)f}\right)^2} \\ 0 \end{bmatrix}, \qquad 0 \leq \gamma \leq \infty\,. \tag{4.28}$$

The rotation matrix $\mathbf{R}_0 = [\mathbf{t}_0 \;\; \mathbf{n}_0 \;\; \mathbf{u}_z]$ contains the listed vectors as its columns. Analogous to the discrete linear SSD in Sec. 4.2.2, a real-valued solution for the rays' direction only exists, if

$$\left| \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0^*,\omega) + \frac{\eta c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0^*)f} \right| \leq 1 \tag{4.29}$$

is fulfilled. For cross-validation of the calculus, a uniformly sampled linear SSD may be chosen as a special case of the convex SSD: $\mathbf{x}_0 = [u,0,0]^{\mathrm{T}}$, $\Delta_u = \Delta_x$, $\mathbf{t}_0 = \mathbf{u}_x$, $\mathbf{n}_0 = \mathbf{u}_y$, $\mathbf{R}_0^* = \mathbf{I}$ (identity matrix), and $\hat{k}_{\cdot,\mathbf{t}_0} = \hat{k}_{\cdot,x}$. Consequently, (4.22) is a special case of (4.28).

4.2.4 Estimation of the Spatial Aliasing Frequency

In the previous sections, the connection between the listening position $\mathbf{x}$, the secondary source position $\mathbf{x}_0$, the sampling distance $\Delta_{\mathbf{x}_0}(\mathbf{x}_0)$, and the temporal frequency f was established for the occurrence of spatial aliasing. This relation will now be used to derive the highest frequency up to which no propagating spatial aliasing artefacts occur a.k.a. the spatial aliasing frequency. Exceeding this frequency can be regarded as a violation of the anti-aliasing criterion. In the following, different aliasing frequencies are derived. For practical relevance, lower bounds for the aliasing frequency covering arbitrary virtual sound fields are explicitly formulated. In the calculus, the asterisk of the stationary phase point for $\mathbf{x}_0$ is skipped for the sake of brevity.

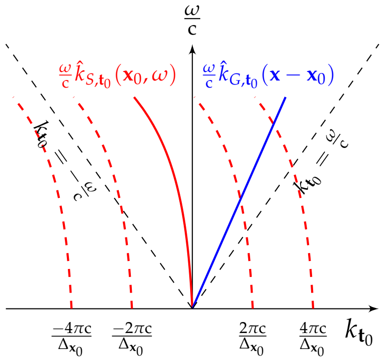


Figure 4.8: Illustration of (4.27).

Aliasing Frequency at the Position x: The quantities involved in the SPA condition of (4.27) are illustrated in Fig. 4.8: Due to the discretisation of the SSD, $\frac{\omega}{c}\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0,\omega)$ (solid red) is repeated at integer multiples of $2\pi/\Delta_{\mathbf{x}_0}(\mathbf{x}_0)$ (dashed red). Although the present approach does not explicitly make use of any spatio-spectral representations, these images can be interpreted as the spectral repetitions discussed within the traditional aliasing model. This is further substantiated by Firtha:[45] He showed, that the direction of the wavenumber vectors locally define the sound fields' spectra in the spatial Fourier domain. As for the traditional model, an intersection between the repetitions and the free-field Green's function (blue) constitutes spatial aliasing. In order to derive the spatial aliasing frequency, the SPA condition in (4.27) has to be solved for f. The involved normalised wavenumber vector $\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)$ of the virtual sound field is generally a function of f. This is illustrated by the bend red lines in Fig. 4.8, which would be straight, if $\hat{\mathbf{k}}_S$ is frequency-independent. Without further assumptions and knowledge about the virtual sound field, an analytic solution to (4.27) is not possible. Fundamental sound fields such as plane, point or line sources exhibit frequency-independent propagation directions, i.e. $\hat{\mathbf{k}}_S(\mathbf{x},\omega) = \hat{\mathbf{k}}_S(\mathbf{x})$. As discussed in Sec. 2.1.1, more complex virtual sources are described by a distribution of monopole sources weighted by the source density $Q_S(\mathbf{x},\omega)$. As long as this density is frequency-dependent also the propagation direction follows this dependency. In the far-field[46] and for high-frequencies[47] these sound fields can be approximated as a single point source weighted with an angular directivity pattern.[48,49] As these assumptions are in agreement with the geometric model, $\hat{\mathbf{k}}_S$ can be regarded as independent of f. The solution to (4.27) yields the frequency

$$f_\eta^{\mathrm{S,WFS}}(\mathbf{x},\mathbf{x}_0) = \frac{\eta c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)\big(\hat{k}_{G,\mathbf{t}_0}(\mathbf{x}-\mathbf{x}_0) - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)\big)}, \tag{4.30}$$

at which the secondary source located at $\mathbf{x}_0$ considerably contributes the η-th aliasing component P_η^{S} to a distinct position $\mathbf{x}$ inside the target region. Since the aliasing frequency defines the bound up to which no aliasing is contributed to $\mathbf{x}$, the mininum of $|f_{\eta\neq 0}^{\mathrm{S,WFS}}(\mathbf{x},\mathbf{x}_0)|$ over all aliasing components has to be considered. For this pair of listening and secondary source positions it reads

$$f^{\mathrm{S,WFS}}(\mathbf{x},\mathbf{x}_0) = \frac{c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)\big|\hat{k}_{G,\mathbf{t}_0}(\mathbf{x}-\mathbf{x}_0) - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)\big|}. \tag{4.31}$$

[45] Firtha, op. cit., Sec. 3.3.3.

[46] The term far-field refers to a large distance from the monopole distribution in comparison to its spatial extent.

[47] The term high-frequency refers to an according short wavelength in comparison to the spatial extent of the monopole distribution

[48] Meyer (1984). "Computer Simulation of Loudspeaker Directivity". In: *J. Aud. Eng. Soc.* 32.5, pp. 294–315.

[49] Ahrens, op. cit., Sec. 2.2.5.

function ALIASINGWFS($\mathcal{S}, \mathbf{x}$)
  $f^{\text{S,WFS}} \leftarrow \infty$
  for $\mathbf{x}_0, \mathbf{x}_0' \leftarrow \partial\Omega$ **do** ▷ (4.24), densely sampled
    if $a_S(\mathbf{x}_0) = 0$ **then** ▷ (4.33)
      continue ▷ next secondary sources
    end if
    $\Delta_{\mathbf{x}_0} \leftarrow \Delta_u |\mathbf{x}_0'|$
    $f \leftarrow \dfrac{c}{\Delta_{\mathbf{x}_0} |\hat{k}_{G,\mathbf{t}_0}(\mathbf{x} - \mathbf{x}_0) - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}$ ▷ (4.31)
    $f^{\text{S,WFS}} \leftarrow \min(f^{\text{S,WFS}}; f)$ ▷ (4.33)
  end for
  return $f^{\text{S,WFS}}$
end function

Figure 4.9: Brute-force search algorithm to determine the aliasing frequency $f^{\text{S,WFS}}(\mathbf{x})$ given by (4.31) and (4.33).

An infinite aliasing frequency is obtained, when $\hat{k}_{G,\mathbf{t}_0}(\mathbf{x} - \mathbf{x}_0)$ and $\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)$ coincide. This is fulfilled, if the direction of $\mathbf{x}$ relative to the secondary sources is aligned with the propagation direction of the virtual sound field $\hat{\mathbf{k}}_S(\mathbf{x}_0)$. It agrees with the work of Firtha[50] showing that aliasing-free synthesis can be achieved for exactly this case.

[50] Firtha, op. cit., Sec. 4.4.2.

As $\hat{k}_{G,\mathbf{t}_0}(\mathbf{x} - \mathbf{x}_0)$ and $\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)$ cannot exceed ± 1, the frequency is lower bounded by

$$f^{\text{S,WFS}}(\mathbf{x}, \mathbf{x}_0) \geq \frac{c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)\left(1 + |\hat{k}_{G,\mathbf{t}_0}(\mathbf{x} - \mathbf{x}_0)|\right)} \geq \frac{c}{2\Delta_{\mathbf{x}_0}(\mathbf{x}_0)}. \tag{4.32}$$

The first inequality defines the lower bound for arbitary virtual sound fields. Additionally, arbitrary positions $\mathbf{x}$ relative to $\mathbf{x}_0$ are included by the second bound. It corresponds to the half-wavelength sampling criterion $\Delta_{\mathbf{x}_0}(\mathbf{x}_0) \leq \lambda/2$,[51] which was already mentioned in conjunction with the traditional aliasing model in Sec. 4.1. The inequalities substantiate, that the geometric model generalises the traditional model towards the listening position and the virtual sound field. The predictions of both models coincide, if these dependencies are unknown or arbitrary.

[51] Van Trees, loc. cit.

The aliasing frequency $f^{\text{S,WFS}}(\mathbf{x})$ for the position $\mathbf{x}$ is defined as the frequency up to which no secondary source contributes any aliasing to $\mathbf{x}$. Hence, the minimum of $f^{\text{S,WFS}}(\mathbf{x}, \mathbf{x}_0)$ over all secondary sources defines this frequency

$$f^{\text{S,WFS}}(\mathbf{x}) = \min_{\mathbf{x}_0 | a_S(\mathbf{x}_0) \neq 0} f^{\text{S,WFS}}(\mathbf{x}, \mathbf{x}_0). \tag{4.33}$$

The minimisation is carried out over the part of the boundary where the secondary source selection criterion $a_S(\mathbf{x}_0)$ is non-zero. Contrary to its definition in (3.5), the criterion is not a function of frequency, since the propagation direction $\hat{\mathbf{k}}_S$ of the virtual sound field is assumed to be frequency-independent. Analytical solutions to the minimisation problem for elementary virtual sound fields $S(\mathbf{x}, \omega)$, e.g. point sources and plane waves, and simple geometries of the SSD are subject to further research. In order to illustrate the principle of the prediction model, it is sufficient to use a brute-force

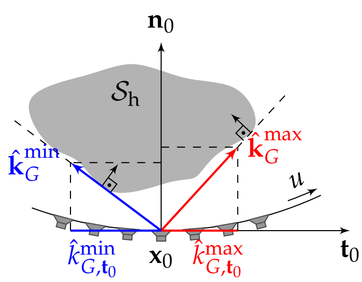


Figure 4.10: Illustration of the involved quantities for the estimation of the aliasing frequency of an extended area $\mathcal{S}_\text{h}$.

function AliasingExtendedWFS($\mathcal{S}, \mathcal{S}_\mathrm{h}$)
  $f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}} \leftarrow \infty$
  for $\mathbf{x}_0, \mathbf{x}_0' \leftarrow \partial\mathcal{S}$ **do** ▷ (4.24), densely sampled
    if $a_S(\mathbf{x}_0) = 0$ **then** ▷ (4.36)
      continue ▷ next secondary sources
    end if
    $\hat{k}^{\min}_{G,\mathbf{t}_0}, \hat{k}^{\max}_{G,\mathbf{t}_0} \leftarrow$ MinMaxWavenumber($\mathcal{S}_\mathrm{h}, \mathbf{x}_0$)
    $f \leftarrow \dfrac{c}{\Delta_{\mathbf{x}_0} \max\left(\left|\hat{k}^{\min}_{G,\mathbf{t}_0} - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)\right|; \left|\hat{k}^{\max}_{G,\mathbf{t}_0} - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)\right|\right)}$
    $f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}} \leftarrow \min(f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}; f)$ ▷ (4.36)
  end for
  return $f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}$
end function

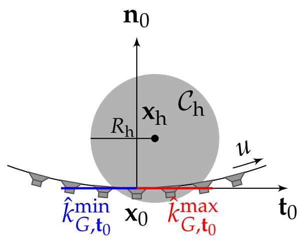
Figure 4.11: Brute-force search algorithm to determine the aliasing frequency $f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}$ for an extended listening area $\mathcal{S}_\mathrm{h}$ given by (4.34) and (4.36). An example of the function MinMaxWavenumber for a circular region is given in Fig. 4.13.

minimisation on a dense grid of $\mathbf{x}_0$. The algorithm used to predict the aliasing-frequency is given in Fig. 4.9. While it is not the most efficient approach to a specific scenario, this method is feasible for the scenarios that are investigated later in Sec. 4.2.5. Moreover, numerical approaches like brute-force or iterative optimisation algorithms might be the only alternative for scenarios with complex virtual sound fields and more sophisticated SSD geometries. For these scenarios, an analytical solution cannot be derived in closed form.

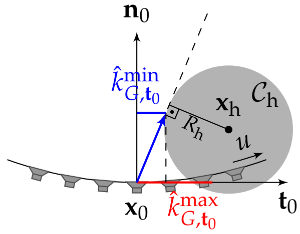
Aliasing Frequency for an Extended Listening Area: So far, the aliasing frequency for a distinct position $\mathbf{x} \in \mathcal{S}$ has been discussed. It is of further interest to find anti-aliasing conditions for the extended area $\mathcal{S}_\mathrm{h} \subseteq \mathcal{S}$. It can be utilised to model, if aliasing affects one or multiple listeners located in this area. Thus, the index h was chosen as an abbreviation for the listener's head. Fig. 4.10 shows an exemplary geometry. As a starting point, the aliasing frequency $f^{\mathrm{S,WFS}}(\mathbf{x}, \mathbf{x}_0)$ for a distinct pair of $\mathbf{x}$ and $\mathbf{x}_0$ is considered, see (4.31). The minimum over all listening positions $\mathbf{x}$ inside $\mathcal{S}_\mathrm{h}$ yields the aliasing frequency $f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}(\mathbf{x}_0)$ for a secondary source not radiating any aliasing components into $\mathcal{S}_\mathrm{h}$. For a convex boundary $\partial\mathcal{S}$, the angle between the normalised wavenumber vector $\hat{\mathbf{k}}_G(\mathbf{x} - \mathbf{x}_0)$ and the tangent vector $\mathbf{t}_0$ is in the range $[0, \pi]$. Hence, the tangential component $\hat{k}_{G,\mathbf{t}_0}(\mathbf{x} - \mathbf{x}_0)$ as the cosine of this angle is a monotonically decreasing function. Searching for the minimum w.r.t. $\mathbf{x}$, only the extremal values $\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ have to be considered, see Fig. 4.10. The aliasing frequency is given by

$$f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}(\mathbf{x}_0) = \min_{\mathbf{x} \in \mathcal{S}_\mathrm{h}} f^{\mathrm{S,WFS}}(\mathbf{x}, \mathbf{x}_0) = \frac{c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)} \cdot \min\left(\frac{1}{\left|\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0) - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)\right|}; \frac{1}{\left|\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0) - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)\right|}\right). \tag{4.34}$$

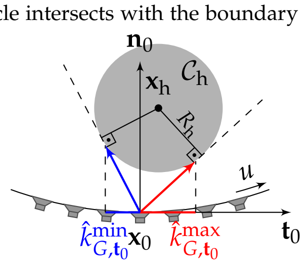
It can be seen that for fixed shape and size of $\mathcal{S}_\mathrm{h}$, the angular distance between $\hat{\mathbf{k}}^{\min}_G(\mathbf{x}_0)$ and $\hat{\mathbf{k}}^{\max}_G(\mathbf{x}_0)$ decreases the further $\mathcal{S}_\mathrm{h}$ is moved



(a) $\mathbf{x}_0$ is part of the circle.



(b) $\mathbf{x}_0$ is not part of the circle, but the circle intersects with the boundary $\partial\mathcal{S}$.



(c) The circle is completely inside $\mathcal{S}$.

Figure 4.12: The three sketches illustrate the three different cases that have to be considered for the computation of $\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ for a circular region $\mathcal{C}_\mathrm{h}$.

function MinMaxWavenumberCircle($\mathcal{C}_\mathrm{h}, \mathbf{x}_0$)
  $\mathbf{x}_\mathrm{h}, R_\mathrm{h} \leftarrow \mathcal{C}_\mathrm{h}$ ▷ centre and radius of circle
  $\varrho_\mathrm{h} \leftarrow {}^{R_\mathrm{h}}/_{|\mathbf{x}_\mathrm{h}-\mathbf{x}_0|}$ ▷ (B.7)
  $\hat{k}_{\mathrm{h},\mathbf{t}_0} \leftarrow \hat{k}_{G,\mathbf{t}_0}(\mathbf{x}_\mathrm{h} - \mathbf{x}_0)$ ▷ (B.10)
  if $\varrho_\mathrm{h} > 1$ **or** $-\sqrt{1-\varrho_\mathrm{h}^2} > \hat{k}_{\mathrm{h},\mathbf{t}_0}$ **then** ▷ (B.12)
    $\hat{k}^{\mathrm{min}}_{G,\mathbf{t}_0} \leftarrow -1$
  else
    $\hat{k}^{\mathrm{min}}_{G,\mathbf{t}_0} \leftarrow \hat{k}_{\mathrm{h},\mathbf{t}_0}\sqrt{1-\varrho_\mathrm{h}^2} - \varrho_\mathrm{h}\sqrt{1-\hat{k}^2_{\mathrm{h},\mathbf{t}_0}}$ ▷ (B.11)
  end if
  if $\varrho_\mathrm{h} > 1$ **or** $+\sqrt{1-\varrho_\mathrm{h}^2} > \hat{k}_{\mathrm{h},\mathbf{t}_0}$ **then** ▷ (B.12)
    $\hat{k}^{\mathrm{max}}_{G,\mathbf{t}_0} \leftarrow +1$
  else
    $\hat{k}^{\mathrm{max}}_{G,\mathbf{t}_0} \leftarrow \hat{k}_{\mathrm{h},\mathbf{t}_0}\sqrt{1-\varrho_\mathrm{h}^2} + \varrho_\mathrm{h}\sqrt{1-\hat{k}^2_{\mathrm{h},\mathbf{t}_0}}$ ▷ (B.11)
  end if
  return $\hat{k}^{\mathrm{min}}_{G,\mathbf{t}_0}, \hat{k}^{\mathrm{max}}_{G,\mathbf{t}_0}$
end function

Figure 4.13: Algorithm to determine the minimum and maximum tangential component of the local wavenumber vector for a secondary source position $\mathbf{x}_0$ and a circular region $\mathcal{C}_\mathrm{h}$ with radius R_h and centre $\mathbf{x}_\mathrm{h}$.

away from $\mathbf{x}_0$. For the limiting case, $k^{\mathrm{min}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $k^{\mathrm{max}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ finally coincide. Analogous to (4.32), the lower bound of $f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}(\mathbf{x}_0)$ for arbitrary virtual sound fields

$$f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}(\mathbf{x}_0) \geq \frac{c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)\left(1+\max\left(|\hat{k}^{\mathrm{max}}_{G,\mathbf{t}_0}(\mathbf{x}_0)|; |\hat{k}^{\mathrm{min}}_{G,\mathbf{t}_0}(\mathbf{x}_0)|\right)\right)} \tag{4.35}$$

is found by inserting the extreme values for $\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)$ into (4.34). A further generalisation towards arbitrary listening areas $\mathcal{S}_\mathrm{h}$ yields a lower bound corresponding to the half-wavelength sampling criterion, again.

The aliasing frequency for $\mathcal{S}_\mathrm{h}$ as the minimum over all active secondary sources reads

$$f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}} = \min_{\mathbf{x}_0 | a_S(\mathbf{x}_0) \neq 0} f^{\mathrm{S,WFS}}_{\mathcal{S}_\mathrm{h}}(\mathbf{x}_0)\,. \tag{4.36}$$

The algorithm for this aliasing frequency is shown in Fig. 4.11. Compared to the baseline algorithm in Fig. 4.9, it is augmented by the function MinMaxWavenumber($\mathcal{S}_\mathrm{h}, \mathbf{x}_0$). It determines $k^{\mathrm{min}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $k^{\mathrm{max}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ for a given secondary source position $\mathbf{x}_0$ and listening area $\mathcal{S}_\mathrm{h}$. For arbitrary shapes of $\mathcal{S}_\mathrm{h}$, this determination is challenging as it requires to find the locations on $\partial\mathcal{S}_\mathrm{h}$ whose tangents are intersecting with $\partial\mathcal{S}$ at $\mathbf{x}_0$.

A circular listening area $\mathcal{C}_\mathrm{h}$—for practical relevance—simplifies the following discussion. Moreover, it is often regarded as an approximation of the listener's head. As shown in Fig. 4.12, three different cases have to be considered for the circular area centred at $\mathbf{x}_\mathrm{h} \in \mathcal{S}$ with radius of R_h: In Fig. 4.12a, the distance between $\mathbf{x}_\mathrm{h}$ and $\mathbf{x}_0$ is smaller than the radius R_h. The secondary source is located inside the circle. No further restriction is applied to the tangential components and they take the according extremal values of ± 1. A similar

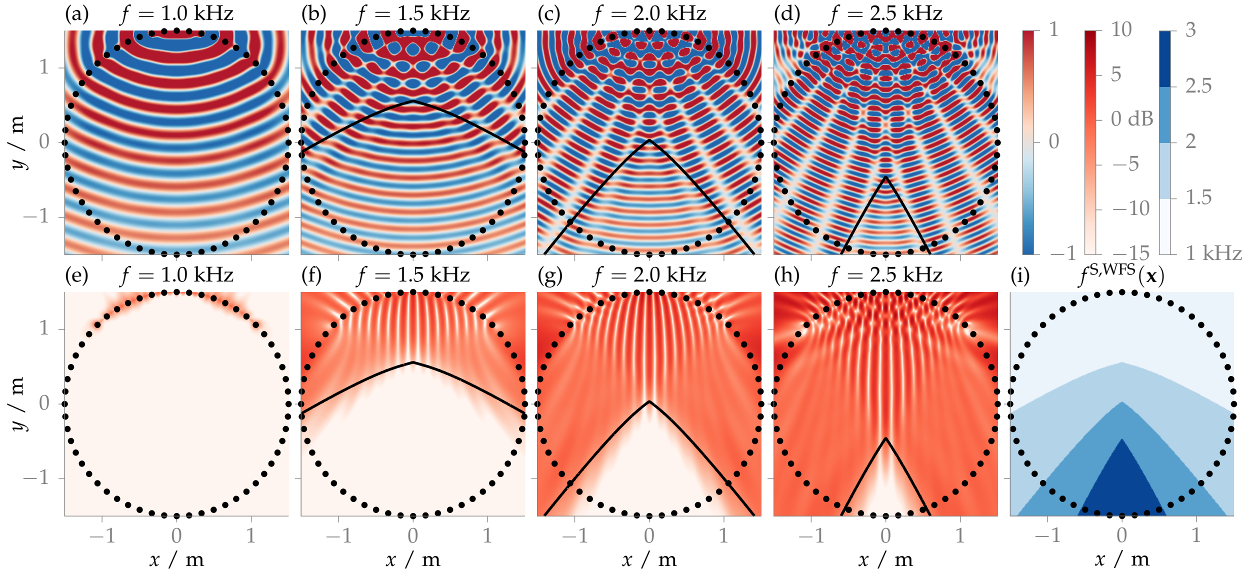


Figure 4.14: The plots (a)-(d) show the real part of a virtual point source located at $\mathbf{x}_{ps} = [0, 2.5, 0]^T$ m synthesised by WFS for different frequencies. The according error $\hat{\varepsilon}(\mathbf{x}, \omega)$ caused by aliasing is plotted in (e)-(h), see (4.37). For the positions above the solid black lines, the predicted anti-aliasing criterion involving $f^{S,WFS}(\mathbf{x})$ is violated. (i) shows the aliasing frequency $f^{S,WFS}(\mathbf{x})$ estimated by the algorithm in Fig. 4.9. A discrete colormap is used for better visibility.

scenario is shown in Fig. 4.12b, where R_h and/or the angle between $\mathbf{x}_h - \mathbf{x}_0$ and the normal vector $\mathbf{n}_0$ are large enough for the circle to be partly outside $\mathcal{S}$. Depending on the halfspace (w.r.t. $\mathbf{n}_0$) in which $\mathbf{x}_h$ is located, either the $k^{min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ or the $k^{max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ component reach its extremal value. The last alternative depicted in Fig. 4.12c covers the case, where the circular area is completely inside $\mathcal{S}$. The derivation of $k^{min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ or $k^{max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ for the three cases is given in Sec. B.2. The resulting algorithm to determine $k^{min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $k^{max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ is given in Fig. 4.13.

4.2.5 Application and Validation

To further study the performance of the model, the predicted aliasing frequency will be compared to numerical simulations of the synthesised sound fields as well as to results of other theoretical treatises in the literature. Hereby, a circular SSD of radius R is assumed. It is chosen to later allow the upcoming comparisons with 2.5D NFCHOA, which is restricted to the circular geometry. All simulations use $R = 1.5$ m, which corresponds to an existing loudspeaker setup at Technische Universität (TU) Berlin, Germany. The setup will also be used for perceptual evaluations presented in Ch. 5 and Ch. 6 in order to allow comparisons with the results of Wierstorf.[52] The secondary source positions are given by $\mathbf{x}_0 = R[\cos\phi_0, \sin\phi_0, 0]^T$. The tangent and normal vector read $\mathbf{t}_0 = [-\sin\phi_0, \cos\phi_0, 0]^T$ and $\mathbf{n}_0 = -[\cos\phi_0, \sin\phi_0, 0]^T$, respectively. The sound field $P(\mathbf{x}, \omega)$ synthesised by the continuous, circular SSD is given by the specialised SLP in (3.13). Its counterpart for the discrete SSD is denoted as $P^S(\mathbf{x}, \omega)$. The normalised error between the synthesised sound fields

$$\hat{\varepsilon}(\mathbf{x}, \omega) = 20 \log_{10} \left| \frac{P^S(\mathbf{x}, \omega) - P(\mathbf{x}, \omega)}{P(\mathbf{x}, \omega)} \right| \tag{4.37}$$

measures the influence of the spatial sampling on the synthesis accuracy. Note that the measure takes the synthesised sound field $P(\mathbf{x}, \omega)$

[52] Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Sec. 5.1 and 5.2.

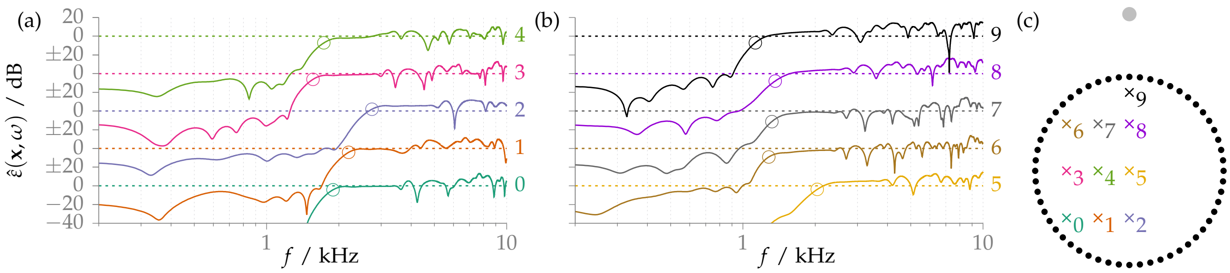


Figure 4.15: In (a) and (b), the solid lines show the error $\hat{\varepsilon}(\mathbf{x},\omega)$ defined in (4.37) for the same synthesis scenario as in Fig. 4.14. The lines have been shifted to enhance visibility. The according 0 dB reference is indicated by horizontal dashed lines of the same colour. The circles mark the estimated aliasing frequencies $f^{\mathrm{S,WFS}}(\mathbf{x})$ given by (4.33). Plot (c) depicts the evaluated positions $\mathbf{x}$ with their corresponding index.

of the continuous SSD instead of the virtual sound field $S(\mathbf{x},\omega)$ as the reference. This intentionally excludes other synthesis artefacts such as diffraction and 2.5D amplitude errors from the evaluation. These were discussed in Ch. 3. If the error decreases considerably below the predicted aliasing frequency, the result of the model can be regarded as reasonable.

Uniformly Discretised SSD: A uniform sampling of the circular SSD yields the sampling distance $\Delta_{\mathbf{x}_0} = 2\pi R/N_0$. For the simulations, $N_0 = 56$ corresponding to mentioned setup at TU Berlin was chosen. The according half-wavelength sampling criterion is then met at approximately 1 kHz for this parametrisation. The sound field reproduced by the discrete SSD reads

$$P^{\mathrm{S}}(\mathbf{x},\omega) = \frac{2\pi R}{N_0}\sum_{\nu=0}^{N_0-1} D(\mathbf{x}_0^{(\nu)},\omega)G(\mathbf{x}-\mathbf{x}_0^{(\nu)},\omega)\,, \tag{4.38}$$

with $\mathbf{x}_0^{(\nu)} = R[\cos(\nu\frac{2\pi}{N_0}),\sin(\nu\frac{2\pi}{N_0}),0]^{\mathrm{T}}$. For a virtual point source, the 2.5D WFS driving function is given by (3.9b) together with Tab. 3.1. The synthesised sound field, the sampling error, and the predicted aliasing frequency are shown in Fig. 4.14. The spatial structure of aliasing with stronger artefacts at positions closer to the virtual point source is in agreement with the plotted sound fields. A significant drop of the sampling error is observable near the predicted boundary between the aliasing-corrupted and aliasing-free region. As the presented ray model is an approximation of the underlying SFS problem, the strict separation between aliasing-free and aliased regions does not reflect the nature of the artefacts gradually reducing with increasing distance to the SSD. In Fig. 4.15, the error is plotted over frequency for ten different positions. Here, the drastic decrease of spatial aliasing artefacts near the predicted frequency (circular markers) becomes even clearer.

The driving signal for a focused point source is given by (3.9b) together with Tab. 3.1. The according synthesised sound field, the sampling error, and the predicted aliasing frequency are shown in Fig. 4.16. An aliasing-free region around the focus point evolves, which narrows with increasing frequency. This phenomenon is correctly predicted by the geometrical model (black lines). For an infinite linear SSD with sampling distance Δ_x, Wierstorf[53] empirically found a formula for the radius of an aliasing-free circular region around the focus point. It is a function of the minimum distance

[53] Ibid., Eq. (3.4).

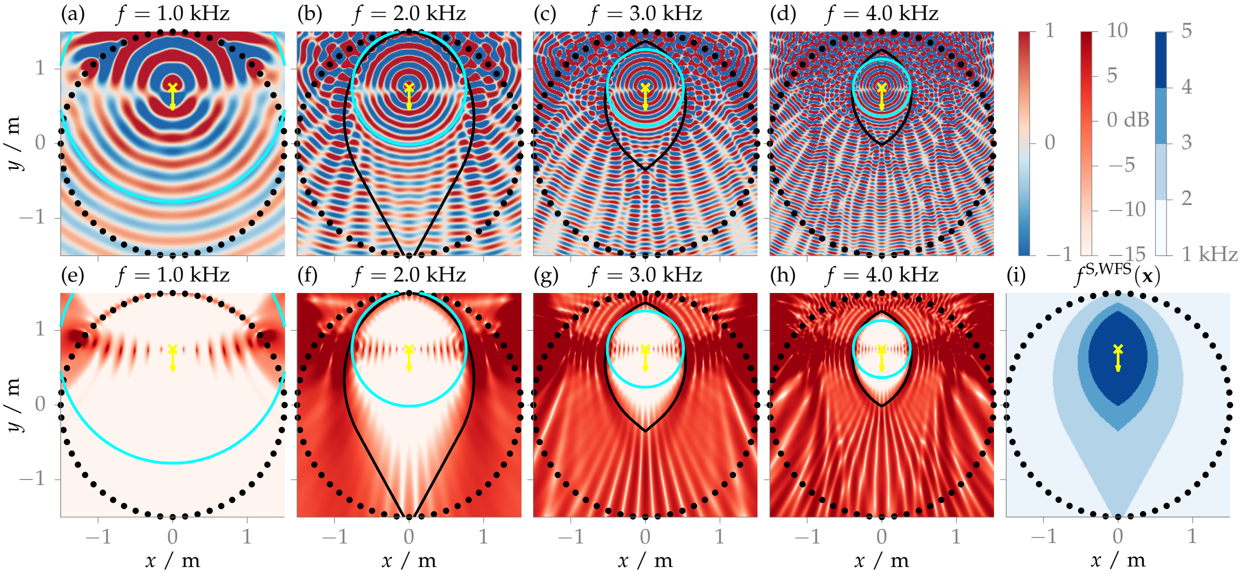


Figure 4.16: The plots show the quantities analogous to Fig. 4.14 for a virtual focused point source located at $\mathbf{x}_{\text{fs}} = [0, 0.75, 0]^{\text{T}}$ m with the orientation $\mathbf{n}_{\text{fs}} = [0, -1, 0]^{\text{T}}$ (yellow cross and arrow). For the positions outside the area surrounded by the black lines, the anti-aliasing criterion involving $f^{\text{S,WFS}}(\mathbf{x})$ is violated, see (4.33). In addition, the cyan circle indicates the aliasing-free area according to (4.39).

d_s between the focus point and the linear SSD. Transferring it to a circular SSD the minimum distance is given by $d_s = R - \rho_{\text{fs}}$, whereas $\rho_{\text{fs}} \leq R$ is the distance of the focal point from the centre of the SSD. The modified formula reads

$$R_{\text{l}} = \frac{d_s c}{f \Delta_x} = \left(1 - \frac{\rho_{\text{fs}}}{R}\right) \frac{N_0 c}{2\pi f} . \tag{4.39}$$

Its results are plotted as the cyan circles in Fig. 4.16: For 1 kHz, the radius underestimates the aliasing-free region. In the remaining plots, the circular region slightly exceeds the predictions of the geometric model. The higher the frequency, the closer the two predictions match. As the formula assumes a circular region, it is not capable of predicting the correct contour of the aliasing-free region. The radius R_{l} in (4.39) may be replaced by the distance of a distinct coordinate $\mathbf{x}$ from the focus point $\mathbf{x}_{\text{fs}}$. Solving the equation for f yields the position-dependent aliasing frequency

$$f^{\text{S,WFS}}(\mathbf{x}) = \frac{N_0 c}{2\pi R} \frac{R - \rho_{\text{fs}}}{|\mathbf{x} - \mathbf{x}_{\text{fs}}|} \tag{4.40}$$

predicted by the model of Wierstorf. For the synthesis scenario under investigation, Wierstorf's prediction yields approximately 2 kHz for the centre position, i.e. $\mathbf{x} = \mathbf{0}$. This also agrees with the plot of Fig. 4.16b, where the cyan circle barely includes the origin for $f = 2$ kHz. The geometric model estimates an aliasing frequency of approximately 4 kHz. The coarse approximation of the aliasing-free region by a circle in Wierstorf's model hence yields to an underestimation of the aliasing frequency by a factor of 2.

Non-Uniformly Discretised SSD: In order to demonstrate the capabilities of the geometric model to incorporate non-standard SSDs, an exponentially spaced circular SSD is chosen. The secondary source positions read $\mathbf{x}_0(u) = R[\cos\phi_0(u), \sin\phi_0(u), 0]^{\text{T}}$ with the azimuth

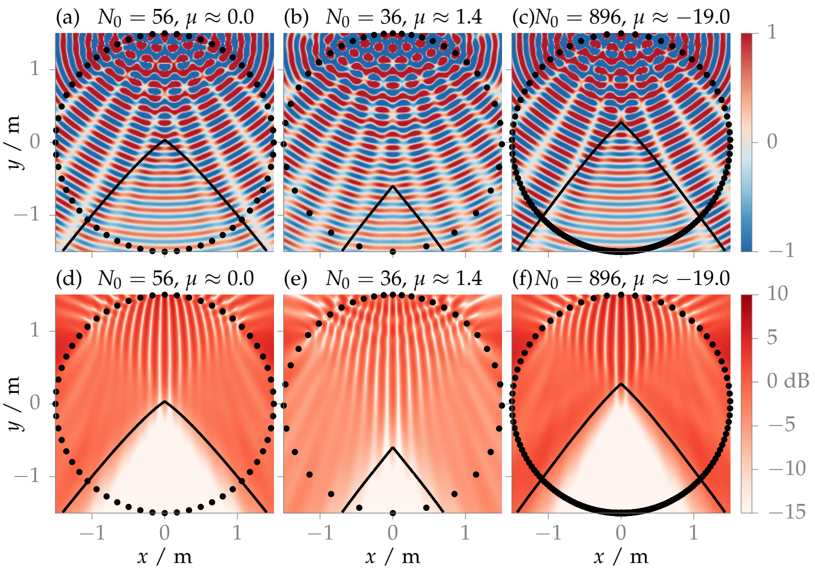


Figure 4.17: The plots (a)-(c) show the real part of a monochromatic ($f = 2.0$ kHz) virtual point source located at $\mathbf{x}_{\mathrm{ps}} = [0, 2.5, 0]^{\mathrm{T}}$ m synthesised with an exponentially sampled circular SSD. The number of secondary sources N_0 and the spacing parameter μ for the exponential sampling are given above each plot. The SSD is driven by WFS. The according error $\hat{\varepsilon}(\mathbf{x}, \omega)$ caused by aliasing is plotted in (d)-(f), see (4.37). For the positions above the solid black lines, the predicted anti-aliasing criterion involving $f^{\mathrm{S,WFS}}(\mathbf{x})$ is violated. The aliasing frequency $f^{\mathrm{S,WFS}}(\mathbf{x})$ is estimated by the algorithm in Fig. 4.9.

angle given as

$$\phi_0(u) = \pi \,\mathrm{sgn}(u) \frac{\mathrm{e}^{\mu|u|} - 1}{\mathrm{e}^{\mu} - 1} + \frac{\pi}{2}, \qquad u \in [-1, 1]\,. \tag{4.41}$$

Depending on whether the spacing parameter μ is negative/positive, the angle between two adjacent secondary sources in-/decreases the closer the secondary source azimuth is to $\pi/2$. Different examples are plotted in Fig. 4.17. The sound field reproduced by the discrete SSD given by (4.38) has to be adjusted to

$$P^{\mathrm{S}}(\mathbf{x}, \omega) = \sum_{\nu=0}^{N_0 - 1} D(\mathbf{x}_0^{(\nu)}, \omega) G(\mathbf{x} - \mathbf{x}_0^{(\nu)}, \omega)\, \Delta_{\mathbf{x}_0}(u^{(\nu)})\,, \tag{4.42}$$

together with $\mathbf{x}_0^{(\nu)} = \mathbf{x}_0(u^{(\nu)})$, $\Delta_{\mathbf{x}_0}(u) = (2\pi R\,\mathrm{e}^{\mu|u|})/(N_0(\mathrm{e}^{\mu}-1))$, and $u^{(\nu)} = (2\nu - N_0)/N_0$. The synthesised sound field and the sampling error are shown in Fig. 4.17 for three different parametrisations of the SSD. The same point source as for uniform case serves as the virtual sound field. The spacing parameter μ and the number of secondary sources N_0 have been chosen such that the number of active secondary sources selected by the selection criterion (3.5) is equal to the uniform case. It allows for comparability as the sound field is always synthesised by same number of secondary sources. For Fig. 4.17b/e, the positive μ leads to a denser SSD around $\pi/2$. Compared to the uniform case in Fig. 4.17a/d, the area of low aliasing error is smaller. A sparser sampling around $\pi/2$ is chosen for Fig. 4.17c/f: Here, an improvement w.r.t. the aliasing error can be observed. For both parametrisations, the predictions of the aliasing frequency by the geometric model (black lines) agree with the error plots. The findings agree with the investigation by Corteel,[54] which indicated a possibly positive impact of irregularly spaced arrays on the aliasing properties.

[54] Corteel (Sept. 2006). “On the use of irregularly spaced loudspeaker arrays for Wave Field Synthesis, potential impact on spatial aliasing frequency”. In: *Proc. of 9th Int. Conf. on Digital Audio Effects (DAFx-06)*. Montreal, Canada, pp. 209–214.

Optimal Discretisation Schemes: The last example showed, that spatial aliasing artefacts in WFS are significantly influenced by the chosen discretisation scheme. It will be demonstrated now, that the geometric model can be used to optimise the sampling on a given SSD contour w.r.t. to the aliasing frequency for specific virtual sound fields. It is assumed, that a (suboptimal) parametrisation $\mathbf{x}_0(u)$ is given and allows to explicitly calculate the secondary source positions. $\mathbf{x}_0(v)$ is the optimal, yet unknown parametrisation. Since $\mathbf{x}_0(u)$ and $\mathbf{x}_0(v)$ describe the same boundary, there is a pair of u and v corresponding to the same secondary source position $\mathbf{x}_0$. Thus, v may be expressed as a function of u, i.e. $v(u)$. The bijectivity requires $v(u)$ to be either strictly increasing or decreasing. Without loss of generality, it is assumed that $v'(u) > 0$ (for increasing $v(u)$) and both parameters share the same support, i.e. $u_{\text{max}} = v(u_{\text{max}})$, $u_{\text{min}} = v(u_{\text{min}})$, and $\Delta_u = \Delta_v$. The task is to find the parametrisation $\mathbf{x}_0(v)$ that maximizes the aliasing frequency. It is shown in Sec. B.3, that

published as Winter et al. (Mar. 2019b). "Array Design for Increased Spatial Aliasing Frequency in Wave Field Synthesis Based on a Geometric Model". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Rostock, Germany, pp. 463–466

$$v'_{\text{opt}}(u) = \frac{(u_{\text{max}} - u_{\text{min}})}{\displaystyle\int_{u_{\text{min}}}^{u_{\text{max}}} \frac{1}{f_u^{\text{S,WFS}}(\mu)} \mathrm{d}\mu} \cdot \frac{1}{f_u^{\text{S,WFS}}(u)} \tag{4.43}$$

defines this optimal relation between v and u. The aliasing frequency for a distinct secondary source with the u-parametrisation is denoted by $f_u^{\text{S,WFS}}(u) = f^{\text{S,WFS}}(\mathbf{x}_0(u))$. The variable can be replaced by e.g. (4.31), or (4.34) in order to optimise the sampling w.r.t. to a specific scenario. Inserting (4.43) into (B.17a) yields the optimal aliasing frequency

$$f_{\text{opt}}^{\text{S,WFS}} = \frac{(u_{\text{max}} - u_{\text{min}})}{\displaystyle\int_{u_{\text{min}}}^{u_{\text{max}}} \frac{1}{f_u^{\text{S,WFS}}(\mu)} \mathrm{d}\mu} . \tag{4.44}$$

The relation basically constitutes the optimal aliasing frequency as the reciprocal of the arithmetic mean of the inverse aliasing frequencies w.r.t. u.

While the relation between u and v is known, an explicit formula for $\mathbf{x}_0(v)$ is not available. In order to perform equidistant sampling w.r.t. v, the samples $u^{(\nu)}$ corresponding to $v^{(\nu)} = \nu \cdot \Delta_u + u_{\text{min}}$ are computed. For this, the equation

$$\nu \cdot \Delta_u = \int_{u_{\text{min}}}^{u^{(\nu)}} v'_{\text{opt}}(\mu) \mathrm{d}\mu \tag{4.45}$$

has to be solved. It can be evaluated by combining numerical integration and root finding algorithms. The resulting positions are given by $\mathbf{x}_0(u^{(\nu)})$.

Fig. 4.18 shows the effect of the optimised SSD discretisation for the synthesis of a virtual point source. The secondary source are distributed on an arc of radius $R = 1.5$ m. The length of the arc is chosen such, that all secondary sources are activated by the selection criterion. The sampling was optimised for the aliasing frequency $f_{\mathcal{C}_\text{h}}^{\text{S,WFS}}$ for a circular listening region $\mathcal{C}_\text{h}$ (red circle) given by (4.36). Compared to the equidistant sampling in Fig. 4.18a, less aliasing

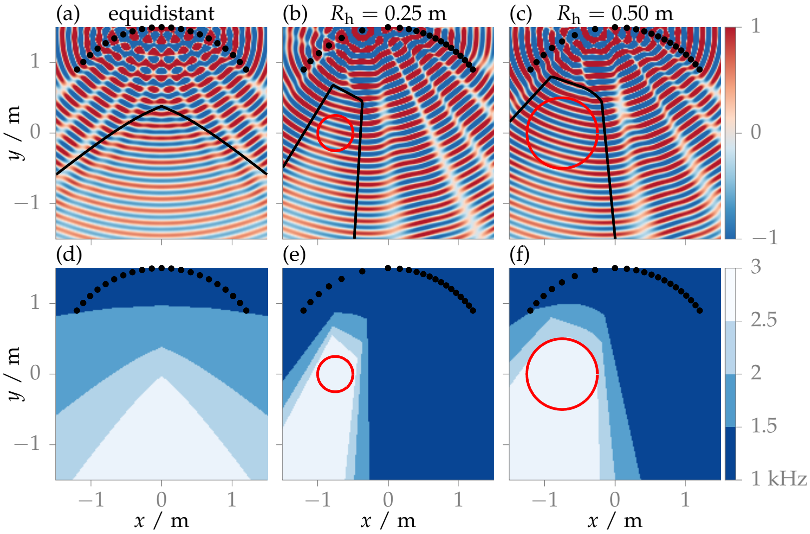


Figure 4.18: The plots in the top row show a monochromatic ($f = 2$ kHz) virtual point source at $\mathbf{x}_{ps} = [0, 2.5, 0]^T$ m synthesised by a circular SSD ($N_0 = 21$, black dots) with different non-uniform spacing patterns. (b) and (c) correspond to optimisation w.r.t. to $f^{S,WFS}_{\mathcal{C}_h}$ for a circular area area $\mathcal{C}_h$ located at $\mathbf{x}_h = [-0.75, 0, 0]^T$ m with $R_h = 0.25$ m and 0.5 m (red circle). For coordinates above the black line, the predicted aliasing frequency $f^{S,WFS}(\mathbf{x})$ defined by (4.33) is lower than 2 kHz. This frequency is shown in the bottom plots in more detail. A discrete colourmap is used for better visibility.

artefacts are present inside $\mathcal{C}_h$ in (b) and (c). This is also reflected by the estimated aliasing frequencies $f^{S,WFS}(\mathbf{x})$ shown Fig. 4.18d-e. Additional simulations using (4.36) reveal, that the $f^{S,WFS}_{\mathcal{C}_h}$ is about 1.62 ($R_l = 0.25$ m) and 1.48 ($R_l = 0.5$ m) times as high as for the equiangular pattern. As rule of thumb, a smaller target area allows for larger increase of the aliasing frequency inside it. For $R_l \to \infty$, the SSD is only optimised w.r.t. to the virtual sound field and $f^{S,WFS}_{\mathcal{C}_h}$ is only 1.16 times as high in comparison to the equidistant case. If the virtual sound field is assumed to be arbitrary or unknown, $f^{S,WFS}_{\mathcal{C}_h}$ would be equivalent to the half-wavelength criterion. The optimal sampling scheme for this criterion is the equi-angular pattern. Again, the presented geometric framework includes the prediction of the traditional model of Sec. 4.1 as a special case.

4.3 Geometric Model for Near-Field-Compensated Higher-Order Ambisonics

It was stated by Ahrens[55] and further discussed by Schultz et al.,[56] that 2.5D WFS is a high-frequency approximation of spatially full-band 2.5D NFCHOA. Thus, the 2.5D driving signal (3.17) is approximated as

$$D^{\mathrm{HOA}}_{2.5\mathrm{D}}(\mathbf{x}_0, \omega) \overset{\omega \to \infty}{\approx} a_S(\mathbf{x}_0) \sqrt{\mathrm{j}\frac{\omega}{c}} \sqrt{8\pi \Delta_S(\mathbf{x}_0)} \langle \hat{\mathbf{k}}_S(\mathbf{x}_0, \omega) | \mathbf{n}_0 \rangle S(\mathbf{x}_0, \omega)\,, \tag{4.46}$$

where the right-hand side of the equation is the 2.5D WFS driving function given by (3.9b). As the high-frequency approximation agrees with the assumptions made for the geometric model for WFS, it can be directly applied to predict the aliasing artefacts of fullband NFCHOA. For the bandwidth-limited counterpart given by (3.18), the effect of the spatial bandwidth limitation on the aliasing properties has to be investigated. This will be done in the upcoming section.

[55] Ahrens, op. cit.

[56] Schultz et al. (Mar. 2019). "On the Connections of High-Frequency Approximated Ambisonics and Wave Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Rostock, Germany, pp. 1446–1449.

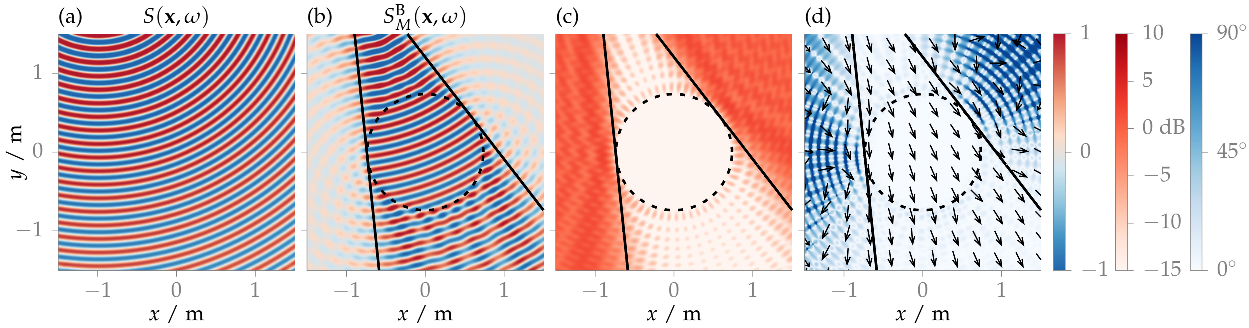


Figure 4.19: (a) shows a monochromatic ($f = 2$ kHz) point source located at $\mathbf{x}_{\mathrm{ps}} = [-1, 2.5, 0]^{\mathrm{T}}$ m. Its bandwidth-limited ($M = 27$) version is plotted in (b). The normalised error

$$20 \lg \left| \frac{S_M^{\mathrm{B}}(\mathbf{x}_0, \omega) - S(\mathbf{x}_0, \omega)}{S(\mathbf{x}_0, \omega)} \right|$$

is given by (c). (d) shows the absolute angle between in the propagation directions $\hat{\mathbf{k}}_S$ and $\hat{\mathbf{k}}_{S_M^{\mathrm{B}}}$ (arrows) of the two sound field. The solid lines indicate the positions $\mathbf{x}$ for which $\frac{\omega}{c}|\mathbf{x} \times \hat{\mathbf{k}}_S(\mathbf{x}, \omega)| = M$. The dashed circle has a radius of $Mc/2\pi f$.

4.3.1 Spatial Bandwidth Limitation

According to (3.17) and (3.18), the bandwidth limitation in the driving function is directly applied to the coefficients of the virtual sound field $\breve{S}_{|m|}^{m}(\omega)$ by weighting them with the modal window $\mathring{w}_m^M$. The weighted coefficients may be interpreted as the coefficients belonging to a bandwidth-limited sound field denoted as $S_M^{\mathrm{B}}(\mathbf{x}, \omega)$. Bandwidth-limited NFCHOA essentially corresponds to a fullband synthesis of this sound field. The high-frequency approximation of the driving signal in (4.46) adapts to

$$D_{2.5\mathrm{D}}^{\mathrm{HOA}}(\mathbf{x}_0, \omega) \approx a_{S_M^{\mathrm{B}}}(\mathbf{x}_0) \sqrt{\mathrm{j}\frac{\omega}{c}} \sqrt{8\pi \Delta_{S_M^{\mathrm{B}}}(\mathbf{x}_0)} \langle \hat{\mathbf{k}}_{S_M^{\mathrm{B}}}(\mathbf{x}_0, \omega) | \mathbf{n}_0 \rangle S_M^{\mathrm{B}}(\mathbf{x}_0, \omega)\,. \tag{4.47}$$

Thus, it is necessary to discuss the properties of $S_M^{\mathrm{B}}(\mathbf{x}, \omega)$ in order to derive a meaningful model for spatial aliasing in NFCHOA. An exemplary point source and its bandwidth-limited version are compared in Fig. 4.19. According to Ahrens,[57] the error between the bandwidth-limited and fullband sound field is negligible inside a circular area around the origin with the radius $\frac{Mc}{\omega}$. This is confirmed by Fig. 4.19c, where the error significantly drops inside the dashed circle. The mentioned criterion is independent of the sound field and of the direction, i.e. the azimuth, of $\mathbf{x}$ relative to the origin. However, it was shown by Hahn and Spors [58] and in Sec. 3.2.1 for bandwidth-limited plane waves, that the error has a dependence on the azimuth relative to the propagation direction of the plane wave. This can also be seen for the point source in Fig. 4.19b, where the sound pressure significantly decreases for positions outside the area contained by the solid black lines. In Fig. 4.19c, these positions exhibit a high error. Inside the fan-shaped area, but outside the dashed circle, the error is mostly caused by amplitude variations. According to Hahn and Spors[59] and based on the analysis in Sec. 3.2.1, these result in fluctuations in temporal frequency spectrum of the sound field. Fig. 4.19d shows, that the propagation direction of the bandwidth-limited sound field is stable within the fan-shaped area. Only minor differences to the fullband sound field can be observed.

The described observations are further substantiated by analytic derivations given in Sec. A.1. A high-frequency approximation of

[57] Ahrens, op. cit.

[58] Hahn and Spors (Mar. 2015a). "Modal Bandwidth Reduction in Data-based Binaural Synthesis including Translatory Head-movements". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 1122–1125

[59] Ibid.

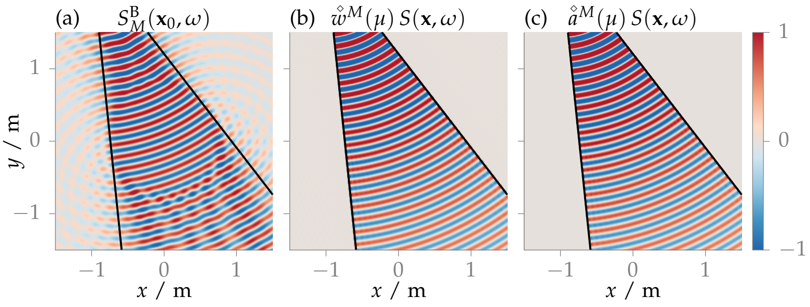


Figure 4.20: (a) shows a bandwidth-limited ($M = 27$) monochromatic ($f = 2$ kHz) point source located at $\mathbf{x}_{\mathrm{ps}} = [-1, 2.5, 0]^{\mathrm{T}}$ m. (b) and (c) plot the high-frequency approximations given by (4.48) and (4.50) with $\mu = \frac{\omega}{c}|\mathbf{x} \times \hat{\mathbf{k}}_S(\mathbf{x},\omega)|$. The black lines indicate the positions $\mathbf{x}$ for which $\frac{\omega}{c}|\mathbf{x} \times \hat{\mathbf{k}}_S(\mathbf{x},\omega)| = M$ holds.

the bandwidth limited sound field is given by

$$S_M^{\mathrm{B}}(\mathbf{x},\omega) \overset{\omega\to\infty}{\approx} \mathring{w}^M\left(\frac{\omega}{c}|\mathbf{x} \times \hat{\mathbf{k}}_S(\mathbf{x},\omega)|\right) S(\mathbf{x},\omega)\,, \tag{4.48}$$

with the sinc[60] interpolation of the modal window $\mathring{w}_m^M$ defined as

$$\mathring{w}^M(\mu) := \sum_{m=-M}^{M} \mathring{w}_m^M \operatorname{sinc}\left((m-\mu)\pi\right)\,. \tag{4.49}$$

[60] Girod et al., op. cit., Eq. (9.21).

An exemplary bandwidth-limited point source and the approximation are shown in Fig. 4.20a and b: Besides the mentioned amplitude variations, the wave fronts of $S_M^{\mathrm{B}}(\mathbf{x},\omega)$ and its approximation coincide within the area contained by the black lines. The sound pressure outside the mentioned area decreases in both plots, whereas the decrease is more abrupt for the approximation. This effect becomes even stronger, the more the temporal frequency f is increasing. Hence, a second high-frequency approximation for $S_M^{\mathrm{B}}(\mathbf{x},\omega)$ may be introduced, where the sound pressure outside the black lines is set to zero. It reads

$$S_M^{\mathrm{B}}(\mathbf{x},\omega) \approx \mathring{a}^M\left(\frac{\omega}{c}|\mathbf{x} \times \hat{\mathbf{k}}_S(\mathbf{x},\omega)|\right) S(\mathbf{x},\omega) \tag{4.50}$$

with

$$\mathring{a}^M(\mu) \approx \begin{cases} \mathring{w}^M(\mu) & \text{, if } \mu \leq M \\ 0 & \text{, otherwise.} \end{cases} \tag{4.51}$$

It is depicted in Fig. 4.20c and exhibits no visible difference to the first approximation in Fig. 4.20b. Since $\mathring{a}^M(\mu)$ is a real-valued function, the phase of $S_M^{\mathrm{B}}(\mathbf{x},\omega)$ in (4.50) coincides with the phase of $S(\mathbf{x},\omega)$. Thus, the local wavenumber vector $\hat{\mathbf{k}}_{S_M^{\mathrm{B}}}(\mathbf{x},\omega)$ and $\hat{\mathbf{k}}_S(\mathbf{x},\omega)$ match for this approximation. The approximation of the 2.5D NFCHOA driving signal in (4.47) may be refined to

$$\begin{aligned} D_{2.5\mathrm{D}}^{\mathrm{HOA}}(\mathbf{x}_0,\omega) \approx\, & a_S(\mathbf{x}_0)\sqrt{\mathrm{j}\frac{\omega}{c}}\sqrt{8\pi\Delta_S(\mathbf{x}_0)}\langle\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)|\mathbf{n}_0\rangle \\ & \cdot \mathring{a}^M\left(\frac{\omega}{c}|\mathbf{x}_0 \times \hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)|\right) S(\mathbf{x}_0,\omega) \end{aligned} \tag{4.52}$$

Compared to the fullband driving signal approximated in (4.47), the factor $\mathring{a}^M$ is the only difference. It is interpreted as an additional secondary source selection criterion, which is frequency dependent and exhibits a lowpass characteristic. Assuming that $\hat{\mathbf{k}}_S(\mathbf{x},\omega)$ is frequency-independent, the frequency

$$f_M^{\mathrm{B}}(\mathbf{x}_0) = \frac{Mc}{2\pi|\mathbf{x}_0 \times \hat{\mathbf{k}}_S(\mathbf{x}_0)|} \tag{4.53}$$

defines the threshold up to which a secondary source is still active. If the aliasing frequency of a single secondary source, e.g. $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,WFS}}(\mathbf{x}_0)$, is greater than $f_M^{\mathrm{B}}(\mathbf{x}_0)$, it does not contribute any aliasing components due to its inactivity. For the secondary source $\mathbf{x}_0$, the NFCHOA aliasing frequency for an extended listening area $\mathcal{S}_\mathrm{h}$ is given by

$$f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}}(\mathbf{x}_0) = \begin{cases} f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,WFS}}(\mathbf{x}_0) & \text{for } f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,WFS}}(\mathbf{x}_0) \leq f_M^{\mathrm{B}}(\mathbf{x}_0)\,, \\ \infty & \text{, otherwise.} \end{cases} \tag{4.54}$$

It is always greater or equal to the WFS aliasing frequency $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,WFS}}(\mathbf{x}_0)$. Hence, the SBL in NFCHOA potentially reduces spatial aliasing artefacts and increases the aliasing frequency. For the fullband case, i.e. $M \to \infty$, the constraint has no effect. This agrees with the prior statement, that WFS and fullband NFCHOA share similar properties. Analogously to the aliasing frequency (4.36) for WFS, the minimum over the secondary sources with non-zero $a_S(\mathbf{x}_0)$ constitutes the overall aliasing frequency

$$f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}} = \min_{\mathbf{x}_0 | a_S(\mathbf{x}_0) \neq 0} f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}}(\mathbf{x}_0)\,. \tag{4.55}$$

An algorithm for its estimation is shown in Fig. 4.21. The NFCHOA aliasing frequency $f^{\mathrm{S,NFCHOA}}(\mathbf{x})$ for a distinct position $\mathbf{x}$ is a special case of $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}}$ for a circular area $\mathcal{S}_\mathrm{h} = \mathcal{C}_\mathrm{h}$ with $\mathbf{x}_\mathrm{h} = \mathbf{x}$ and $R_\mathrm{h} = 0$. It is estimated with the same algorithm. It can be deduced from the presented calculus, that the predicted aliasing frequencies depend on the chosen modal bandwidth M but are independent of the actual shape of the modal window $\mathring{w}_m^M$. Potential effects of the shape on the aliasing properties are not covered by the developed model.

4.3.2 Application and Validation

Influence of SBL: For 2.5D synthesis scenarios, NFCHOA states the explicit solution to the SLP defined in (3.13). Thus, only a circular SSD can be used for the evaluation. For comparability, the same uniformly discretised SSD as for WFS is used, i.e. $R = 1.5$ m, $N_0 = 56$, and $\Delta_{\mathbf{x}_0} = {}^{2\pi R}/_{N_0}$. It synthesises a virtual point source located at $\mathbf{x}_{\mathrm{ps}} = [0, 2.5, 0]^{\mathrm{T}}$ m. The driving signal is given by (3.18) and Tab. 3.2. In order to quantify the synthesis accuracy, the normalised error used for WFS is re-defined to

$$\hat{\varepsilon}(\mathbf{x}, \omega) := 20 \log_{10} \left| \frac{P^{\mathrm{S}}(\mathbf{x}, \omega) - P_{\mathrm{ref}}(\mathbf{x}, \omega)}{P(\mathbf{x}, \omega)} \right| \,. \tag{4.56}$$

As a reference, $P_{\mathrm{ref}}(\mathbf{x}, \omega)$ denotes the sound field reproduced by a continuous SSD without spatial bandwidth limitation. Thus, the error incorporates the combined effects of spatial aliasing and the SBL.

function AliasingExtendedNFCHOA($\mathcal{S}, \mathcal{S}_\mathrm{h}, M$)
  $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}} \leftarrow \infty$
  for $\mathbf{x}_0, \mathbf{x}_0' \leftarrow \partial\mathcal{S}$ **do** ▷ (4.24), densely sampled
    if $a_S(\mathbf{x}_0) = 0$ **then** ▷ (4.55)
      continue ▷ next secondary sources
    end if
    $\Delta_{\mathbf{x}_0} \leftarrow \Delta_u |\mathbf{x}_0'|$
    $\hat{k}_{G,\mathbf{t}_0}^{\min}, \hat{k}_{G,\mathbf{t}_0}^{\max} \leftarrow$ MinMaxWavenumber($\mathcal{S}_\mathrm{h}, \mathbf{x}_0$)
    $f \leftarrow \dfrac{c}{\Delta_{\mathbf{x}_0} \max\left(|\hat{k}_{G,\mathbf{t}_0}^{\min} - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|; |\hat{k}_{G,\mathbf{t}_0}^{\max} - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|\right)}$
    $f_M^\mathrm{B} \leftarrow \dfrac{Mc}{2\pi |\mathbf{x}_0 \times \hat{\mathbf{k}}_S(\mathbf{x}_0)|}$ ▷ (4.53)
    if $f \leq f_M^\mathrm{B}$ **then** ▷ (4.54)
      $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}} \leftarrow \min(f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}}; f)$ ▷ (4.55)
    end if
  end for
  return $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}}$
end function

Figure 4.21: Brute-force search algorithm to determine the NFCHOA aliasing frequency $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,NFCHOA}}$ for an extended area $\mathcal{S}_\mathrm{h}$ given by (4.54). An example of the function MinMaxWavenumber for a circular region is given in Fig. 4.9.

Fig. 4.22 shows the synthesised sound field and the according error for different modal windows. As already discussed in Sec. 3.2.1, the max-$\mathbf{r}_E$ weighting function leads to less amplitude fluctuations and smoother wave fronts compared to the rectangular window of same M. The estimated frequencies $f_M^\mathrm{B}(\mathbf{x})$ (dashed) and $f^{\mathrm{S,NFCHOA}}(\mathbf{x})$ (solid) do not take the window type into account. The geometric model is not capable to explain this phenomenon. For $M = 20$, the error is mainly caused by SBL for both window types. This is predicted correctly by using the frequency $f_M^\mathrm{B}(\mathbf{x})$. Increasing the bandwidth to $M = 34$, increases the area which is free of SBL artefacts. Larger values of M cause more spatial aliasing as the artefacts-free area is now dominantly restricted by the solid black lines symbolising the estimated aliasing frequency $f^{\mathrm{S,NFCHOA}}(\mathbf{x})$. For $M = 300$, the SBL has no effect. The spatial structure of the sound field and the error are very similar to WFS.[61] This agrees with the prior statement, that fullband NFCHOA exhibits similar aliasing properties as WFS. Overall, the results confirm, that spatial aliasing increased with larger value of M. For a low spatial bandwidth, the artefacts of the SBL are dominating.

[61] see Fig. 4.14c/g

Optimal Modal Bandwidth: It was shown in the previous example that a small modal bandwidth reduces spatial aliasing but has the drawback of a small listening region with high synthesis accuracy. In the literature,[62] the optimal modal bandwidth w.r.t. a trade-off between spatial aliasing and bandwidth limitation is given by $M = \lfloor (N_0-1)/2 \rfloor$. This value will be compared with the predictions of the model: A circular listening region $\mathcal{C}_\mathrm{h}$ with radius R_h concentric to

[62] Ahrens, op. cit., Eq. (4.26).

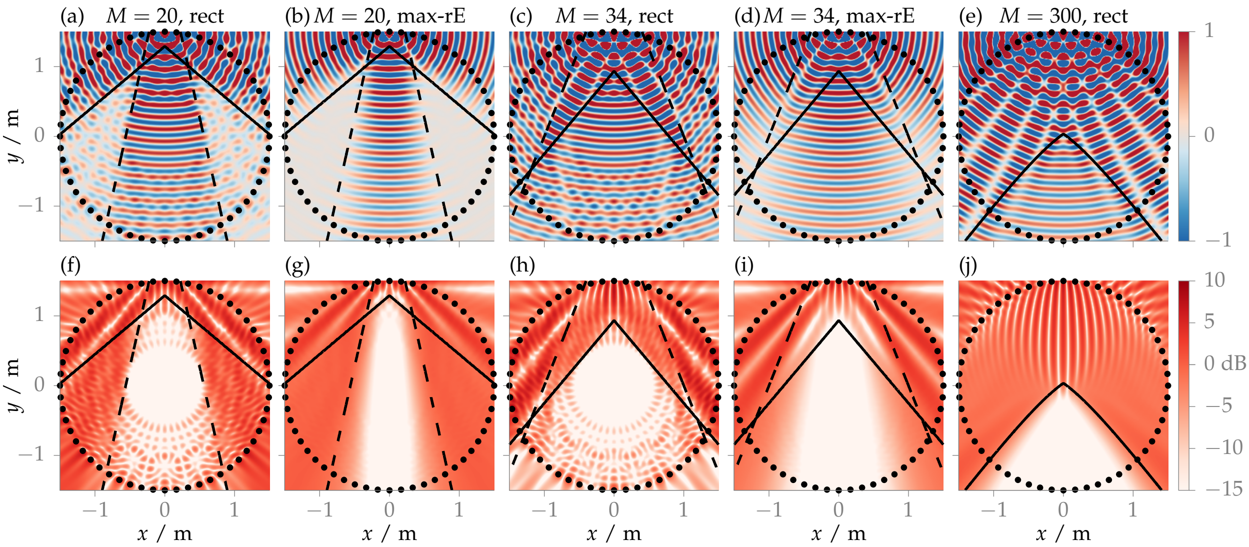


Figure 4.22: The plots (a)-(e) show the real part of a monochromatic ($f = 2$ kHz) virtual point source located at $\mathbf{x}_{\mathrm{ps}} = [0, 2.5, 0]^{\mathrm{T}}$ m synthesised by NF-CHOA for different modal windows $\mathring{w}_m^M$. The according error $\hat{\varepsilon}(\mathbf{x}, \omega)$ caused by aliasing and SBL is plotted in (f)-(j), see (4.56). For the positions $\mathbf{x}$ above the solid black lines, the predicted anti-aliasing criterion involving $f^{\mathrm{S,NFCHOA}}(\mathbf{x})$ is violated, see (4.55). It is estimated using the algorithms in Fig. 4.13 and Fig. 4.21 for $\mathcal{S}_{\mathrm{h}} = \mathcal{C}_{\mathrm{h}}$ with $R_{\mathrm{h}} = 0$ and $\mathbf{x}_{\mathrm{h}} = \mathbf{x}$. For positions, which are not contained by the dashed black lines, $f_M^{\mathrm{B}}(\mathbf{x}) < f$, see (4.53).

the circular SSD is considered. The SBL frequency is given as

$$f_M^{\mathrm{B}}(\mathbf{x}) = \frac{Mc}{2\pi|\mathbf{x} \times \hat{\mathbf{k}}_S(\mathbf{x})|} \geq \frac{Mc}{2\pi|\mathbf{x}|} \geq \frac{Mc}{2\pi R_{\mathrm{h}}} \quad , \forall \mathbf{x} \in \mathcal{C}_{\mathrm{h}} \,. \tag{4.57}$$

The first inequality results from the length of the cross-product which takes its maximum value, if the involved vectors are perpendicular. In this case, it is equal to the product of the individual lengths of the two vectors. It results in a lower bound for $f_M^{\mathrm{B}}(\mathbf{x})$ for an arbitrary virtual sound field, i.e. arbitrary $\hat{\mathbf{k}}_S(\mathbf{x})$. The second inequality considers the maximum for $|\mathbf{x}|$ inside the circular area.

For the aliasing frequency of the area, (4.54) and (4.34) have to be considered. The necessary values for $\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ for a circular area are given by (B.12). Due to the circular symmetry of the scenario, $\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ can be further simplified to

$$\hat{k}^{\{\min,\max\}}_{G,\mathbf{t}_0}(\mathbf{x}_0) = \begin{cases} \mp\frac{R_{\mathrm{h}}}{R} & \text{if } R_{\mathrm{h}} \leq R\,, \\ \mp 1 & \text{if } R_{\mathrm{h}} > R\,, \end{cases} \tag{4.58}$$

where the upper and lower option for $\mp$ applies for $\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$, respectively. Inserting the values together with $\Delta_{\mathbf{x}_0} = {}^{2\pi R}/_{N_0}$ into (4.34) results in

$$f^{\mathrm{S}}_{\mathcal{C}_{\mathrm{h}}}(\mathbf{x}_0) = \begin{cases} \frac{cN_0}{2\pi R(1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|)} & \text{if } R_{\mathrm{h}} > R\,, \\ \frac{cN_0}{2\pi(R_{\mathrm{h}}+R|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|)} & \text{if } R_{\mathrm{h}} \leq R\,. \end{cases} \tag{4.59}$$

According to (4.57), the radius R_{h} which is free of artefacts caused by SBL is lower bounded by $\frac{Mc}{2\pi f}$. Inserting the radius into (4.59) leads to an implicit formulation w.r.t. $f^{\mathrm{S}}_{\mathcal{C}_{\mathrm{h}}}(\mathbf{x}_0)$. Its solution reads

$$f^{\mathrm{S}}_{\mathcal{C}_{\mathrm{h}}}(\mathbf{x}_0) = \begin{cases} \frac{N_0 c}{2\pi R(1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|)} & \text{if } M > \frac{N_0}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}\,, \\ \frac{(N_0-M)c}{2\pi R|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|} & \text{if } M \leq \frac{N_0}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}\,. \end{cases} \tag{4.60}$$

In order to have an infinite aliasing frequency $f^{\mathrm{S}}_{\mathcal{C}_\mathrm{h}|M}(\mathbf{x}_0)$, see (4.54), $f^{\mathrm{B}}_{M}(\mathbf{x}_0)$ has to be smaller than $f^{\mathrm{S}}_{\mathcal{C}_\mathrm{h}}(\mathbf{x}_0)$ for all secondary sources. Taking the definition of $f^{\mathrm{B}}_{M}(\mathbf{x}_0)$ in (4.53) into account, the condition is formulated via

$$\frac{cM}{2\pi R|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|} \overset{!}{<} \begin{cases} \frac{cN_0}{2\pi R(1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|)} & \text{if } M > \frac{N_0}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}\,, \\ \frac{c(N_0-M)}{2\pi R|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|} & \text{if } M \leq \frac{N_0}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}\,. \end{cases} \tag{4.61}$$

Hereby, it was exploited that $|\mathbf{x}_0 \times \hat{\mathbf{k}}_S(\mathbf{x}_0)| = R|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|$ holds for circular SSDs. The condition is further simplified resulting in two (independent) constrained conditions:

$$M \overset{!}{<} N_0 \frac{|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|} \quad , \text{if } M > \frac{N_0}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}\,, \text{or} \tag{4.62a}$$

$$M \overset{!}{<} \frac{N_0}{2} \quad , \text{if } M \leq \frac{N_0}{1+|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|}\,, \tag{4.62b}$$

Since $|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|$ is always positive and less or equal 1, there is no value for M, that fulfils the first condition without violating its constraint. All M fulfilling the second condition automatically comply with the according constraint independent of $|\hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|$. Thus, it is necessary for M to be smaller than $N_0/2$. Under the premise that the size of the available listening area is supposed to be as large as possible, the best choice for M is the largest integer fulfilling this criterion. The prediction of the geometric model for optimal modal bandwidth in NFCHOA reads $M = \lfloor (N_0-1)/2 \rfloor$ and agrees with the result from Sec. 4.1 and from the literature.[63] According to the model, it corresponds to a circular region with $R_\mathrm{h} = \frac{Mc}{2\pi f}$ around the SSD centre, which is free of artefacts caused by spatial aliasing and the SBL.

[63] Ibid., Eq. (4.26).

4.4 Geometric Model for Local Wave Field Synthesis using Spatial Bandwidth Limitation

An appropriate choice for the spatial bandwidth M can be utilised to reduce spatial aliasing. However, without further modification NFCHOA does not allow for shifting the area of high synthesis accuracy from the array centre. LWFS-SBL uses the same mechanism as NFCHOA to avoid spatial aliasing. In this method, it is possible to adjust the expansion centre $\mathbf{x}_c$ of the Circular Harmonics representation of the virtual sound field and shift the location of the area with high synthesis accuracy to this position. This was demonstrated in Sec. 3.4. In Sec. 4.4.1, the geometric model for NFCHOA will be generalised to incorporate bandwidth-limited sound fields, which were expanded around positions other than the coordinates' origin. It enables the prediction of the spatial aliasing frequency in LWFS-SBL.

The practical realisation of LWFS-SBL was discussed in Sec. 3.4.2: The CHT of the desired sound field is converted into a discrete PWD.

This discretisation constitutes a second sampling process in addition to the spatial sampling caused by the discrete SSD. Contrary to the secondary sources, the resolution of the PWD is only constrained by the computational effort and thus less critical. A discussion on the effects of the discrete PWD is presented in Sec. 4.4.2. It should be noted, that the frequency crossover between WFS and LWFS-SBL necessary to synthesise point sources is not taken into account, here.

4.4.1 Continuous Plane Wave Decomposition

For a continuous PWD, the 2.5D LWFS-SBL driving signal is given by (3.25). For convenience, it is given here again:

$$D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL}}(\mathbf{x}_0,\omega) = \frac{1}{2\pi}\int_0^{2\pi} \bar{S}(\phi_{\mathrm{pw}},\mathbf{x}_\mathrm{c},\omega) D_{2.5\mathrm{D,pw}}^{\mathrm{WFS}}(\mathbf{x}_0-\mathbf{x}_\mathrm{c}|\mathbf{n}_{\mathrm{pw}},\omega)\,\mathrm{d}\phi_{\mathrm{pw}}\,. \tag{4.63}$$

The integral describes the weighted superposition of the 2.5D WFS driving signal $D_{2.5\mathrm{D,pw}}^{\mathrm{WFS}}$ for virtual plane waves with their propagation directions $\mathbf{n}_{\mathrm{pw}}$ distributed on the unit circle. The weights are given by the plane wave coefficients $\bar{S}$ of the virtual sound field expanded around $\mathbf{x}_\mathrm{c}$. A high-frequency approximation of the LWFS-SBL driving signal based on the SPA is derived in Sec. A.2.1. It reads

$$D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL}}(\mathbf{x}_0,\omega) \overset{\omega\to\infty}{\approx} \sqrt{\mathrm{j}\frac{\omega}{c}}\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|}\, a_S(\mathbf{x}_0)\langle\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)|\mathbf{n}_0\rangle S(\mathbf{x}_0,\omega)\,. \tag{4.64}$$

Despite of the different distance correction factors, the approximation is very similar to the high-frequency approximation of the bandwidth-limited NFCHOA driving signal given by (4.47). Thus, similar steps are taken to incorporate the SBL into the approximation. The expansion around $\mathbf{x}_\mathrm{c}$ instead of coordinates' origin $\mathbf{0}$ is handled by a shift of the coordinate frame. Analogous to the discussion in Sec. 4.3.1, the driving signal is further approximated to

$$D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL}}(\mathbf{x}_0,\omega) \approx \sqrt{\mathrm{j}\frac{\omega}{c}}\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|}a_S(\mathbf{x}_0)\langle\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)|\mathbf{n}_0\rangle \cdot \mathring{a}^M\left(\frac{\omega}{c}|(\mathbf{x}_0-\mathbf{x}_\mathrm{c})\times\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)|\right) S(\mathbf{x}_0,\omega)\,. \tag{4.65}$$

$\mathring{a}^M$ is defined in (4.51). As the major difference to NFCHOA, the argument of $\mathring{a}^M$ contains the shifted secondary source position $\mathbf{x}_0 - \mathbf{x}_\mathrm{c}$. The SBL frequency (4.53) is adopted to

$$f_M^{\mathrm{B}}(\mathbf{x}_0,\mathbf{x}_\mathrm{c}) = \frac{Mc}{2\pi|(\mathbf{x}_0-\mathbf{x}_\mathrm{c})\times\hat{\mathbf{k}}_S(\mathbf{x}_0)|}\,. \tag{4.66}$$

The aliasing frequency for a single secondary source given by (4.54) is adjusted according to the shift and reads

$$f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}}(\mathbf{x}_0) = \begin{cases} f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,WFS}}(\mathbf{x}_0) & \text{for } f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,WFS}}(\mathbf{x}_0) \le f_M^{\mathrm{B}}(\mathbf{x}_0,\mathbf{x}_\mathrm{c})\,, \\ \infty & \text{, otherwise.} \end{cases} \tag{4.67}$$

function AliasingExtendedLWFS-SBL$(\mathcal{S}, \mathcal{S}_\mathrm{h}, M, \mathbf{x}_\mathrm{c})$
  $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}} \leftarrow \infty$
  for $\mathbf{x}_0, \mathbf{x}'_0 \leftarrow \partial\mathcal{S}$ **do** ▷ (4.24), densely sampled
    if $a_S(\mathbf{x}_0) = 0$ **then** ▷ (4.55)
      continue ▷ next secondary sources
    end if
    $\Delta_{\mathbf{x}_0} \leftarrow \Delta_u |\mathbf{x}'_0|$
    $\hat{k}_{G,\mathbf{t}_0}^{\min}, \hat{k}_{G,\mathbf{t}_0}^{\max} \leftarrow$ MinMaxWavenumber$(\mathcal{S}_\mathrm{h}, \mathbf{x}_0)$
    $f \leftarrow \dfrac{c}{\Delta_{\mathbf{x}_0} \max\left(|\hat{k}_{G,\mathbf{t}_0}^{\min} - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|; |\hat{k}_{G,\mathbf{t}_0}^{\max} - \hat{k}_{S,\mathbf{t}_0}(\mathbf{x}_0)|\right)}$
    $f_M^\mathrm{B} \leftarrow \dfrac{Mc}{2\pi|(\mathbf{x}_0 - \mathbf{x}_\mathrm{c}) \times \hat{\mathbf{k}}_S(\mathbf{x}_0)|}$ ▷ (4.53)
    if $f \leq f_M^\mathrm{B}$ **then** ▷ (4.54)
      $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}} \leftarrow \min\left(f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}}; f\right)$ ▷ (4.55)
    end if
  end for
  return $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}}$
end function

Figure 4.23: Brute-force search algorithm to determine the LWFS-SBL aliasing frequency $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}}$ for an extended area $\mathcal{S}_\mathrm{h}$ given by (4.54). An example of the function MinMaxWavenumber for a circular region is given in Fig. 4.9.

Thus, the aliasing frequency for LWFS-SBL is estimated by modifying the algorithm for NFCHOA in Fig. 4.21 such that it incorporates $\mathbf{x}_\mathrm{c}$ for f_M^B. Again, the aliasing frequency $f^{\mathrm{S,LWFS-SBL}}(\mathbf{x})$ for a single position $\mathbf{x}$ is a special case of $f_{\mathcal{S}_\mathrm{h}}^{\mathrm{S,LWFS-SBL}}$ where $\mathcal{S}_\mathrm{h}$ collapses to a single point in space. The algorithm is given in Fig. 4.23.

4.4.2 Discrete Plane Wave Decomposition

It was discussed in Sec. 3.4.2, that the chosen practical realisation of LWFS-SBL requires the discretisation of the involved PWD. The continuous PWD is approximated via a summation over discrete angles $\phi_\mathrm{pw}^{(l)} = \frac{2\pi}{N_\mathrm{pw}} l$. In the frequency domain, the resulting driving signal reads

$$D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL,S}}(\mathbf{x}_0, \omega) = \frac{1}{N_\mathrm{pw}} \sum_{l=0}^{N_\mathrm{pw}-1} \bar{S}(\phi_\mathrm{pw}^{(l)}, \mathbf{x}_\mathrm{c}, \omega) D_{2.5\mathrm{D,pw}}^{\mathrm{WFS}}(\mathbf{x}_0 - \mathbf{x}_\mathrm{c} | \mathbf{n}_\mathrm{pw}^{(l)}, \omega) \,. \tag{4.68}$$

Similar to the discretised linear SSD in Sec. 4.2.2, the sampling is modelled as the multiplication of the continuous PWD $\bar{S}$ with a Dirac comb.[64] Using the Fourier series of the Dirac comb,[65] the discretised plane wave coefficients can be separated into different aliasing components indexed by ζ. The according aliasing components of the LWFS-SBL driving signal are defined as

$$D_{2.5\mathrm{D},\zeta}^{\mathrm{LWFS-SBL,S}}(\mathbf{x}_0, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \bar{S}(\phi_\mathrm{pw}, \mathbf{x}_\mathrm{c}, \omega) \mathrm{e}^{-\mathrm{j}\zeta N_\mathrm{pw} \phi_\mathrm{pw}} D_{2.5\mathrm{D,pw}}^{\mathrm{WFS}}(\mathbf{x}_0 - \mathbf{x}_\mathrm{c} | \mathbf{n}_\mathrm{pw}, \omega) \, \mathrm{d}\phi_\mathrm{pw} \,. \tag{4.69}$$

It should be emphasised, that $D_{2.5\mathrm{D},\zeta}^{\mathrm{LWFS-SBL,S}}$ describes the aliasing components w.r.t. the discrete PWD. They shall not be confused with

[64] see (4.17)
[65] Williams, loc. cit.

the aliasing components of the driving signal caused by the discrete SSD. These are indexed by η. Summing all the components over ζ leads to the continuous driving signal as given in (4.63). It is derived in Sec. A.2.2, that the aliasing components of the driving function can be high-frequency approximated via

$$D^{\mathrm{LWFS-SBL,S}}_{\mathrm{2.5D},\zeta}(\mathbf{x}_0,\omega) \approx \sqrt{\mathrm{j}\frac{\omega}{\mathrm{c}}}\,\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|}\,a_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0)\langle\hat{\mathbf{k}}_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0,\omega)|\mathbf{n}_0\rangle S^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega)\,. \tag{4.70}$$

The derivation is analogous to the approximation for the continuous case which resulted in (4.64). The sound field $S^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega)$ defines the ζ-th aliasing component of the virtual sound field originating from a discretised PWD.[66] $a_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0)$ and $\hat{\mathbf{k}}_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0,\omega)$ denote the secondary source selection criterion and the normalised local wavenumber vector according to the aliasing component $S^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega)$. The propagation direction $\hat{\mathbf{k}}_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0,\omega)$ is necessary to derive the spatial aliasing frequency from (4.27). A meaningful expression is found in Sec. A.2.2. The normalised wavenumber vector can be related to the original virtual sound field via

[66] see (A.30a)

$$\hat{\mathbf{k}}_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0,\omega) \approx \hat{\mathbf{k}}_S(\mathbf{x}^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega),\omega) \tag{4.71}$$

with $\mathbf{x}^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega)$ fulfilling the implicit equation (A.34)

$$\zeta N_{\mathrm{pw}}\frac{\mathrm{c}}{\omega} \overset{!}{=} \left\langle \mathbf{x}^{\mathrm{S}}_\zeta - \mathbf{x}_0 \middle| \mathbf{R}_{\frac{\pi}{2}}\hat{\mathbf{k}}_S(\mathbf{x}^{\mathrm{S}}_\zeta,\omega)\right\rangle\,. \tag{4.72}$$

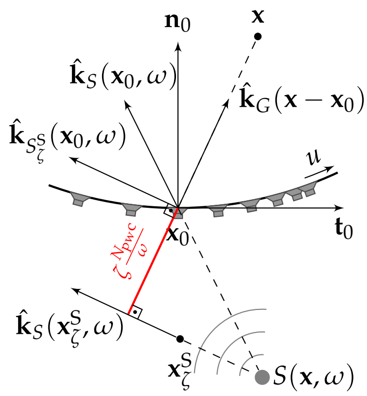


Figure 4.24: The sketch shows a synthesis scenario for LWFS-SBL with a discrete convex SSD (black arc) and discrete PWD. The grey dot illustrates an exemplary virtual point source.

$\mathbf{R}_{\frac{\pi}{2}}$ denotes a rotation matrix causing a counter-clockwise rotation of $\hat{\mathbf{k}}_S(\mathbf{x}^{\mathrm{S}}_\zeta,\omega)$ about $\pi/2$. A geometric interpretation of the condition is depicted in Fig. 4.24. The signed distance (red line) between $\mathbf{x}_0$ and the ray defined by $\mathbf{x}^{\mathrm{S}}_\zeta$ and $\hat{\mathbf{k}}_S(\mathbf{x}^{\mathrm{S}}_\zeta,\omega)$ is equal to $\zeta N_{\mathrm{pw}}\frac{\mathrm{c}}{\omega}$. It can also be seen, that the resulting $\hat{\mathbf{k}}_{S^{\mathrm{S}}_\zeta}(\mathbf{x}_0,\omega)$ differs from the propagation direction of the original virtual source field $\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)$. As the condition is implicit w.r.t. $\mathbf{x}^{\mathrm{S}}_\zeta$, the equation cannot be solved without additional knowledge about the virtual sound field. Solutions to specific virtual sound fields remain for future research. Assuming the solution to be known, the derivation is carried on by inserting $\hat{\mathbf{k}}_S(\mathbf{x}^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega),\omega)$ into (4.27):

$$\langle\,\mathbf{t}_0\,|\,\hat{\mathbf{k}}_S(\mathbf{x}^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega),\omega)\rangle + \frac{\eta\mathrm{c}}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)\,f} \overset{!}{=} \langle\,\mathbf{t}_0\,|\,\hat{\mathbf{k}}_G(\mathbf{x}-\mathbf{x}_0)\,\rangle\,. \tag{4.73}$$

It was stated in Sec. 4.2.4, that an explicit expression for the aliasing frequency in WFS can be given by (4.30), if the propagation direction of the virtual sound field is independent of f, i.e. $\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega) = \hat{\mathbf{k}}_S(\mathbf{x}_0)$. However, this assumption does not allow to solve Eq. (4.73) for f, since the position $\mathbf{x}^{\mathrm{S}}_\zeta(\mathbf{x}_0,\omega)$ remains as a function of frequency. Without knowledge of the specific virtual sound field, no analytic solution is possible for the aliasing frequency in LWFS-SBL with a discrete PWD.

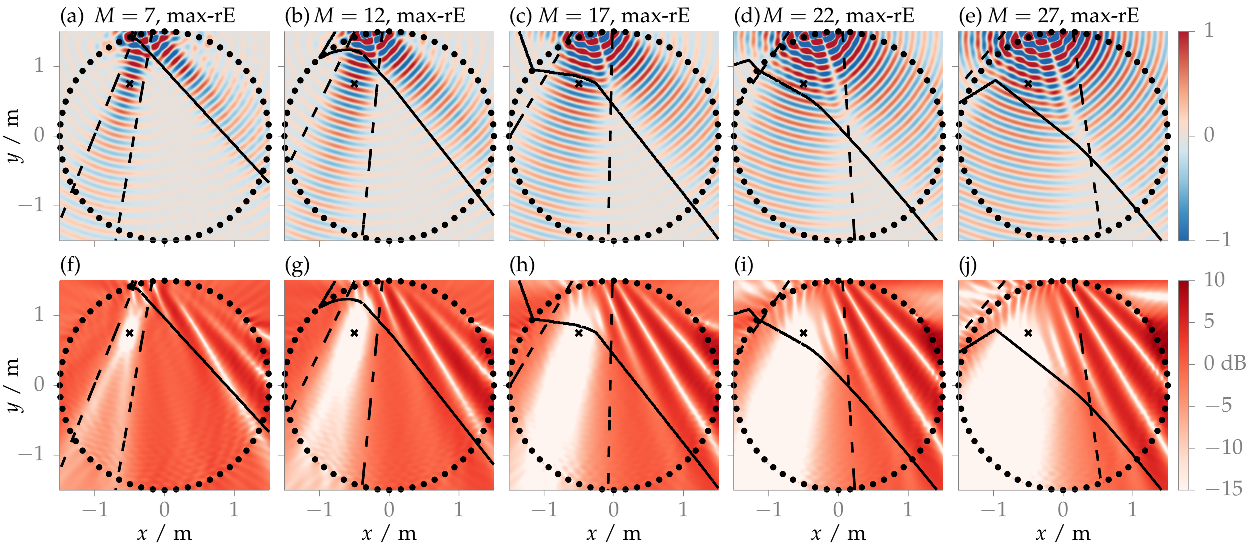


Figure 4.25: The plots (a)-(e) show the real part of a monochromatic ($f = 2$ kHz) virtual point source located at $\mathbf{x}_{\mathrm{ps}} = [0, 2.5, 0]^{\mathrm{T}}$ m synthesised with LWFS for different M. The expansion centre is set to $\mathbf{x}_{\mathrm{c}} = [-0.5, 0.75, 0]^{\mathrm{T}}$ m (cross). The plots (f)-(j) below show the error according $\hat{\varepsilon}(\mathbf{x}, \omega)$, see (4.56). For the positions $\mathbf{x}$ above the solid black lines, the predicted anti-aliasing criterion involving $f^{\mathrm{S}}_{M}(\mathbf{x})$ is violated, see (4.67). It is estimated using the algorithms in Fig. 4.13 and Fig. 4.21 for $\mathcal{S}_{\mathrm{h}} = \mathcal{C}_{\mathrm{h}}$ with $R_{\mathrm{h}} = 0$ and $\mathbf{x}_{\mathrm{h}} = \mathbf{x}$ with a coordinate frame shifted about $\mathbf{x}_{\mathrm{c}}$. For positions, which are not contained by the dashed black lines, $f^{\mathrm{B}}_{M}(\mathbf{x} - \mathbf{x}_{\mathrm{c}}) < f$, see (4.53).

4.4.3 Application and Validation

With this example, the predictions of the geometric model extended towards LWFS-SBL are compared to numeric simulations. In addition, the influence of spatial bandwidth M and the expansion centre $\mathbf{x}_{\mathrm{c}}$ on the predicted aliasing frequency is examined. As for NFCHOA, an equi-angularly sampled circular SSD with $N_0 = 56$ loudspeakers and $R = 1.5$ m radius is considered. Fig. 4.25 shows the synthesised sound fields, and the normalised error[67] for $\mathbf{x}_{\mathrm{c}} = [-0.5, 0.75, 0]^{\mathrm{T}}$ m and different bandwidths. The plots show a dominance of the SBL artefacts (dashed lines) for low M. Spatial aliasing (solid lines) increases with higher M. For the circular SSD under investigation ($N_0 = 56$), $M = 27$ is optimal for the centre position in NFCHOA. It can be seen in Fig. 4.25e/j, that $\mathbf{x}_{\mathrm{c}}$ is already corrupted by spatial aliasing for this spatial bandwidth. Obviously, the optimal M w.r.t. a trade off between spatial aliasing and SBL is a function of $\mathbf{x}_{\mathrm{c}}$.

The aliasing frequency as a function of M is shown in Fig. 4.26 for ten different positions $\mathbf{x}_{\mathrm{c}}$. The plot can also be used to find the optimal, i.e. aliasing-free, M for a given frequency. For comparison, the aliasing frequency for $M \to \infty$ corresponding to conventional WFS is also shown. A circular listening region $\mathcal{C}_{\mathrm{h}}$ with $R_{\mathrm{h}} = 8.5$ cm was chosen for the simulation to approximate the human head.[68] For all positions, the aliasing frequencies increase with decreasing M. The observable discontinuities can be explained as follows: The aliasing frequency is computed as the minimum over all active secondary sources. Thus, it is possible, that a single secondary source determines $f^{\mathrm{S}}_{\mathcal{C}_{\mathrm{h}}|M}$ for a wide range of values for M. If this source is deactivated due to the SBL, the aliasing frequency increases drastically.

Low values for M have the drawback that SBL artefacts have to be taken into account. The red-shaded area in Fig. 4.26a marks the SBL frequency $f^{\mathrm{B}}_{\mathcal{C}_{\mathrm{h}}|M}$ for a circular region $\mathcal{C}_{\mathrm{h}}$. The optimal trade-off between the two artefacts is given by the intersection points of $f^{\mathrm{B}}_{\mathcal{C}_{\mathrm{h}}|M}$

[67] see Eq. (4.56)

[68] Algazi et al. (2001). "Elevation localization and head-related transfer function analysis at low frequencies". In: *J. Acoust. Soc. Am.* 109.3, pp. 1110–1122, Sec. IV.F.

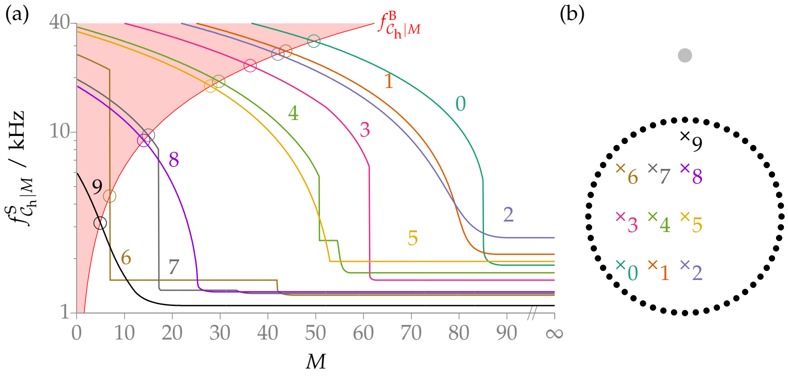


Figure 4.26: (a) shows the estimated aliasing frequency $f^{\mathrm{S}}_{\mathcal{C}_\mathrm{h}|M}$, see (4.67) as a function of M. The aliasing frequency was computed using the algorithms in Fig. 4.21 and Fig. 4.13. The parameters are $R_\mathrm{h} = 8.5$ cm, and $\mathbf{x}_\mathrm{h} = \mathbf{x}_\mathrm{c}$. The area shaded in red indicates $f^{\mathrm{S}}_{\mathcal{C}_\mathrm{h}|M} < f^{\mathrm{B}}_{\mathcal{C}_\mathrm{h}|M} = \frac{Mc}{2\pi R_\mathrm{h}}$. (b) depicts the positions $\mathbf{x}_\mathrm{h} = \mathbf{x}_\mathrm{c}$ with their index corresponding to the left plot.

and $f^{\mathrm{S}}_{\mathcal{C}_\mathrm{h}|M}$ (coloured circles). The corresponding frequency will be referred to as the maximum artefact-free frequency and will be used for the upcoming discussion: At the positions 0 to 5, the artefact-free frequency is between ≈ 9.3 (pos. 5) and ≈ 17.3 (pos. 0) times as high as the aliasing frequency of WFS ($M = \infty$). With at least 17.9 kHz (pos. 5), the frequency reaches to the end of the audible range. For the positions 6 to 9, which are closer to the virtual point source, the aliasing frequency is comparatively low: A very small value for M has to be chosen to avoid spatial aliasing. Compared to the positions discussed before, the gain w.r.t. spatial aliasing frequency is relatively small. At pos. 9, the artefact-free frequency (black circle) is about 2.9-times as high as the frequency for WFS.

4.5 Geometric Model for Local Wave Field Synthesis using Virtual Secondary Sources

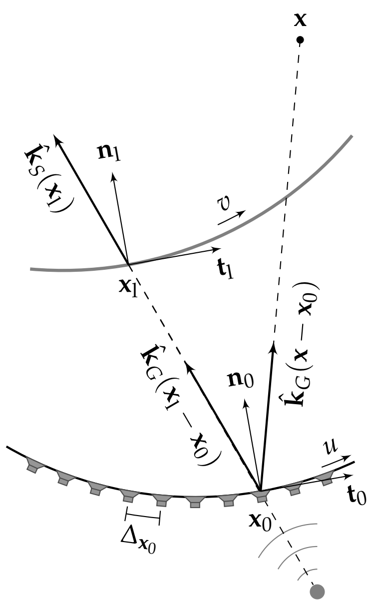


Figure 4.27: The sketch shows a synthesis scenario for a discrete convex SSD (loudspeaker symbols) and a continuous convex virtual SSD (grey line).

In the following, the geometric model for WFS will be generalised towards LWFS-VSS. As presented in Sec. 3.5, LWFS-VSS utilises focused sources as Virtual Secondary Sources (VSSs) which are distributed on the boundary of the target listening region $\partial\mathcal{S}_\mathrm{l}$. The geometrical model will be applied to LWFS-VSS with a continuous virtual SSD in Sec. 4.5.1. In a practical implementation of LWFS-VSS, the virtual SSD has to be sampled. Similar to PWD in the previous section, the resolution of this discretisation is mainly constrained by the computational effort and is not systemically limited. Its impact on the spatial aliasing will be discussed within the geometrical framework in Sec. 4.5.2.

4.5.1 Continuous Virtual Secondary Source Distribution

The 2.5D LWFS-VSS driving signal given in (3.36) states a superposition of the driving signals $D^{\mathrm{WFS}}_{\mathrm{2.5D,fs}}(\mathbf{x}_0|\mathbf{x}_\mathrm{l}, \mathbf{n}_\mathrm{l}, \omega)$ to synthesise a distinct focused source filtered by the driving signal $D^{\mathrm{WFS}}_{\mathrm{2.5D}}(\mathbf{x}_\mathrm{l}, \omega)$ for the focused source to synthesise the virtual sound field. It is rewritten as the line integral

$$D_{2.5D}^{LWFS-VSS}(\mathbf{x}_0,\omega) = \int_{v_{min}}^{v_{max}} D_{2.5D,fs}^{WFS}(\mathbf{x}_0|\mathbf{x}_l(v),\mathbf{n}_l(v),\omega) D_{2.5D}^{WFS}(\mathbf{x}_l(v),\omega)\,|\mathbf{x}_l'(v)|\,\mathrm{d}v\,. \tag{4.74}$$

Analogous to the real SSD, the boundary $\partial\mathcal{S}_l$ of the local target region is described as a curve $\mathbf{x}_l(v)$ depending on the parameter $v \in [v_{min}, v_{max}]$, see Fig. 4.27. The component-wise derivative of $\mathbf{x}_l$ w.r.t. v, the unit tangent vector, and inward pointing boundary normal vector are accordingly given by $\mathbf{x}_l'$, $\mathbf{t}_l$, and $\mathbf{n}_l$. The integral is approximated by $D_{2.5D,SPA}^{LWFS-VSS}$ using the SPA defined in Gleichung A.1. For the stationary VSS $\mathbf{x}_l^* = \mathbf{x}_l(v^*)$, the condition

$$0 \overset{!}{=} \left.\frac{\partial\Phi_S(\mathbf{x}_l(v),\omega) + \Phi_{fs}(\mathbf{x}_0 - \mathbf{x}_l(v),\omega)}{\partial v}\right|_{v=v^*} \tag{4.75}$$

has to hold. Similar to the derivation for WFS, Φ_S is the phase of the virtual sound field included in $D_{2.5D}^{WFS}$. The phase of $D_{2.5D,fs}^{WFS}$ is denoted by Φ_{fs}. Following the definition in Tab. 3.1 its phase is the negated phase of the free-field Green's function, i.e. $\Phi_{fs}(\mathbf{x}_l) = -\Phi_G(\mathbf{x}_l - \mathbf{x}_0)$. The SPA constraint for the involved normalised wavenumber vector is given by

$$\hat{\mathbf{k}}_S(\mathbf{x}_l^*,\omega) \overset{!}{=} \hat{\mathbf{k}}_G(\mathbf{x}_l^* - \mathbf{x}_0)\,. \tag{4.76}$$

Its solution $\mathbf{x}_l^* = \mathbf{x}_l^*(u) = \mathbf{x}_l(u, v^*)$ is generally a function of u due to the involved $\mathbf{x}_0(u)$. As shown in Fig. 4.27, the direction of $\mathbf{x}_l^*$ relative to $\mathbf{x}_0$ has to be aligned with the propagation direction of the virtual sound field at $\mathbf{x}_l^*$. The high-frequency approximation states, that the LWFS-VSS driving signal for $\mathbf{x}_0$ is mainly determined by the $\mathbf{x}_l^*$ fulfilling this condition. Vice versa, $\mathbf{x}_l$ is the stationary VSS for all secondary sources along the ray defined as $\mathbf{x}_0(\mathbf{x}_l,\omega) = \mathbf{x}_l - \gamma\hat{\mathbf{k}}_S(\mathbf{x}_l,\omega)$ with $\gamma > 0$. Thus, $\mathbf{x}_0$ corresponds to the intersection point of the ray and the SSD boundary $\partial\mathcal{S}$. With $\mathbf{x}_l \in \mathcal{S}$ and $\partial\mathcal{S}$ being convex, $\mathbf{x}_0$ is uniquely defined by (4.76).

The SPA of the LWFS-VSS driving signal is inserted into the SLP for the convex boundaries defined by (4.24). The result reads

$$P(\mathbf{x},\omega) \approx \int_{u_{min}}^{u_{max}} D_{2.5D,SPA}^{LWFS-VSS}(\mathbf{x}_0(u),\omega) G(\mathbf{x}-\mathbf{x}_0(u),\omega)\,|\mathbf{x}_0'(u)|\,\mathrm{d}u. \tag{4.77}$$

After incorporating spatial sampling of the SSD, the SPA condition w.r.t. u is formulated as

$$\left.\frac{\partial\Phi_S(\mathbf{x}_l^*(u),\omega) + \Phi_{fs}(\mathbf{x}_0(u) - \mathbf{x}_l^*(u),\omega) + \Phi_G(\mathbf{x}-\mathbf{x}_0(u))}{\partial u}\right|_{u=u^*} \overset{!}{=} \eta\frac{2\pi}{\Delta_u}. \tag{4.78}$$

The chain rule of differentiation, i.e. $\frac{\partial}{\partial u} = \frac{\partial}{\partial v^*}\frac{\partial v^*}{u}$, and (4.75) are used to derive the equivalent condition for the normalised wavenumber vectors as

$$\langle \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x}_l^*(u^*) - \mathbf{x}_0(u^*))\rangle + \frac{\eta c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0)\, f} \overset{!}{=} \langle\, \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x} - \mathbf{x}_0(u^*))\,\rangle\,. \tag{4.79}$$

function ALIASINGEXTENDEDLWFS-VSS($\mathcal{S}, \mathcal{S}_\mathrm{h}, \mathcal{S}_\mathrm{l}$) ▷ $\mathcal{S}$ and $\mathcal{S}_\mathrm{l}$ are assumed to be convex
  $f^{\mathrm{S,LWFS-VSS}}_{\mathcal{S}_\mathrm{h}} \leftarrow \infty$
  for $\mathbf{x}_\mathrm{l}, \mathbf{x}'_\mathrm{l} \leftarrow \partial\mathcal{S}_\mathrm{l}$ **do** ▷ (4.74), densely sampled
    if $a_S(\mathbf{x}_\mathrm{l}) = 0$ **then** ▷ (4.82)
      continue ▷ next virtual secondary source
    end if
    $\mathbf{x}_0, \mathbf{x}'_0 \leftarrow$ FINDINTERSECTION($\mathbf{x}_\mathrm{l}, \hat{\mathbf{k}}_S(\mathbf{x}_\mathrm{l}), \mathcal{S}$) ▷ intersection of ray $\mathbf{x}_\mathrm{l} - \gamma\hat{\mathbf{k}}_S(\mathbf{x}_\mathrm{l})$ and $\partial\mathcal{S}$
    $\Delta_{\mathbf{x}_0} \leftarrow \Delta_u |\mathbf{x}'_0|$
    $\hat{k}^{\min}_{G,\mathbf{t}_0}, \hat{k}^{\max}_{G,\mathbf{t}_0} \leftarrow$ MINMAXWAVENUMBER($\mathcal{S}_\mathrm{h}, \mathbf{x}_0$)
    $f \leftarrow \dfrac{c}{\Delta_{\mathbf{x}_0} \max\left(\left|\hat{k}^{\min}_{G,\mathbf{t}_0} - \hat{k}_{G,\mathbf{t}_0}(\mathbf{x}_\mathrm{l} - \mathbf{x}_0)\right|; \left|\hat{k}^{\max}_{G,\mathbf{t}_0} - \hat{k}_{G,\mathbf{t}_0}(\mathbf{x}_\mathrm{l} - \mathbf{x}_0)\right|\right)}$
    $f^{\mathrm{S,LWFS-VSS}}_{\mathcal{S}_\mathrm{h}} \leftarrow \min(f^{\mathrm{S,LWFS-VSS}}_{\mathcal{S}_\mathrm{h}}; f)$
  end for
  return $f^{\mathrm{S,LWFS-VSS}}_{\mathcal{S}_\mathrm{h}}$
end function

Figure 4.28: Generic brute-force search algorithm to determine the aliasing frequency $f^{\mathrm{S,LWFS-VSS}}_{\mathcal{S}_\mathrm{h}}$ given by (4.82). The aliasing frequency $f^{\mathrm{S,LWFS-VSS}}(\mathbf{x})$ for a single position can be computed by using a circular region $\mathcal{S}_\mathrm{h}$ with zero radius. An example of the function MINMAXWAVENUMBER for a circular region is given in Fig. 4.13.

Since $\mathbf{x}_0$ is uniquely defined for a given $\mathbf{x}^*_\mathrm{l}$ by the SPA condition (4.76) of the first integral, (4.79) is expressed in terms of $\mathbf{x}^*_\mathrm{l}$ via

$$\langle \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x}^*_\mathrm{l} - \mathbf{x}_0(\mathbf{x}^*_\mathrm{l}, \omega)) \rangle + \frac{\eta c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0(\mathbf{x}^*_\mathrm{l}, \omega))\, f} \stackrel{!}{=} \langle \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x} - \mathbf{x}_0(\mathbf{x}^*_\mathrm{l}, \omega)) \rangle. \tag{4.80}$$

Under the assumption, that the $\hat{\mathbf{k}}_S$ is not a function of frequency, $\mathbf{x}_0(\mathbf{x}^*_\mathrm{l}, \omega)$ is also independent of ω. Thus, the aliasing frequency for a pair $\mathbf{x}$ and $\mathbf{x}_\mathrm{l}$ reads

$$f^{\mathrm{S,LWFS-VSS}}(\mathbf{x}, \mathbf{x}_\mathrm{l}) = \frac{c}{\Delta_{\mathbf{x}_0}(\mathbf{x}_0(\mathbf{x}_\mathrm{l}))\left|\hat{k}_{G,\mathbf{t}_0}(\mathbf{x} - \mathbf{x}_0(\mathbf{x}_\mathrm{l})) - \hat{k}_{G,\mathbf{t}_0}(\mathbf{x}_\mathrm{l} - \mathbf{x}_0(\mathbf{x}_\mathrm{l}))\right|}\,. \tag{4.81}$$

The frequency for a single position $\mathbf{x}$ is given by

$$f^{\mathrm{S,LWFS-VSS}}(\mathbf{x}) = \min_{\mathbf{x}_0 | a_S(\mathbf{x}_\mathrm{l}) \neq 0} f^{\mathrm{S,LWFS-VSS}}(\mathbf{x}, \mathbf{x}_\mathrm{l})\,. \tag{4.82}$$

The aliasing frequency is further generalised towards extended listening areas $\mathcal{S}_\mathrm{h}$ as it was presented for WFS in Sec. 4.2.4. An algorithm to estimate the corresponding frequency $f^{\mathrm{S,LWFS-VSS}}_{\mathcal{S}_\mathrm{h}}$ is shown in Fig. 4.28. The major differences to the WFS algorithm in Fig. 4.11 are the minimisation over $\partial\mathcal{S}_\mathrm{l}$ instead of the $\partial\mathcal{S}$ and the function FINDINTERSECTION to determine $\mathbf{x}_0$ for a given $\mathbf{x}_\mathrm{l}$.

4.5.2 Discrete Virtual Secondary Source Distribution

As shown in Fig. 4.29, the virtual SSD (grey dots) is now spatially discretised. The sampling distance in the v-domain is denoted as Δ_v. The resulting distance in the Cartesian space is defined as $\Delta_{\mathbf{x}_\mathrm{l}}(\mathbf{x}_\mathrm{l}) := \Delta_v \mathbf{x}'_\mathrm{l}$. The integration over the continuous virtual SSD in (4.74) is approximated as summation over an equidistant grid w.r.t. v. As for the LWFS-SBL in Sec. 4.4.2, this discretisation results in additional aliasing components which shall not be confused with the components stemming from the SSD sampling. The components w.r.t. v

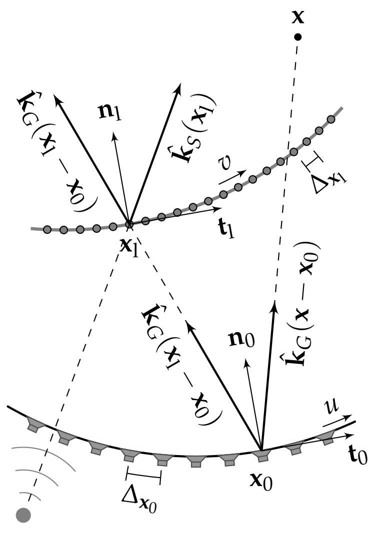


Figure 4.29: The sketch shows a synthesis scenario for a discrete convex SSD (loudspeaker symbols) and a discrete convex virtual SSD (grey dots).

read

$$D^{\mathrm{LWFS-VSS,S}}_{\mathrm{2.5D},\zeta}(\mathbf{x}_0,\omega) = \int_{v_{\min}}^{v_{\max}} D^{\mathrm{WFS}}_{\mathrm{2.5D,fs}}(\mathbf{x}_0|\mathbf{x}_\mathrm{l}(v),\mathbf{n}_\mathrm{l}(v),\omega) D^{\mathrm{WFS}}_{\mathrm{2.5D}}(\mathbf{x}_\mathrm{l}(v),\omega) \mathrm{e}^{-\mathrm{j}\zeta\frac{2\pi}{\Delta_v}v} \,|\mathbf{x}'_\mathrm{l}(v)|\,\mathrm{d}v\,. \tag{4.83}$$

For the SPA of this integral, the additional phase term $\mathrm{j}\zeta\frac{2\pi}{\Delta_v}v$ has to be considered. The condition for the continuous virtual SSD in (4.76) is modified to

$$\langle \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_S(\mathbf{x}_\mathrm{l}^*,\omega)\rangle + \frac{\zeta c}{\Delta_{\mathbf{x}_\mathrm{l}}(\mathbf{x}_\mathrm{l}^*)\, f} \stackrel{!}{=} \langle\, \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{l}^* - \mathbf{x}_0)\,\rangle\,. \tag{4.84}$$

The ray equation, which defines the secondary source position $\mathbf{x}_0$ for which $\mathbf{x}_\mathrm{l}^*$ is all stationary VSS is given analogously to (4.28) by

$$\mathbf{x}_0 = \mathbf{x}_\mathrm{l}^* + \gamma\, \mathbf{R}_\mathrm{l}^* \begin{bmatrix} \hat{k}_{S,\mathbf{t}_\mathrm{l}}(\mathbf{x}_\mathrm{l}^*,\omega) + \frac{\eta c}{\Delta_{\mathbf{x}_\mathrm{l}}(\mathbf{x}_\mathrm{l}^*,\omega))f} \\ \sqrt{1-\left(\hat{k}_{S,\mathbf{t}_\mathrm{l}}(\mathbf{x}_\mathrm{l}^*,\omega) + \frac{\eta c}{\Delta_{\mathbf{x}_\mathrm{l}}(\mathbf{x}_\mathrm{l}^*)f}\right)^2} \\ 0 \end{bmatrix}, \qquad 0 \le \gamma \le \infty \tag{4.85}$$

The rotation matrix $\mathbf{R}_\mathrm{l} = [\mathbf{t}_\mathrm{l}\;\; \mathbf{n}_\mathrm{l}\;\; \mathbf{u}_z]$ contains the listed vectors as its columns. It can be seen from this equation, that even with the assumption, that $\hat{\mathbf{k}}_S$ is frequency independent, $\mathbf{x}_0$ still remains a function of frequency for $\zeta \neq 0$. Inserting, $\mathbf{x}_0$ into the SPA condition for the discrete SSD in (4.79) will lead to an implicit relation w.r.t. f. Similar to the discussions in Sec. 4.5.2 on the discrete PWD for LWFS-SBL, the resulting equation cannot be solved without further assumptions. However, further insight into the mathematical structure of the resulting aliasing frequency is gained, if (4.79) and (4.85) are rearranged to

$$\zeta\frac{c}{f} \stackrel{!}{=} \Delta_{\mathbf{x}_\mathrm{l}}(\mathbf{x}_\mathrm{l}^*)\langle\, \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{l}^* - \mathbf{x}_0) - \hat{\mathbf{k}}_S(\mathbf{x}_\mathrm{l}^*,\omega)\,\rangle, \text{ and} \tag{4.86a}$$

$$\eta\frac{c}{f} \stackrel{!}{=} \Delta_{\mathbf{x}_0}(\mathbf{x}_0^*)\langle\, \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x} - \mathbf{x}_0^*) - \hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{l}^* - \mathbf{x}_0^*)\,\rangle\,. \tag{4.86b}$$

The right-hand sides of both equations are integer multiples of the wavelength $\lambda = c/f$. Vice versa, λ is a common divisor of both right-hand sides. The aliasing frequency for a triple $\mathbf{x}_0$, $\mathbf{x}_\mathrm{l}$, and $\mathbf{x}$ as the smallest f fulfilling both equations is determined by the Greatest Common Divisor (GCD) of the right-hand sides. The numerical computation of the GCD is challenging as the involved quantities are in general real-valued. Furthermore, the aliasing frequency for an extended area $\mathcal{S}_\mathrm{h}$ cannot be derived as beforehand, where simply the extremal values of $\langle \mathbf{t}_0 \,|\, \hat{\mathbf{k}}_G(\mathbf{x}-\mathbf{x}_0)\rangle$ had to be considered. Distinct combinations of $\mathbf{x}_0$, $\mathbf{x}_\mathrm{l}$, and $\mathbf{x} \in \mathcal{S}_\mathrm{h}$ might share a larger GCD and, thus, a lower aliasing frequency than for the extremal values.

4.5.3 Application and Validation

A circular SSD with $N_0 = 56$ equi-angularly spaced loudspeakers and a radius of $R = 1.5$ m is used for the numerical simulations.

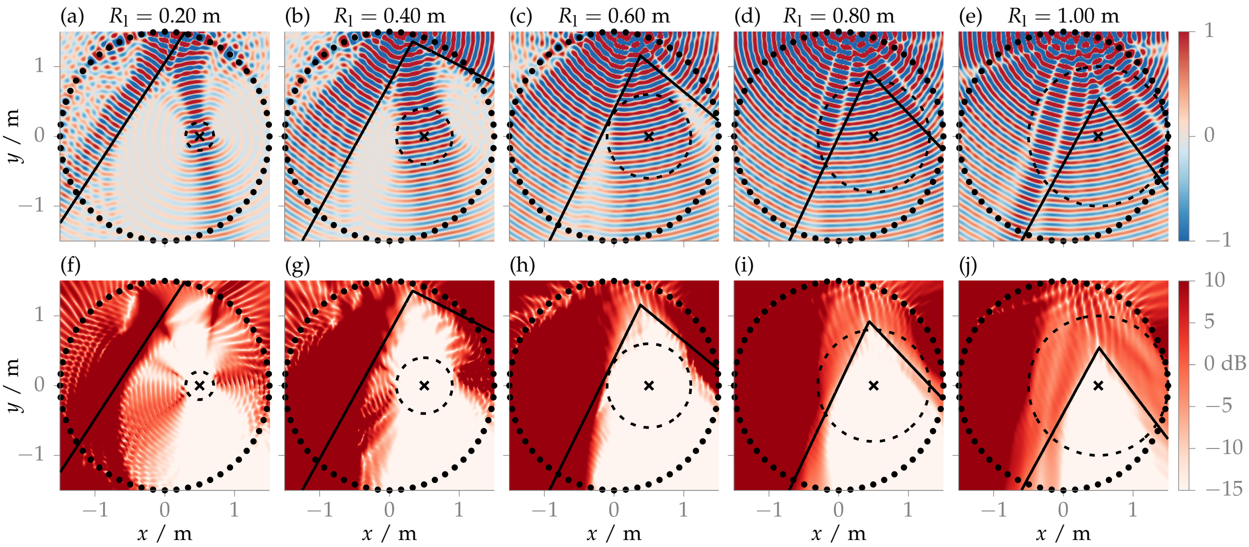


Figure 4.30: The plots (a)-(e) show the real part of the synthesised sound field for a monochromatic ($f = 2.5$ kHz) virtual point source located at $\mathbf{x}_{ps} = [0, 2.5, 0]^T$ m. It is synthesised with LWFS-VSS using a circular virtual SSD of different radii R_l (dashed) and centre around $\mathbf{x}_l = [0.5, 0, 0]^T$ m (cross). The plots (f)-(j) below show the according error $\hat{\varepsilon}(\mathbf{x}, \omega)$, see (4.37). For the positions $\mathbf{x}$ above the solid black lines, the predicted anti-aliasing criterion involving $f^{S,LWFS-VSS}(\mathbf{x})$ is violated, see (4.82). It is estimated using the algorithms in Fig. 4.13 and Fig. 4.28 for $\mathcal{S}_h = \mathcal{C}_h$ with $R_h = 0$ and $\mathbf{x}_h = \mathbf{x}$.

A continuous virtual SSD on a circular boundary $\mathcal{C}_l$ with radius R_l and centre $\mathbf{x}_l$ is assumed. The connection between these parameters on the aliasing frequency is investigated. The error defined in (4.37) measures the influence of the spatial sampling on the synthesis accuracy. It does not incorporate diffraction and near-field artefacts of the focused sound sources. These phenomena are not covered by the geometric model. Sound field simulations for a varying radius R_l are shown in Fig. 4.30: As expected, the aliasing artefacts become stronger, the larger R_l gets: Especially for the scenarios in Fig. 4.30d/e, synthesis error inside $\mathcal{C}_l$ (dashed circle) are clearly visible. The geometric model correctly predicts the increased errors shown in Fig. 4.30i/j.

The aliasing frequency as a function of R_l is shown Fig. 4.31 for ten different positions $\mathbf{x}_l$. A circular listening region $\mathcal{C}_h$ with $R_h = 8.5$ cm and $\mathbf{x}_h = \mathbf{x}_l$ was chosen for the simulation to approximate the head[69] of a human listener located inside $\mathcal{C}_l$. For the different positions, the evaluated range of R_l is individually chosen in order to prevent an intersection of $\mathcal{C}_l$ with the SSD. In general, an increase of the aliasing frequency with decreasing R_l can be observed. The red-shaded area in Fig. 4.31 marks radii $\mathbf{x}_l$ which are smaller than the head radius. The coloured circles mark the aliasing frequencies for $\mathbf{x}_h = \mathbf{x}_l$: At the positions 0-6, the aliasing frequency is between ≈ 17 (pos. 5) and ≈ 33 (pos. 0) times as high as the aliasing frequency of WFS. Despite of position 9, the aliasing frequencies reach to the end of the audible range with a least 18.0 kHz (pos. 8). A trend similar to LWFS-SBL[70] can be observed, where the gain w.r.t. the aliasing frequency decreases for positions which are closer to the virtual point source. At pos. 9, the artefact-free frequency (black circle) is about 5.4-times as high as the frequency for WFS. In general, LWFS-VSS reaches higher frequencies in comparison to LWFS-SBL. This can be explained by comparing the two mechanisms which are used in the two methods to achieve LSFS: In LWFS-VSS, the size of target area is explicitly

[69] Ibid., Sec. IV.F.

[70] see Sec. 4.4.3

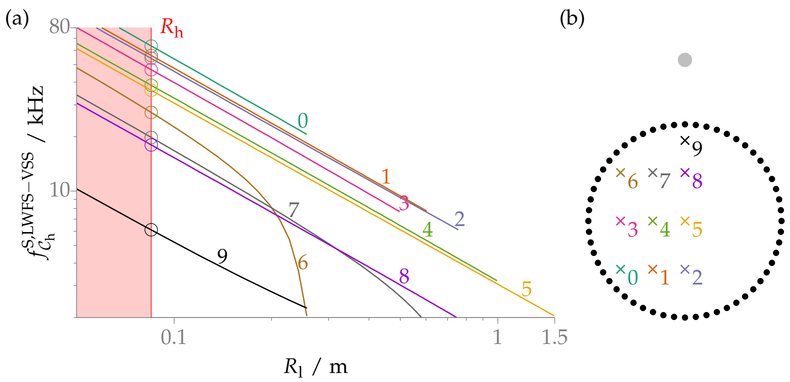


Figure 4.31: (a) shows the estimated aliasing frequency $f_{\mathcal{C}_h}^{S,LWFS-VSS}$, see (4.82) as a function of the radius R_l. The aliasing frequency was computed using the algorithms in Fig. 4.13 and Fig. 4.28. The parameters are $R_h = 8.5$ cm, and $\mathbf{x}_h = \mathbf{x}_l$. The area shaded in red indicates $R_l < R_h$. (b) depicts the positions $\mathbf{x}_h = \mathbf{x}_l$ with their index corresponding to the left plot.

defined by $R_l = R_h$ and it is independent of the temporal frequency f. For LWFS-SBL, the spatial bandwidth M leads to an frequency-dependent size of the target region, which may be unnecessarily larger than R_h at a distinct frequency. This potentially activates more secondary sources contributing additional spatial aliasing to the listening region. A frequency-dependent choice of $M(f)$ to establish a constant radius or a secondary source specific $M(\mathbf{x}_0)$ to individually avoid aliasing might be suitable extension to LWFS-SBL. A similar approach was taken by Firtha[71] with a position-dependent low-pass filtering of the driving signals to locally avoid spatial aliasing.

[71] Firtha, loc. cit.

4.6 Summary

This chapter presented a ray-based approximation of SFS allowing the description of spatial aliasing artefacts. From this, anti-aliasing conditions for different scenarios and synthesis methods were derived. In general, the estimated aliased regions agree with the spatial structure of the aliasing artefacts observed in the numerically simulated sound fields. Other than the traditional approach[72] and prior approaches, the resulting criteria incorporate the specific virtual sound field and geometric parameters such as the listening position or the SSD geometry. It was further shown that the derived aliasing frequencies are lower bounded by the half-wavelength criterion, if the mentioned parameters are unknown or arbitrary. The criterion demands the distance between adjacent secondary sources to be half the wavelength to avoid aliasing frequency. As this bound is also the outcome of the traditional model, the presented geometric model includes the traditional approach as a special case. With the presented framework, a generic tool to predict the spatial aliasing frequency without the actual simulation of the synthesised sound field emerged.

[72] see Sec. 4.1

For WFS, the model correctly predicts the influence of non-uniform discretisation of the SSD. As an outstanding capability, the geometric framework allowed to derive an analytic solution to optimise the sampling schemes in order to maximise the aliasing frequency for a given SSD contour.

The model was further generalised to incorporate NFCHOA and the two LWFS approaches. Here, the geometric model was used to find optimal parametrisations of the SFS techniques. For NFCHOA and LWFS-SBL, the frequency f_M^{B} was additionally introduced to quantify the artefacts arising from the SBL. As a trade-off between SBL and spatial aliasing, the optimal spatial bandwidth M in NF-CHOA was analytically derived within the framework. The result agrees with prior treatises in the literature. In general, the model correctly predicted that LSFS is capable of increasing the spatial aliasing frequency. LSFS becomes less effective the closer the target region is located to the active loudspeakers.

As the geometric model is based upon high-frequency approximations, some phenomena cannot be predicted. The impact of different modal windows in NFCHOA and LWFS-SBL and the near-field artefacts of the focused sources in LWFS-VSS are not covered by the model. The additional discretisation of the involved PWD and the virtual SSD in LWFS-SBL and LWFS-VSS was discussed. The concatenation of two sampling processes did however not allow for analytic expressions for the aliasing frequency.

5 Spatial Perception: Azimuthal Localisation

In Sec. 1.1, SFS was motivated by the assumption that a physically perfect synthesis of a desired sound field guarantees perceptual authenticity. The required physical accuracy is, however, very unlikely to be achieved in practical SFS systems. This became especially clear in the last chapter, where spatial aliasing was revisited as a major artefact in SFS. Hence, the impact of the physical artefacts on the perception of the synthesised sound fields has to be investigated. As LSFS is capable of enhancing the synthesis accuracy around the listener's position, the potential reduction of perceptual impairments constitutes an interesting research item.

Rumsey et al.[1] investigated in how far spatial fidelity contributes to the overall audio quality in home cinema surround sound reproduction: With approximately thirty percent, its contribution is significant and has to be regarded in investigations on sound quality. As an important aspect of spatial fidelity in SFS, the auditory localisation caused by the synthesised sound field is supposed to align with the position of the virtual sound sources relative to the listener. To clarify the difference between physics and perception, the terms sound event and auditory event are commonly used.[2] In the literature,[3] auditory localisation is split into the three dimensions of a head centred spherical coordinate system: The localisation of the azimuthal/horizontal direction is mainly determined by the binaural cues Interaural Level Difference (ILD) and Interaural Time Difference (ITD). The displacement of the auditory event is only influenced by the ITD in the fine structure of signal components below 1.6 kHz.[4] Due to the diffraction of the head and torso, the ILD becomes more prominent at higher frequencies. Above the mentioned frequency threshold, it is the dominant mechanism together with the ITD between the signal envelopes. The reliability of the individual cues is attributed to the Interaural Coherence (IC).[5] For the frontal direction, humans are able to notice a change of approximately one degree in sound source azimuth[6] a.k.a. the Minimum Audible Angle (MAA).[7] Spectral cues are used in the median plane to localise the direction of elevated sources. As most practical setups for SFS restrict the synthesis to the horizontal plane, i.e. 2.5D synthesis, elevated virtual sources are not supposed to be synthesised. Although artefacts

[1] Rumsey et al. (2005). "On the relative importance of spatial and timbral fiedelities in judgements of degraded multichannel audio quality". In: *J. Acoust. Soc. Am.* 118.2, pp. 968–976.

[2] Blauert (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press, pp. 2–3.

[3] Ibid., Cha. 2.

[4] Ibid., p. 173.

[5] Faller and Merimaa (2004). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence". In: *J. Acoust. Soc. Am.* 116.5, pp. 3075–89.

[6] Blauert, op. cit., Tab. 2.1.

[7] Mills (1958). "On the Minimum Audible Angle". In: *J. Acoust. Soc. Am.* 30.4, pp. 237–246.

of SFS might result in a perceivable elevation of the corresponding auditory event, azimuthal localisation is expected to be the dominating phenomenon. Human distance perception is less accurate and hence less critical for SFS than directional localisation. Zahorik et al.[8] reported in their research summary, that many studies ascertain an underestimation of the distance to far sources and an overestimation for close sources. Moreover, human distance perception heavily depends on environmental parameters determining cues like the direct-to-reverberant energy ratio.[9] As the most critical dimension, the azimuthal localisation is selected as the target of investigation in order to assess the spatial fidelity in SFS.

In order to compare different studies with each other, consistent definitions of the terms describing the localisation accuracy are necessary. The MAA is measured within a two-alternative forced choice test, where subjects had to indicate, whether an auditory event occurred to the left or to the right of a reference event.[10] It is defined as the difference in angle between the corresponding sound events, where 75 percent of the subjects' responses were correct. Blauert defines the MAA under the term localisation blur.[11] The term blur is however used by various authors[12,13] within the context of a source-identification task: The subjects had to indicate the azimuth of the auditory event using a suitable pointing method. Although differently defined, the blur generally quantifies the spread in azimuth around the average location (bias) as a kind of standard deviation. This ambiguous usage is partly[14] caused by the fact that the MAA and standard deviation in the source-identification task were reported as close to equal.[15] MAA and standard deviation are however not equivalent as they are defined w.r.t. different experimental setups. Later research by Hartmann and Rakerd[16] involving decision theory moreover showed, that the MAA is interpreted incorrectly leading to an even larger deviation between the two definitions of the localisation blur. Within this chapter, the term blur is omitted completely. The accuracy in the conducted source-identification experiment will be quantified by the bias and the standard deviation.

Various localisation experiments for WFS conducted at Delft University[17,18] showed a standard deviation below 1.5° for linear arrays with a loudspeaker distance of 11 cm, which is slightly increased by approximately 0.5° for 22 cm. For the two distances, an MAA below 1° and 2° was measured by Start.[19] Wierstorf et al.[20,21] showed in their studies, that even with a large distance between the loudspeakers (≈ 67 cm) the localisation bias in WFS is below 5° for all investigated listening positions. Since WFS does not actively avoid spatial aliasing, this type of artefact does not seem to have a major influence on the localisation accuracy. Wierstorf et al. argued that for the tested conditions, the ITD cues below the aliasing frequency resemble the ones of the target sound field. Thus, humans focus on the unimpaired cues leading to the desired spatial perception. Moreover, low-frequency ITD cues exhibit a dominant role in the localisation.[22,23] As an alternative explanation, WFS accurately syn-

[8] Zahorik et al. (2005). "Auditory Distance Perception in Humans: A Summary of Past and Present Research". In: *Acta Acustica united with Acustica* 91.3, pp. 409–420, Sec. 2.

[9] Ibid., Sec. 3.1.2.

[10] Hartmann (1983). "Localization of sound in rooms". In: *J. Acoust. Soc. Am.* 74.5, pp. 1380–1391, Appendix C.

[11] Blauert, op. cit., Sec. 2.1.

[12] Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Sec. 5.1.

[13] Verheijen (1997). "Sound Reproduction by Wave Field Synthesis". PhD thesis. Delft University of Technology, Sec. 6.2.

[14] Ibid., p. 144.

[15] Hartmann, loc. cit.

[16] Hartmann and Rakerd (1989). "On the minimum audible angle—A decision theory approach". In: *J. Acoust. Soc. Am.* 85.5, pp. 2031–2041.

[17] Vogel (1993). "Application of wave field synthesis in room acoustics". PhD thesis. Delft University of Technology, Sec. 4.6.

[18] Verheijen, op. cit., Sec. 6.2.

[19] Start (1997). "Direct Sound Enhancement by Wave Field Synthesis". PhD thesis. Delft University of Technology, Sec. 6.5.

[20] Wierstorf, loc. cit.

[21] Wierstorf et al. (2017). "Assessing localization accuracy in sound field synthesis". In: *J. Acoust. Soc. Am.* 141.2, pp. 1111–1119.

[22] Wightman and Kistler (1992). "The dominant role of low-frequency interaural time differences in sound localization". In: *J. Acoust. Soc. Am.* 91.3, pp. 1648–1661.

[23] Macpherson (June 2013). "Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation". In: *Proc. of Meetings on Acoustics* 19.1.

thesises the first wave front which is then followed by the additional aliasing contributions.[24] This might trigger the precedence effect,[25] where the first wave front dominates the perceived direction. Ahrens also mentions, that the time intervals between the different wave fronts are so small that they should trigger summing localisation,[26] instead.

[24] Ahrens (2012). *Analytic Methods of Sound Field Synthesis*. T-Labs Series in Telecommunication Services. Berlin Heidelberg, Germany: Springer-Verlag, Sec. 4.4.4.2.

[25] Litovsky et al. (1999). "The precedence effect". In: *J. Acoust. Soc. Am.* 106.4, pp. 1633–1654.

[26] Blauert, op. cit., Sec. 3.1.

[27] see Sec. 3.3.1

[28] Wierstorf et al., op. cit.

For NFCHOA with a rectangular modal window,[27] different loudspeaker distances and spatial bandwidths M were investigated by Wierstorf et al.:[28] For listening positions near the array centre, localisation was very close to transparent. For off-centre positions, however, a significant impairment of localisation and even source splitting, i.e. the perception of multiple auditory events, were observed. The artefacts became stronger the lower the modal order M. This can be explained by the limited region of high synthesis accuracy around the centre caused by the SBL in NFCHOA. Its size increases with M. The findings are supplemented by experiments for HOA without near-field-compensation. Frank et al.[29] investigated first, third, and fifth-order Ambisonics for an irregular loudspeaker setup in an echoic environment: The localisation accuracy generally increased with M and is highest for the central listening position. The max-$\mathbf{r}_E$ window outperformed the rectangular one for the off-centre listening position. Frank[30] further showed for an off-centre listening position in a circular 8-channel array, that source splitting occurs in third-order Ambisonics with a rectangular window. Stitt[31] concluded in his study on first and third-order Ambisonics for a circular array of 2.2 m radius, that the localisation error is largely influenced by the spatial relation between the off-centre position and the target source direction. The error is maximum, if the listener is shifted away from the centre into a direction which is perpendicular to the source direction.

[29] Frank et al. (Nov. 2008). "Localization experiments using different 2D Ambisonics decoders". In: *Proc. of 25th Tonmeistertagung – VDT Internation Convention*. Leipzig, Germany.

[30] Frank (June 2013a). "Phantom Sources using Multiple Loudspeakers in the Horizontal Plane". PhD thesis. Graz, Austria: Institute of Electronic Music and Acoustics.

[31] Stitt (June 2015). "Ambisonics and Higher-Order Ambisonics for Off-Centre Listeners: Evaluation of Perceived and Predicted Image Direction". PhD thesis. Belfast, UK: Queen's University Belfast, Sec. 5.5.

This chapter extends the work of Wierstorf et al. towards LWFS-SBL and LWFS-VSS and compares their localisation attributes with the conventional SFS techniques. As the key research hypothesis, the potential improvement of the localisation accuracy is investigated. The chapter is structured as follows: The evaluation method for azimuthal localisation is revisited in Sec. 5.1. Its validation experiment including the results is presented in Sec. 5.2. The main study comparing the localisation of the SFS approaches is comprised in Sec. 5.3. A summary is given afterwards.

5.1 Evaluation Method[32]

[32] This section has been published in a modified version in Winter et al. (May 2017b). "Improvement of the reporting method for closed-loop human localization experiments". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany.

The realisation of listening tests with different physically existing loudspeaker setups including a varying, possibly very high number of loudspeakers and different geometries is infeasible in practice. Furthermore the localisation shall not be influenced by the properties of the listening room and the directional characteristics of real loudspeakers, as the focus of this work lies on the artefacts introduced by the synthesis method. In order to investigate different listening positions, listeners would have to be positioned in a reproducible manner. The awareness of being moved may bias the listener not least because of visual effects. Moreover, randomised test designs are difficult to implement taking the mentioned aspects into account. It is thus sensible to use dynamic binaural synthesis to simulate SFS under free-field conditions over headphones. The evaluation method is very similar to the one of Wierstorf,[33] which was validated[34] and successfully used in localisation experiments for SFS. In the following subsections the details of the evaluation method and its difference to the one of Wierstorf are presented. If not stated otherwise, the descriptions apply for both setups. After the evaluation method has been described, potential drawbacks and sources of error introduced by the apparatus are discussed.

[33] Wierstorf, op. cit., Ch. 4.

[34] Wierstorf et al. (Sept. 2012). "Perception and evaluation of sound fields". In: *59th Open Seminar on Acoustics*. Boszkowo, Poland.

5.1.1 Dynamic Binaural Synthesis

Fundamental Principle: A Head-Related Impulse Response (HRIR) represents the acoustic free-field transmission path from a sound source to the outer ears. Its Fourier transform[35] is commonly referred to as the Head-Related Transfer Function (HRTF). Both terms will be used equivalently. The characteristics of the HRTFs are exploited by the human auditory system in order to deduce spatial information. HRTFs differ amongst individuals due to variations w.r.t. anatomy. They depend on the orientation and position of the listener relative to the source. The head-above-torso orientation has to be considered as a additional degree of freedom.[36] It mostly effects spectral cues and its influence on ITD and ILD is below the audible threshold.[37,38] Thus, for the investigation of azimuthal localisation it is sufficient to assume that the HRTF $H_{\{L,R\}}(\mathbf{x}_s^\dagger, \omega)$ only depends on the source position $\mathbf{x}_s^\dagger$ in the xy-plane. Fig. 5.1 shows, that the source position is defined in a shifted and rotated coordinate frame relative to the listener's position and head orientation ϕ_h. Dynamic binaural synthesis utilising HRTFs is a common approach to auralise a virtual sound source. The desired ear signals are synthesised by filtering the dry source signal $\tilde{S}(\omega)$ emitted by a sound source with the according impulse response. The generated ear signals are played back over headphones. The orientation of the listener's head is tracked simultaneously and the impulse responses are switched according to the apparent source position. The transmission path from the headphones to the ear drums has to be compensated using a suitable Headphone Compensation Filter (HPCF). Since dynamic binaural synthesis simulates the ear signals corresponding to a given acoustic environment it may also be used to simulate a whole loudspeaker array driven by SFS. An example is illustrated in Fig. 5.2: A virtual point source emitting the signal $\tilde{S}(\omega)$ is synthesised by a circular loudspeaker array using a suitable SFS method. Each loudspeaker is treated as a individual source which is supposed to be simulated over headphones. The individual driving functions $D(\mathbf{x}_0, \omega)$ are filtered by the HRTF $H_{\{L,R\}}(\mathbf{x}_0^\dagger, \omega)$ corresponding to the apparent

ir_generic.m

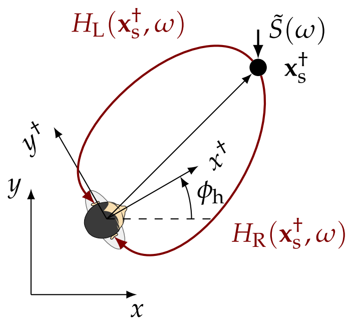


Figure 5.1: A sound source at $\mathbf{x}_s$ is simulated by convolving its signal $\tilde{S}(\omega)$ with $H_{\{L,R\}}(\mathbf{x}_s^\dagger, \omega)$ corresponding to the apparent source position $\mathbf{x}_s^\dagger$.

[35] see Eq. (2.2)

[36] Brinkmann et al. (Sept. 2014). "Audibility of head-above-torso orientation in head-related transfer functions". In: *Proc. of Forum Acusticum*. Kraków, Poland.

[37] Popko (Sept. 2013). "Zur Hörbarkeit und Interpolation von Kopf-über-Torso-Orientierungen bei Aufnahmen mit einem Kopf-und-Torso-Simulator". Bachelor's Thesis. Technische Universität Berlin, Sec. 4.3.

[38] Brinkmann et al. (Aug. 2015). "Audibility and Interpolation of Head-Above-Torso Orientation in Binaural Technology". In: *IEEE J. Sel. Topics Signal Process.* 9.5, pp. 931–942, Sec. III.A.2.

loudspeaker positions $\mathbf{x}_0^\dagger$ relative to the listener. The superposition of all loudspeakers results in the ear signals

$$B_{\{L,R\}}(\mathbf{x},\phi_h,\omega) = \tilde{S}(\omega) \underbrace{\sum_{\mathbf{x}_0} D(\mathbf{x}_0,\omega) H_{\{L,R\}}(\mathbf{x}_0^\dagger,\omega)}_{H^{\mathrm{SFS}}_{\{L,R\}}(\mathbf{x},\phi_h,\omega)} \,. \tag{5.1}$$

The result of the summation may be interpreted as the Binaural Transfer Functions (BTFs) $H^{\mathrm{SFS}}_{\{L,R\}}(\mathbf{x},\phi_h,\omega)$ of the loudspeaker array for a given SFS method (including its parameters), listening position and head orientation. The BTFs may be computed offline for a given resolution w.r.t. ϕ_h and fed directly into a binaural renderer for auralisation.

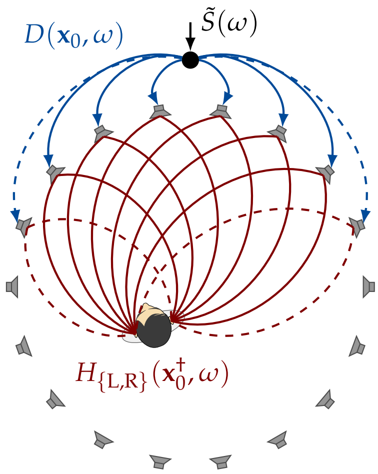


Figure 5.2: Binaural Synthesis of a circular loudspeaker array which is driven by SFS in order to reproduce a virtual point source emitting $\tilde{S}(\omega)$ (black dot).

Implementation: The basic principle of binaural synthesis as a tool for dynamically generating the necessary stimuli is illustrated in Fig. 5.3. The head tracker provides the horizontal orientation of the listener's head which is fed into the convolution core of the system. Based on the current head orientation the corresponding impulse response is selected from the current BTF dataset. All BTF datasets had a resolution of one degree. The input signal which is supposed to be emitted by the virtual sound source is convolved with selected impulse response in a block-wise manner. Each block is 1024 samples long. Possible changes in head orientation are handled by convolving the current signal block separately with the old and new impulse response and cosine-shaped cross-fading the results within the duration of one block. The SoundScape Renderer[39] was utilised as the convolution core. The input signal for the SoundScape Renderer was provided by Pure Data[40] which allows to root the dry source signal into different convolution instances of the SoundScape Renderer. Each instance contains the BTFs corresponding to a specific condition, i.e. SFS method and listening position. This means that the system was able to instantaneously switch between different conditions without having to restart the playback of the dry audio signal. All components operated at a sampling frequency of 44.1 kHz.

[39] Geier and Spors (Nov. 2012). "Spatial Audio with the SoundScape Renderer". In: *Proc. of 27th Tonmeistertagung – VDT Internation Convention*. Cologne, Germany.

[40] Puckette (Sept. 1996). "Pure Data: another integrated computer music environment". In: *Proc. of the Second Intercollege Computer Music Concerts*. Tachikawa, Japan, pp. 37–41.

Dry Source Signal: A Gaussian white noise pulse train of 100 s length was used as the signal emitted by the sound source. Each pulse had a duration of 700 ms followed by a pause of 300 ms. The noise signals of each pulse were statistically independent. A cosine-shaped fade-in/fade-out of 20 ms length was applied at the begin and the end of each pulse. The signal was bandpass filtered with a fourth order Butterworth filter between 125 Hz and 20000 Hz. In the experiment, the signal was played back in a loop and was filtered by the current BTF for binaural reproduction.

Head-Related Transfer Functions: The HRTF dataset used to create the desired BTFs was measured in an anechoic chamber with a Head and Torso Simulator (HATS). A sound source was placed in the

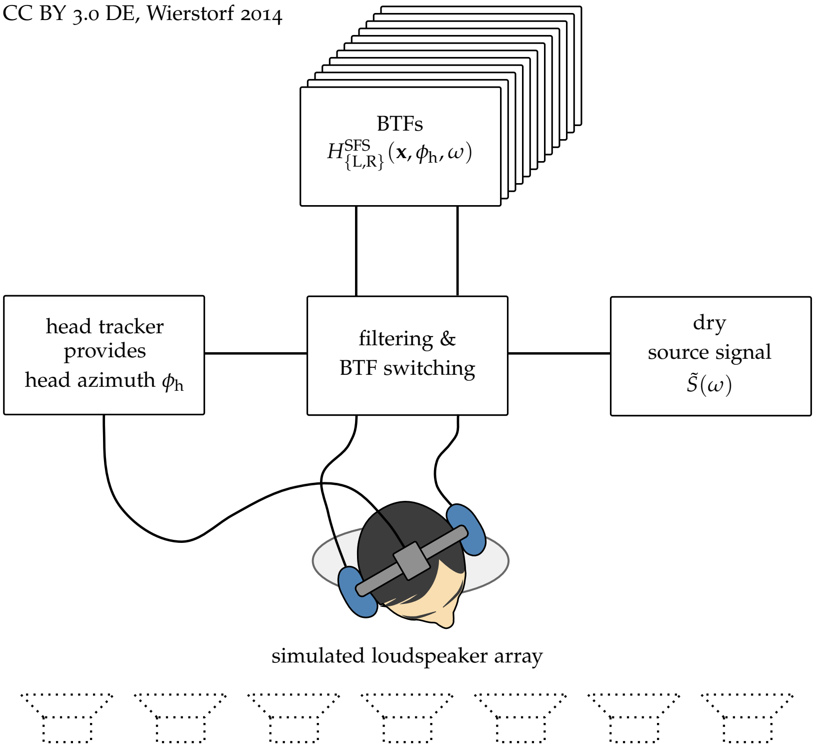


Figure 5.3: Basic system components for dynamic Binaural Synthesis. The illustration was originally published in Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Fig. 4.1

horizontal plane (at height of the ears) with a distance of 3 m and an azimuth varying from 0° to 359° with 1° resolution. Details about the measurement procedure and involved equipment can be found in the original publication.[41] For non-measured source directions, the HRTFs were linearly interpolated using the two nearest measured HRTFs. For distances smaller or larger than the measured 3 m the delay and the amplitude of the HRTFs were adjusted according to the speed of sound and the free-field distance attenuation, respectively. The HPCF for the AKG 601 headphones provided with the dataset is applied to the BTFs.

Minor modifications to the interpolation strategies were later applied for the main study in Sec. 5.3. They were not regarded in the validation study as no additional modifications to the original study should be made. Instead of the 3 m distance, the HRTFs for 2 m were chosen from the same dataset. The 2 m distance is closer to the distances between the listener and the loudspeakers occurring in the used evaluation setup. For non-measured source directions, a linear interpolation scheme was separately applied to the magnitude and phase spectrum of the HRTFs. It results in a better preservation of the impulse shape in the interpolated HRIRs. This required a low-frequency correction of the dataset.[42]

[41] Wierstorf et al. (May 2011). "A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances". In: *Proc. of 130th Aud. Eng. Soc. Conv.* London, UK.

[42] Erbes et al. (May 2017). "Free Database of Low Frequency Corrected Head-Related Transfer Functions and Headphone Compensation Filters". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany.

5.1.2 Location and Hardware

The listening test by Wierstorf et al. took place in a 83 m^3 acoustically damped listening room (room Calypso in the Telefunken building of TU Berlin, Germany). The listeners sat on a heavy rotatable chair,

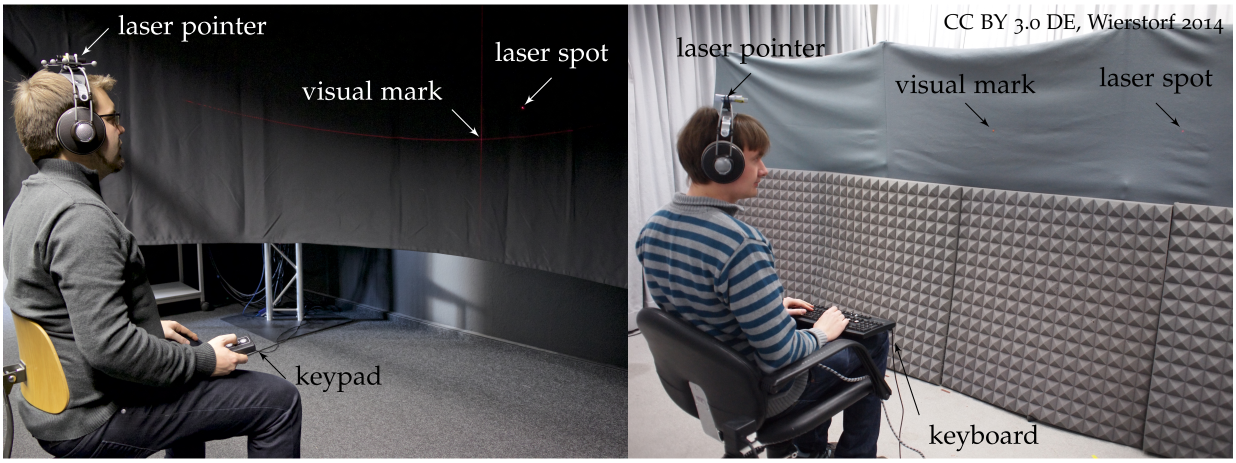


Figure 5.4: Listener during the localisation experiment in the current study (left) and for the investigations of Wierstorf et al. (right). The rooms were dark during the experiment. The right picture was originally published in Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Fig. 4.3

1.5 m in front of a straight curtain. They wore open headphones (AKG K601) with an attached head tracker (Polhemus Fastrak). The head tracker had an update rate of 120 Hz, but due to further data processing the effective update rate was 60 Hz. Its measured tracking accuracy is around 1°. As shown in Fig. 5.4 (right), the listeners had a keyboard for entering the response. In a separate room, a computer equipped with a multichannel sound card including D/A converters (RME Hammerfall DSP MADI) played back all sounds. The signals travelled through a headphone amplifier (Behringer Powerplay Pro-XL HA 4700) and an analogue cable to the headphones in the listening room, a distance of approximately 5 m.

The present listening tests were conducted in a 86 m^3 acoustically damped room (Audio laboratory at the Institute of Communications Engineering, University of Rostock, Germany). The listeners sat on a rotatable chair and were surrounded by a circular curtain with a radius of approximately 1.5 m. They wore open headphones (AKG K601) with six optical markers attached to it, which form a trackable rigid body. The head tracking was achieved with an optical tracking system using eight infra-red cameras (NaturalPoint OptiTrack). The tracking system has an update rate of 120 Hz. The listeners had a keypad in their hands for entering the response.[43] In a separate room, a computer equipped with a stereo sound card (Focusrite Scarlett 2i2, 1st Generation) was used for audio playback. The signals travelled through an analogue cable of approximately 6 m length to the head phones inside the listening room.

[43] see Fig. 5.4 (left)

5.1.3 Procedure

Various conditions are presented successively to listeners via the headphones using the technique for dynamic binaural synthesis described in Sec. 5.1.1. The participants are instructed to determine the horizontal direction of the perceived auditory event, while the vertical position should be ignored. They are explicitly instructed to select the dominant event, if multiple auditory events are perceived. A pointing method similar to the one of Makous and Middlebrooks[44]

[44] Makous and Middlebrooks (1990). "Two-dimensional sound localization by human listeners". In: *J. Acoust. Soc. Am.* 87.5, pp. 2188–2200.

was used, where the listeners were asked to point into the direction using the laser pointer while the sound event is present. The laser pointer was mounted on the headphone. This has the advantage that the listener is directly facing the source, a region in which the MAA is the smallest.[45] The curtain served as a projection surface for the laser. If the listeners were sure to point into the correct direction, they pressed a key on the input device. The localisation result was calculated as the arithmetic mean of 10 values obtained from the head tracker. For the respective update rate, this corresponds to a time of 167 ms (Wierstorf et al.) and 83 ms (current). After the key press, the next condition started instantaneously. In an a-priori calibration phase, the listener was indicated to point towards a given visual mark on the curtain. Wierstorf et al. pasted a small permanent mark on the curtain. In current study, a steady laser cross was projected onto the curtain and switched off after the calibration stage. The room was darkened after calibration.

[45] Mills, op. cit.

5.1.4 Discussion on Potential Sources of Error

Dynamic binaural synthesis itself cannot be regarded as a transparent reproduction method: Even with individual HRTFs including individual HPCFs it is likely to be distinguishable from the acoustic scenario it simulates, especially for broadband noise signals.[46] Moreover, the measurement effort to acquire the individual HRTFs is considerably large. While non-individual HRTFs decrease this effort, it adds additional artefacts to the binaural synthesis: Anthropomorphic differences, e.g. the shape of the pinna, between the listener and the HATS used for the measurements are likely to cause deviations in the magnitude spectrum of the HRTFs. Same holds for the non-individual HPCFs which are only available for the HATS. The neglected dependency of the HRTF on the head-above-torso orientation potentially causes additional spectral deviations.[47] As human localisation in the median plane relies on spectral cues, such distortions may add elevation to the perceived direction of the auditory event. As the distance between ears of the HATS might also deviate from the one of the listener, the ITD cues are distorted and may cause artefacts in the azimuthal localisation for lateral sound sources. As the listener has the task to turn to the perceived direction of the auditory event, the ITD should be close to zero, if the sound source is finally in front of listener's head. This is independent of individual or non-individual HRTFs. However, if multiple sources as in SFS are superimposed, it is unknown how this distorted ITDs effect the azimuthal localisation.

[46] Brinkmann et al. (2017). "On the authenticity of individual dynamic binaural synthesis". In: *J. Acoust. Soc. Am.* 142.4, pp. 1784–1795, Sec. IV.

[47] Brinkmann et al., loc. cit.

An additional source of error is the mounting of the pointing device on the listener's head, which is illustrated in Fig. 5.5. It has been already discussed for the original study,[48] that the relative location of the laser pointer on the headphones might vary among the listeners (and sessions). This can be caused by e.g. undesired contact, switching on/off the pointer, or changing the batteries. Moreover,

[48] Wierstorf et al., "Perception and evaluation of sound fields".

the position of the headphones is different each time it is mounted on the listener's head. Consequently, the orientation of the pointing device and the listener's median plane do not necessarily align. During the calibration phase, the listener has to point to the visual calibration mark using the laser pointer. The resulting orientation is calibrated as zero degree head azimuth. If now the binaurally simulated scenario is perceived directly in front, i.e. in the median plane, the listener is forced to turn the head in order to align the laser beam with this direction. Simultaneously, the perceived event is not in front any more due to the head motion. Hence, a bias is potentially introduced to the localisation result for each session.

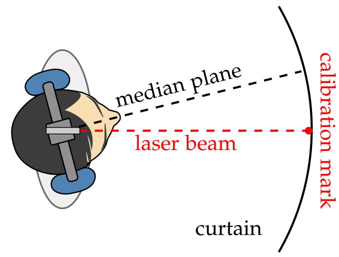


Figure 5.5: Sketch of an exaggeratedly misplaced laser pointer leading to a deviation of the listener's median plane and the direction of the laser pointer.

Being aware of the particular drawbacks of using binaural synthesis for investigating azimuthal localisation in SFS, a particular question arises: How small can the effects of the possible artefacts in SFS on the azimuthal localisation be, so that they are still detectable with the presented evaluation method? In other words, how accurate are listeners able to localise in the horizontal plane using the presented method? Different studies were conducted, that compared the perceived direction of a real loudspeaker and its binaural synthesis. As long as head tracking was applied, the localisation errors were usually in the range of 1° to 5°.[49,50,51] One reason for the varying results for the localisation performance found in the literature is the fact that such experiments are critical regarding the utilised pointing method. In the study of Wierstorf et al.,[52] the localisation accuracy was around 1° for real as for the simulated loudspeaker, but only if the loudspeakers were not positioned more than 30° to the side. For loudspeaker positioned further to the side, an undershoot in the reported angle occurred in both cases: the test subjects tend to localise lateral sources closer to the front. It is suspected by the author, that the finite projection plane, i.e. the straight curtain, was the main reason for this observation. As it is replaced by a circular curtain in the current study, a comparison between both studies is reasonable.

49 Makous and Middlebrooks, op. cit.

50 Bronkhorst (1995). "Localization of real and virtual sound sources". In: *J. Acoust. Soc. Am.* 98.5, pp. 2542–53.

51 Seeber (2003). "Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode". PhD thesis. Technischen Universität München.

52 Wierstorf et al., op. cit.

5.2 Validation of Evaluation Method[53]

53 This section has been published in a modified version in Winter et al., op. cit.

This section describes the details of the localisation experiment to validate the evaluation method presented in Sec. 5.1. As already stated, the method is very similar to the one used by Wierstorf et al..[54] Besides some minor modification such as the used hardware, the shape of the curtain for the pointing method is considered to be a major difference between the two experiments, see Fig. 5.6. Therefore, it is obligatory to re-validate the new apparatus and compare the results of both studies. Note, that the aim in the original study was to compare human localisation of real sound sources with their respective simulation via anechoic binaural synthesis or binaural synthesis with room reflections. The focus is now shifted towards localisation in anechoic binaural synthesis only and how the reporting method can be improved.

54 Wierstorf et al., op. cit.

5.2.1 Conditions

The experiment contained 11 unique listening conditions[55] where a single sound source emitting the source signal described in Sec. 5.1.1 was simulated. The positions of the sound sources are indicated in Fig. 5.6. Each listener had to pass each condition six times leading to 66 trials in total. The order of presentation was randomised with respect to repetitions and condition, while the first 11 trials where meant for training and contained each unique condition exactly once. In the experiment of Wierstorf et al.,[56] the remaining 55 trials were split into two sessions with 22 and 33 trials containing each unique condition exactly two and three times, respectively.

[55] Due to the additional presentation techniques, i.e. binaural room simulation and real loudspeakers, the experiment by Wierstorf et al. originally contained 33 different conditions.

[56] Ibid.

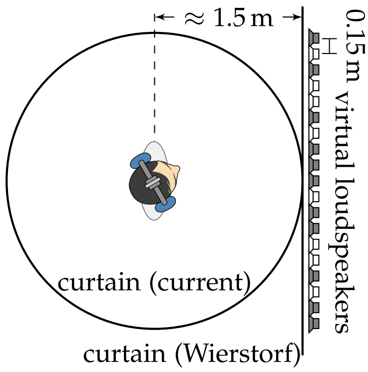


Figure 5.6: Sketch of experimental setup and the linear array consisting of the 19 virtual sound sources (loudspeaker symbols) with a spacing of 0.15 m. The eleven source positions used in the experiment are shaded dark.

5.2.2 Participants

11 listeners were recruited for both experiments. The age of the participants ranged from 21 to 33 years for the study of Wierstorf et al. and from 26 to 60 in the current study with a respective average of 28.6 and 38 years. 4 and 2 of the listeners had prior experience with listening tests.

5.2.3 Methods for Data Analysis

This section presents the statistical methods used to evaluate and compare the acquired data. As a result of each listening experiment the four-dimensional dataset $\phi^b_{ls}(\phi_c)$ describes the perceived azimuths. The index l corresponds to one of the L listeners. The listening condition and respective ground truth source azimuth are denoted by c and ϕ_c, respectively. The total number of conditions is $C = 11$. As each condition is presented B times to each listener, these repetitions are indicated by b. In the study of Wierstorf et al.,[57] the experiment was split into two sessions, which is considered via the index s. It is assumed, that all samples $\phi^b_{ls}(\phi_c)$ are statistically independent due to the randomisation of the presentation order. The signed localisation error is given as

[57] Ibid.

$$\Delta^b_{ls}(\phi_c) = \phi^b_{ls}(\phi_c) - \phi_c\,. \tag{5.2}$$

As already discussed in Sec. 5.1.4, two systemic artefacts in the evaluation method were identified: First, the discrepancy between the direction of the laser point and the listeners' median plane introduces a direction independent bias to the results. It varies among the listeners (and sessions). Second, a localisation undershoot was reported by Wierstorf et al.[58,59] and may also be present in the current study. The undershoot would manifest itself as a systematic dependency between the localisation error and the azimuth of the loudspeaker. A linear mixed-effects model[60] is fit to both datasets separately in order to investigate both artefacts. The term mixed hereby refers to a combination of fixed and random effects. The

[58] Ibid., Sec. 3.

[59] Wierstorf, op. cit., Sec. 4.2.8.

[60] Henderson (1975). "Best Linear Unbiased Estimation and Prediction under a Selection Model". In: *Biometrics* 31.2, pp. 423–447, Eq. (1).

model specification reads

$$\Delta^b_{ls}(\phi_c) = \beta_0 + \gamma_{0ls} + (\beta_1 + \gamma_{1ls})\phi_c + \epsilon_{lcbs} \tag{5.3a}$$

$$\begin{bmatrix} \gamma_{0ls} \\ \gamma_{1ls} \end{bmatrix} \underset{ls}{\overset{\text{i.i.d.}}{\sim}} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right) \tag{5.3b}$$

$$\epsilon_{lcbs} \underset{lcbs}{\overset{\text{i.i.d.}}{\sim}} \mathcal{N}(0, \sigma) \tag{5.3c}$$

The model assumes a fixed linear dependency described by the slope β_1. If there is a localisation undershoot present in the data, this parameter would be significantly smaller than zero. β_0 denotes a fixed intercept. The Best Linear Unbiased Estimates (BLUEs)[61,62,63] of the fixed effects are denoted as $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. The significance of both effects will be tested via a one-sample, two-tailed t-test with $\nu = CLB - 2$ degrees of the freedom.[64,65] To incorporate differences of intercept and slope among listeners and sessions, the random effects γ_{0ls} and γ_{1ls} are introduced. They are normally distributed with the according variances σ_0^2 and σ_1^2 and the covariance σ_{01}. The Best Linear Unbiased Predictions (BLUPs)[66,67,68] of the random effects are denoted as $\hat{\gamma}_{0ls}$ and $\hat{\gamma}_{1ls}$. The significance of the random effects can be tested via a Likelihood-ratio test[69] between a null model omitting these random effects and the model defined by (5.3). A significant result indicates a strong variability among the listeners, which does not allow to pool the data along this dimension without further treatment. The normally distributed residual error is denoted as ϵ_{lcbs}. With the BLUEs and BLUPs of the effects it is possible to compensate the artefacts of the evaluation method. The corrected localisation azimuth and signed localisation error read

$$\tilde{\phi}^b_{ls}(\phi_c) = \frac{\phi^b_{ls}(\phi_c) - \hat{\beta}_0 - \hat{\gamma}_{0ls}}{(1 + \hat{\beta}_1 + \hat{\gamma}_{1ls})} \text{ and} \tag{5.4a}$$

$$\tilde{\Delta}^b_{ls}(\phi_c) = \tilde{\phi}^b_{ls}(\phi_c) - \phi_c \text{ , respectively.} \tag{5.4b}$$

Note, that the compensation is done separately for each listener and session. The remaining standard deviation of the corrected localisation error can be interpreted as the combined localisation uncertainty of the evaluation method and the listeners. An F-test[70] with $\nu_1 = \nu_2 = CLB - 1$ degrees of freedom is conducted to compare the standard deviations of both experiments.

5.2.4 Results and Discussion[71]

It turned out during data analysis, that the standard deviation of the localisation error for one listener in each study was approximately twice as high compared to the maximum among the other participants. These participants were excluded from the analysis resulting into $L = 10$ subjects per study. The signed localisation error together with its corrected counterpart are shown in Fig. 5.7: For the experiment by Wierstorf et al.,[72] the regression revealed a significant dependency between the localisation error and the ground truth azimuth (see top left, black line): A significant[73] slope of $\hat{\beta}_1 \approx -0.047$

[61] Ibid., Eq. (5).

[62] McLean et al. (Feb. 1991). "A Unified Approach to Mixed Linear Models". In: *The American Statistician* 45.1, p. 54, Eq. (2).

[63] Robinson (Feb. 1991). "That BLUP is a Good Thing: The Estimation of Random Effects". In: *Statistical Science* 6.1, pp. 15–32, Eq. (1.2).

[64] Howell (2010). *Statistical Methods for Psychology*. 7th ed. Wadsworth, Cengage Learning, Sec. 9.11.

[65] Bortz and Schuster (2010). *Statistik für Human-und Sozialwissenschaftler*. 7th ed. Springer, Sec. 11.2.2.

[66] Henderson, loc. cit.

[67] McLean et al., loc. cit.

[68] Robinson, loc. cit.

[69] Hox (2010). *Multilevel Analysis: Techniques and Applications*. New York, USA: Routledge, Sec. 3.2.2.

[70] Bortz and Schuster, op. cit., Sec. 8.6.1.

[71] The raw data is published as Winter et al. (2017a). *Improvement of reporting method for closed-loop human localization experiments – Data*. DOI: 10.5281/zenodo.245826.

[72] Wierstorf et al., op. cit.

[73] $p \approx 0.003$, two-tailed t-test

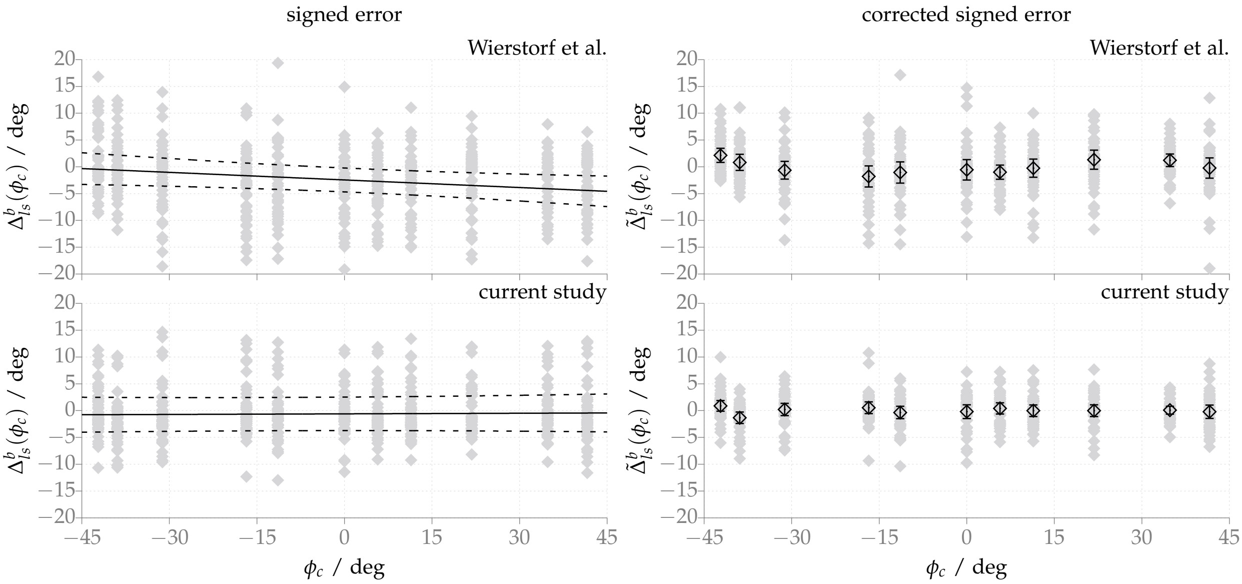


Figure 5.7: The top and the bottom row show the results for the study of Wierstorf et al. and the current study, respectively. The left column shows the signed localisation error $\Delta^b_{ls}(\phi_c)$ as a function of the ground truth source azimuth (grey). The solid black line illustrates the relation $\hat{\beta}_0 + \hat{\beta}_1\phi_c$ with the estimated intercept $\hat{\beta}_0$ and slope $\hat{\beta}_1$. The dashed lines indicate the 99%-confidence interval of the estimate. The corrected error $\tilde{\Delta}^b_{ls}(\phi_c)$ is shown in the right column (grey). The sample mean per condition together with the 99%-confidence interval of the according population mean is plotted black.

is observed. This can be interpreted as the localisation undershoot since the listeners tend to localise towards the middle, i.e. 0°. Surprisingly, a significant[74] intercept of $\hat{\beta}_0 \approx -2.447°$ is also present. For the current experiment, no significant[75] fixed effects were found. This can be observed in Fig. 5.7 (bottom left), as the confidence interval of the regression line includes 0°. For both experiments, highly significant[76] random effects could be found. The corrected localisation error is shown on the right side of Fig. 5.7. The overall standard deviations for Wierstorf et al.[77] and the current study are 4.5° and 2.9°, respectively. The former is also significantly higher[78] than the latter.

The main reason for the absence of the localisation undershoot in the current study is most probably the circular shape of the curtain establishing a close to rotationally invariant projection plane for the pointing method. As depicted in Fig. 5.4, the ends of the straight curtain in the study of Wierstorf et al. define a clearly visible limit of projection plane. Even in a dark room these limits are observable due to the change of the reflection pattern of the laser pointer between the curtain and the adjacent wall. Being aware of these limits might have forced the participants to localise towards the centre of the curtain. A reason for the decrease in standard deviation between the two studies might be the increased update rate of head tracker.[79] As a constant number of values have been captured from the head tracker for averaging, the listeners had to keep their head still for a shorter time frame.

Both studies revealed significant random effects on the localisation error among listeners (and sessions). In order to meaningfully combine the localisation results of individual listeners, the effects had to be compensated for each listener as it was done by (5.4). In the subsequent experiments, the azimuthal localisation for SFS techniques is supposed to be evaluated using the presented appara-

[74] $p \approx 0.005$, two-tailed t-test
[75] $p > 0.6$, two-tailed t-test
[76] $p \lll 0.0001$, Likelihood-ratio Test
[77] Ibid.
[78] $p \lll 0.0001$, F-test
[79] see Sec. 5.1.3

tus. Here, the random effects of the binaural synthesis add up to possible artefacts introduced by the SFS methods. While the latter are part of the investigation, the random effects have to be corrected. Additional to the SFS conditions, single sound sources similar to the 11 listening conditions used within this study have to be presented to the listener within each session. The same model as of (5.3) is fitted to these calibration conditions. The listener/session specific intercept and slope is compensated in localisation results for the SFS conditions.

5.3 Comparison of (Local) Sound Field Synthesis Methods

After the (re)-validation of the dynamic binaural synthesis as a measurement tool, the different techniques for (L)SFS are finally compared with respect to their azimuthal localisation. The investigation focuses on the following aspects: (i) In order to conduct a fair comparison between the (L)SFS methods, the impact of their parametrisation on the localisation accuracy is to be investigated. (ii) LWFS is compared to conventional SFS methods focusing on the question, whether the extended approaches lead to a better azimuthal localisation than the non-local methods. (iii) It is evaluated, whether a transparent azimuthal localisation can be achieved. For this, the localisation accuracy for the individual SFS method should be indistinguishable from the one for the reference condition.

5.3.1 Conditions

As the reference/calibration condition, a binaurally simulated point source positioned at $[0, 2.5, 0]^{\mathrm{T}}$ m with the listener at the coordinates' origin was used, see Fig. 5.8. The point source emits the dry source signal described in Sec. 5.1.1. A binaurally simulated, circular array of 56 equiangularly spaced loudspeakers centred at the coordinates' origin with a radius of 1.5 m was employed to synthesise this point source. The setup was chosen to have maximum comparability with the experiments of Wierstorf.[80] It also correspond to an existing loudspeaker array at TU Berlin, Germany. The array is driven by WFS, NFCHOA, LWFS-SBL, and LWFS-VSS with different parametrisations. The listeners were positioned at ten different listening positions and are initially oriented along the positive-y direction. The order of presentation was randomised w.r.t. repetitions and conditions.

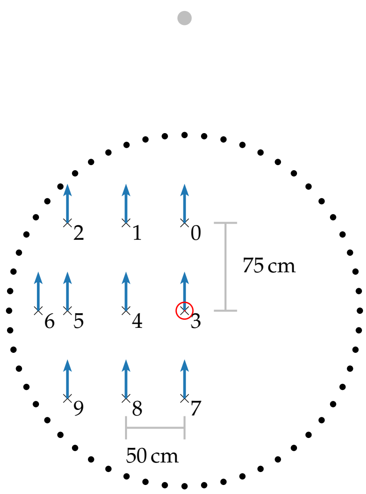


Figure 5.8: A virtual point source (grey dot) is synthesised by a circular array of 56 loudspeakers (black dots). The investigated listening position with the initial listener orientation are plotted as black crosses and blue arrows, respectively. The according digit serves as an identifier. The position for the reference/calibration condition is indicated by the red circle.

[80] Wierstorf, op. cit., Sec. 5.1.

The study was split into two experiments: As already mentioned in the introduction of this chapter,[81] the human localisation performance in NFCHOA heavily depends on modal bandwidth M and the modal window $\mathring{w}_m^M$. In the first experiment,[82] NFCHOA using a rectangular and the max-$\mathbf{r}_E$ window with $M = 27$, 13, and 6 are investigated. It was stated by Ahrens,[83] that WFS can be regarded as a high-frequency approximation of NFCHOA for very high modal

[81] see Ch. 5

[82] The data of this experiment was acquired in coorperation with Paul Frenkel as part of his Bachelor thesis, see Frenkel (Apr. 2018). “Human Localisation in Near-Field-Compensated Higher-Order Ambisonics”. Bachelor's Thesis. University of Rostock.

[83] Ahrens, op. cit.

bandwidths. The rectangular window with $M = 300$ is additionally considered in order to check whether both methods lead to similar results. Together with the calibration condition which was presented ten times, each listeners had to pass 80 stimuli in total.

The second experiment focused on the LWFS approaches and their parametrisations. For LWFS-SBL, the modal bandwidths 27 and 3 combined with a rectangular and max-$\mathbf{r}_E$ window were investigated. A larger number of plane waves $N_{\mathrm{pw}} = 1024$ is used for the discrete PWD to avoid additional spatial aliasing. The expansion centre $\mathbf{x}_{\mathrm{c}}$ was set to the position of the listener. For LWFS-VSS, a quasi-continuous[84] circular virtual SSD was used. Its centre $\mathbf{x}_{\mathrm{l}}$ was located at the listening position. Three different radii R_{l} (15, 30, and 45 cm) were investigated. The reference position $\mathbf{x}_{\mathrm{ref}}$ was set to $\mathbf{x}_{\mathrm{l}}$. In addition, WFS and NFCHOA with a rectangular window of $M = 27$ were added. The reference position $\mathbf{x}_{\mathrm{ref}}$ in WFS is fixed to the coordinates' origin. The listeners had to pass 100 stimuli including 10 times the reference/calibration condition.

[84] the term refers to a discrete virtual SSD with $N_{\mathrm{fs}} = 1024$, which does not lead to additional spatial aliasing artefacts

For each stimuli, a fixed offset was added to the head tracking data resulting in a circular shift of the corresponding BTF in head azimuth. The offset was pseudo-randomly picked for each listener and each condition from a discrete uniform distribution between $\pm 30^\circ$ degrees with 5° step size. A similar approach was taken by Wierstorf[85] to ensure a broader distribution of the sound events. Moreover, the sudden shift between consecutive conditions helps the listener to recognise, that the next stimulus is presented.

[85] Wierstorf, op. cit., Sec. 5.1.1.

To avoid loudness differences as an additional cue among the conditions, a loudness model[86,87] was used to adjust the loudness of all conditions to the calibration condition. The implementation of the model is part of the GENESIS loudness toolbox.[88] For the loudness estimation, the dry source signal was filtered by the BTF for the initial listening orientation, i.e. the positive-y direction. The estimated loudness was averaged across both ears. The difference between the condition and the reference was then compensated for the whole BTF set. After the compensation no severe loudness differences between the conditions were noticed during informal listening tests. Moreover, the randomisation of the order of presentation and the random offset added to the head tracking allows the assumption, that listeners were not able to use loudness as an additional cue.

[86] *Procedure for the Computation of Loudness of Steady Sounds* (2007). Standard ANSI S3.4-2007. New York, USA: American National Standards Institute.

[87] Moore et al. (1997). "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness". In: *J. Aud. Eng. Soc.* 45.4, pp. 224–240.

[88] GENESIS (Jan. 2010). *Loudness Toolbox 1.0*.

The current study is augmented by selected results from the experiments of Wierstorf.[89] Among other parametrisations, he investigated the same synthesis scenario for WFS and NFCHOA with $M = 28$. The collected data[90] will be analysed with the same statistical methods as the current study, if applicable.

[89] Wierstorf, op. cit., Sec. 5.1.

[90] available under Wierstorf (June 2016). *Listening test results for sound field synthesis localization experiment*. DOI: 10.5281/zenodo.55439.

5.3.2 Methods for Data Analysis

The results are of similar nature as the ones of the validation experiment in Sec. 5.2. The data is contained in the four-dimensional dataset $\phi^{b}_{lm}(\mathbf{x})$. m denotes one of the synthesis methods including the SFS approaches and the calibration condition m_c. Again, b and l describe the repetition and the listener, respectively. The listening position is given by $\mathbf{x}$. The number of repetitions B_m, listeners L_m and positions X_m differ across the methods. The ground truth azimuth $\phi_{lm}^{\mathrm{gt},b}(\mathbf{x})$ is the sum of the virtual point source azimuth relative to the listening position and the random offset added to tracking data. Using the localisation results for the calibration condition, the mixed effects model

$$\phi_{lm_c}^{b}(\mathbf{0}) = \beta_0 + \gamma_{0l} + (\beta_1 + \gamma_{1l})\phi_{lm_c}^{\mathrm{gt},b}(\mathbf{0}) + \epsilon_{lcb} \tag{5.5a}$$

$$\begin{bmatrix}\gamma_{0l}\\ \gamma_{1l}\end{bmatrix} \overset{i.i.d.}{\underset{l}{\sim}} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\sigma_0^2 & \sigma_{01}\\ \sigma_{01} & \sigma_1^2\end{bmatrix}\right) \tag{5.5b}$$

$$\epsilon_{lcb} \overset{i.i.d.}{\underset{lcb}{\sim}} \mathcal{N}(0,\sigma) \tag{5.5c}$$

is fit for each experiment, separately. The estimated fixed effects $\hat{\beta}_0$ and $\hat{\beta}_1$ and the predicted random effects $\hat{\gamma}_{0l}$ and $\hat{\gamma}_{1l}$ are used to correct the localisation results. The corrected signed error reads

$$\tilde{\Delta}_{lm}^{b}(\mathbf{x}) = \frac{\phi_{lm}^{b}(\mathbf{x}) - \hat{\beta}_0 - \hat{\gamma}_{0l}}{\hat{\beta}_1 + \hat{\gamma}_{1l}} - \phi_{lm}^{\mathrm{gt},b}(\mathbf{x}) = \tilde{\phi}_{lm}^{b}(\mathbf{x}) - \phi_{lm}^{\mathrm{gt},b}(\mathbf{x})\,. \tag{5.6}$$

whereby $\tilde{\phi}_{lm}^{b}(\mathbf{x})$ denotes the corrected localisation azimuth. For the results of Wierstorf,[91] the raw azimuth measurements are not available since the stored data was already shifted about the random offset. The calibration condition does always have the same ground-truth $\phi_{lm_c}^{\mathrm{gt},b}(\mathbf{0})$, which does not allow to estimate the slopes β_1 and γ_{1l}. For this case, the calibration is restricted to the intercept.

[91] Ibid.

For the descriptive statistics, the mean signed localisation error

$$\bar{\Delta}_{\cdot m}^{\cdot}(\mathbf{x}) = \frac{1}{L_m B_m}\sum_{b,l}\tilde{\Delta}_{lm}^{b}(\mathbf{x}) \tag{5.7}$$

is introduced, which is an unbiased estimator for the true localisation bias $\mu_m(\mathbf{x})$ of an SFS method m at the position $\mathbf{x}$. The according symmetric confidence interval for $\mu_m(\mathbf{x})$ with a confidence level of $(1-\alpha) = 0.99$ is computed via the t-distribution and $L_m B_m - 1$ degrees of freedom.[92] To quantify the overall localisation performance for an SFS method, the Root-Mean-Square Error (RMSE)

[92] Howell, op. cit., Sec. 7.3.

$$\mathrm{RMSE}_m = \sqrt{\mathrm{MSE}_m} = \sqrt{\frac{1}{L_m B_m X_m}\sum_{b,l,\mathbf{x}}(\tilde{\Delta}_{lm}^{b}(\mathbf{x}))^2} \tag{5.8a}$$

$$= \sqrt{\frac{1}{X_m}\sum_{\mathbf{x}}(\bar{\Delta}_{\cdot m}^{\cdot}(\mathbf{x}))^2 + \frac{1}{L_m B_m X_m}\sum_{b,l,\mathbf{x}}(\tilde{\Delta}_{lm}^{b}(\mathbf{x}) - \bar{\Delta}_{\cdot m}^{\cdot}(\mathbf{x}))^2} \tag{5.8b}$$

is used. The Mean-Square Error (MSE) is an unbiased estimator of $\sigma_m^2 + \tau_m^2$ with $\tau_m^2 := \frac{1}{X_m}\sum_{\mathbf{x}}\mu_m^2(\mathbf{x})$. Thus, it incorporates the systematic deviations a.k.a. the bias for each listening position via τ_m, and also stochastic variations via the standard deviation σ_m. In order to test for the transparency of an SFS method, the RMSE of m is compared to the RMSE of the calibration condition m_c via a one-tailed F-test.[93] It uses the ratio $\frac{\mathrm{MSE}_m}{\mathrm{MSE}_{m_c}}$ as its test statistic. The null hypothesis H_0 states equal standard deviations i.e. $\sigma_m^2 = \sigma_{m_c}^2$ and zero bias i.e. $\tau_m^2 = 0$. It is hereby assumed, that the localisation error for the calibration condition does not exhibit any bias after the data correction.

[93] Bortz and Schuster, loc. cit.

In the following, considerations about the involved variables and the underlying distributions for the F-test are presented. They allow for a sophisticated test design and a better interpretation of the results, later. For the assumed normal distributed corrected signed error $\tilde{\Delta}_{lm}^b(\mathbf{x})$, the ratio

$$\frac{\mathrm{MSE}_m}{\mathrm{MSE}_{m_c}} \frac{\sigma_{m_c}^2}{\sigma_m^2} \tag{5.9}$$

follows the non-central F-distribution[94] with the degrees of freedom $\nu_1 = L_m B_m X_m$ and $\nu_2 = L_{m_c} B_{m_c}$ and non-centrality parameter $\delta = \nu_1 \frac{\tau_m^2}{\sigma_m^2}$. For the F-test, the null hypothesis H_0 is rejected, if the test statistic exceeds a critical value f_{crit}. A type-I error occurs, if the test falsely rejects the null hypothesis. Its probability is quantified via

[94] Howell, op. cit., Sec. 11.12.

$$P\left(\frac{\mathrm{MSE}_m}{\mathrm{MSE}_{m_c}} > f_{\mathrm{crit}} \middle| H_0 : \sigma_{m_c}^2 = \sigma_m^2, \tau_m^2 = 0\right) = 1 - F(f_{\mathrm{crit}}, \nu_1, \nu_2, 0)\,, \tag{5.10}$$

where F denotes the cumulative distribution function of the non-central F-distribution with $\delta = 0$. Note, that the variances $\sigma_{m_c}^2$ and σ_m^2 in (5.9) cancel out, if the null hypothesis is true. Thus, the ratio $\frac{\mathrm{MSE}_m}{\mathrm{MSE}_{m_c}}$ follows the F-distribution described in (5.10). For the scenario under investigation, a type-I error corresponds to falsely detecting a difference in localisation although it does not exist. The probability of a type-II error reads

$$P\left(\frac{\mathrm{MSE}_m}{\mathrm{MSE}_{m_c}} < f_{\mathrm{crit}} \middle| \frac{\sigma_m^2}{\sigma_{m_c}^2}, \frac{\tau_m^2}{\sigma_m^2}\right) = F\left(f_{\mathrm{crit}} \frac{\sigma_{m_c}^2}{\sigma_m^2}, \nu_1, \nu_2, \frac{\nu_1 \tau_m^2}{\sigma_m^2}\right) \tag{5.11}$$

and quantifies the chance of falsely stating transparency although there exists an effect. It depends on the assumed effect sizes, which are quantified by $\sigma_m^2/\sigma_{m_c}^2$ and τ_m^2/σ_m^2. As mentioned by Leventhal[95] and Brinkmann et al.,[96] significance tests are usually designed to achieve a small type-I error probability. However, this generally reduces the power of the test i.e. its capability to detect differences: It comes at the cost of an increased type-II error probability. In the context of showing authenticity of audio presentation methods, the mentioned design criterion is misleading as it favours judgements towards authenticity. The Brinkmann et al. suggest to choose a specific minimum effect size, which can be regarded as negligible. They then choose a balanced test design, where both error types are equally probable for this effect size. This strategy is adapted to the current study: The SFS method is supposed to not further increase the localisation uncertainty. Thus a small increase of variance about 10 percent is accepted, i.e. $\sigma_m^2 = 1.1 \cdot \sigma_{m_c}^2$. For the localisation bias, its variation among the listening positions should not exceed the localisation uncertainty, i.e. $\tau_m^2 = \sigma_m^2$. Together with these relations,

[95] Leventhal (1986). "Type 1 and Type 2 Errors in the Statistical Analysis of Listening Tests". In: *J. Aud. Eng. Soc.* 34.6, pp. 437–453.

[96] Brinkmann et al., op. cit., Sec. II.E.

the equality of (5.10) and (5.11) for the balanced test design results in

$$1 - F(f_{\text{crit}}, \nu_1, \nu_2, 0) = F(f_{\text{crit}}/1.1, \nu_1, \nu_2, \nu_1) \,. \tag{5.12}$$

It is numerically solved for f_{crit} via a root finding algorithm. The results for the three experiments are listed in Tab. 5.1. Since a significance test is conducted for each SFS method (18 in total), error accumulation has to be considered. For the worst-case, the total type-I and type-II error probabilities are given by the sum of the individual errors. With approximately 0.0321, the error probability is below the widely accepted 0.05 for the type-I error.

Exp.	ν_1	ν_2	f_{crit}	Prob.
1	200	200	1.5055	0.0020
2	200	200	1.5055	0.0020
3-1	600	360	1.5108	10^{-5}
3-2	450	270	1.5113	0.0001

Table 5.1: Parameters for the F-tests with balanced type-I and type-II error probability for the experiments under investigation. For the study of Wierstorf, two different sample sizes were used for WFS (3-1) and NFCHOA (3-2).

5.3.3 Results[97]

A summary of the results is given in Fig. 5.9: In general, a close-to-zero bias is found for all methods at the centre line, i.e. $x = 0$. For NFCHOA ($M = 6$) with the rectangular window[98], strong localisation artefacts can be observed for all positions which are not on the centre line. The localisation bias increases, the further the listener is positioned towards the negative y-direction. The RMSE of 60° is reduced by increasing the modal bandwidth M to 13 and further to 27.[99] The results from both experiments for $M = 27$ and the results from Wierstorf for $M = 28$ are very similar.[100] All three show a significant increase of the RMSE compared to the calibration condition. Increasing the modal bandwidth further to $M = 300$ only leads to a slight improvement of localisation accuracy.[101] A similar yet reduced relation between the modal bandwidth and the localisation accuracy bias is present for the max-$\mathbf{r}_E$ window:[102] Starting with a RMSE of 23.4° for $M = 6$, it decreases to 7.3° for $M = 13$ and further to 4.5° for $M = 27$. As the only condition for NFCHOA, $M = 27$ with the max-$\mathbf{r}_E$ window can be regarded as transparent, because the RMSE is not significant. In summary, the results for NFCHOA show, that an increase of the modal bandwidth and the usage of the max-$\mathbf{r}_E$ window has a positive effect on the localisation accuracy. The improvements decrease for higher the modal bandwidths M.

The human localisation performance in WFS and NFCHOA ($M = 300$) can be regarded as very similar:[103] In all three data collections localisation bias does not exceed 3.5°. The results from Wierstorf show narrower confidence intervals, since the sample size was larger in comparison to the current study. All three methods exhibit a significant increase of the RMSE w.r.t. the calibration condition.

For LWFS-VSS[104] with $R_l = 0.15$ cm and 0.30 cm, listeners at positions off the centre line, i.e. $x \neq 0$, tend to localise the virtual source further towards the centre. The effect is less pronounced for the larger radius leading to a smaller RMSE. For $R_l = 0.45$ cm, an artefact is observable for the "top left" position at $\mathbf{x} = [-1.0, 0.75, 0]^{\mathrm{T}}$ m (red arrow): Here, the radius is large enough so that the virtual SSD intersects with loudspeaker array and a subset of the focused sources outside is not synthesised. With this outlier excluded from the RMSE calculations, this parametrisation of LWFS-VSS is transparent. It is

[97] The raw data is published as Winter and Frenkel (2019). *Data of Listening Experiments for Azimuthal Localisation in (Local) Sound Field Synthesis*. DOI: 10.5281/zenodo.3252380.

[98] see Fig. 5.9a

[99] see Fig. 5.9a/c/e

[100] see Fig. 5.9e-g

[101] see Fig. 5.9e/i

[102] see Fig. 5.9b/d/g

[103] see Fig. 5.9i-k

[104] see Fig. 5.9l-n

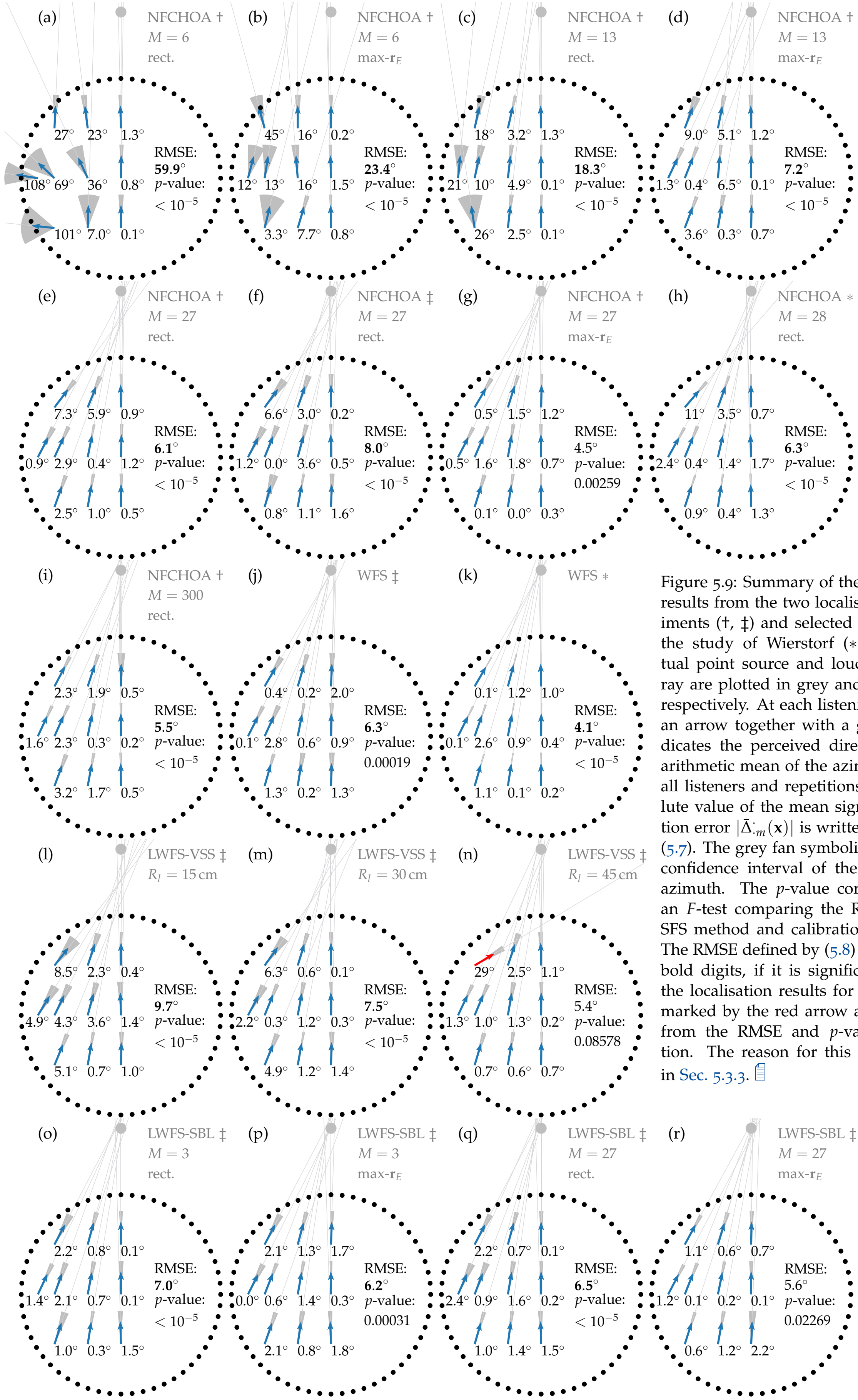


Figure 5.9: Summary of the localisation results from the two localisation experiments (†, ‡) and selected results from the study of Wierstorf (∗). The virtual point source and loudspeaker array are plotted in grey and black dots, respectively. At each listening position, an arrow together with a grey line indicates the perceived direction as the arithmetic mean of the azimuth among all listeners and repetitions. The absolute value of the mean signed localisation error $|\bar{\Delta}_{\cdot m}(\mathbf{x})|$ is written below, see (5.7). The grey fan symbolises the 99%-confidence interval of the localisation azimuth. The p-value corresponds to an F-test comparing the RMSE of the SFS method and calibration condition. The RMSE defined by (5.8) is written in bold digits, if it is significant. In (n), the localisation results for the position marked by the red arrow are excluded from the RMSE and p-value calculation. The reason for this is explained in Sec. 5.3.3.

however not applicable for all listening positions due to the mentioned artefact.

All LWFS-SBL parametrisations[105] exhibit a localisation bias that is below 3° for all listening positions. The RMSE values for $M = 27$ are slightly smaller in comparison to $M = 3$ for both window types. The localisation accuracy for the rectangular window is inferior to the max-$\mathbf{r}_E$ weighting. However, the pairwise difference in RMSE does not exceed 1° for both parameters. Transparency is reached for $M = 27$ with the max-$\mathbf{r}_E$ window.

[105] see Fig. 5.90-r

5.3.4 Discussion

The results for NFCHOA agree with prior studies: The positive effect of large modal bandwidths on the accuracy was already reported by Wierstorf.[106] The superiority of the max-$\mathbf{r}_E$ window agrees with the findings of Frank et al.[107] for HOA. Stitt[108] showed for HOA, that the localisation accuracy is heavily influenced by the position of the virtual source relative to the off-centre positions. These findings agree with the fact, that the accuracy for the centre line $x = 0$ was nearly independent of the parametrisation of NFCHOA. Along the centre line, the listener only moves towards or away from the virtual point source. This was reported as the best-case by Stitt. For $x \neq 0$, the listener is also shifted perpendicular to the source direction. The impaired accuracy, especially for low orders, in this cases also confirms the statements of Stitt. For low modal bandwidths, a wide confidence interval of the localisation bias caused by a large standard deviation at positions close to the loudspeaker array was observed in the current study. It aligns with Wierstorf,[109] who showed for a low modal bandwidth, that the listeners likely perceive multiple sources which is an artefact of NFCHOA. Due to the evaluation paradigm used in the current study they are forced to choose one of the sources. The number of a measurements in the current study is not sufficient to identify the individual locations of the split sources. The source splitting phenomenon was also reported by Frank.[110]

[106] Wierstorf, "Perceptual Assessment of sound field synthesis", Sec. 5.1.

[107] Frank et al., op. cit.

[108] Stitt, loc. cit.

The results for WFS and NFCHOA with $M = 300$ further support the theoretical findings, that both SFS methods share very similar properties for high modal bandwidths.

[109] Wierstorf, loc. cit.

[110] Frank, op. cit., Sec. 3.3.

Independent of its parametrisation, the localisation bias in LWFS-SBL is comparatively low at all investigated listening positions. The increased RMSE for low orders and the rectangular weighting function is mostly attributed to an increased standard deviation. This agrees with experiments of Frank,[111,112] where the perceived source width for HOA increases, the smaller the length of the $\mathbf{r}_E$-vector. Since larger modal bandwidths and the max-$\mathbf{r}_E$ weighting increase the length, the smaller localisation blur for the corresponding conditions observed in the current study seems reasonable.

[111] Frank (2013b). "Source Width of Frontal Phantom Sources: Perception, Measurement, and Modeling". In: *Archives of Acoustics* 38.3, pp. 311–319, Sec. 4.

[112] Zotter and Frank (2019). *Ambisonics*. Springer International Publishing, Sec. 2.3.1.

In LWFS-VSS, the largest radius of the virtual SSD led to the best localisation performance, despite the mentioned artefact for one listening position. From a physical standpoint, the employed focused

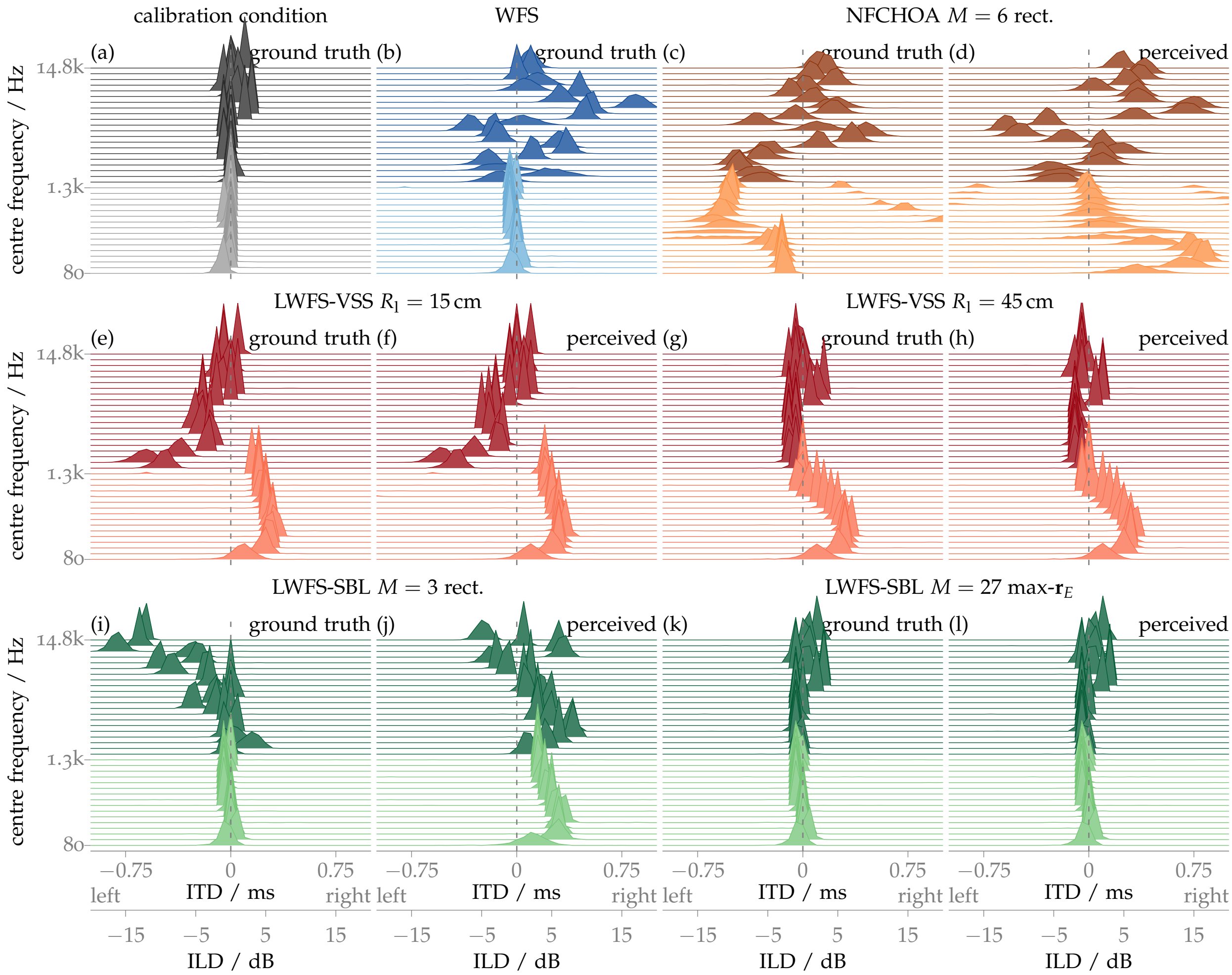


Figure 5.10: The plots show IC-weighted histograms of the ITD for auditory bands with centre frequencies below 1.4 kHz. The ILD is shown for frequencies above this threshold. (a) shows the calibration condition with the listener facing the source as a reference. For the remaining plots, the listener was located at $[1.25, 0, 0]^T$ m, see position 6 in Fig. 5.8. The listener was either oriented towards the virtual point source (ground truth) or in the average localisation direction from the listening experiment (perceived). For WFS in (b), only one plot is shown, as the both directions differ less then 1°.

sources approximate the desired point sources best for high frequencies and large distances to the listener. This was shown in Sec. 3.2.1. Thus, near-field artefacts are a possible explanation for the inferior performance of smaller radii. Moreover, off-centre listening positions result in different distances of the individual focused sources from the loudspeaker array. This results into heterogeneous artefacts for each of the focused sources.

In the following, the results are further discussed in the context of psychoacoustics involving analysis on the invoked ITDs and ILDs. Afterwards a potential connection between the predictions of the geometric model introduced in Ch. 4 and the localisation accuracy are investigated.

Relation to Psychoacoustic Phenomena: The low localisation bias at positions with $x = 0$ for all SFS methods is not further surprising, as the synthesis scenario is completely symmetric at these position. Possible artefacts of the SFS methods such as additional wavefronts approach the listener likewise from the positive and negative x-halfspace with the same delay and amplitude. Summing localisation[113] is triggered, where the combined auditory event is perceived "between" the individual sound events. Since the virtual point source

[113] Blauert, op. cit.

is positioned on the y-axis, the invoked auditory localisation matches its position.

The results for listening positions off the centre line are further discussed in conjunction with the invoked binaural cues relevant for humans' azimuthal localisation. The procedure is very similar to the discussions by Wierstorf et al.:[114] The cues are estimated by a slightly modified version of the binaural model published by May et al.[115,116] The required binaural signal is generated by filtering the dry source signal with the BTF $H^{\mathrm{SFS}}_{\{\mathrm{L,R}\}}(\mathbf{x}, \phi_{\mathrm{h}}, \omega)$ corresponding to the exemplary listening position $\mathbf{x} = [1.25, 0, 0]^{\mathrm{T}}$ m and selected SFS methods from the experiment. The head orientation ϕ_{h} is either set to the relative direction of the virtual point source as the physical ground truth or to the perceived direction shown in Fig. 5.9. The left and right ear signals are separately filtered by second-order Butterworth Bandpass filter with its passband between 500 Hz and 2 kHz to approximate the transfer function of the middle ear.[117] The resulting signals are processed by a fourth-order IIR Allpole Gammatone filterbank[118] in order to mimic the frequency selectivity of the human cochlea. The centre frequencies of the filters are equidistantly distributed on an Equivalent Rectangular Bandwidth (ERB) scale[119] with 1 ERB distance. This results in 46 auditory channels per ear within the range of 80 Hz to 16 kHz. The half-wave rectification and 1 kHz lowpass filtering is applied afterwards to model neural transduction in the inner hair cells.[120] The binaural cues are individually estimated for each auditory band on segments of 20 ms length with an overlap of 10 ms. A Hann a.k.a. Hanning window[121] was applied to each segment. The IC and ITD correspond to the maximum value and the corresponding time of the normalised interaural cross-correlation. The ILD results from the ratio of the time-averaged signal power between both ears measured in decibel. ITD and ILD are defined such, that positive values correspond to a lateralisation to the right. The estimated values are aggregated over time to form histograms for each auditory band. The bin sizes are set to 50 μs and 1 dB for ITD and ILD, respectively. Each sample is weighted by the corresponding IC value as a measure of reliability.[122]

For reference, the histograms of the calibration condition are plotted in Fig. 5.10a. The minor deviations from zero in ITD and ILD are most likely caused by asymmetries of the HATS used for measuring the employed HRTF dataset. It was already found by Wierstorf et al.,[123] that WFS correctly reconstructs the ITD.[124] The ILD is affected by the spatial aliasing causing large deviations from the reference. The broad width and small height of the distributions in the ILD histograms further indicate a small reliability. As the perceived direction follows the ITD cues, the results for WFS agree with the work of Faller and Merimaa.[125] The authors showed, that the selection of the relevant binaural cues based on their reliability is beneficial for localisation.

In addition to the deviations in ILD, the low modal bandwidth in the selected NFCHOA condition causes less reliable and inconsistent

[114] Wierstorf et al., "Assessing localization accuracy in sound field synthesis", Sec. IV.

[115] May et al. (Jan. 2011). "A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End". In: *IEEE Trans. Audio, Speech, Language Process.* 19.1, pp. 1–13, Sec. II.

[116] the implementation of the model is part of Two!Ears Team (Oct. 2018). *Two!Ears Auditory Model 1.5*. DOI: 10.5281/zenodo.1458420.

[117] Puria (2003). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions". In: *J. Acoust. Soc. Am.* 113.5, pp. 2773–2789.

[118] Lyon (1997). "All-pole models of auditory filtering". In: *Diversity in Auditory Mechanics*, pp. 205–211.

[119] Glasberg and Moore (1990). "Derivation of auditory filter shapes from notched-noise data". In: *Hearing Research* 47.1, pp. 103–138, Eq. (4).

[120] Dau et al. (1996). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure". In: *J. Acoust. Soc. Am.* 99.6, pp. 3615–3622, Sec. B.1.

[121] Harris (Jan. 1978). "On the use of windows for harmonic analysis with the discrete Fourier transform". In: *Proceedings of the IEEE* 66.1, pp. 51–83, Eq. (27a).

[122] Faller and Merimaa, op. cit.

[123] Wierstorf et al., loc. cit.

[124] see Fig. 5.10b

[125] Faller and Merimaa, op. cit.

ITD cues:[126] The ITD values are clustered w.r.t. the centre frequencies. Lower auditory bands near the 80 Hz limit exhibit similar ITDs, while bands near the 1.4 kHz limit also share a common yet different set of ITDs. The broad distribution of the ITDs in between these clusters can be interpreted as a kind of transition zone, where the ITDs of the two groups are superimposed, leading to unreliable estimates. The observations further substantiate the already discussed source splitting phenomenon, where the different ITD groups are perceived as individual auditory events.

[126] see Fig. 5.10c/d

For LWFS-VSS, the smallest radius of $R_\mathrm{l} = 15$ cm leads to contradicting cues.[127] For the perceived direction, the ILD and ITD indicate a lateralisation to the left and to the right, respectively. Most of the distributions are all relatively narrow stating a high reliability which can be attributed to the reduced spatial aliasing. As a potential interpretation, the auditory event is perceived as a non-trivial combination of ITD and ILD resulting in a direction which is "in between" the directions corresponding to the individual cues. It is worth noting, that ILD cues for high frequencies coincide with the reference, which agrees with the focused sources being most accurate in this frequency region. For the larger radius $R_\mathrm{l} = 45$ cm, almost all ILD cues match the reference, which can be explained by the decreased near-field artefacts.[128] Although the deviation between ITD and ILD is less pronounced, the ITDs still show a tendency towards the right side. For this condition, localisation is dominated by the ILD cues.

[127] see Fig. 5.10e/f

[128] see Fig. 5.10g/h

Both parametrisations of LWFS-SBL reliably reconstruct the ITD cues.[129] Due to the stronger SBL for $M = 3$, the ILDs at high frequencies are corrupted. Moreover, their reliability is inferior compared to the second LWFS-SBL condition, where all cues almost perfectly match the reference condition. Despite this visible difference, the localisation performance in both parametrisations is very similar. This serves as another example, where the perceived direction is dominated by the most reliable cues.

[129] see Fig. 5.10i-l

Relation to the Geometric Model: In Ch. 4, the geometric model led to two fundamental frequencies describing the trade-off between spatial aliasing and the limitation of the available listening area. The aliasing frequency $f^\mathrm{S}(\mathbf{x})$ and the SBL frequency $f^\mathrm{B}_M(\mathbf{x})$ mark the spectral bound up to which no considerable synthesis artefacts are present in proximity to the listeners head. The absolute value of the mean signed localisation error $\bar{\Delta}_{:m}(\mathbf{x})$ is put in relation to both frequencies in Fig. 5.11. For all four tested SFS methods, the aliasing frequency has nearly no influence on the localisation accuracy. The bias is mainly determined by $f^\mathrm{B}_M(\mathbf{x})$. Below the threshold of approximately 1.5 kHz, the localisation bias increases significantly. This establishes a link to the psychoacoustics interpretation as the ITD is the dominant cue in the described frequency range. The lower the $f^\mathrm{B}_M(\mathbf{x})$, the more auditory bands for which the ITD is to be estimated are corrupted.

It was already discussed in Ch. 4, that near-field artefacts of the fo-

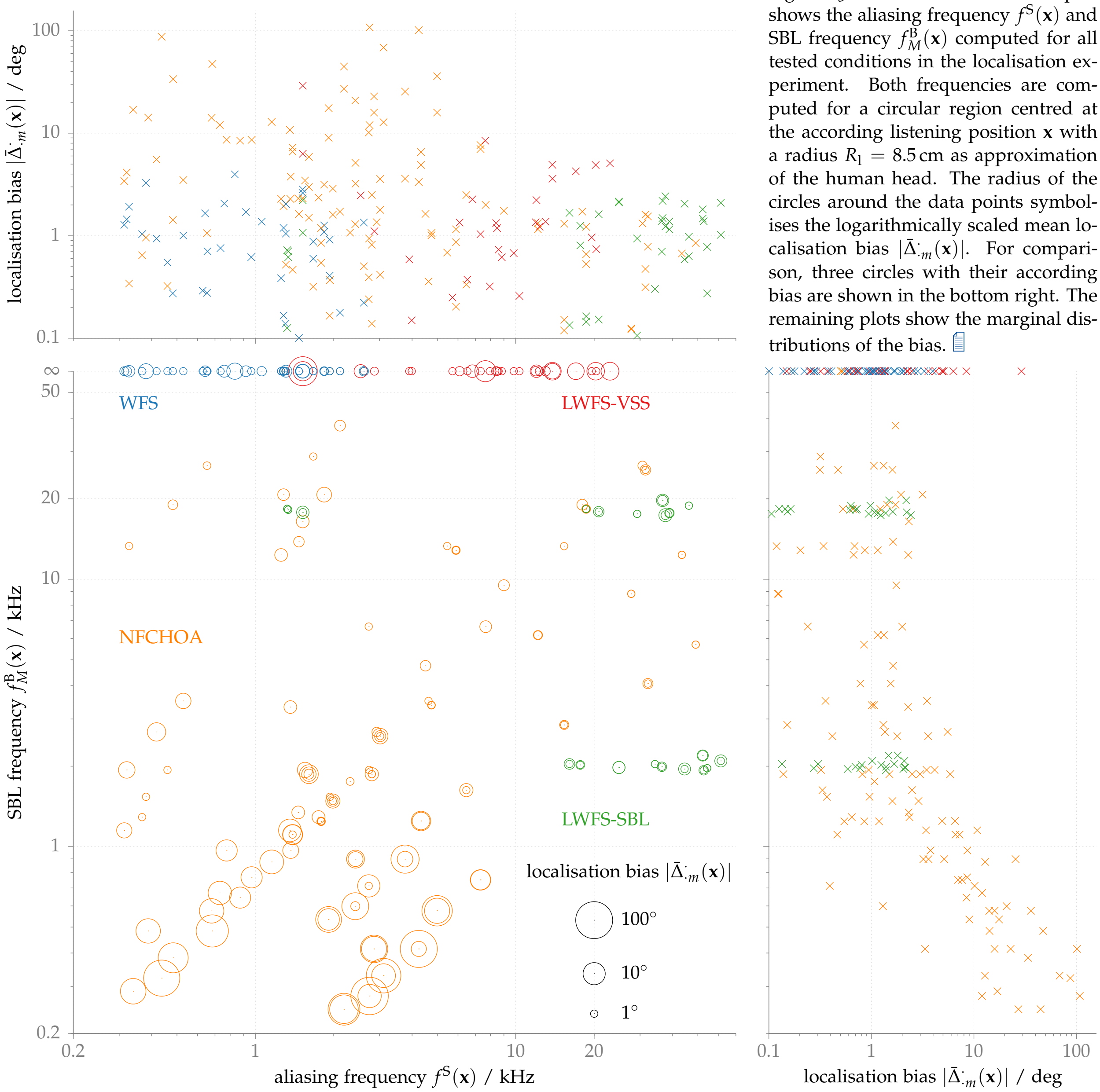


Figure 5.11: The bottom left plot shows the aliasing frequency $f^{S}(\mathbf{x})$ and SBL frequency $f^{B}_{M}(\mathbf{x})$ computed for all tested conditions in the localisation experiment. Both frequencies are computed for a circular region centred at the according listening position $\mathbf{x}$ with a radius $R_l = 8.5\,\mathrm{cm}$ as approximation of the human head. The radius of the circles around the data points symbolises the logarithmically scaled mean localisation bias $|\bar{\Delta}_{\cdot m}(\mathbf{x})|$. For comparison, three circles with their according bias are shown in the bottom right. The remaining plots show the marginal distributions of the bias.

cused sources are not considered by the ray approximations involved in the geometric model. The increased localisation bias for some LWFS-VSS conditions caused by those artefacts cannot be explained by the model. For some data points of NFCHOA and LWFS-SBL, two concentric circles are plotted in the diagram. These represent the localisation bias for the rectangular and the max-$\mathbf{r}_E$ window with the same spatial bandwidth and listening position. The geometric model does not regard different modal window types and is not capable of explaining these variations in localisation bias.

5.4 Summary

In this chapter, the selected analytic methods for (L)SFS were evaluated w.r.t. its azimuthal localisation. Prior to the actual listening tests, the used evaluation method was (re-)validated and compared to the study of Wierstorf et al.[130] in a separate experiment. As the main difference to the original apparatus, a circular curtain as the projection surface for the pointing method was used instead of a straight curtain. Besides the rotational invariance, this setup also allows for listening tests with sound sources located 360° around the listener. Other than in the original experiment, no systematic dependency of the signed localisation error and physical ground-truth azimuth of the source was found. Thus, no localisation undershoot is present. The statistical evaluation however revealed, that a session-dependent bias is present in the localisation results. It is mainly attributed to anthropomorphic differences and to an offset between the listener's median plane and the direction of the pointing device. As a consequence for the subsequent experiment on the SFS methods, calibration conditions had to be added in order to compensate for this.

[130] Wierstorf et al., "Perception and evaluation of sound fields".

Four SFS methods were investigated for a circular array of 56 loudspeakers and 1.5 m radius synthesising a virtual point source. For all four approaches, parametrisations could be found, that led to transparent or close-to transparent azimuthal localisation. For these, the localisation bias was below 3°. Prior findings regarding the good localisation properties of WFS were confirmed. The accuracy in NFCHOA heavily depends on the listening position, the used modal bandwidth and the weighting function. As a general rule of thumb, max-$\mathbf{r}_E$ weighting with a reasonably high bandwidth leads to the most accurate and homogeneous localisation results over the whole listening area. Again, it is confirmed that NFCHOA of very high order shares similar properties to WFS. In LWFS-VSS, near-field artefacts introduced by the employed focused source have to be considered. Best results were found for the largest radius of the virtual SSD. This has the drawback, that the supported listening positions have to be further restricted, as the virtual SSD must not intersect with the loudspeaker array. Contrary to NFCHOA, the localisation bias in LWFS-SBL did not show a strong dependency on the modal bandwidth and weighting function. It can be attributed to the expansion centre of the Circular Harmonics representation, which is shifted to the listening position.

6 Perception of Timbre: Colouration

As already outlined in Sec. 1.1, timbral fidelity contributes approximately seventy percent to the overall quality humans perceive in the context of surround sound.[1] Timbre has been standardised[2] as the attribute that "*enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar*". Colouration describes the difference of two sounds in timbre, whereas one is considered to be the uncoloured reference.[3] It may be presented explicitly to the listener or has been built up internally by prior listening experience. These options directly correspond to the concepts of authenticity and plausibility discussed in the context of quality assessment in Ch. 1. According to the extensive overview of Wierstorf[4] on different definitions of timbre, it is a multidimensional percept and the underlying metric measuring the colouration between two sounds is unknown and non-trivial.

For WFS, it has been shown that spatial aliasing leads to a perceivable colouration of the reproduced sound field compared to the desired sound field as a reference.[5,6,7] Since the spatial aliasing increases for larger distances between the loudspeakers, the perceived colouration increases. Investigations on colouration in NFCHOA are not known to the author. Numerical simulations of Solvang[8] for HOA with rectangular weighting showed, that the colouration likely increases for off-centre positions, if the number of loudspeakers is higher than $2M + 1$. The spatial bandwidth or Ambisonics order is denoted by M. Frank[9] investigated the colouration for moving sources in HOA: The max-$\mathbf{r}_E$ window led to less artefacts w.r.t. timbre than the rectangular weighting for all investigated orders, number of loudspeakers and listening positions. Further, colouration is less pronounced for 3rd-order with 8 loudspeakers than for 7th-order with 16 loudspeakers.

As LSFS enhances the synthesis accuracy around the listener's position, the question arises, whether the perceived colouration can be reduced by such techniques as well. Within this chapter, the results of two listening experiments comparing the colouration introduced by WFS, NFCHOA, LWFS-VSS and LWFS-SBL are presented. The study focuses on the following aspects:

1. The impact of different parametrisations of the LWFS methods on the perceived colouration is investigated.
2. The influence of the audio content, i.e. the source signal emitted

This chapter is published in major parts as Winter et al. (Oct. 2018a). "Colouration in Local Wave Field Synthesis". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 26.10, pp. 1913–1924

[1] Rumsey et al. (2005). "On the relative importance of spatial and timbral fiedelities in judgements of degraded multichannel audio quality". In: *J. Acoust. Soc. Am.* 118.2, pp. 968–976.

[2] *Acoustical Terminology* (1994). Standard ANSI S1.1-1994. New York, USA: American National Standards Institute.

[3] Wierstorf et al. (Aug. 2014). "Coloration in Wave Field Synthesis". In: *Proc. of 55th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Helsinki, Finland, Sec. 2.

[4] Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Sec. 5.2.

[5] Wittek et al. (Oct. 2007). "On the Sound Color Properties of Wavefield Synthesis and Stereo". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA.

[6] Wierstorf et al., "Coloration in Wave Field Synthesis".

[7] Wierstorf, loc. cit.

[8] Solvang (2008). "Spectral Impairment of Two-Dimensional Higher Order Ambisonics". In: *J. Aud. Eng. Soc.* 56.4, pp. 267–279.

[9] Frank (June 2013a). "Phantom Sources using Multiple Loudspeakers in the Horizontal Plane". PhD thesis. Graz, Austria: Institute of Electronic Music and Acoustics, Sec. 5.3.

by the virtual sound field, is analysed.

3. LWFS methods are compared to conventional SFS methods focusing on the question, whether the extended approaches lead to a less coloured reproduction than the non-local methods.
4. It is evaluated, whether transparent reproduction can be achieved, that is, indistinguishable from a reference.

This chapter is organised as follows: The details of the evaluation method are presented in Sec. 6.1. The main study comparing the colouration of different SFS is comprised in Sec. 6.2. A summary is given afterwards.

6.1 Evaluation Method

For the evaluation of colouration, binaural synthesis is used to simulate SFS under free-field conditions over headphones. The reasons for employing binaural synthesis have been presented in conjunction with the localisation experiments in Sec. 5.1. For rating the colouration a modified Multiple Stimulus with Hidden Reference and Anchor (MUSHRA)[10] paradigm is used. The method was validated[11] and successfully used in colouration experiments for SFS.[12,13,14,15] In the following subsections the details of the evaluation method are presented.

[10] *Method for the subjective assessment of intermediate quality level of audio systems* (2015). Standard ITU-R BS.1534-3. International Telecommunication Union Radiocommunication Assembly.

[11] Wierstorf, op. cit., Sec. 4.4.

[12] Wittek et al., "On the Sound Color Properties of Wavefield Synthesis and Stereo", p. 3.2.1.

[13] Wittek (2007). "Perceptual differences between wavefield synthesis and stereophony". PhD thesis. University of Surrey, Sec. 8.2.2.

[14] Wierstorf et al., "Coloration in Wave Field Synthesis".

[15] Wierstorf, "Perceptual Assessment of sound field synthesis", Sec. 5.2.

[16] Ibid., Sec. 5.2.1.

6.1.1 Static Binaural Synthesis

The study was performed using an approach for non-individual binaural synthesis which is similar to the one for localisation experiments in Ch. 5. For the details, the reader is referred to Sec. 5.1.1. The originally used head tracking to dynamically adjust the apparent source azimuth to the head orientation is deactivated to avoid changes in timbre due to head rotation.[16] Thus, only one BTF per SFS method and listening position is necessary as the head azimuth is fixed. The BTFs are generated with the same HRTFs dataset and interpolation strategies as for the main localisation experiment.

The first dry source signal $\tilde{S}(\omega)$ was a pink noise pulse train with a pulse duration of 900 ms (including cosine-shaped fade-in/fade-out of 50 ms, each) and a pause of 500 ms. It was double-checked with the published experimental data,[17] that this is the same stimulus as in the study of Wierstorf.[18] He motivated the choice of this stimulus by stating that it already has been used by Wittek.[19] However, the definitions of Wittek[20] regarding the pulse duration and fade-in/fade-out suggest, that the burst length for his experiments was only 800 ms. Here, linguistic ambiguities whether the fade-in/fade-out length is already included in the pulse duration could not be finally resolved. The second signal was a female speech sample of eight seconds duration. A music stimulus was used as a third signal for training purposes. In the experiment, the signal was seamlessly looped and filtered by the current BTF for binaural reproduction.

[17] Wierstorf and Hohnerlein (Nov. 2016). *Coloration of a point source in Wave Field Synthesis – data*. DOI: 10.5281/zenodo.164589.

[18] Wierstorf, op. cit., Sec. 5.2.

[19] Wittek, "Perceptual differences between wavefield synthesis and stereophony", Sec. 8.2.

[20] Ibid., Sec. 7.3.2.

6.1.2 Location and Hardware

The experiments were conducted separately at two different facilities. At the University of Rostock, it took place in a 86 m^3 acoustically damped room (Audio laboratory, R8202, Institute of Communications Engineering). At the TU Berlin, the experiment was conducted in a 54 m^3 acoustically damped listening room (room Pinta, Telefunken Building). In both cases, the listeners wore open headphones (AKG K601). In a separate room, a computer equipped with a sound card (Focusrite Scarlett 2i2, 1st Gen. in Rostock and RME Hammerfall DSP MADI + Behringer HA4700 Powerplay Pro-XL in Berlin) was used for audio playback. The signals were transmitted via an analogue cable of approximately 6 m length to the headphones inside the listening room.

6.1.3 Procedure

This study used a modified MUSHRA test paradigm.[21] The original quality scale is replaced by a continuous scale ranging from *no difference* (0) to *very different* (1) and a different lower anchor is introduced. The term "modified" is however skipped in the following for brevity. The subjects were asked to use a Graphical User Interface (GUI) with one slider per condition to assess the respective colouration compared to an explicitly given reference stimulus. The numerical values (0,1) of the scale were not shown to the subjects. Within each run, the respective conditions (including differently parametrised SFS methods), the hidden reference and a lower anchor had to be rated. The latter condition is a intentionally degraded version of the reference, which is supposed to be rated as *very different* by all test participants. The current MUSHRA standard[22] also defines mid anchor, which is chosen such that it is rated in between the other conditions. Due to the exploratory nature of the current study, no sensible choice for the mid anchor in MUSHRA colouration experiments is known to the author. It is thus omitted. The order of runs and the arrangement of the conditions in the GUI were randomised. An additional run had to be passed beforehand for training. During a single run, the listener could switch instantaneously between the conditions as often as desired.

[21] *Method for the subjective assessment of intermediate quality level of audio systems*.

[22] Ibid., Sec. 5.1.

6.1.4 Discussion

It was already discussed for the localisation experiments, that binaural synthesis cannot be regarded as fully transparent.[23] Wierstorf[24] supported his choice of binaural synthesis by a study of Olive et al.[25] where preference ratings for different loudspeakers were independent between real loudspeakers and binaural simulations. He further validated binaural synthesis as an evaluation tool by measurements with a second HATS to create a mismatch to the employed HRTF dataset: A single loudspeaker and different array configurations driven by WFS to synthesise a virtual point source were used.

[23] Brinkmann et al. (2017). "On the authenticity of individual dynamic binaural synthesis". In: *J. Acoust. Soc. Am.* 142.4, pp. 1784–1795, Sec. IV.

[24] Wierstorf, "Perceptual Assessment of sound field synthesis", Sec. 4.4.

[25] Olive et al. (May 2007). "Listener Loudspeaker Preference Ratings Obtained In Situ Match those Obtained Via a Binaural Room Scanning Measurement and Playback System". In: *Proc. of 122nd Aud. Eng. Soc. Conv.* Vienna, Austria.

The spectral deviations of the measured ear signals resulting from the real loudspeakers or their non-individual binaural simulation are nearly independent of the configuration, if the single loudspeaker or the virtual point source is in front of the listener. Thus, the spectral distortions introduced by non-individual synthesis can be regarded as independent of the SFS setup. Wierstorf did not further discuss, that the distortions lead to a frequency-dependent amplification or attenuation of up to 15 dB. Thus, binaural synthesis has a filtering effect and can emphasise, diminish or even mask colouration artefacts in distinct frequency ranges. It has to be assumed for the upcoming evaluations, that the mentioned distortions averaged over all test participants are approximately zero due to the individual anthropomorphic differences. Thus, the average colouration ratings converge to the "true" colouration introduced by the SFS methods.

Besides the mentioned drawbacks, the binaural synthesis resolves some issues a listening experiment conducted with real loudspeaker arrays would introduce. A discussion on the benefits was already given in Sec. 5.1.

6.2 Comparison of (Local) Sound Field Synthesis Methods

6.2.1 Conditions

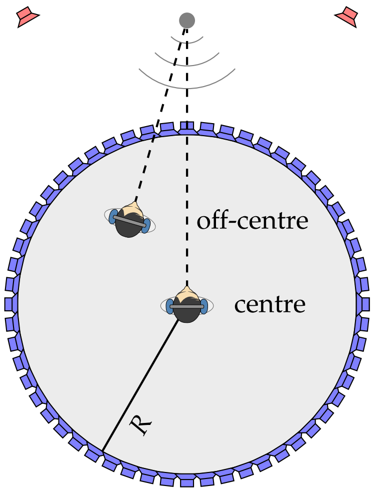


Figure 6.1: The blue and the red loudspeakers illustrate the circular loudspeaker array and the stereophonic setup, respectively. The grey bullet symbolises the virtual point source. The two manikins are positioned at the two listening positions.

As the reference condition, a binaurally simulated point source positioned at $[0, 2.5, 0]^{\mathrm{T}}$ m was used. It is depicted in Fig. 6.1. This reference condition filtered by a 2nd-order Butterworth high pass with a cutoff-frequency of 5 kHz served as the lower anchor. A high-pass is chosen due to the nature of spatial aliasing, which adds energy to the reproduced sound field at high frequencies. A binaurally simulated, circular array of 56 loudspeakers with $R = 1.5$ m, centred at the coordinates' origin was employed to synthesise this point source using WFS, NFCHOA, LWFS-SBL, and LWFS-VSS. For NFCHOA, the spatial bandwidth $M = 27$ and a rectangular window is used.[26] The reference position $\mathbf{x}_{\mathrm{ref}}$ in WFS is set to the coordinates' origin.[27] For LWFS-SBL, the expansion centre $\mathbf{x}_{\mathrm{c}}$ was set to the position of the listener. For LWFS-VSS, a quasi-continuous circular virtual SSD was used. Its centre $\mathbf{x}_{\mathrm{l}}$ was located at the listening position. Additionally, a stereophonic setup with the loudspeakers positioned at $[\pm 1.4, 2.5, 0]^{\mathrm{T}}$ m and the phantom source panned to the centre between both loudspeakers was included.

[26] see (3.18) and Tab. 3.3

[27] see (3.9a) and Tab. 3.1

As shown in Fig. 6.1, two listening positions were tested: The centre position is co-located with the coordinates' origin. The off-centre position was set to $[-0.5, 0.75, 0]^{\mathrm{T}}$ m. The positions correspond to position 3 and 1 of the localisation experiments.[28] The listener was always oriented towards the virtual point source except for the off-centre stereo condition, where the head pointed to the nearest loudspeaker. It ensured that colouration was the only perceivable change between conditions and the relative source position was fixed.

[28] see Fig. 5.8

Investigated Aspect	Source Signal	Listening Position	Exp.
number of plane waves in LWFS-SBL	noise	centre	1
number of plane waves in LWFS-SBL	noise	off-centre	1
modal bandwidth in LWFS-SBL	noise	off-centre	1
modal windows in LWFS-SBL	noise	off-centre	2
radius of local region in LWFS-VSS	noise	centre	2
radius of local region in LWFS-VSS	noise	off-centre	2
influence of the source signal	speech	centre	1
influence of the source signal	speech	off-centre	1
comparison of all SFS methods	noise	centre	2
comparison of all SFS methods	noise	off-centre	2

Table 6.1: Summary of study

In order to avoid loudness differences as an additional cue among the conditions, a loudness model[29] was used to adjust the loudness of all conditions to the reference condition. For the loudness estimation, the noise stimulus was used as a dry source signal. The implementation of the model is part of the GENESIS Toolbox.[30]

[29] *Procedure for the Computation of Loudness of Steady Sounds* (2007). Standard ANSI S3.4-2007. New York, USA: American National Standards Institute; Moore et al. (1997). "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness". In: *J. Aud. Eng. Soc.* 45.4, pp. 224–240.

[30] GENESIS (Jan. 2010). *Loudness Toolbox 1.0*.

As already outlined in the introduction, the study was split into two experiments. Each experiment consisted of five MUSHRA runs. Tab. 6.1 summarises the runs together with the investigated aspects as well as the experiments they belong to. To investigate the influence of the discrete PWD involved in LWFS-SBL, the number of plane waves N_{pw} is varied from 64 to 1024. A base-2 logarithmic scale was chosen in order to cover a large range of values. A rectangular modal window with $M = 27$ was used. WFS, NFCHOA, and Stereo were added as baseline conditions. Together with the reference and the anchor, this resulted in ten conditions. The evaluation is carried out for the centre and the off-centre position in two separate MUSHRA runs. The influence of the modal bandwidth M and the modal window were investigated in two additional runs: for the first, M was varied from 3 to 27 on a linear scale with a step size of 4 for the rectangular modal weighting function. A comparison between the rectangular and the max-$\mathbf{r}_E$ window is performed in the second run. For both runs, the off-centre position is used and WFS was added as a baseline. The number of plane waves N_{pw} is set to 1024.

For the parametrisation of LWFS-VSS, the radius R_{l} of the local region was varied on a linear scale ranging from 15 to 45 cm in 17.5 cm steps. The number of focused sources N_{fs} was not further investigated as the same effects as for N_{pw} in LWFS-SBL are to be expected. The investigation is repeated for the centre and off-centre position.

In order to study the impact of the source signal on the perceived colouration, the first two runs for LWFS-SBL are repeated for the female speech stimulus. The results are compared with the colouration ratings for the noise stimulus.

Finally, the last two runs compare all four SFS approaches for both listening positions. Two parametrisations for each of the LWFS methods were included.

6.2.2 Participants

The first experiment was conducted separately at the two facilities mentioned in Sec. 6.1.2. 11 and 9 listeners were recruited for the experiment in Rostock and Berlin, respectively. The age of the participants ranged from 19 to 60 years with an average of approximately 34 years. The second experiment was conducted exclusively in Rostock. 21 listeners were recruited. The age of the participants ranged from 22 to 60 years, with an average of approximately 34 years. All test participants self-reported normal hearing.

6.2.3 Methods for Data Analysis

For each of the ten runs the results can be summarized in a two-dimensional dataset m_l^c of MUSHRA ratings, where l and c correspond to one of the L listeners and to one of the listening conditions. The pairwise difference of the ratings for two conditions c_1 and c_2 by the same listener is defined as $\Delta_l^{c_1,c_2} := m_l^{c_1} - m_l^{c_2}$. As the number of participants is relatively small and the data is bounded to a finite interval, normal distribution of the data cannot be assumed ruling out several parametric statistic methods for data analysis.[31] Sporer et al. further argue, that even for trained expert listeners, an equivalent interpretation of the MUSHRA scale cannot be guaranteed. Moreover, doubts regarding the normality and statistical independence of the ratings were raised by Mendonça and Delikaris-Manias.[32] Hence, non-parametric approaches assuming as little as possible about the underlying distribution are used: for a given pair of conditions, the rating differences are ordered in ascending order to get the respective order statistics, with $\Delta_{(i)}^{c_1,c_2}$ denoting the ith smallest rating.[33] The sample median $\tilde{\Delta}^{c_1,c_2} := \frac{1}{2}(\Delta_{(\lfloor L/2 \rfloor)}^{c_1,c_2} + \Delta_{(\lceil L/2 \rceil)}^{c_1,c_2})$ serves as a good point estimator for the respective population median $\mu_\Delta^{c_1,c_2}$. For its 95%-confidence interval, the distribution-free method given by Hahn and Meeker[34] is used, which also utilises order statistics.

To investigate whether a condition is perceived as coloured w.r.t. the reference, Null Hypothesis Significance Testing (NHST) with $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \leq 0$ is used. Differences between c_1 and the reference are considered to be significant for p-values below 0.001. The test statistic is computed via the non-parametric Wilcoxon signed-rank test,[35] which was recently recommended for MUSHRA.[36] The test had the following settings: For tied values, the average of the ranks spanned by them is assigned.[37] The ranks of differences being zero are regarded in the negative and positive rank-sum with the weight of one half. An analogous procedure is carried out to analyse a potential improvement, i.e. reduced colouration, by LWFS compared to the conventional SFS methods: The null hypotheses $H_0 : \mu_\Delta^{c_1,\mathrm{WFS}} \geq 0$ and $H_0 : \mu_\Delta^{c_1,\mathrm{NFCHOA}} \geq 0$ are considered to investigate the difference

[31] Sporer et al. (Oct. 2009). "Statistics of MUSHRA Revisited". In: *Proc. of 127th Aud. Eng. Soc. Conv.* New York, USA, Sec. 1.

[32] Mendonça and Delikaris-Manias (May 2018). "Statistical Tests with MUSHRA Data". In: *Proc. of 144th Aud. Eng. Soc. Conv.* Milan, Italy.

[33] Hahn and Meeker (1991). *Statistical Intervals: A Guide for Practitioners*. Vol. 392. John Wiley & Sons, p. 75/76.

[34] Ibid., p. 82/83.

[35] Wilcoxon (1945). "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6, pp. 80–83.

[36] Mendonça and Delikaris-Manias, op. cit., Sec. 4.2.

[37] Siegel (1956). *Nonparametric statistics for the behavioral sciences*. McGraw-hill, p. 83.

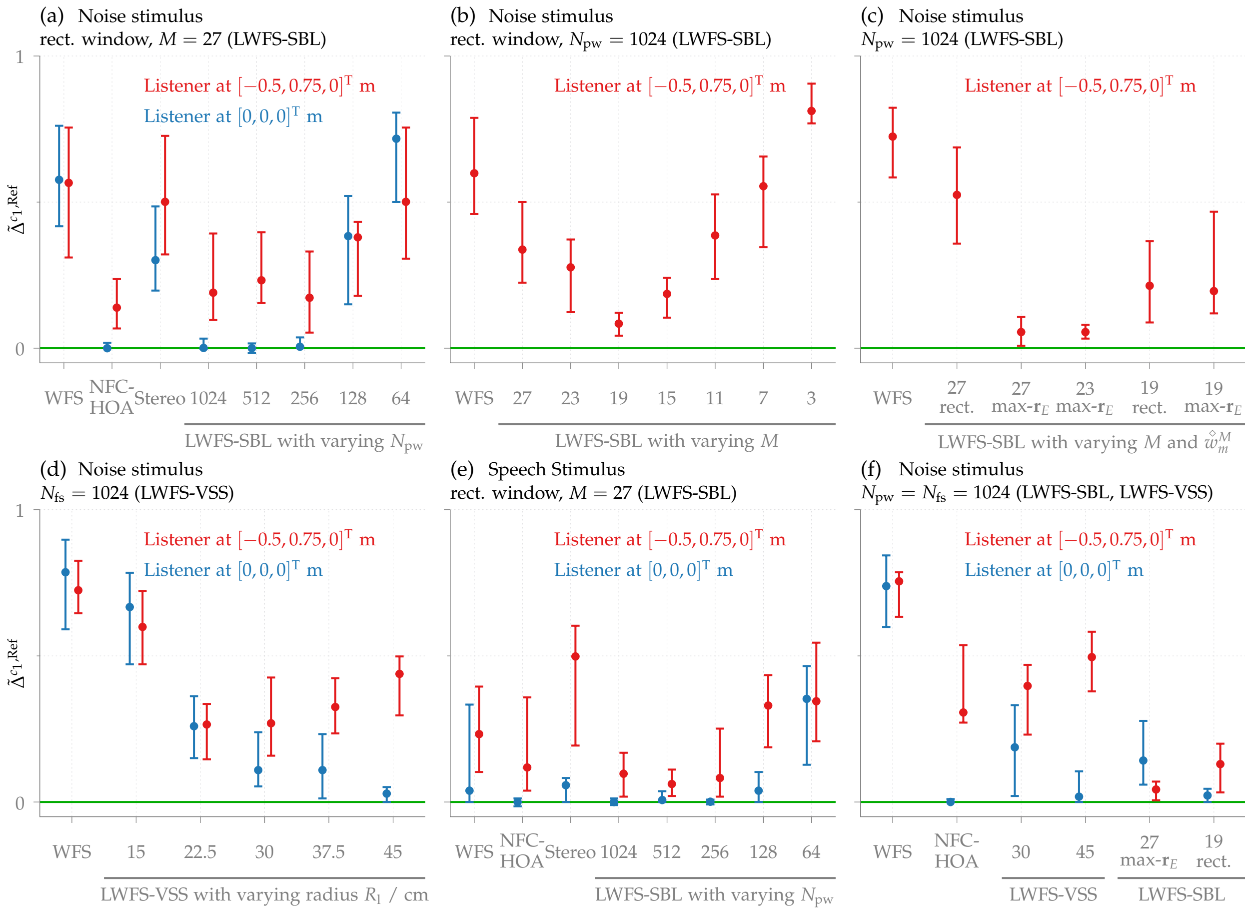


Figure 6.2: The plots show the sample median $\tilde{\Delta}^{c_1,\mathrm{Ref}}$ (bullets) and the 95% Confidence Interval (errorbars) for all ten MUSHRA runs. Both values were computed over all listeners from the difference $\Delta_l^{c_1,\mathrm{Ref}}$ between their individual ratings for the SFS method and the reference condition.

between c_1 and the respective SFS method.

Effect sizes provide a convenient way to illustrate differences between conditions in a general way and thus complement NHST. Here, the Vargha-Delaney A (VDA)[38] is used, as it does not require the data to be normal distributed. Actual effect sizes are interpreted according to the authors' original recommendation: 0.50, 0.56, 0.64, and 0.71 correspond to negligible, small, medium and large effects, respectively.[39]

For the comparison of the results for the two source signals the quantity $\Gamma_l^{c_1}$ as the difference between the two colouration ratings for noise and speech made by the same listener l for the same condition c_1 is introduced. Hereby, a positive value corresponds to a larger rating for the noise stimulus. Again, NHST for the population median $\mu_\Gamma^{c_1}$ and the VDA is utilised to investigate the impact of the source signal.

[38] Vargha and Delaney (2000). "A Critique and Improvement of the "CL" Common Language Effect Size Statistics of McGraw and Wong". In: *Journal of Educational and Behavioral Statistics* 25.2, pp. 101–132, Eq. (14).

[39] Ibid., Tab. 1.

6.2.4 Results

The results of all ten MUSHRA runs are publicly available.[40] For the upcoming discussions, the sample medians $\tilde{\Delta}^{c_1,\mathrm{Ref}}$ of the pairwise differences versus the reference condition are plotted in Fig. 6.2. For the centre listening position, the perceived colouration of LWFS-SBL decreases as the number of plane waves N_{pw} increases.[41] Together

[40] Winter et al. (2018b). *Colouration in Local Wave Field Synthesis – Data*. DOI: 10.5281/zenodo.1158027.

[41] see Fig. 6.2a (blue)

	$H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \le 0$								$H_0 : \mu_\Gamma^{c_1} \le 0$			
	noise				speech							
	centre		off-centre		centre		off-centre		centre		off-centre	
condition c_1	p	VDA	p	VDA	p	VDA	p	VDA	p	VDA	p	VDA
WFS	$< .001$	.96	$< .001$	.99	.013	.74	$< .001$	.91	$< .001$	.89	$< .001$	.80
NFCHOA	.378	.51	$< .001$	.91	.594	.54	$< .001$	.84	.551	.51	.676	.53
Stereo	$< .001$	.93	$< .001$	.99	.024	.70	$< .001$	.97	$< .001$	.87	.131	.57
LWFS-SBL $N_{\mathrm{pw}} = 1024$	.174	.56	$< .001$	.92	.636	.53	$< .001$	.78	.053	.60	$< .001$	.72
LWFS-SBL $N_{\mathrm{pw}} = 512$	.649	.52	$< .001$	.91	.066	.59	$< .001$	.79	.877	.47	.001	.78
LWFS-SBL $N_{\mathrm{pw}} = 256$	.115	.59	$< .001$	.91	.285	.59	$< .001$	.83	.273	.55	.013	.64
LWFS-SBL $N_{\mathrm{pw}} = 128$	$< .001$	.90	$< .001$	.95	.004	.71	$< .001$	.95	$< .001$	.85	.324	.53
LWFS-SBL $N_{\mathrm{pw}} = 64$	$< .001$	.94	$< .001$	.98	$< .001$	.89	$< .001$	.93	$< .001$	.79	.007	.67

Table 6.2: Comparison of p-values and effect sizes (VDA) for the noise and speech stimulus as the dry source signal. For $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \le 0$, p-values below 0.001 value suggests that there is a significant difference in the perceived timbre between c_1 and the reference. For $H_0 : \mu_\Gamma^{c_1} \le 0$, statistical significance is interpreted as a stronger colouration for the noise than for the speech stimulus.

with NFCHOA, a transparent presentation is achieved for $N_{\mathrm{pw}} = 1024, 512$, and 256 as only non-significant small effects were found.[42] The corresponding results for the off-centre listening position are shown in Fig. 6.2a (red): previously transparent presentations also suffer from colouration, now. Significant large differences from the reference were found for all methods.[43]

The influence of M for a rectangular window can be observed in Fig. 6.2b: Beginning from $M = 27$ colouration decreases until an optimum at $M = 19$ is reached. If the modal bandwidth is further decreased, the colouration ratings increase, again. Even for the optimal value of 19, LWFS-SBL is likely to be perceived as coloured at the off-centre position.[44]

The results for the shape of the modal window are shown in Fig. 6.2c: with the max-$\mathbf{r}_E$ window function, the colouration can be further reduced for the off-centre listening position. However, many listeners were still able to distinguish between the reference and the reproduction: For all tested conditions, a significant difference with a large effect of at least 0.85 was observed.[45]

For the centre listening position, the colouration of LWFS-VSS decreases with an increasing radius R_1.[46] The results for the off-centre position are plotted in Fig. 6.2d (red) revealing no obvious dependency on R_1. For both listening positions, none of the parametrisations reached transparent reproduction.[47]

Considering the results displayed in Fig. 6.2a/e for the centre listening position (blue), the perceived colouration for the speech signal is generally shifted towards zero compared to the noise signal. Despite for methods, which have already been rated very close to the reference for the noise stimulus (NFCHOA and LWFS-SBL with $N_{\mathrm{pw}} = 1024, 512, 256$) NHST shows a significant large decrease of the perceived colouration for the speech stimulus.[48] For the mentioned conditions also the effect size for the difference to the reference decreases.[49] A similar yet reduced effect can be observed in Fig. 6.2a/e for the off-centre position (red). The colouration ratings show only a slight shift towards the reference. For both, noise and speech signal, all methods have a significant difference from reference for the off-centre position.[50] Although still being large, the effects for the speech signal (VDA > 0.77) are smaller than for the noise signal

[42] $p > 0.114$, VDA < 0.60, see Tab. 6.2, $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \le 0$, noise, centre

[43] $p < 0.001$, VDA > 0.90, see Tab. 6.2, $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \le 0$, noise, off-centre

[44] $p < 0.001$, VDA ≈ 0.87

[45] $p < 0.001$, VDA > 0.85

[46] see Fig. 6.2d (blue)

[47] $p < 0.001$, VDA > 0.80

[48] $p < 0.001$, VDA > 0.78, see Tab. 6.2, $H_0 : \mu_\Gamma^{c_1} \le 0$, centre

[49] see Tab. 6.2, $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \le 0$, compare noise and speech for centre

	$H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \leq 0$				$H_0 : \mu_\Delta^{c_1,\mathrm{WFS}} \geq 0$				$H_0 : \mu_\Delta^{c_1,\mathrm{NFCHOA}} \geq 0$			
	centre		off-centre		centre		off-centre		centre		off-centre	
condition c_1	p	VDA	p	VDA	p	VDA	p	VDA	p	VDA	p	VDA
WFS	$< .001$	.98	$< .001$	.97					1.000	.99	1.000	.93
NFCHOA	.187	.55	$< .001$	.99	$< .001$	.99	$< .001$	.93				
LWFS-VSS $\rho_l = 30$ cm	$< .001$	.88	$< .001$	.99	$< .001$	.89	$< .001$	.92	1.000	.90	.527	.50
LWFS-VSS $\rho_l = 45$ cm	.004	.77	$< .001$	.97	$< .001$	.96	$< .001$	.86	1.000	.76	.994	.69
LWFS-SBL $M = 27$ max-$\mathbf{r}_E$	$< .001$	.89	$< .001$	.85	$< .001$	.89	$< .001$	.96	1.000	.91	$< .001$	.95
LWFS-SBL $M = 19$ rect.	.018	.68	$< .001$	.93	$< .001$	.97	$< .001$	.95	.994	.67	$< .001$	.86

Table 6.3: p-values and effect sizes (VDA) for the comparison of different synthesis methods. For $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \leq 0$, p-values below 0.001 value suggests that there is a significant difference in the perceived timbre between c_1 and the reference. For $H_0 : \mu_\Delta^{c_1,\mathrm{WFS}} \geq 0$ and $H_0 : \mu_\Delta^{c_1,\mathrm{NFCHOA}} \geq 0$, statistical significance is interpreted as a stronger colouration for the respective SFS method than for c_1.

(VDA > 0.90).

The results of the comparison between the (L)SFS methods are plotted in Fig. 6.2f. Visual inspection as well as the p-values for $H_0 : \mu_\Delta^{c_1,\mathrm{WFS}} \geq 0$ listed in Tab. 6.3 suggest that all parametrisations of LWFS-SBL and LWFS-VSS lead to a significant improvement of the synthesis w.r.t. perceived colouration compared to conventional WFS. This holds for both listening positions, since large effects (VDA > 0.86) could be observed. For the centre listening position, no method outperforms NFCHOA, as non-significant differences were observed.[51] For the off-centre position, parametrisations for LWFS-SBL were found which are perceived significantly less coloured than NFCHOA with a large effect.[52] For LWFS-VSS, none of the parametrisations achieved this improvement. Transparent synthesis is not likely to be achieved by any SFS method for the off-centre position as significant differences with large effects were found.[53]

6.2.5 Discussion

The discussion follows the same structure as for the localisation experiments in Sec. 5.3.4. In the following, the results are further analysed in the context of psychoacoustics with the invoked cues. Afterwards, a potential connection between the predictions of the geometric model introduced in Ch. 4 and the perceived colouration is investigated.

Relation to Psychoacoustic Phenomena: The Composite Loudness Level (CLL)[54] essentially models the perceived binaural loudness level as a function of frequency incorporating the selectivity of humans. Differences in the CLL to a reference have been shown to correspond to the perceived colouration.[55,56] The concept has been successfully applied – partly in modified versions – to discuss or predict colouration in stereo panning[57], VBAP[58], HOA,[59] and WFS.[60] The model used to extract CLL follows the descriptions by Pulkki et al.[61] with minor modifications: The dry source signal filtered by the BTF $H^{\mathrm{SFS}}_{\{\mathrm{L,R}\}}(\mathbf{x}, \phi_\mathrm{h}, \omega)$ constitutes the input to the model. The binaural signal is further processed with the same bandpass filter and Gammatone filterbank as described in Sec. 5.3.4 in order to model the middle ear and the frequency selectivity of the human cochlea. For each auditory band the time-averaged signal power is computed. In the original publication, the values are compressed by $\sqrt[4]{\cdot}$ resulting in

51 $p > 0.994$, see Tab. 6.3, $H_0 : \mu_\Delta^{c_1,\mathrm{NFCHOA}} \geq 0$, centre

52 $p > 0.994$, VDA > 0.85 see Tab. 6.3, $H_0 : \mu_\Delta^{c_1,\mathrm{NFCHOA}} \geq 0$, off-centre

53 $p < 0.001$, VDA > 0.84, see Tab. 6.3, $H_0 : \mu_\Delta^{c_1,\mathrm{Ref}} \leq 0$, off-centre

54 Pulkki et al. (1999). "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model". In: *J. Aud. Eng. Soc.* 47.4, pp. 203–217

55 Ono et al. (Nov. 2001). "Binaural Modeling of Multiple Sound Source Perception: Methodology and Coloration Experiments". In: *Proc. of 111th Aud. Eng. Soc. Conv.* New York, USA

56 Ono et al. (Apr. 2002). "Binaural Modeling of Multiple Sound Source Perception: Coloration of Wideband Sound". In: *Proc. of 112th Aud. Eng. Soc. Conv.* Munich, Germany

57 Pulkki et al., op. cit.

58 Pulkki (May 2001). "Coloration of Amplitude-Panned Virtual Sources". In: *Proc. of 110th Aud. Eng. Soc. Conv.* Amsterdam, The Netherlands

59 Frank, op. cit., Cha. 5.

60 Wierstorf et al. (Mar. 2015). "Klangverfärbung in der Wellenfeldsynthese - Experimente und Modellierung". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 490–493.

61 Pulkki et al., op. cit.

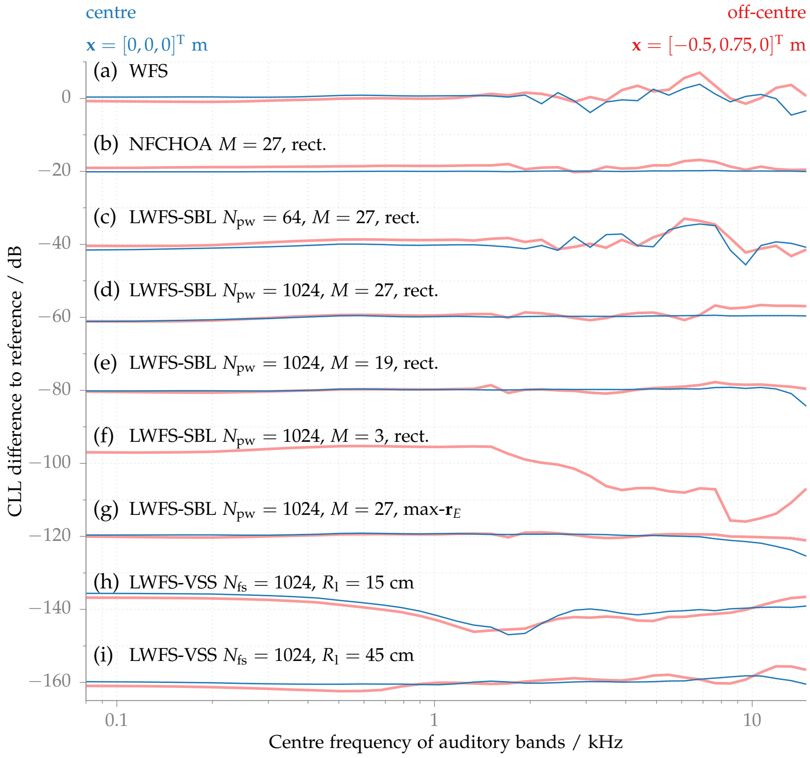


Figure 6.3: The plot shows the CLL difference between selected conditions and the reference. The noise stimulus was used as the dry source signal. The centre and the off-centre listening positions are marked in blue and red, respectively. Each condition was incrementally shifted about −20 dB to enhance visibility.

the loudness measured in sone.[62] The loudness of both ears is added for each band and converted to a loudness level spectra measured in phon.[63] Since, however, no calibrated measurement system was used in the present experiments, the values of the CLL are given in decibel. Moreover, only differences in CLL w.r.t. the reference condition are of interest.

[62] Ibid., Eq. (2).

[63] Ibid., Eq. (3).

The CLL differences for selected conditions are shown in Fig. 6.3. The CLL were computed for the noise stimulus as the dry source signal. It should be noted, that the overall loudness of the BTFs was adjusted to the overall loudness for the colouration experiments. Thus, the showed CLL differences directly correspond to the ear signals provided to the test subjects.

For WFS[64], the CLL difference show strong fluctuations at high frequencies caused by spatial aliasing. They lead to a perceivable difference w.r.t. the reference for both listening positions. The findings agree with the results from Wierstorf[65] where all investigated synthesis setups driven by WFS led to colouration.

[64] see Fig. 6.3a

[65] Wierstorf, op. cit., Sec. 5.2.

For the centre listening position, no CLL difference over the whole frequency range can be observed for NFCHOA.[66] This is also reflected by the colouration ratings in the experiment, where no difference to the reference was present. It is further in agreement with prior expectations, since NFCHOA achieves high synthesis accuracy around the centre of the loudspeaker array and is expected to perform at least as well as the LSFS techniques for this listening position. Furthermore, it was shown in Sec. 4.3.2, that $M = 27$ corresponds to the best trade-off between spatial aliasing and SBL artefacts for the

[66] see Fig. 6.3b

employed loudspeaker setup.

For LWFS-SBL, strong fluctuations at high frequencies are present, if the number of plane waves N_{pw} is too low.[67] As these artefacts are strongly attenuated for $N_{\mathrm{pw}} = 1024$,[68] they can be attributed to spatial aliasing caused by the coarse discretisation of the involved PWD. For the off-centre position, CLL differences are however still present. It was shown in Sec. 4.4.3, that the chosen spatial bandwidth $M = 27$ is too high for this listening position.[69] Although diminished, spatial aliasing is not completely prevented. The colouration ratings agree with these findings, as this parameterisation of LWFS-SBL was found transparent for the centre and still coloured for the off-centre listening position.

[67] see Fig. 6.3c

[68] see Fig. 6.3d

[69] see pos. 7 in Fig. 4.26

The influence of the spatial bandwidth M on the colouration was presented in combination with Fig. 6.2b and can be explained as follows: as already discussed for the geometric model for LWFS-SBL, the optimal choice M postulates a trade-off between spatial aliasing and SBL artefacts. For values of M below 19, colouration increases due to lowpass filtering of the reproduced sound field, which is an artefact of the SBL. For $M = 3$, the CLL difference is plotted in Fig. 6.3f and confirms the described level loss at high frequencies. Even for the optimal $M = 19$, deviations in the CLL spectrum shown in Fig. 6.3d (red) are observable. Thus, this condition is still perceived as coloured. For M above 19, colouration increases again due to significant aliasing contributions.[70] The effect of the max-$\mathbf{r}_E$ weighting on the CLL difference is shown in Fig. 6.3g: For the off-centre position, the spectrum is close to flat. Although still distinguishable from the reference, this parametrisation led to a reduction of colouration compared to the conventional SFS techniques and for the off-centre position. It can, however, be deduced from the corresponding blue line and the colouration ratings in Fig. 6.2f that it is likely to be perceived coloured for the centre position: A level loss at high frequencies due to the SBL is present. In total, a single position-independent choice for the spatial bandwidth M and the modal window does not lead to optimal timbral properties for all listening positions. Again, this agrees with the investigations in Sec. 4.4.3 on spatial aliasing and SBL.

[70] see e.g. Fig. 6.3d (red)

For LWFS-VSS, the small radius $R_{\mathrm{l}} = 15$ cm of the virtual SSD causes large deviations in CLL at the lower frequencies.[71] Consequently, its colouration was rated as very different. Similar to the discussion on localisation in Sec. 5.3.4, the artefacts can be explained by the low-frequency behaviour of the involved focused sources leading to undesired interferences. The effect can be diminished by increasing the radius, i.e. $R_{\mathrm{l}} = 45$ cm.[72] For the centre, CLL differences are only present at high frequencies. Thus, this condition was rated very close to the reference. For the off-centre listening position, spatial aliasing is more pronounced and the colouration generally increases.

[71] see Fig. 6.3h

[72] see Fig. 6.3i

Generally, it is not surprising that the speech stimulus led to less colouration as its spectrum exhibits less energy at high frequencies, where most of the artefacts introduced by the different SFS tech-

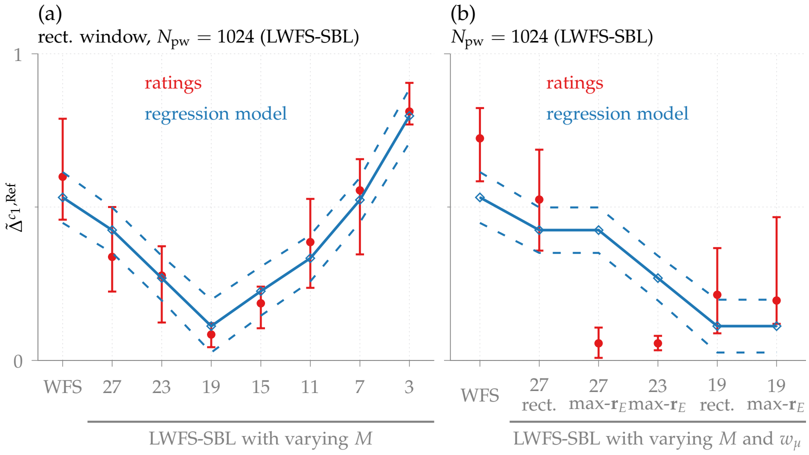


Figure 6.4: The red plots show the sample median $\tilde{\Delta}^{c_1,\mathrm{Ref}}$ (bullets) and the 95% Confidence Interval (errorbars) for the two MUSHRA runs already plotted in Fig. 6.2b and 6.2c The prediction of the linear mixed effects model described by (6.1) is plotted in solid blue with 95% Confidence Interval (dashed). The model was fit to the data in (a).

niques are present. As the spatial aliasing increases for the off-centre listening position, colouration is more likely to be perceived even for the speech stimulus.

Relation to the Geometric Model: The geometric model in Ch. 4 is able to predict the aliasing frequency f^{S} and the SBL frequency f^{B}. These estimate the artefact-free frequency range. Contrary to the localisation results, the colouration is not measured on an absolute scale. As already outlined in Sec. 6.2.3, equivalent interpretation of the MUSHRA scale among listeners and runs cannot be guaranteed.[73] A pooled analysis of the acquired data as for the localisation experiments is thus not sensible. As a proof of concept, the relation between colouration ratings and the mentioned frequency will be exemplarily discussed for two of the MUSHRA runs. The selected results are, again, shown in Fig. 6.4. The linear mixed effects model

$$\Delta_l^{c_1,\mathrm{Ref}} = \beta_{00} + \beta_{10} f^{\mathrm{S}} + \beta_{01} f^{\mathrm{B}} + \beta_{11} f^{\mathrm{S}} f^{\mathrm{B}} + \gamma_l + \epsilon_{lc_1} \tag{6.1a}$$

$$\gamma_l \overset{\text{i.i.d.}}{\underset{l}{\sim}} \mathcal{N}(0, \sigma_0) \tag{6.1b}$$

$$\epsilon_{lc_1} \overset{\text{i.i.d.}}{\underset{lc_1}{\sim}} \mathcal{N}(0, \sigma) \tag{6.1c}$$

for the pairwise difference between the reference and a condition c_1 is fit to the data shown in Fig. 6.4a. The fixed effects β describe a multiplicative interaction between the two frequencies. A random intercept γ_l is added to model shifts in the MUSHRA scale between different listeners. It partly models the effect of differently interpreted scales, but does not cover skewing effects reported by Zielinski et al.[74] For WFS the infinite SBL frequency was truncated to 20 kHz. The result of the regression is plotted in blue. All fixed effects are significant.[75] As stated by Kruschke,[76] care has to be taken when interpreting models with interaction terms. A possible interpretation of the fixed effects considers the derivatives w.r.t. each of the independent variables. With the BLUEs of the fixed effects $\hat{\beta}$,

[73] Zielinski et al. (Oct. 2007). "Potential Biases in MUSHRA Listening Tests". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA.

[74] Ibid., Sec. 4.5.

[75] $p \lll 10^{-5}$, t-test

[76] Kruschke (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* 2nd ed. Academic Press, Sec. 11.2.

they read

$$\frac{\partial \Delta_l^{c_1,\mathrm{Ref}}}{\partial f^{\mathrm{S}}} \approx \hat{\beta}_{10} + \hat{\beta}_{11} f^{\mathrm{B}} \approx 9.2 \cdot 10^{-5}\mathrm{Hz}^{-1} - 5.9 \cdot 10^{-9}\mathrm{Hz}^{-2} \cdot f^{\mathrm{B}} \tag{6.2a}$$

$$\frac{\partial \Delta_l^{c_1,\mathrm{Ref}}}{\partial f^{\mathrm{B}}} \approx \hat{\beta}_{01} + \hat{\beta}_{11} f^{\mathrm{S}} \approx 6.6 \cdot 10^{-5}\mathrm{Hz}^{-1} - 5.9 \cdot 10^{-9}\mathrm{Hz}^{-2} \cdot f^{\mathrm{S}} \tag{6.2b}$$

An interesting question is the weighting of the two phenomena, namely SBL and spatial aliasing, for the judgement of colouration. If both derivatives are equal, a change of f^{S} or f^{B} about the same amount has the same impact on the colouration rating. This leads to the relation

$$f^{\mathrm{S}} \approx \frac{\hat{\beta}_{10} - \hat{\beta}_{01}}{\hat{\beta}_{11}} + f^{\mathrm{B}} \approx -4.4\mathrm{kHz} + f^{\mathrm{B}}\,. \tag{6.3}$$

The spatial aliasing frequency can be about 4.4 kHz smaller than the SBL frequency is order to have the same (local) effect on the colouration. Thus, the SBL artefacts are more critical then spatial aliasing artefacts. It is however emphasised, that the regression was applied to one single MUSHRA run and general statements about the relation are not to be expected. The model fitted to the data in Fig. 6.4a is used to predict the ratings for Fig. 6.4b: A general offset between ratings and prediction can be observed. The general trend of an increasing colouration with increasing M is correctly predicted. However, the positive effect of the modal weighting function is not covered. This was already mentioned in the description of the geometric model in Sec. 4.3.2. The regression shows, that the results are not easily generalised towards other MUSHRA experiments.

6.3 Summary

In this chapter, the selected analytic methods (L)SFS were evaluated w.r.t. their timbral fidelity. The influence of the different LWFS-SBL parameters on the colouration agrees well with the expectations raised by prior analysis of physical properties in Ch. 3 and Ch. 4. In the case of LWFS-VSS, the near-field artefacts introduced by the utilised focused sources play an important role: For the centre listening position, the largest of the tested local target region radii led to the smallest colouration rating, which contradicts the prediction of the geometric model. For the off-centre position, a smaller radius has to be chosen due to the increased spatial aliasing artefacts, which dominate the near-field artefacts for this listening position.

The frequency spectrum of human speech comprises most energy at frequencies below 4 kHz. Colouration caused by spatial aliasing and SBL which is occurring at high frequencies is not evoked by a signal which does not have components at these frequencies. The comparison of the colouration ratings for broadband noise and speech confirmed this.

For the given loudspeaker setup and the desired point source, both LWFS techniques are able to decrease perceived colouration compared to conventional WFS. For the tested off-centre position, only LWFS-SBL outperformed NFCHOA, if parametrised accordingly. Hereby, the max-$\mathbf{r}_E$ window function reduced colouration. However, for the investigated off-centre listening position, no (L)SFS parametrisation leads to a fully transparent presentation of the desired point source.

This chapter has shown that a more transparent reproduction compared to conventional SFS is possible with LWFS. A comparison to the predictions of the geometric model indicated that the upper frequency limit due to SBL might be more critical w.r.t. colouration than the spatial aliasing frequency. This has, however, to be confirmed on a larger experimental scale, since only a single MUSHRA run was investigated as a proof of concept.

7 Conclusion

7.1 Summary

This thesis analysed the physical and perceptual properties of selected approaches for LSFS and compared them to conventional SFS. As a central research question, the potential improvement of physical and perceptual fidelity with LSFS was investigated.

In Ch. 2, the fundamentals of linear acoustics were revisited. A different view on the integral formulation of the linearised wave equation was presented, where the sound field was separated into a homogeneous and an inhomogeneous part with respect to a given boundary. The resulting formulation of SFS coincides with well established alternatives, namely the equivalent scattering problem and the simple source formulation.[1,2]

In addition to WFS and NFCHOA for conventional SFS, Ch. 3 introduced the mathematical foundations of LWFS-SBL and LWFS-VSS as the selected approaches to LSFS. Special attention was drawn to the according discrete-time implementation and consequences on the synthesis results. As already done for LWFS-VSS by Spors and Ahrens,[3] a discrete-time implementation for LWFS-SBL is proposed, which concatenates existing components for WFS and NFCHOA. This is of special interest, as audio rendering software with real-time capability are already available for the mentioned conventional SFS approaches.[4] The involved PWD for virtual point sources was stabilised using a dual-band approach with a frequency crossover between WFS and LWFS-SBL.

A geometric model to predict spatial aliasing as a function of the virtual sound field, the target region, and the loudspeaker array geometry was presented in Ch. 4. In particular, a trade-off between the available listening area and spatial aliasing as a function of position was shown. As a general guideline, LSFS becomes less effective the closer the target region is located to the active loudspeakers. The impact of some parameters on the synthesis accuracy cannot be predicted by the geometric model, e.g. the modal window in NFCHOA and LWFS-SBL. The high-frequency approximation made to derive the model does not incorporate the near-field/low-frequency behaviour of the focused sources used in LWFS-VSS. Besides its prediction capabilities, the model can also be applied to design optimal spacing patterns for the employed loudspeaker array for a given vir-

[1] Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, Sec. 8.7.

[2] Ahrens (2010). "The Single-layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources". PhD thesis. Berlin, Germany: Technische Universität Berlin, Sec. 2.6.2.

[3] Spors and Ahrens (Oct. 2010b). "Local Sound Field Synthesis by Virtual Secondary Sources". In: *Proc. of 40th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Tokyo, Japan.

[4] Geier and Spors (Nov. 2012). "Spatial Audio with the SoundScape Renderer". In: *Proc. of 27th Tonmeistertagung – VDT Internation Convention*. Cologne, Germany.

tual sound field and listening area. This was exemplarily shown for WFS.

The four (L)SFS methods were perceptually evaluated in Ch. 5 and Ch. 6 regarding their spatial and timbral fidelity for a circular loudspeaker array. The findings for WFS confirmed prior research[5] as a good localisation accuracy is preserved and severe colouration is caused by spatial aliasing. Results from the literature[6,7] on localisation accuracy in HOA panning agree with the behaviour of NFCHOA. In general, a higher order and the max-$\mathbf{r}_E$ weighting leads to more accurate localisation especially for off-centre positions. With the tested parametrisation, NFCHOA is only transparent w.r.t. timbre in the centre position. LWFS-SBL turned out as very easy to handle, since the localisation bias is only slightly affected by the chosen parameters. Best accuracy could be achieved with the max-$\mathbf{r}_E$ weighting for high orders. For colouration, the trade-off between the SBL and spatial aliasing could be accurately explained by the frequencies predicted by the geometric model. LWFS-VSS turned out to be more unpredictable regarding the effect of its parametrisation. This is mostly due to the already mentioned characteristics of the employed focused sources, which have to be additionally considered. The introduced artefacts contradict the central paradigm of LSFS, where the synthesis accuracy increases with a smaller target region. Thus, its behaviour is also less accurately predicted by the geometric model. For the chosen parameters, LWFS-VSS leads to inferior spatial and especially timbral fidelity compared to LWFS-SBL. It was, however, possible with suitable parametrisation to improve colouration and preserve good localisation in comparison to its baseline algorithm WFS.

[5] Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin, Sec. 5.1 and 5.2.

[6] Frank (June 2013a). "Phantom Sources using Multiple Loudspeakers in the Horizontal Plane". PhD thesis. Graz, Austria: Institute of Electronic Music and Acoustics, Cha. 3.

[7] Stitt (June 2015). "Ambisonics and Higher-Order Ambisonics for Off-Centre Listeners: Evaluation of Perceived and Predicted Image Direction". PhD thesis. Belfast, UK: Queen's University Belfast, Sec. 5.5.

7.2 Outlook

In both investigated LWFS methods, intermediate representations, namely the PWD and the VSSs, were necessary to compute the driving signals. These representations had to be discretised, which potentially led to additional spatial aliasing as discussed in Ch. 4. Additional effort has to be spent in order to find mathematical solutions to circumvent the intermediate representations not only to avoid artefacts but also to reduce computational complexity. Promising results were published by Hahn et al.[8] for virtual plane waves in LWFS-SBL. Moreover, the high-frequency approximations for bandwidth-limited sound fields derived in Sec. A.1 may be directly fed into WFS to achieve LSFS. Here, the impact of the involved approximations on the synthesis accuracy has to be further analysed.

[8] Hahn et al. (Aug. 2017). "Synthesis of a spatially band-limited plane wave in the time-domain using wave field synthesis". In: *Proc. of 25th European Signal Processing Conference (EUSIPCO)*. Kos Island, Greece, pp. 673–677.

So far, LSFS was only investigated for static scenarios. Since a major feature of LSFS is the optimised synthesis for a given location, a dynamically tracked listener has to be considered. This requires to revisit not only the time-domain implementations of the LSFS approaches but also the perceptual attributes. In addition, dynamic scenarios may include moving sound sources, which were recently investigated WFS by Firtha[9] on theoretical level.

[9] Firtha (2019). "A Generalized Wave Field Synthesis Framework with Application for Moving Virtual Sources". PhD thesis. Budapest University of Technology and Economics.

As a proof of concept, the geometric model was applied to design loudspeaker arrays in WFS. This functionality should be further extended towards other (L)SFS approaches. An interesting application is the array design for large-scale sound reinforcement, where the virtual sound field is fixed for certain scenarios.[10] For this, future models need to support the frequency-dependent directivity of the employed loudspeakers including individual drivers for different frequency ranges.

[10] Schultz (2016). "Sound Field Synthesis for Line Source Array Applications in Large-Scale Sound Reinforcement". PhD thesis. University of Rostock.

Within this thesis, particularly important aspects of the overall sound quality[11] were investigated. Assessing the preference of the listeners[12] for different (L)SFS approaches could be another step towards an overall quality rating. It was shown in the cited publication, that the preparation of the audio material a.k.a. the mix largely influences the ratings. Thus, experienced staff to master such scenes for (L)SFS is needed. In additional, more sophisticated software tools for mastering in LSFS similar to the ones used by the Ambisonics community[13] have to be developed.

[11] Rumsey et al. (2005). "On the relative importance of spatial and timbral fiedelities in judgements of degraded multichannel audio quality". In: *J. Acoust. Soc. Am.* 118.2, pp. 968–976.

[12] Wierstorf et al. (2018). "Listener Preference for Wave Field Synthesis, Stereophony, and Different Mixes in Popular Music". In: *J. Aud. Eng. Soc.* 66.5, pp. 385–396.

[13] Zotter and Frank (2019). *Ambisonics*. Springer International Publishing, Sec. 5.9.

The localisation results in Ch. 5 were discussed in conjunction with the relevant binaural cues predicted by an auditory model. As an additional step, models might directly predict the perceived direction. This was already done by Wierstorf[14] for WFS and NFCHOA using the ITD model of Dietz et al.[15] As mentioned by Frank,[16] more sophisticated probabilistic models are necessary to predict effects like the source splitting in NFCHOA. Promising approaches for auditory localisation of multiple sources including head movements are available.[17,18,19]

[14] Wierstorf, op. cit., Ch. 6.

[15] Dietz et al. (2011). "Auditory model based direction estimation of concurrent speakers from binaural signals". In: *Speech Communication* 53.5, pp. 592–605.

[16] Frank, op. cit., Sec. 3.4.

[17] May et al. (Apr. 2015). "Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues". In: *Proc. of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brisbane, Australia, pp. 2679–2683.

[18] Schymura et al. (Sept. 2015). "Binaural Sound Source Localisation and Tracking using a dynamic Spherical Head Model". In: *Proc. of 16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Dresden, Germany, pp. 165–169.

[19] Ma and Brown (Sept. 2016). "Speech Localisation in a Multitalker Mixture by Humans and Machines". In: *Proc. of 17th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. San Francisco, USA, pp. 3359–3363.

In Sec. 6.2.5, the connection between the colouration ratings and the prediction of the geometric model could only be discussed to a limited extent. This was partly caused by the used MUSHRA test paradigm making comparisons between different runs difficult. Further research has to augment the findings in the present work by using alternative test paradigms.

The thesis did not take any effects of the playback room into account as the theory and experiments assumed free-field conditions. It was shown by Erbes and Spors[20] for WFS, that room reflections mitigate the spectral fluctuations caused by spatial aliasing. Thus, colouration due to the aliasing might be less critical. Additional experiments for the timbral fidelity of (L)SFS in rooms have to be conducted.

[20] Erbes and Spors (Mar. 2017). "Influence of the Listening Room on Spectral Properties of Wave Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Kiel, Germany, pp. 1057–1060.

A High-Frequency Approximations

As a central concept of the upcoming derivations, the SPA is used as a integral approximation. Given a complex-valued function $F(u) = A_F(u)e^{+\mathrm{j}\phi_F(u)}$ with its phase term rapidly oscillating compared to its slowly changing amplitude, the following approximation of the integral[1,2,3]

$$\int_a^b F(u)\mathrm{d}u \approx \sum_{u^*\in[a,b]} F(u^*)\sqrt{\frac{2\pi}{|\phi_F''(u^*)|}}\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\mathrm{sgn}(\phi_F''(u^*))} \tag{A.1}$$

holds. It constitutes the summation over the stationary points u^* in the interval $[a,b]$, for which the first-order derivative of the phase $\phi_F'(u^*)$ vanishes and the second-order derivative $\phi_F''(u^*)$ is non-zero. The integral is approximately zero, if no stationary point is present in $[a,b]$. The approximation is based upon the idea that the integration over a complex sinusoid with a rapidly changing phase yields zero except for the contributions from u^* and its neighbourhood. Rigorous derivations of the approximation are given in the mentioned publications.[4,5,6]

[1] Wong (2001). *Asymptotic Approximations of Integrals*. Vol. 34. Classics in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), Eq. (3.2).

[2] Bleistein and Handelsman (1986). *Asymptotic Expansions of Integrals*. New York, USA: Dover Publications, Eq. (6.5.1).

[3] Bleistein (1984). *Mathematical Methods for Wave Phenomena*. Orlando, USA: Academic Press, Eq. (2.7.18).

[4] Wong, op. cit., Sec. II.3.

[5] Bleistein and Handelsman, op. cit., Sec. 6.1.

[6] Bleistein, op. cit., Sec. 2.7.

A.1 Bandwidth-Limited Sound Fields

The goal of this section is to derive a high-frequency approximation of the sound field

$$S_M^{\mathrm{B}}(\mathbf{x},\omega) = \sum_{m=-\infty}^{\infty} \mathring{w}_m^M \mathring{S}_m(\rho,\omega)\mathrm{e}^{+\mathrm{j}m\phi} \tag{A.2}$$

bandwidth-limited in the Circular Harmonics domain by multiplying the coefficients $\mathring{S}_m(\rho,\omega)$ of the original sound field $S(\mathbf{x},\omega)$ with the modal window $\mathring{w}_m^M$. It will relate the bandwidth-limited version to the original sound field in the spatial domain. In order to apply the SPA to the summation, an integral representation of the ICHT is derived in Sec. A.1.1. The resulting integrals are approximated in Sec. A.1.2 and Sec. A.1.3 in order to derive the desired relation.

A.1.1 Integral Representation of Inverse Circular Harmonics Transform

In general, the ICHT of a 2π-periodic function $F(\phi)$ is given as the Fourier series

$$F(\phi) = \sum_{m=-\infty}^{\infty} \mathring{F}_m e^{+jm\phi} \tag{A.3}$$

with the coefficients $\mathring{F}_m$. They can can be expressed as samples of a continuous function denoted as $\mathring{F}(\mu)$. The sampling is described by the multiplication of $\mathring{F}(\mu)$ with a Dirac comb.[7] Using the sifting property of the Dirac delta distribution,[8,9] the sound field can be expressed as

$$F(\phi) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \delta(\mu - m)\mathring{F}(\mu) e^{+j\mu\phi}\, d\mu\,. \tag{A.4}$$

The Dirac Comb is expressed by its Fourier series,[10] which results in

$$F(\phi) = \sum_{\eta=-\infty}^{\infty} \underbrace{\int_{-\infty}^{\infty} \mathring{F}(\mu) e^{+j\mu(\phi - 2\pi\eta)}\, d\mu}_{=F_{\text{cont}}(\phi - 2\pi\eta)}\,. \tag{A.5}$$

$F_{\text{cont}}(\phi)$ can be interpreted as the non-periodic part of $F(\phi)$, which is equal to $F(\phi)$ only for $\phi \in [0, 2\pi]$ and zero, otherwise. $F(\phi)$ is expressed as the sum of shifted versions $F_{\text{cont}}(\phi)$ to preserve its 2π-periodicity. The principle is analogous to the spectral repetitions occurring in the temporal frequency domain, when sampling is applied to a continuous time signal.[11] The corresponding CHT for the continuous coefficients reads

$$\mathring{F}(\mu) = \frac{1}{2\pi} \int_0^{2\pi} F(\phi) e^{-j\mu\phi}\, d\phi \tag{A.6}$$

Inserting the right-hand-side of (A.3) for $F(\phi)$, the continuous coefficients are related to the discrete coefficients via

$$\mathring{F}(\mu) = \frac{1}{2\pi} \int_0^{2\pi} \sum_{m=-\infty}^{\infty} \mathring{F}_m e^{+jm\phi} e^{-j\mu\phi}\, d\phi = \sum_{m=-\infty}^{\infty} \mathring{F}_m \operatorname{sinc}\left((m-\mu)\pi\right)\,. \tag{A.7}$$

$\mathring{F}(\mu)$ states the interpolation of $\mathring{F}_m$ using the sinus cardinalis[12] as the interpolation kernel.

[7] Girod et al. (2001). *Signal and Systems*. Wiley, Sec. 11.3.1.

[8] Ibid., Eq. (8.15).

[9] Gel'fand and Shilov (1964). *Generalized functions: Vol. 1. Properties and operations*. New York, London: Academic Press, p. 4.

[10] Girod et al., op. cit., Eq. (11.12).

[11] Ibid., Sec. 11.3.

[12] Ibid., Eq. (9.21).

A.1.2 Approximation of Circular Harmonics Transform

Applying (A.6) to the sound field $S(\mathbf{x}, \omega)$ yields the continuous CHT coefficients

$$\mathring{S}(\mu, \rho, \omega) = \frac{1}{2\pi} \int_0^{2\pi} S(\mathbf{x}, \omega) e^{-j\mu\phi}\, d\phi\,. \tag{A.8}$$

The amplitude-phase notation introduced in Sec. 2.3 is used to express the sound field as $A_S(\mathbf{x}, \omega) e^{+j\Phi_S(\mathbf{x}, \omega)}$. Thus, the phase term involved in the integral and its corresponding first- and second-order derivative read

$$\Phi(\phi) = \Phi_S(\mathbf{x}, \omega) - \mu\phi\,, \tag{A.9a}$$

$$\frac{\partial \Phi(\phi)}{\partial \phi} = \Phi'_{S,\phi}(\mathbf{x}, \omega) - \mu\,, \tag{A.9b}$$

$$\frac{\partial^2 \Phi(\phi)}{\partial \phi^2} = \Phi''_{S,\phi\phi}(\mathbf{x}, \omega)\,. \tag{A.9c}$$

The first- and second-order derivative of Φ_S w.r.t. ϕ are denoted by $\Phi'_{S,\phi}$ and $\Phi''_{S,\phi\phi}$, respectively. In order to find the stationary point ϕ^*, the first-order derivative in (A.9b) is set to zero. The resulting SPA condition reads

$$\mu \stackrel{!}{=} \Phi'_{S,\phi}(\mathbf{x}, \omega) = \left\langle \frac{\partial \mathbf{x}}{\partial \phi} \middle| \nabla \Phi_S(\mathbf{x}, \omega) \right\rangle = -\left\langle \frac{\partial \mathbf{x}}{\partial \phi} \middle| \mathbf{k}_S(\mathbf{x}, \omega) \right\rangle\,. \tag{A.10}$$

The second equality is established by applying the chain-rule of differentiation. The gradient is replaced by the local wavenumber vector $\mathbf{k}_S(\mathbf{x}, \omega)$ defined in (2.53) for the last equality. The solution to the condition is the stationary phase point $\phi^* = \phi^*(\rho, \mu) \in [0, 2\pi]$ with the corresponding coordinate is denoted as $\mathbf{x}^* = \mathbf{x}^*(\mu)$. The derivative of (A.10) w.r.t. μ with the stationary phase point inserted reads

$$1 = \frac{\partial \Phi'_{S,\phi}(\mathbf{x}^*(\mu), \omega)}{\partial \mu} = \frac{\partial \phi^*}{\partial \mu} \Phi''_{S,\phi\phi}(\mathbf{x}^*, \omega)\,. \tag{A.11}$$

It is important for later derivations. Finally, the SPA of (A.8) reads

$$\mathring{S}(\mu, \rho, \omega) \approx \frac{\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\operatorname{sign}(\Phi''_{S,\phi\phi}(\mathbf{x}^*,\omega))}}{\sqrt{2\pi |\Phi''_{S,\phi\phi}(\mathbf{x}^*, \omega)|}} \underbrace{A_S(\mathbf{x}^*, \omega) \mathrm{e}^{+\mathrm{j}\Phi_S(\mathbf{x}^*,\omega)}}_{S(\mathbf{x}^*,\omega)} \mathrm{e}^{-\mathrm{j}\mu\phi^*}\,. \tag{A.12}$$

A.1.3 Approximation of Inverse Circular Harmonics Transform

Utilizing (A.5) for the bandwidth-limited sound field $S_M^{\mathrm{B}}(\mathbf{x}, \omega)$ yields the expression

$$S_M^{\mathrm{B}}(\mathbf{x}, \omega) = \sum_{\eta=-\infty}^{\infty} \int_{-\infty}^{\infty} \mathring{w}^M(\mu)\, \mathring{S}(\mu, \rho, \omega) \mathrm{e}^{+\mathrm{j}\mu(\phi - 2\pi\eta)}\, \mathrm{d}\mu\,. \tag{A.13}$$

The SPA of the sound field's coefficients (A.12) is inserted into (A.13) for $\mathring{S}(\mu, \rho, \omega)$. The resulting relation reads

$$S_M^{\mathrm{B}}(\mathbf{x}, \omega) \approx \sum_{\eta=-\infty}^{\infty} \int_{-\infty}^{\infty} \mathring{w}^M(\mu) \frac{\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\operatorname{sign}(\Phi''_{S,\phi\phi}(\mathbf{x}^*,\omega))}}{\sqrt{2\pi |\Phi''_{S,\phi\phi}(\mathbf{x}^*, \omega)|}} \cdot A_S(\mathbf{x}^*, \omega) \mathrm{e}^{+\mathrm{j}\Phi_S(\mathbf{x}^*,\omega)} \mathrm{e}^{+\mathrm{j}\mu(\phi - 2\pi\eta - \phi^*)}\, \mathrm{d}\mu\,. \tag{A.14}$$

Note, that $\mathbf{x}$ and $\mathbf{x}^*$ have to share the same radius ρ, since the SPA derived in Sec. A.1.2 does only involve the azimuth angle. The phase term involved in the integral and its according derivatives read

$$\Phi(\mu) = \Phi_S(\mathbf{x}^*,\omega) + \mu(\phi - 2\pi\eta - \phi^*) \tag{A.15a}$$

$$\Phi'_\mu(\mu) = \frac{\partial\phi^*}{\partial\mu}\underbrace{\left(\Phi'_{S,\phi}(\mathbf{x}^*,\omega) - \mu\right)}_{=0\text{ , see (A.9b)}} + (\phi - 2\pi\eta - \phi^*) \tag{A.15b}$$

$$\Phi''_{\mu\mu}(\mu) = \frac{\partial^2\phi^*}{\partial\mu^2}\underbrace{\left(\Phi'_{S,\phi}(\mathbf{x}^*,\omega) - \mu\right)}_{=0\text{ , see (A.9b)}} + \underbrace{\frac{\partial\phi^*}{\partial\mu}\left(\frac{\partial\phi^*}{\partial\mu}\Phi''_{S,\phi\phi}(\mathbf{x}^*,\omega) - 2\right)}_{=-\left[\Phi''_{S,\phi\phi}(\mathbf{x}^*,\omega)\right]^{-1}\text{ , see (A.11)}}. \tag{A.15c}$$

The second-order derivative in (A.15c) is of special interest, as it relates $\Phi''_{\mu\mu}(\mu)$ to the second-order derivative $\Phi''_{S,\phi\phi}(\mathbf{x}^*,\omega)$ of the SPA in Sec. A.1.2. By setting $\Phi'_\mu(\mu)$ to zero, the SPA condition reads

$$\phi^*(\rho,\mu) \overset{!}{=} \phi - 2\pi\eta\,. \tag{A.16}$$

Its solution is the stationary phase point $\mu^* = \mu^*(\phi,\rho)$. According to the SPA condition for the CHT given by (A.10), $\phi^*(\rho,\mu)$ must not exceed $[0,2\pi]$ in order to have a non-zero approximation. For a given ϕ, there exists only a single η in (A.16) for which $\phi^*(\rho,\mu)$ lies in this interval. Thus, the summation w.r.t. η in (A.14) vanishes since only one non-zero contribution is present. The approximation reads

$$S^{\mathrm{B}}_M(\mathbf{x},\omega) \approx \overbrace{\left[\frac{\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\left[\mathrm{sign}(\Phi''_{S,\phi\phi}(\mathbf{x}^*(\mu^*),\omega)) + \mathrm{sign}(\Phi''_{\mu\mu}(\mu^*))\right]}}{\sqrt{|\Phi''_{S,\phi\phi}(\mathbf{x}^*(\mu^*),\omega)\Phi''_{\mu\mu}(\mu^*)|}}\right]}^{=1\text{, see (A.15c)}} \mathring{w}^M(\mu^*) \tag{A.17}$$
$$\cdot \underbrace{A_S(\mathbf{x}^*(\mu^*),\omega)\mathrm{e}^{+\mathrm{j}\Phi_S(\mathbf{x}^*(\mu^*),\omega)}}_{S(\mathbf{x}^*(\mu^*),\omega)} \underbrace{\mathrm{e}^{+\mathrm{j}\mu^*(\phi - 2\pi\eta - \phi^*(\mu^*))}}_{=1\text{, see (A.15b)}}.$$

As both $\mathbf{x}$ and $\mathbf{x}^*(\mu)$ share the same radius ρ and are periodic w.r.t. their azimuth angle, (A.16) results in $\mathbf{x}^*(\mu^*) = \mathbf{x}$ holds. Using (A.10), μ^* is substituted by $\Phi'_{S,\phi}(\mathbf{x},\omega)$. This further simplifies the SPA to

$$S^{\mathrm{B}}_M(\mathbf{x},\omega) \approx \mathring{w}^M(\Phi'_{S,\phi}(\mathbf{x},\omega))S(\mathbf{x},\omega)\,. \tag{A.18}$$

The modal windows $\mathring{w}^M_m$ considered in this thesis are even symmetric, i.e. $\mathring{w}^M_m = \mathring{w}^M_{-m}$. Considering (A.7), their continuous counterparts are also even symmetric. Consequentially, $\mathring{w}^M(\mu) = \mathring{w}^M(|\mu|)$ holds. The absolute value of $\Phi'_{S,\phi}(\mathbf{x},\omega)$ corresponds to the norm of the vector product $\mathbf{x} \times \mathbf{k}_S$, see (A.10). The relation between the bandwidth-limited and the fullband sound field is finally given by

$$S^{\mathrm{B}}_M(\mathbf{x},\omega) \approx \mathring{w}^M(|\mathbf{x} \times \mathbf{k}_S|)\, S(\mathbf{x},\omega)\,. \tag{A.19}$$

A.2 Local Wave Field Synthesis using Spatial Bandwidth Limitation

In this section, high-frequency approximations for the 2.5D LWFS-SBL driving functions introduced in Sec. 3.4 are derived. Continuous and discretised PWD are considered in Sec. A.2.1 and Sec. A.2.2, respectively.

A.2.1 Approximation for Continuous Plane Wave Decomposition

Using the definition of the PWD in (2.49) and the amplitude-phase notation for the plane wave coefficients $\bar{S}(\phi_{\mathrm{pw}}, \omega)$, the virtual sound field $S(\mathbf{x}, \omega)$ evaluated at $\mathbf{x}_0$ reads

$$S(\mathbf{x}_0, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \underbrace{A_{\bar{S}}(\phi_{\mathrm{pw}}, \omega) \mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi_{\mathrm{pw}}, \omega)}}_{=\bar{S}(\phi_{\mathrm{pw}}, \omega)} \mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}} \rangle} \, \mathrm{d}\phi_{\mathrm{pw}} \,. \tag{A.20a}$$

It is hereby assumed without loss of generality, that the expansion centre $\mathbf{x}_\mathrm{c}$ coincides with the coordinates' origin. The driving function of LWFS-SBL is given by (3.25). Inserting the amplitude-phase notation for the plane wave coefficients and the WFS driving function for the plane waves[13] yields

[13] see (3.9b) and Tab. 3.1

$$\begin{aligned} D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL}}(\mathbf{x}_0, \omega) = \frac{1}{2\pi} \sqrt{\mathrm{j}\frac{\omega}{c}} \sqrt{8\pi |\mathbf{x}_0 - \mathbf{x}_{\mathrm{ref}}|} \int_0^{2\pi} a_{\mathrm{pw}}(\mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}) \\ \cdot \langle \mathbf{n}_0 | \mathbf{n}_{\mathrm{pw}} \rangle A_{\bar{S}}(\phi_{\mathrm{pw}}, \omega) \mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi_{\mathrm{pw}}, \omega)} \mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}} \rangle} \, \mathrm{d}\phi_{\mathrm{pw}} \end{aligned} \tag{A.20b}$$

with the secondary source selection criterion $a_{\mathrm{pw}}(\mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}})$ for a distinct plane wave given by Tab. 3.1. In both integrals, the involved phase term and the according first- and second-order derivatives read

$$\Phi(\phi_{\mathrm{pw}}, \omega) = \Phi_{\bar{S}}(\phi_{\mathrm{pw}}, \omega) - \frac{\omega}{c} \rho_0 \cos(\phi_0 - \phi_{\mathrm{pw}}) \,, \tag{A.21a}$$

$$\frac{\partial \Phi(\phi_{\mathrm{pw}}, \omega)}{\partial \phi_{\mathrm{pw}}} = \Phi'_{\bar{S},\phi}(\phi_{\mathrm{pw}}, \omega) - \frac{\omega}{c} \rho_0 \sin(\phi_0 - \phi_{\mathrm{pw}}) \,, \tag{A.21b}$$

$$\frac{\partial^2 \Phi(\phi_{\mathrm{pw}}, \omega)}{\partial \phi_{\mathrm{pw}}^2} = \Phi''_{\bar{S},\phi\phi}(\phi_{\mathrm{pw}}, \omega) + \frac{\omega}{c} \rho_0 \cos(\phi_0 - \phi_{\mathrm{pw}}) \,. \tag{A.21c}$$

The first- and second-order derivative of $\Phi_{\bar{S}}$ w.r.t. ϕ are denoted by $\Phi'_{\bar{S},\phi}$ and $\Phi''_{\bar{S},\phi\phi}$, respectively. Since both integrals share the phase term, the common SPA condition reads

$$\Phi'_{\bar{S},\phi}(\phi_{\mathrm{pw}}, \omega) \stackrel{!}{=} \frac{\omega}{c} \rho_0 \sin(\phi_0 - \phi_{\mathrm{pw}}) \,. \tag{A.22}$$

The solution to it is the stationary phase point $\phi_{\mathrm{pw}}^* = \phi_{\mathrm{pw}}^*(\mathbf{x}_0, \omega)$. The corresponding propagation direction is denoted as $\mathbf{n}_{\mathrm{pw}}^* = \mathbf{n}_{\mathrm{pw}}^*(\mathbf{x}_0, \omega)$. The SPA of the virtual sound field reads

$$\begin{aligned} S(\mathbf{x}_0, \omega) \approx \frac{\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\operatorname{sign}(\Phi''_{\bar{S},\phi\phi}(\phi_{\mathrm{pw}}^*, \omega) + \frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle)}}{\sqrt{2\pi |\Phi''_{\bar{S},\phi\phi}(\phi_{\mathrm{pw}}^*, \omega) + \frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle|}} \\ \cdot A_{\bar{S}}(\phi_{\mathrm{pw}}^*, \omega) \mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi_{\mathrm{pw}}^*, \omega)} \mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle} \end{aligned} \tag{A.23}$$

and the SPA of the LWFS-SBL driving function is given by

$$\begin{aligned} D_{2.5\mathrm{D}}^{\mathrm{LWFS-SBL}}(\mathbf{x}_0, \omega) \approx \sqrt{\mathrm{j}\frac{\omega}{c}} \sqrt{8\pi |\mathbf{x}_0 - \mathbf{x}_{\mathrm{ref}}|} a_{\mathrm{pw}}(\mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^*) \langle \mathbf{n}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle \\ \underbrace{\frac{\mathrm{e}^{+\mathrm{j}\frac{\pi}{4}\operatorname{sign}(\Phi''_{\bar{S},\phi\phi}(\phi_{\mathrm{pw}}^*, \omega) + \frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle)}}{\sqrt{2\pi |\Phi''_{\bar{S},\phi\phi}(\phi_{\mathrm{pw}}^*, \omega) + \frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle|}} A_{\bar{S}}(\phi_{\mathrm{pw}}^*, \omega) \mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi_{\mathrm{pw}}^*, \omega)} \mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle \mathbf{x}_0 | \mathbf{n}_{\mathrm{pw}}^* \rangle}}_{\approx S(\mathbf{x}_0, \omega),\ \text{see (A.23)}} . \end{aligned} \tag{A.24}$$

Since the approximate equality holds in (A.23), also the phase of the left-hand and the right-hand side is approximately equal. The phase term of virtual sound field reads

$$\Phi_S(\mathbf{x}_0,\omega) \approx \Phi_{\bar{S}}(\phi^*_{\mathrm{pw}},\omega) - \frac{\omega}{c}\langle \mathbf{x}_0|\mathbf{n}^*_{\mathrm{pw}}\rangle\,. \tag{A.25}$$

The relation can be used to express the local wavenumber vector[14] of the virtual sound field as

[14] see (2.53)

$$\mathbf{k}_S(\mathbf{x}_0,\omega) \approx \frac{\omega}{c}\mathbf{n}^*_{\mathrm{pw}} - \nabla_{\mathbf{x}_0}\phi^*_{\mathrm{pw}} \underbrace{\left.\frac{\partial\Phi(\phi_{\mathrm{pw}},\omega)}{\partial\phi_{\mathrm{pw}}}\right|_{\phi_{\mathrm{pw}}=\phi^*_{\mathrm{pw}}}}_{=0,\text{ see (A.21b)}}\,. \tag{A.26}$$

Using this equality together with the local dispersion relation in (2.54), the stationary plane wave direction $\mathbf{n}^*_{\mathrm{pw}}$ in (A.24) is expressed by the normalised local wavenumber vector $\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)$. According to Tab. 3.1 and (3.5), $a_{\mathrm{pw}}(\mathbf{x}_0|\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega))$ is equal to the virtual sound field specific selection criterion $a_S(\mathbf{x}_0)$. The final approximation reads

$$D^{\mathrm{LWFS-SBL}}_{2.5\mathrm{D}}(\mathbf{x}_0) \approx \sqrt{\mathrm{j}\frac{\omega}{c}}\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|}a_S(\mathbf{x}_0)\langle\mathbf{n}_0|\hat{\mathbf{k}}_S(\mathbf{x}_0,\omega)\rangle S(\mathbf{x}_0,\omega)\,. \tag{A.27}$$

A.2.2 Approximation for Discrete Plane Wave Decomposition

The continuous PWD is approximated via a summation of over discrete angles $\phi^{(l)}_{\mathrm{pw}} = \frac{2\pi}{N_{\mathrm{pw}}}l$. The resulting sound field reads

$$S^{\mathrm{S}}(\mathbf{x}_0,\omega) = \frac{1}{N_{\mathrm{pw}}}\sum_{l=0}^{N_{\mathrm{pw}}-1} A_{\bar{S}}(\phi^{(l)}_{\mathrm{pw}})\mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi^{(l)}_{\mathrm{pw}})}\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle\mathbf{x}_0|\mathbf{n}^{(l)}_{\mathrm{pw}}\rangle} \tag{A.28}$$

The propagation direction belonging to $\phi^{(l)}_{\mathrm{pw}}$ is denoted as $\mathbf{n}^{(l)}_{\mathrm{pw}}$. By inserting the amplitude-notation for $\bar{S}(\phi_{\mathrm{pw}},\omega)$ and the WFS driving function for a plane wave into (4.68), the LWFS-SBL driving function is analogously given as

$$D^{\mathrm{LWFS-SBL,S}}_{2.5\mathrm{D}}(\mathbf{x}_0,\omega) = \frac{1}{N_{\mathrm{pw}}}\sqrt{\mathrm{j}\frac{\omega}{c}}\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|} \tag{A.29}$$
$$\sum_{l=0}^{N_{\mathrm{pw}}-1} a_{\mathrm{pw}}(\mathbf{x}_0|\mathbf{n}^{(l)}_{\mathrm{pw}})\langle\mathbf{n}_0|\mathbf{n}^{(l)}_{\mathrm{pw}}\rangle A_{\bar{S}}(\phi^{(l)}_{\mathrm{pw}})\mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi^{(l)}_{\mathrm{pw}})}\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle\mathbf{x}_0|\mathbf{n}^{(l)}_{\mathrm{pw}}\rangle}\,.$$

The sampling in the PWD domain is modelled as a multiplication of the continuous PWD with a Dirac comb.[15] With its Fourier series,[16] the sound field and the driving function are split up into individual aliasing components indexed by ζ. They read

[15] Girod et al., op. cit., Sec. 11.3.1.

[16] Ibid., Eq. (11.12).

$$S^{\mathrm{S}}_{\zeta}(\mathbf{x}_0,\omega) = \tag{A.30a}$$
$$\frac{1}{2\pi}\int_0^{2\pi} A_{\bar{S}}(\phi_{\mathrm{pw}},\omega)\mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi_{\mathrm{pw}},\omega)}\mathrm{e}^{-\mathrm{j}\zeta N_{\mathrm{pw}}\phi_{\mathrm{pw}}}\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle\mathbf{x}_0|\mathbf{n}_{\mathrm{pw}}\rangle}\,\mathrm{d}\phi_{\mathrm{pw}}$$

and

$$D^{\mathrm{LWFS-SBL,S}}_{2.5\mathrm{D},\zeta}(\mathbf{x}_0,\omega) = \frac{1}{2\pi}\sqrt{\mathrm{j}\frac{\omega}{c}}\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|}\int_0^{2\pi} a_{\mathrm{pw}}(\mathbf{x}_0|\mathbf{n}_{\mathrm{pw}}) \tag{A.30b}$$
$$\cdot\langle\mathbf{n}_0|\mathbf{n}_{\mathrm{pw}}\rangle A_{\bar{S}}(\phi_{\mathrm{pw}},\omega)\mathrm{e}^{+\mathrm{j}\Phi_{\bar{S}}(\phi_{\mathrm{pw}},\omega)}\mathrm{e}^{-\mathrm{j}\zeta N_{\mathrm{pw}}\phi_{\mathrm{pw}}}\mathrm{e}^{-\mathrm{j}\frac{\omega}{c}\langle\mathbf{x}_0|\mathbf{n}_{\mathrm{pw}}\rangle}\,\mathrm{d}\phi_{\mathrm{pw}}\,.$$

Besides the additional exponential $\mathrm{e}^{-\mathrm{j}\zeta N_{\mathrm{pw}}\phi_{\mathrm{pw}}}$, both integrals match their counterpart in (A.20). The same derivation steps lead to

$$D_{2.5\mathrm{D},\zeta}^{\mathrm{LWFS-SBL,S}}(\mathbf{x}_0,\omega) \approx \sqrt{\mathrm{j}\frac{\omega}{\mathrm{c}}}\sqrt{8\pi|\mathbf{x}_0-\mathbf{x}_{\mathrm{ref}}|}\, a_{S_\zeta^{\mathrm{S}}}(\mathbf{x}_0)\langle\mathbf{n}_0|\mathbf{k}_{S_\zeta^{\mathrm{S}}}(\mathbf{x}_0,\omega)\rangle S_\zeta^{\mathrm{S}}(\mathbf{x}_0,\omega) \tag{A.31}$$

as the connection between the sound field and the aliasing component of the driving function. The SPA condition

$$\Phi'_{\bar{S},\phi}(\phi_{\mathrm{pw}},\omega) \stackrel{!}{=} \frac{\omega}{\mathrm{c}}\rho_0\sin(\phi_0-\phi_{\mathrm{pw}})+\zeta N_{\mathrm{pw}} \tag{A.32}$$

accordingly differs from (A.22) about an additional term depending on ζ. The solution to it is the stationary phase point $\phi^*_{\mathrm{pw},\zeta} = \phi^*_{\mathrm{pw},\zeta}(\mathbf{x}_0,\omega)$. The corresponding propagation direction is denoted as $\mathbf{n}^*_{\mathrm{pw},\zeta} = \mathbf{n}^*_{\mathrm{pw},\zeta}(\mathbf{x}_0,\omega)$. Analogous to (A.26), $\hat{\mathbf{k}}_{S_\zeta^{\mathrm{S}}}(\mathbf{x}_0,\omega) \approx \mathbf{n}^*_{\mathrm{pw},\zeta}$ holds.

For the geometric model in Sec. 4.4.2, the local wavenumber vector of aliasing component $\mathbf{k}_{S_\zeta^{\mathrm{S}}}$ is key to predict spatial aliasing. Although it is unknown, a connection to the original sound field can be established by considering an additional position $\mathbf{x}_\zeta^{\mathrm{S}}$. It is assumed that the stationary plane wave direction $\mathbf{n}^*_{\mathrm{pw}}(\mathbf{x}_\zeta^{\mathrm{S}},\omega)$ for the continuous driving function in (A.22) coincides with $\mathbf{n}^*_{\mathrm{pw},\zeta}(\mathbf{x}_0,\omega)$. Considering (A.26), the local wavenumber vectors are approximately equal, i.e. $\hat{\mathbf{k}}_{S_\zeta^{\mathrm{S}}}(\mathbf{x}_0,\omega) \approx \hat{\mathbf{k}}_S(\mathbf{x}_\zeta^{\mathrm{S}},\omega)$. Combining both SPA conditions (A.22) and (A.32) allows to formulate

$$\frac{\omega}{\mathrm{c}}\rho_0\sin(\phi_0-\phi^*_{\mathrm{pw}})+\zeta N_{\mathrm{pw}} \stackrel{!}{=} \frac{\omega}{\mathrm{c}}\rho_\zeta^{\mathrm{S}}\sin(\phi_\zeta^{\mathrm{S}}-\phi^*_{\mathrm{pw}})\,. \tag{A.33}$$

Together with (A.26), the condition is rearranged to

$$\zeta\frac{\mathrm{c}}{f} \stackrel{!}{=} \frac{2\pi}{N_{\mathrm{pw}}}\left\langle \mathbf{x}_\zeta^{\mathrm{S}}-\mathbf{x}_0 \middle| \mathbf{R}_{\frac{\pi}{2}}\hat{\mathbf{k}}_S(\mathbf{x}_\zeta^{\mathrm{S}},\omega)\right\rangle\,, \tag{A.34}$$

where $\mathbf{R}_{\frac{\pi}{2}}$ denotes a rotation matrix rotating the normalised wavenumber vector $\hat{\mathbf{k}}_S(\mathbf{x}_0^\dagger)$ about $\pi/2$.

B Miscellaneous Derivations

B.1 Relation between 2.5D and 3D Near-Field-Compensated Higher-Order Ambisonics

The completeness relations of the Circular Harmonics and Spherical Harmonics[1] allow to postulate

$$\frac{1}{2\pi}\sum_{m=-\infty}^{\infty} \mathrm{e}^{+\mathrm{j}m\phi} = \delta(\phi),\ \text{and} \tag{B.1}$$

$$\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} \Psi_n^{-m}\left(\frac{\pi}{2},0\right)\Psi_n^m(\vartheta,\phi) = \delta(\cos(\vartheta))\delta(\phi)\,. \tag{B.2}$$

The Dirac delta distribution $\delta(\phi)$ in (B.2) is replaced by the left-hand side of (B.1). In addition, $\Psi_n^m(\vartheta,\phi) = \Psi_n^m(\vartheta,0)\mathrm{e}^{+\mathrm{j}m\phi}$ from Eq. (2.34) is used resulting in

$$\sum_{m=-\infty}^{\infty}\sum_{n=|m|}^{\infty} \Psi_n^{-m}\left(\frac{\pi}{2},0\right)\Psi_n^m(\vartheta,0)\mathrm{e}^{+\mathrm{j}m\phi} = \frac{\delta(\cos(\vartheta))}{2\pi}\sum_{m=-\infty}^{\infty} \mathrm{e}^{+\mathrm{j}m\phi}\,. \tag{B.3}$$

The orthogonality of the Circular Harmonics allows to directly compare the coefficients for each m, which yields

$$\sum_{n=|m|}^{\infty} \Psi_n^{-m}\left(\frac{\pi}{2},0\right)\Psi_n^m(\vartheta,0) = \frac{\delta(\cos(\vartheta))}{2\pi}\,. \tag{B.4}$$

Again, $\Psi_n^m(\vartheta,\phi) = \Psi_n^m(\vartheta,0)\mathrm{e}^{+\mathrm{j}m\phi}$ from Eq. (2.34) is incorporated to formulate Eq. (3.16). An alternative derivation is given by Poletti.[2]

[1] Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, Eq. (1.26) and Eq. (6.47).

[2] Poletti (Mar. 2017). "Spherical coordinate descriptions of cylindrical and spherical Bessel beams". In: *J. Acoust. Soc. Am.* 141.3, pp. 2069–2078, App. A.

B.2 Extremal Values of Tangential Components of the Local Wavenumber Vector for a Circle

The tangential component of the local wavenumber vector for the free-field Green's function is given by

$$\hat{k}_{G,\mathbf{t}_0}(\mathbf{x}-\mathbf{x}_0) = \langle \mathbf{t}_0 | \hat{\mathbf{k}}_G(\mathbf{x}-\mathbf{x}_0)\rangle = \left\langle \mathbf{t}_0 \middle| \frac{\mathbf{x}-\mathbf{x}_0}{|\mathbf{x}-\mathbf{x}_0|} \right\rangle\,. \tag{B.5}$$

If the circle $\mathcal{C}_\mathrm{h}$ is completely inside Ω, see Fig. B.1c, the outer bounds of the circle define the extremal values for the normalised local wavenumber vector of the free-field Green's function. They are given by

$$\hat{\mathbf{k}}_G^{\{\min,\max\}}(\mathbf{x}_0) = \begin{bmatrix} \cos\phi_\mathcal{C} & \mp\sin\phi_\mathcal{C} \\ \pm\sin\phi_\mathcal{C} & \cos\phi_\mathcal{C} \end{bmatrix} \hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{h}-\mathbf{x}_0)\,, \tag{B.6}$$

which defines a rotation of the normalised local wavenumber vector $\hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{h} - \mathbf{x}_0)$ clockwise or counter-clockwise about the angle

$$\phi_\mathcal{C} = \arcsin\left(\frac{R_\mathrm{h}}{|\mathbf{x}_\mathrm{h} - \mathbf{x}_0|}\right) = \arcsin\left(\varrho_\mathrm{h}\right) . \tag{B.7}$$

For brevity, the ratio ϱ_h of the circle radius R_h and the distance between $\mathbf{x}_0$ and the centre $\mathbf{x}_\mathrm{h}$ is introduced. If the ratio is larger than 1, the circle includes the $\mathbf{x}_0$ and the $\arcsin(\cdot)$ has no real solution. Hence, the case shown in Fig. B.1a applies and $\hat{k}^{\mathrm{min,max}}_{G,\mathbf{t}_0}(\mathbf{x}_0) = \mp 1$ holds. Using the trigonometric identities, the bounding vectors can be expressed by

$$\hat{\mathbf{k}}^{\{\mathrm{min,max}\}}_G(\mathbf{x}_0) = \begin{bmatrix} \sqrt{1-\varrho_\mathrm{h}^2} & \mp\varrho_\mathrm{h} \\ \pm\varrho_\mathrm{h} & \sqrt{1-\varrho_\mathrm{h}^2} \end{bmatrix} \hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{h} - \mathbf{x}_0) . \tag{B.8}$$

Their according tangential components are given by evaluating the vector-matrix-vector multiplication

$$\begin{aligned} \hat{k}^{\{\mathrm{min,max}\}}_{G,\mathbf{t}_0}(\mathbf{x}_0) &= \langle \mathbf{t}_0 | \hat{\mathbf{k}}^{\{\mathrm{min,max}\}}_G(\mathbf{x}_0) \rangle \\ &= \sqrt{1-\varrho_\mathrm{h}^2}\left(t_{0,x}\hat{k}_{G,x}(\mathbf{x}_\mathrm{h} - \mathbf{x}_0) + t_{0,y}\hat{k}_{G,y}(\mathbf{x}_\mathrm{h} - \mathbf{x}_0)\right) \\ &\quad \mp \varrho_\mathrm{h}\left(-t_{0,y}\hat{k}_{G,x}(\mathbf{x}_\mathrm{h} - \mathbf{x}_0) + t_{0,x}\hat{k}_{G,y}(\mathbf{x}_\mathrm{h} - \mathbf{x}_0)\right) . \end{aligned} \tag{B.9}$$

The first bracket constitutes the scalar product of the tangent vector $\mathbf{t}_0$ and the normalised wavenumber vector $\hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{h} - \mathbf{x}_0)$, which is the tangential component

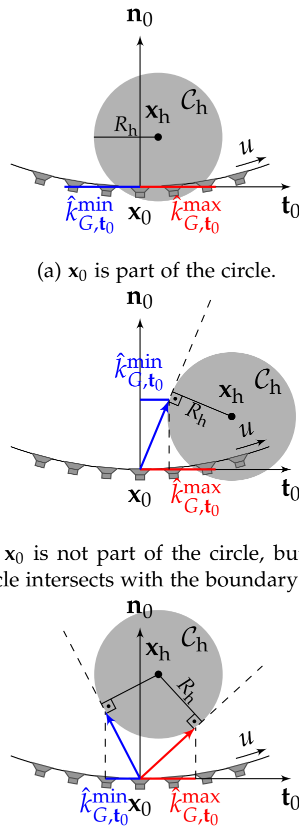
$$\hat{k}_{\mathrm{h},\mathbf{t}_0} := \hat{k}_{G,\mathbf{t}_0}(\mathbf{x}_\mathrm{h} - \mathbf{x}_0) = \Big\langle \mathbf{t}_0 \Big| \hat{\mathbf{k}}_G(\mathbf{x}_\mathrm{h} - \mathbf{x}_0) \Big\rangle . \tag{B.10}$$

For the second bracket, the tangent vector is rotated about $\pi/2$, which is equivalent to the normal vector $\mathbf{n}_0$. As $\mathbf{x}_\mathrm{h}$ lies inside the convex Ω, the normal component is always positive. It can hence be expressed by $\sqrt{1-\hat{k}^2_{\mathrm{h},\mathbf{t}_0}}$. This yields

$$\hat{k}^{\{\mathrm{min,max}\}}_{G,\mathbf{t}_0}(\mathbf{x}_0) = \hat{k}_{\mathrm{h},\mathbf{t}_0}\sqrt{1-\varrho_\mathrm{h}^2} \mp \varrho_\mathrm{h}\sqrt{1-\hat{k}^2_{\mathrm{h},\mathbf{t}_0}} . \tag{B.11}$$

As the remaining task, the case depicted in Fig. B.1b has to be detected. Here, the circle overlaps with the boundary $\partial\Omega$. This is done by inserting the ∓ 1 for $\hat{k}^{\{\mathrm{min,max}\}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and solving the equation for $\hat{k}_{\mathrm{h},\mathbf{t}_0}$. As a result, $\hat{k}_{\mathrm{h},\mathbf{t}_0} = \mp\sqrt{1-\varrho_\mathrm{h}^2}$ marks the critical value below/above which $\hat{k}^{\{\mathrm{min,max}\}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ has to be assigned to ± 1. A conditional expression covering all three cases of Fig. B.1 is given by

$$\hat{k}^{\{\mathrm{min,max}\}}_{G,\mathbf{t}_0}(\mathbf{x}_0) = \begin{cases} \mp 1 & \text{if } \varrho_\mathrm{h} > 1 \text{ , else} \\ \mp 1 & \text{if } \hat{k}_{\mathrm{h},\mathbf{t}_0} \lessgtr \sqrt{1-\varrho_\mathrm{h}^2} , \\ \hat{k}_{\mathrm{h},\mathbf{t}_0}\sqrt{1-\varrho_\mathrm{h}^2} \mp \varrho_\mathrm{h}\sqrt{1-\hat{k}^2_{\mathrm{h},\mathbf{t}_0}} & \text{otherwise,} \end{cases} \tag{B.12}$$



(a) $\mathbf{x}_0$ is part of the circle.

(b) $\mathbf{x}_0$ is not part of the circle, but the circle intersects with the boundary $\partial\mathcal{S}$.

(c) The circle is completely inside $\mathcal{S}$.

Figure B.1: The three sketches illustrate the three different cases that have to be considered for the computation of $\hat{k}^{\mathrm{max}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\mathrm{min}}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ for a circular region $\mathcal{C}_\mathrm{h}$.

where the upper and lower option for $\mp$ and $\gtrless$ applies for $\hat{k}^{\min}_{G,\mathbf{t}_0}(\mathbf{x}_0)$ and $\hat{k}^{\max}_{G,\mathbf{t}_0}(\mathbf{x}_0)$, respectively.

For the special case of $\mathcal{C}_\mathrm{h}$ being in the centre of a circular SSD with radius R, $\hat{k}_{\mathrm{h},\mathbf{t}_0} = 0$ and $\varrho_\mathrm{h} = {}^{R_\mathrm{h}}\!/\!{}_R$ holds for all $\mathbf{x}_0$. This yields

$$\hat{k}^{\{\min,\max\}}_{G,\mathbf{t}_0}(\mathbf{x}_0) = \begin{cases} \mp 1 & \text{if } \frac{R_\mathrm{h}}{R} > 1 \text{ , else} \\ \mp\frac{R_\mathrm{h}}{R} & \text{otherwise.} \end{cases} \tag{B.13}$$

B.3 Optimal Sampling Scheme for Wave Field Synthesis

All aliasing frequencies derived in Sec. 4.2.4 exhibit the mathematical structure w.r.t. $\mathbf{x}_0$ and are, thus, jointly discussed. Equidistant sampling w.r.t. u and v leads to the aliasing frequencies

$$f^{\mathrm{S,WFS}}_u(u) = \frac{c}{|\mathbf{x}'_0(u)| \cdot \Delta_u \cdot \gamma(\mathbf{x}_0(u))}\,, \tag{B.14}$$

$$f^{\mathrm{S,WFS}}_v(v) = \frac{c}{|\mathbf{x}'_0(v)| \cdot \Delta_v \cdot \gamma(\mathbf{x}_0(v))}\,. \tag{B.15}$$

The scenario-dependent function is denoted as γ and includes the remaining terms given in e.g. (4.31), or (4.34), which do only depend on $\mathbf{x}_0$, but no the actual parametrisation u and v. $\mathbf{x}_0(u) = \mathbf{x}_0(v)$, $\mathbf{x}'_0(u) = v'(u)\mathbf{x}'_0(v)$, and $\Delta_u = \Delta_v$ allows to formulate the relation between the aliasing frequencies for the two parametrisations. Since $v'(u) > 0$, it reads

$$f^{\mathrm{S,WFS}}_v(v) = v'(u) f^{\mathrm{S,WFS}}_u(u)\,. \tag{B.16}$$

Since the aliasing frequency is calculated as the minimum over the SSD, the resulting optimisation problem is formulated as

$$\underset{v'(u)}{\text{maximise}} \quad \min_v f^{\mathrm{S,WFS}}_v(v) = \min_u \left[v'(u) f^{\mathrm{S,WFS}}_u(u)\right] \tag{B.17a}$$

$$\text{subject to} \quad u_{\max} - u_{\min} = \int_{u_{\min}}^{u_{\max}} v'(\mu)\mathrm{d}\mu\,. \tag{B.17b}$$

The condition in (B.17b) results from the fact, that u and v share the same support. The solution to the problem is given by

$$v'_{\mathrm{opt}}(u) = \frac{(u_{\max} - u_{\min})}{\int_{u_{\min}}^{u_{\max}} \frac{1}{f^{\mathrm{S,WFS}}_u(\mu)}\mathrm{d}\mu} \frac{1}{f^{\mathrm{S,WFS}}_u(u)}\,. \tag{B.18}$$

This can be proven via reductio ad absurdum: If $v'_{\mathrm{opt}}(u)$ is not the optimal solution, there has to exist a function $w(u)$, such that $v'_{\mathrm{opt}}(u) + w'(u)$ leads to larger minimum (B.17a) and still fulfils the condition (B.17b). This leads to

$$0 < w'(u) f^{\mathrm{S,WFS}}_u(u) \qquad \forall u \in [u_{\min}, u_{\max}] \text{ , and} \tag{B.19a}$$

$$0 = \int_{u_{\min}}^{u_{\max}} w'(\mu)\mathrm{d}\mu = w(u_{\max}) - w(u_{\min})\,. \tag{B.19b}$$

Since the aliasing frequency $f^{\mathrm{S}}_u(u)$ is always positive, the first condition is reformulated to $w'(u) > 0$. Thus, $w(u)$ has to be a strictly increasing function, which violates the second condition. $v'_{\mathrm{opt}}(u)$ has to be the optimal solution.

Acronyms

2.5D	2½-dimensional 3, 25, 27–29, 32–34, 37, 38, 42–44, 48–51, 60, 61, 65, 67, 68, 72, 76, 83, 127, 131
2D	two-dimensional 18–20, 24, 25, 27, 29, 33, 37, 42, 50, 54
3D	three-dimensional 3, 4, 10–12, 19, 20, 24–27, 32, 42, 44, 50, 52, 131
a.k.a.	also known as 10, 17, 24, 33, 34, 36, 44, 48, 49, 56, 83, 97, 103, 123
BLUE	Best Linear Unbiased Estimate 93, 118
BLUP	Best Linear Unbiased Prediction 93
BTF	Binaural Transfer Function 87, 88, 96, 103, 108, 115, 116
CHT	Circular Harmonics Transform 3, 18, 48, 71, 125, 127
CIIM	Corrected Impulse Invariance Method 37
CLL	Composite Loudness Level 115–117
D/A	digital-to-analog 5, 29, 89
DFT	Discrete Fourier Transform 29
DTFT	Discrete Time Fourier Transform 29
e.g.	exempli gratia 5, 8, 17, 22, 23, 26, 30, 35, 39, 45, 57, 64, 68, 90, 117, 121, 133
ERB	Equivalent Rectangular Bandwidth 103
FD	Fractional Delay 31
FIR	Finite Impulse Response 29–31, 39
GCD	Greatest Common Divisor 79
GUI	Graphical User Interface 109
HATS	Head and Torso Simulator 87, 90, 103, 109
HIE	Helmholtz-Integral-Equation 15, 16, 24
HOA	Higher Order Ambisonics 31, 85, 101, 107, 115, 122
HPCF	Headphone Compensation Filter 86, 88, 90
HRIR	Head-Related Impulse Response 86, 88
HRTF	Head-Related Transfer Function 86–88, 90, 103, 108, 109
i.e.	id est 8, 14–16, 19, 21, 24, 25, 28, 36, 38, 40, 47, 48, 56, 62, 64, 66, 68, 70, 74, 75, 77, 83, 85, 87, 91, 92, 94, 96, 98, 99, 107, 112, 117, 130
IC	Interaural Coherence 83, 102, 103
ICHT	Inverse Circular Harmonics Transform 3, 34, 124–126
IDFT	Inverse Discrete Fourier Transform 36, 39, 40
IFFT	Inverse Fast Fourier Transform 36, 37, 39, 45
IIR	Infinite Impulse Response 29, 30, 36, 37, 39, 40, 45, 103
ILD	Interaural Level Difference 83, 86, 102–104
ITD	Interaural Time Difference 83, 84, 86, 90, 102–104, 123
ITU	International Telecommunication Union 4

LR	Linkwitz-Riley 39–41, 45
(L)SFS	(Local) Sound Field Synthesis 2, 6–8, 44, 95–105, 110–120, 122, 123, 148, 149
LSFS	Local Sound Field Synthesis 5–7, 22–24, 46, 80, 82, 83, 107, 116, 121–123, 148, 149
LTSI	linear time and space invariant 12
LWFS	Local Wave Field Synthesis 1–3, 7, 23, 37–45, 47, 71–82, 85, 95, 96, 99, 101, 104–108, 110–115, 117, 119–122, 127–130, 135
LWFS-SBL	Local Wave Field Synthesis using Spatial Bandwidth Limitation 1–3, 7, 23, 37–43, 45, 71–75, 78–82, 85, 95, 96, 101, 104–107, 110, 111, 113–115, 117, 119–122, 127–130
LWFS-VSS	Local Wave Field Synthesis using Virtual Secondary Sources 1, 2, 7, 23, 41–45, 76–80, 82, 85, 95, 96, 99, 101, 104–107, 110, 111, 114, 115, 117, 119, 121, 122
MAA	Minimum Audible Angle 83, 84, 90
MSE	Mean-Square Error 97
MUSHRA	Multiple Stimulus with Hidden Reference and Anchor 108, 109, 111–113, 118–120, 123
NFCHOA	Near-Field-Compensated Higher-Order Ambisonics 1–3, 5, 6, 23, 25, 31–42, 44–49, 60, 65–73, 75, 82, 85, 95, 96, 99, 101, 103, 105–107, 110, 111, 114–116, 120–123, 131
NHST	Null Hypothesis Significance Testing 112–114
PWD	Plane Wave Decomposition 3, 20, 23, 37–39, 44, 45, 47, 71–74, 76, 79, 82, 96, 111, 117, 121, 122, 127–129
QoE	Quality of Experience 6
RMSE	Root-Mean-Square Error 97–101
SBL	Spatial Bandwidth Limitation 1–3, 7, 23, 34, 35, 37–45, 47, 49, 66, 68–75, 78–82, 85, 95, 96, 101, 104–107, 110, 111, 113–122, 127–130, 135
SFA	Sound Field Analysis 22, 39
SFS	Sound Field Synthesis 5, 6, 10, 14, 17, 20, 22–24, 28, 31, 40, 46–48, 50, 81–87, 90, 91, 94, 95, 97–104, 106, 108–113, 115, 117, 120, 121, 148, 149
SLP	Single Layer Potential 24, 25, 27, 29, 32, 42, 44, 47, 50, 54, 60, 68, 77
SPA	Stationary Phase Approximation 27, 44, 51–54, 56, 72, 77–79, 124, 126–128, 130
SSD	Secondary Source Distribution 2, 22, 26, 28–32, 35, 38, 41–48, 50–58, 60–65, 68, 70–82, 96, 99, 101, 106, 110, 117, 133
TU	Technische Universität 60, 61, 88, 95, 109
VBAP	Vector Based Amplitude Panning 4, 6, 115
VDA	Vargha-Delaney A 113
VSS	Virtual Secondary Source 1, 2, 7, 23, 41–45, 76–80, 82, 85, 95, 96, 99, 101, 104–107, 110, 111, 114, 115, 117, 119, 121, 122, 135
w.r.t.	with respect to 6–9, 11, 13, 14, 16, 18, 27, 30, 33, 36, 39, 41, 47, 48, 51, 54, 58, 60, 63–65, 69, 70, 73–80, 84, 86, 87, 95, 99, 104, 105, 107, 112, 115, 116, 118–120, 122, 126–128, 133
WFS	Wave Field Synthesis 1, 3, 5, 6, 23, 25–31, 34–47, 50–65, 68, 69, 72, 74–78, 80, 81, 84, 95, 96, 99, 101–103, 106, 107, 109–111, 115, 116, 118, 120–123, 128, 129, 133

Bibliography

Abramowitz and Stegun (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. 55. Courier Corporation (cit. on pp. 13, 18).

Acoustical Terminology (1994). Standard ANSI S1.1-1994. New York, USA: American National Standards Institute (cit. on p. 107).

Advanced sound system for programme production (July 2018). Standard ITU-R BS.2051-2. International Telecommunication Union Radiocommunication Assembly (cit. on p. 4).

Ahrens (2010). "The Single-layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources". PhD thesis. Berlin, Germany: Technische Universität Berlin. DOI: 10.14279/depositonce-2663 (cit. on pp. 4, 23, 26, 32, 121).

Ahrens (2012). *Analytic Methods of Sound Field Synthesis*. T-Labs Series in Telecommunication Services. Berlin Heidelberg, Germany: Springer-Verlag. DOI: 10.1007/978-3-642-25743-8 (cit. on pp. 5, 18–20, 31, 32, 35, 46–50, 56, 65, 66, 69, 71, 85, 95).

Ahrens and Spors (2008). "An Analytical Approach to Sound Field Reproduction using Circular and Spherical Loudspeaker Distributions". In: *Acta Acustica united with Acustica* 94.6, pp. 988–999. DOI: 10.3813/AAA.918115 (cit. on pp. 32, 44).

Ahrens and Spors (Aug. 2009). "An Analytical Approach to 2.5D Sound Field Reproduction Employing Circular Distributions of Non-Omnidirectional Loudspeakers". In: *Proc. of 17th European Signal Processing Conference (EUSIPCO)*. Glasgow, UK (cit. on pp. 24, 31).

Ahrens, Geier, and Spors (May 2010). "Perceptual Assessment of Delay Accuracy and Loudspeaker Misplacement in Wave Field Synthesis". In: *Proc. of 128th Aud. Eng. Soc. Conv.* London, UK (cit. on p. 30).

Algazi, Avendano, and Duda (2001). "Elevation localization and head-related transfer function analysis at low frequencies". In: *J. Acoust. Soc. Am.* 109.3, pp. 1110–1122. DOI: 10.1121/1.1349185 (cit. on pp. 75, 80).

Arfken and Weber (2005). *Mathematical Methods for Physicists*. 6th ed. Elsevier (cit. on pp. 8, 9, 21).

Bamford (1995). "An Analysis of Ambisonics Sound Systems of First and Second Order". MA thesis. Waterloo, Canada: University of Waterloo (cit. on p. 31).

Baumgartner (2011). "Time Domain Fast-Multipole Translation for Ambisonics". MA thesis. Graz, Austria: Institute of Electronic Music and Acoustics (cit. on p. 37).

Berkhout (1988). "A Holographic Approach to Acoustic Control". In: *J. Aud. Eng. Soc.* 36.12, pp. 977–995 (cit. on pp. 5, 23, 25).

Berkhout, de Vries, and Vogel (1993). "Acoustic control by wave field synthesis". In: *J. Acoust. Soc. Am.* 93.5, pp. 2764–2778. DOI: 10.1121/1.405852 (cit. on p. 25).

Blackstock (2000). *Fundamentals of physical acoustics*. John Wiley & Sons (cit. on p. 10).

Blauert (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press (cit. on pp. 4, 83–85, 102).

Blauert and Jekosch (Aug. 2003). “Concepts Behind Sound Quality: Some Basic Considerations”. In: *Proc. of International Congress and Exposition on Noise Control Engineering (INTERNOISE)*. Seogwipo, Korea (cit. on p. 6).

Bleistein (1984). *Mathematical Methods for Wave Phenomena*. Orlando, USA: Academic Press (cit. on p. 124).

Bleistein and Handelsman (1986). *Asymptotic Expansions of Integrals*. New York, USA: Dover Publications (cit. on p. 124).

Blumlein (1931). *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems*. British Patent Specification 394,325 (cit. on p. 4).

Bortz and Schuster (2010). *Statistik für Human-und Sozialwissenschaftler*. 7th ed. Springer. DOI: 10.1007/978-3-642-12770-0 (cit. on pp. 93, 98).

Bregman (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT press (cit. on p. 4).

Brewer (2005). *Designing Better Maps: A Guide for GIS Users*. New York, USA: ESRI Press (cit. on p. 8).

Brinkmann, Roden, Lindau, and Weinzierl (Aug. 2015). “Audibility and Interpolation of Head-Above-Torso Orientation in Binaural Technology”. In: *IEEE J. Sel. Topics Signal Process.* 9.5, pp. 931–942. DOI: 10.1109/JSTSP.2015.2414905 (cit. on pp. 86, 90).

Brinkmann, Roden, Lindau, and Weinzierl (Sept. 2014). “Audibility of head-above-torso orientation in head-related transfer functions”. In: *Proc. of Forum Acusticum*. Kraków, Poland (cit. on p. 86).

Brinkmann, Lindau, and Weinzierl (2017). “On the authenticity of individual dynamic binaural synthesis”. In: *J. Acoust. Soc. Am.* 142.4, pp. 1784–1795. DOI: 10.1121/1.5005606 (cit. on pp. 90, 98, 109).

Bronkhorst (1995). “Localization of real and virtual sound sources”. In: *J. Acoust. Soc. Am.* 98.5, pp. 2542–53. DOI: 10.1121/1.413219 (cit. on p. 91).

Cherry (1953). “Some Experiments on the Recognition of Speech, with One and with Two Ears”. In: *J. Acoust. Soc. Am.* 25.5, pp. 975–979. DOI: 10.1121/1.1907229 (cit. on p. 4).

Coleman, Jackson, Olik, Møller, Olsen, and Abildgaard Pedersen (2014). “Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array”. In: *J. Acoust. Soc. Am.* 135.4. DOI: 10.1121/1.4866442 (cit. on p. 23).

Colton and Kress (2013). *Inverse acoustic and electromagnetic scattering theory*. 3rd ed. Vol. 93. Applied Mathematical Sciences. Springer Science & Business Media. DOI: 10.1007/978-1-4614-4942-3 (cit. on pp. 15, 25).

Corteel (Sept. 2006). “On the use of irregularly spaced loudspeaker arrays for Wave Field Synthesis, potential impact on spatial aliasing frequency”. In: *Proc. of 9th Int. Conf. on Digital Audio Effects (DAFx-06)*. Montreal, Canada, pp. 209–214 (cit. on p. 63).

Corteel, Kuhn-Rahloff, and Pellegrini (May 2008). “Wave Field Synthesis Rendering with Increased Aliasing Frequency”. In: *Proc. of 124th Aud. Eng. Soc. Conv.* Amsterdam, The Netherlands (cit. on p. 46).

Daniel (2001). “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia”. PhD thesis. Université Paris 6 (cit. on p. 31).

Daniel (May 2003). “Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format”. In: *Proc. of 23rd Intl. Aud. Eng. Soc. Conf. on Signal Processing in Audio Recording and Reproduction*. Copenhagen, Denmark (cit. on pp. 5, 23, 31, 36).

Daniel, Rault, and Polack (Sept. 1998). “Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions”. In: *Proc. of 105th Aud. Eng. Soc. Conv.* San Francisco, USA (cit. on p. 34).

Dau, Püschel, and Kohlrausch (1996). “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure”. In: *J. Acoust. Soc. Am.* 99.6, pp. 3615–3622. DOI: 10.1121/1.414959 (cit. on p. 103).

Dietz, Ewert, and Hohmann (2011). “Auditory model based direction estimation of concurrent speakers from binaural signals”. In: *Speech Communication* 53.5, pp. 592–605. DOI: 10.1016/j.specom.2010.05.006 (cit. on p. 123).

Donley, Ritz, and Kleijn (June 2018). "Multizone Soundfield Reproduction With Privacy- and Quality-Based Speech Masking Filters". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 26.6, pp. 1041–1055. DOI: 10.1109/TASLP.2018.2798804 (cit. on p. 46).

Eckert and Hornung (July 2018). *Fake Science – Die Lügenmacher*. Das Erste. television documentary. URL: https://www.daserste.de/information/reportage-dokumentation/dokus/videos/exclusiv-im-ersten-fake-science-die-luegenmacher-video-102.html (cit. on p. 7).

Erbes and Spors (Mar. 2017). "Influence of the Listening Room on Spectral Properties of Wave Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Kiel, Germany, pp. 1057–1060 (cit. on p. 123).

Erbes, Geier, Wierstorf, and Spors (May 2017). "Free Database of Low Frequency Corrected Head-Related Transfer Functions and Headphone Compensation Filters". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany (cit. on p. 88).

Faller and Merimaa (2004). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence". In: *J. Acoust. Soc. Am.* 116.5, pp. 3075–89. DOI: 10.1121/1.1791872 (cit. on pp. 83, 103).

Fazi (2010). "Sound field reproduction". PhD thesis. University of Southampton (cit. on pp. 31, 47).

Fazi and Nelson (2013). "Sound field reproduction as an equivalent acoustical scattering problem". In: *J. Acoust. Soc. Am.* 134.5, pp. 3721–3729. DOI: 10.1121/1.4824343 (cit. on pp. 16, 17, 32, 48).

Feng, Yang, and Yang (2018). "Time-domain sound field reproduction using the group Lasso". In: *J. Acoust. Soc. Am.* 143.2, EL55–EL60. DOI: 10.1121/1.5022280 (cit. on p. 23).

Fink (Sept. 1992). "Time reversal of ultrasonic fields. I. Basic principles". In: *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 39.5, pp. 555–566. DOI: 10.1109/58.156174 (cit. on p. 28).

Firtha, Fiala, Schultz, and Spors (2017). "Improved Referencing Schemes for 2.5D Wave Field Synthesis Driving Functions". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 25.5, pp. 1117–1127. DOI: 10.1109/TASLP.2017.2689245 (cit. on pp. 6, 20, 25, 27, 28).

Firtha (2019). "A Generalized Wave Field Synthesis Framework with Application for Moving Virtual Sources". PhD thesis. Budapest University of Technology and Economics (cit. on pp. 8, 21, 25–27, 46, 50, 56, 57, 81, 123).

Franck (2011). "Efficient Algorithms for Arbitrary Sample Rate Conversion with Application to Wave Field Synthesis". PhD thesis. Technische Universität Ilmenau, p. 269 (cit. on p. 30).

Franck, Graefe, Korn, and Strauss (Sept. 2007). "Reproduction of Moving Sound Sources by Wave Field Synthesis: An Analysis of Artifacts". In: *Proc. of 32nd Intl. Aud. Eng. Soc. Conf. on DSP For Loudspeakers*. Hilleroed, Danmark (cit. on p. 30).

Frank (June 2013a). "Phantom Sources using Multiple Loudspeakers in the Horizontal Plane". PhD thesis. Graz, Austria: Institute of Electronic Music and Acoustics (cit. on pp. 34, 85, 101, 107, 115, 122, 123).

Frank (2013b). "Source Width of Frontal Phantom Sources: Perception, Measurement, and Modeling". In: *Archives of Acoustics* 38.3, pp. 311–319 (cit. on p. 101).

Frank, Zotter, and Sontacchi (Nov. 2008). "Localization experiments using different 2D Ambisonics decoders". In: *Proc. of 25th Tonmeistertagung – VDT Internation Convention*. Leipzig, Germany (cit. on pp. 85, 101).

Frenkel (Apr. 2018). "Human Localisation in Near-Field-Compensated Higher-Order Ambisonics". Bachelor's Thesis. University of Rostock (cit. on p. 95).

Geier and Spors (Nov. 2012). "Spatial Audio with the SoundScape Renderer". In: *Proc. of 27th Tonmeistertagung – VDT Internation Convention*. Cologne, Germany (cit. on pp. 29, 44, 87, 121).

Gel'fand and Shilov (1964). *Generalized functions: Vol. 1. Properties and operations*. New York, London: Academic Press (cit. on pp. 11, 13, 33, 48, 125).

GENESIS (Jan. 2010). *Loudness Toolbox 1.0*. URL: http://genesis-acoustics.com/en/loudness_online-32.html (cit. on pp. 96, 111).

Gerzon (1973). "Periphony: With-Height Sound Reproduction". In: *J. Aud. Eng. Soc.* 21.1, pp. 2–10 (cit. on pp. 4, 6, 31).

Girod, Rabenstein, and Stenger (2001). *Signal and Systems*. Wiley (cit. on pp. 10, 11, 29, 33, 47–49, 52, 67, 125, 129).

Glasberg and Moore (1990). "Derivation of auditory filter shapes from notched-noise data". In: *Hearing Research* 47.1, pp. 103–138. DOI: 10.1016/0378-5955(90)90170-T (cit. on p. 103).

Gradshteyn and Ryzhik (2007). *Tables of Integrals, Series, and Products*. Ed. by Alan Jeffrey and Daniel Zwillinger. 7th ed. Academic Press (cit. on p. 14).

Grosswald (1978). *Bessel Polynomials*. Vol. 698. Lecture Notes in Mathematics. Berlin Heidelberg, Germany: Springer-Verlag. DOI: 10.1007/BFb0063135 (cit. on p. 36).

Gumerov and Duraiswami (2004). *Fast multipole methods for the Helmholtz equation in three dimensions*. Oxford, UK: Elsevier Science (cit. on pp. 11, 17–19, 37).

Hahn and Meeker (1991). *Statistical Intervals: A Guide for Practitioners*. Vol. 392. John Wiley & Sons (cit. on p. 112).

Hahn and Spors (Mar. 2015a). "Modal Bandwidth Reduction in Data-based Binaural Synthesis including Translatory Head-movements". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 1122–1125 (cit. on pp. 35, 66).

Hahn and Spors (May 2015b). "Sound Field Synthesis of Virtual Cylindrical Waves Using Circular and Spherical Loudspeaker Arrays". In: *Proc. of 138th Aud. Eng. Soc. Conv.* Warsaw, Poland (cit. on p. 19).

Hahn and Spors (May 2017). "Further Investigations on the Design of Radial Filters for the Driving Functions of Near-Field Compensated Higher-Order Ambisonics". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany (cit. on pp. 36, 37).

Hahn, Winter, and Spors (June 2016). "Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain". In: *Proc. of 140th Aud. Eng. Soc. Conv.* Paris, France (cit. on pp. 7, 18, 19, 23, 37).

Hahn, Winter, and Spors (Aug. 2017). "Synthesis of a spatially band-limited plane wave in the time-domain using wave field synthesis". In: *Proc. of 25th European Signal Processing Conference (EUSIPCO)*. Kos Island, Greece, pp. 673–677. DOI: 10.23919/EUSIPCO.2017.8081292 (cit. on p. 122).

Hannemann and Donohue (2008). "Virtual Sound Source Rendering Using a Multipole-Expansion and Method-of-Moments Approach". In: *J. Aud. Eng. Soc.* 56.6, pp. 473–481 (cit. on p. 23).

Harris (Jan. 1978). "On the use of windows for harmonic analysis with the discrete Fourier transform". In: *Proceedings of the IEEE* 66.1, pp. 51–83. DOI: 10.1109/PROC.1978.10837 (cit. on p. 103).

Hartmann (1983). "Localization of sound in rooms". In: *J. Acoust. Soc. Am.* 74.5, pp. 1380–1391. DOI: 10.1121/1.390163 (cit. on p. 84).

Hartmann and Rakerd (1989). "On the minimum audible angle—A decision theory approach". In: *J. Acoust. Soc. Am.* 85.5, pp. 2031–2041. DOI: 10.1121/1.397855 (cit. on p. 84).

Hawley, Litovsky, and Culling (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer". In: *J. Acoust. Soc. Am.* 115.2, pp. 833–843. DOI: 10.1121/1.1639908 (cit. on p. 4).

Henderson (1975). "Best Linear Unbiased Estimation and Prediction under a Selection Model". In: *Biometrics* 31.2, pp. 423–447. DOI: 10.2307/2529430 (cit. on pp. 92, 93).

Howell (2010). *Statistical Methods for Psychology*. 7th ed. Wadsworth, Cengage Learning (cit. on pp. 93, 97, 98).

Hox (2010). *Multilevel Analysis: Techniques and Applications*. New York, USA: Routledge (cit. on p. 93).

Ifeachor and Jervis (2002). *Digital Signal Processing: A Practical Approach*. 2nd ed. Pearson Education (cit. on p. 37).

Immohr (Mar. 2017). "Zuhörerverfolgung für lokale Schallfeldsynthese". Bachelor's Thesis. University of Rostock (cit. on p. 44).

Ince, Hatton, and Graham-Cumming (Feb. 2012). "The case for open computer programs". In: *Nature* 482.7386, pp. 485–488. DOI: 10.1038/nature10836 (cit. on p. 7).

Ise (1999). "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems". In: *Acta Acustica united with Acustica* 85.1, pp. 78–87 (cit. on p. 23).

Jackson (Oct. 2000). "A correction to impulse invariance". In: *IEEE Signal Process. Lett.* 7.10, pp. 273–275. DOI: 10.1109/97.870677 (cit. on p. 37).

Jensen, Kuperman, Porter, and Schmidt (2011). *Computational Ocean Acoustics*. 2nd ed. Modern Acoustics and Signal Processing. New York, USA: Springer. DOI: 10.1007/978-1-4419-8678-8 (cit. on p. 11).

Jin and Kleijn (Dec. 2015). "Theory and Design of Multizone Soundfield Reproduction Using Sparse Methods". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 23.12, pp. 2343–2355. DOI: 10.1109/TASLP.2015.2479037 (cit. on p. 23).

Junger and Feit (1986). *Sound, Structures, and Their Interaction*. Cambridge, USA: MIT Press (cit. on p. 14).

Kennedy, Sadeghi, Abhayapala, and Jones (2007). "Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields". In: *IEEE Trans. Signal Process.* 55.6, pp. 2542–2556. DOI: 10.1109/TSP.2007.893738 (cit. on pp. 7, 22).

Kinsler, Frey, Coppens, and Sanders (Dec. 1999). *Fundamentals of Acoustics*. 4th ed. John Wiley & Sons, Inc. (cit. on p. 21).

Kirkeby and Nelson (1993). "Reproduction of plane wave sound fields". In: *J. Acoust. Soc. Am.* 94.5. DOI: 10.1121/1.407330 (cit. on p. 22).

Kirkeby, Nelson, Orduna-Bustamante, and Hamada (1996). "Local sound field reproduction using digital signal processing". In: *J. Acoust. Soc. Am.* 100.3. DOI: 10.1121/1.416060 (cit. on p. 22).

Kolundžija, Faller, and Vetterli (2011). "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion". In: *J. Aud. Eng. Soc.* 59.10, pp. 721–734 (cit. on p. 22).

Kruschke (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. 2nd ed. Academic Press (cit. on p. 118).

Kuntz (2009). "Wave Field Analysis Using Virtual Circular Microphone Arrays". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg (cit. on pp. 20, 22).

Laakso, Valimaki, Karjalainen, and Laine (Jan. 1996). "Splitting the unit delay [FIR/all pass filters design]". In: *IEEE Signal Process. Mag.* 13.1, pp. 30–60. DOI: 10.1109/79.482137 (cit. on p. 30).

Lazer et al. (2018). "The science of fake news". In: *Science* 359.6380, pp. 1094–1096. DOI: 10.1126/science.aao2998 (cit. on p. 7).

Leventhal (1986). "Type 1 and Type 2 Errors in the Statistical Analysis of Listening Tests". In: *J. Aud. Eng. Soc.* 34.6, pp. 437–453 (cit. on p. 98).

Linkwitz (1976). "Active Crossover Networks for Noncoincident Drivers". In: *J. Aud. Eng. Soc.* 24.1, pp. 2–8 (cit. on p. 39).

Lipshitz and Vanderkooy (1983). "A Family of Linear-Phase Crossover Networks of High Slope Derived by Time Delay". In: *J. Aud. Eng. Soc.* 31.1/2, pp. 2–20 (cit. on p. 40).

Litovsky, Colburn, Yost, and Guzman (1999). "The precedence effect". In: *J. Acoust. Soc. Am.* 106.4, pp. 1633–1654. DOI: 10.1121/1.427914 (cit. on p. 85).

Lösler and Zotter (Mar. 2015). "Comprehensive radial filter design for practical higher-order Ambisonic recording". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 452–455 (cit. on p. 39).

Lyon (1997). "All-pole models of auditory filtering". In: *Diversity in Auditory Mechanics*, pp. 205–211 (cit. on p. 103).

Ma and Brown (Sept. 2016). "Speech Localisation in a Multitalker Mixture by Humans and Machines". In: *Proc. of 17th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. San Francisco, USA, pp. 3359–3363 (cit. on p. 123).

Macpherson (June 2013). "Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation". In: *Proc. of Meetings on Acoustics* 19.1. DOI: 10.1121/1.4799913 (cit. on p. 84).

Makous and Middlebrooks (1990). "Two-dimensional sound localization by human listeners". In: *J. Acoust. Soc. Am.* 87.5, pp. 2188–2200. DOI: 10.1121/1.399186 (cit. on pp. 89, 91).

Manolakis and Ingle (June 2011). *Applied Digital Signal Processing: Theory and Practice*. Cambridge, USA: Cambridge University Press. DOI: 10.1017/CBO9780511835261 (cit. on pp. 29, 31, 36, 39, 48, 49).

May, van de Par, and Kohlrausch (Jan. 2011). "A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End". In: *IEEE Trans. Audio, Speech, Language Process.* 19.1, pp. 1–13. DOI: 10.1109/TASL.2010.2042128 (cit. on p. 103).

May, Ma, and Brown (Apr. 2015). "Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues". In: *Proc. of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brisbane, Australia, pp. 2679–2683. DOI: 10.1109/ICASSP.2015.7178457 (cit. on p. 123).

McClellan and Parks (2005). "A personal history of the Parks-McClellan algorithm". In: *IEEE Signal Process. Mag.* 22.2, pp. 82–86. DOI: 10.1109/MSP.2005.1406492 (cit. on p. 31).

McLean, Sanders, and Stroup (Feb. 1991). "A Unified Approach to Mixed Linear Models". In: *The American Statistician* 45.1, p. 54. DOI: 10.2307/2685241 (cit. on p. 93).

Mendonça and Delikaris-Manias (May 2018). "Statistical Tests with MUSHRA Data". In: *Proc. of 144th Aud. Eng. Soc. Conv.* Milan, Italy (cit. on p. 112).

Method for the subjective assessment of intermediate quality level of audio systems (2015). Standard ITU-R BS.1534-3. International Telecommunication Union Radiocommunication Assembly (cit. on pp. 108, 109).

Meyer (1984). "Computer Simulation of Loudspeaker Directivity". In: *J. Aud. Eng. Soc.* 32.5, pp. 294–315 (cit. on p. 56).

Mills (1958). "On the Minimum Audible Angle". In: *J. Acoust. Soc. Am.* 30.4, pp. 237–246. DOI: 10.1121/1.1909553 (cit. on pp. 83, 90).

Miyoshi and Kaneda (Feb. 1988). "Inverse filtering of room acoustics". In: *IEEE Trans. Acoust., Speech, Signal Process.* 36.2, pp. 145–152. DOI: 10.1109/29.1509 (cit. on p. 22).

Moore, Glasberg, and Baer (1997). "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness". In: *J. Aud. Eng. Soc.* 45.4, pp. 224–240 (cit. on pp. 96, 111).

Multichannel stereophonic sound system with and without accompanying picture (Aug. 2012). Standard ITU-R BS.775-3. International Telecommunication Union Radiocommunication Assembly (cit. on p. 4).

Möser (2009). *Engineering Acoustics*. Springer. DOI: 10.1007/978-3-540-92723-5 (cit. on p. 13).

Oldfield (Apr. 2013). "The analysis and improvement of focused source reproduction with wave field synthesis". PhD thesis. Salford, UK: University of Salford (cit. on p. 46).

Olive, Welti, and Martens (May 2007). "Listener Loudspeaker Preference Ratings Obtained In Situ Match those Obtained Via a Binaural Room Scanning Measurement and Playback System". In: *Proc. of 122nd Aud. Eng. Soc. Conv.* Vienna, Austria (cit. on p. 109).

Ono, Pulkki, and Karjalainen (Nov. 2001). "Binaural Modeling of Multiple Sound Source Perception: Methodology and Coloration Experiments". In: *Proc. of 111th Aud. Eng. Soc. Conv.* New York, USA (cit. on p. 115).

Ono, Pulkki, and Karjalainen (Apr. 2002). "Binaural Modeling of Multiple Sound Source Perception: Coloration of Wideband Sound". In: *Proc. of 112th Aud. Eng. Soc. Conv.* Munich, Germany (cit. on p. 115).

Pierce (2014). "Basic Linear Acoustics". In: *Springer Handbook of Acoustics*. Ed. by Thomas D Rossing. 2nd ed. Springer, pp. 29–115. DOI: 10.1007/978-1-4939-0755-7 (cit. on p. 10).

Poletti (Mar. 2017). "Spherical coordinate descriptions of cylindrical and spherical Bessel beams". In: *J. Acoust. Soc. Am.* 141.3, pp. 2069–2078. DOI: 10.1121/1.4978787 (cit. on p. 131).

Poletti (2005). "Three-dimensional surround sound systems based on spherical harmonics". In: *J. Aud. Eng. Soc.* 11.53, pp. 1004–1025 (cit. on p. 23).

Pomberger (2008). "Angular and Radial Directivity Control for Spherical Loudspeaker Arrays". Diploma Thesis. University of Music and Dramatic Arts, Graz, Austria (cit. on p. 36).

Popko (Sept. 2013). "Zur Hörbarkeit und Interpolation von Kopf-über-Torso-Orientierungen bei Aufnahmen mit einem Kopf-und-Torso-Simulator". Bachelor's Thesis. Technische Universität Berlin (cit. on p. 86).

Powell and Chau (Nov. 1991). "A technique for realizing linear phase IIR filters". In: *IEEE Trans. Signal Process.* 39.11, pp. 2425–2435. DOI: 10.1109/78.97998 (cit. on p. 41).

Procedure for the Computation of Loudness of Steady Sounds (2007). Standard ANSI S3.4-2007. New York, USA: American National Standards Institute (cit. on pp. 96, 111).

Puckette (Sept. 1996). "Pure Data: another integrated computer music environment". In: *Proc. of the Second Intercollege Computer Music Concerts*. Tachikawa, Japan, pp. 37–41 (cit. on p. 87).

Pulkki (1997). "Virtual Sound Source Positioning Using Vector Base Amplitude Panning". In: *J. Aud. Eng. Soc.* 45.6, pp. 456–466 (cit. on pp. 4, 6).

Pulkki (May 2001). "Coloration of Amplitude-Panned Virtual Sources". In: *Proc. of 110th Aud. Eng. Soc. Conv.* Amsterdam, The Netherlands (cit. on p. 115).

Pulkki, Karjalainen, and Huopaniemi (1999). "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model". In: *J. Aud. Eng. Soc.* 47.4, pp. 203–217 (cit. on pp. 115, 116).

Puria (2003). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions". In: *J. Acoust. Soc. Am.* 113.5, pp. 2773–2789. DOI: 10.1121/1.1564018 (cit. on p. 103).

Raake and Wierstorf (Sept. 2016). "Assessment of audio quality and experience using binaural-hearing models". In: *Proc. of 22nd Int. Congress on Acoustics (ICA)*. Buenos Aires, Argentina (cit. on p. 6).

Rafaely (2015). *Fundamentals of Spherical Array Processing*. Heidelberg, Germany: Springer. DOI: 10.1007/978-3-662-45664-4 (cit. on p. 22).

Robinson (Feb. 1991). "That BLUP is a Good Thing: The Estimation of Random Effects". In: *Statistical Science* 6.1, pp. 15–32. DOI: 10.1214/ss/1177011926 (cit. on p. 93).

Rumsey (2001). *Spatial Audio*. Focal Press (cit. on p. 4).

Rumsey, Zieliński, Kassier, and Bech (2005). "On the relative importance of spatial and timbral fiedelities in judgements of degraded multichannel audio quality". In: *J. Acoust. Soc. Am.* 118.2, pp. 968–976. DOI: 10.1121/1.1945368 (cit. on pp. 6, 83, 107, 123).

Salvador (May 2010). "Wave Field Synthesis Using Fractional Order Filters and Fractional Delays". In: *Proc. of 120th Aud. Eng. Soc. Conv.* London, UK (cit. on p. 29).

Schneider and Wierstorf (June 2014). *Gnuplot-colorbrewer: ColorBrewer color schemes for gnuplot*. DOI: 10.5281/zenodo.10282 (cit. on p. 8).

Schultz, Erbes, Spors, and Weinzierl (Mar. 2013). "Derivation of IIR-pre-filters for soundfield synthesis using linear secondary source distributions". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Meran, Italy (cit. on p. 30).

Schultz (2016). "Sound Field Synthesis for Line Source Array Applications in Large-Scale Sound Reinforcement". PhD thesis. University of Rostock. DOI: 10.18453/rosdok_id00001765 (cit. on pp. 13, 15, 24, 25, 28, 29, 48–50, 123).

Schultz and Spors (2014). "Comparing Approaches to the Spherical and Planar Single Layer Potentials for Interior Sound Field Synthesis". In: *Acta Acustica united with Acustica* 100.5, pp. 900–911. DOI: 10.3813/AAA.918769 (cit. on pp. 16, 32).

Schultz, Firtha, Winter, and Spors (Mar. 2019). "On the Connections of High-Frequency Approximated Ambisonics and Wave Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Rostock, Germany, pp. 1446–1449 (cit. on p. 65).

Schymura, Winter, Kolossa, and Spors (Sept. 2015). "Binaural Sound Source Localisation and Tracking using a dynamic Spherical Head Model". In: *Proc. of 16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Dresden, Germany, pp. 165–169 (cit. on p. 123).

Seeber (2003). "Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode". PhD thesis. Technischen Universität München (cit. on p. 91).

Siegel (1956). *Nonparametric statistics for the behavioral sciences*. McGraw-hill (cit. on p. 112).

Solvang (2008). "Spectral Impairment of Two-Dimensional Higher Order Ambisonics". In: *J. Aud. Eng. Soc.* 56.4, pp. 267–279 (cit. on p. 107).

Sporer, Liebetrau, and Schneider (Oct. 2009). "Statistics of MUSHRA Revisited". In: *Proc. of 127th Aud. Eng. Soc. Conv.* New York, USA (cit. on p. 112).

Spors (Mar. 2008). "Investigation of spatial aliasing artifacts of wave field synthesis in the temporal domain". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Dresden, Germany (cit. on p. 54).

Spors (2006). "Active listening room compensation for spatial sound reproduction systems". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg. (cit. on pp. 12, 14).

Spors (Oct. 2007). "Extension of an Analytic Secondary Source Selection Criterion for Wave Field Synthesis". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA (cit. on p. 26).

Spors and Ahrens (Oct. 2008). "A comparison of wave field synthesis and higher-order Ambisonics with respect to physical properties and spatial sampling". In: *Proc. of 125th Aud. Eng. Soc. Conv.* San Fransisco, USA (cit. on p. 46).

Spors and Ahrens (Oct. 2009). "Spatial Sampling Artifacts of Wave Field Synthesis for the Reproduction of Virtual Point Sources". In: *Proc. of 126th Aud. Eng. Soc. Conv.* Munich, Germany (cit. on p. 46).

Spors and Ahrens (May 2010a). "Analysis and Improvement of Pre-equalization in 2.5-dimensional Wave Field Synthesis". In: *Proc. of 128th Aud. Eng. Soc. Conv.* London, UK (cit. on p. 30).

Spors and Ahrens (Oct. 2010b). "Local Sound Field Synthesis by Virtual Secondary Sources". In: *Proc. of 40th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Tokyo, Japan (cit. on pp. 5, 7, 23, 41, 43, 121).

Spors and Rabenstein (May 2006). "Spatial Aliasing Artifacts Produced by Linear and Circular Loudspeaker Arrays used for Wave Field Synthesis". In: *Proc. of 120th Aud. Eng. Soc. Conv.* Paris, France (cit. on pp. 46–48).

Spors, Rabenstein, and Ahrens (May 2008). "The theory of Wave Field Synthesis revisited". In: *Proc. of 124th Aud. Eng. Soc. Conv.* Amsterdam, The Netherlands (cit. on p. 25).

Spors, Kuscher, and Ahrens (Oct. 2011). "Efficient realization of model-based rendering for 2.5-dimensional near-field compensated higher order Ambisonics". In: *Proc. of 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, USA. DOI: 10.1109/ASPAA.2011.6082325 (cit. on pp. 22, 36, 37).

Start (1997). "Direct Sound Enhancement by Wave Field Synthesis". PhD thesis. Delft University of Technology (cit. on pp. 25, 50, 84).

Stitt (June 2015). "Ambisonics and Higher-Order Ambisonics for Off-Centre Listeners: Evaluation of Perceived and Predicted Image Direction". PhD thesis. Belfast, UK: Queen's University Belfast (cit. on pp. 85, 101, 122).

Teutsch (2006). "Wavefield Decomposition Using Microphone Arrays and Its Application to Acoustic Scene Analysis". PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg (cit. on p. 22).

Torick (1998). "Highlights in the History of Multichannel Sound". In: *J. Aud. Eng. Soc.* 46.1/2, pp. 27–31 (cit. on p. 4).

Tufte (May 2006). *Beautiful Evidence*. 1st ed. Cheshire, Connecticut: Graphics Press, LLC (cit. on p. 8).

Two!Ears Team (Oct. 2018). *Two!Ears Auditory Model 1.5*. DOI: 10.5281/zenodo.1458420 (cit. on pp. 8, 103).

Van Trees (2004). *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons (cit. on pp. 5, 49, 57).

Vargha and Delaney (2000). "A Critique and Improvement of the "CL" Common Language Effect Size Statistics of McGraw and Wong". In: *Journal of Educational and Behavioral Statistics* 25.2, pp. 101–132. DOI: 10.3102/10769986025002101 (cit. on p. 113).

Verheijen (1997). "Sound Reproduction by Wave Field Synthesis". PhD thesis. Delft University of Technology (cit. on pp. 25, 26, 84).

Vogel (1993). "Application of wave field synthesis in room acoustics". PhD thesis. Delft University of Technology (cit. on pp. 25, 84).

de Vries and Berkhout (1981). "Wave theoretical approach to acoustic focusing". In: *J. Acoust. Soc. Am.* 70.3. DOI: 10.1121/1.386911 (cit. on p. 28).

Wierstorf (2014). "Perceptual Assessment of sound field synthesis". PhD thesis. Technische Universität Berlin. DOI: 10.14279/depositonce-4310 (cit. on pp. 5–8, 22, 60, 61, 84, 86, 88, 89, 92, 95, 96, 99, 101, 107–110, 116, 122, 123).

Wierstorf (June 2016). *Listening test results for sound field synthesis localization experiment*. DOI: 10.5281/zenodo.55439 (cit. on pp. 96, 97).

Wierstorf and Hohnerlein (Nov. 2016). *Coloration of a point source in Wave Field Synthesis – data*. DOI: 10.5281/zenodo.164589 (cit. on p. 108).

Wierstorf and Spors (Apr. 2012). "Sound Field Synthesis Toolbox". In: *Proc. of 132nd Aud. Eng. Soc. Conv.* Budapest, Hungary (cit. on pp. 8, 29).

Wierstorf, Geier, and Spors (May 2011). "A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances". In: *Proc. of 130th Aud. Eng. Soc. Conv.* London, UK (cit. on p. 88).

Wierstorf, Spors, and Raake (Sept. 2012). "Perception and evaluation of sound fields". In: *59th Open Seminar on Acoustics*. Boszkowo, Poland (cit. on pp. 86, 88–94, 106).

Wierstorf, Hohnerlein, Spors, and Raake (Aug. 2014). "Coloration in Wave Field Synthesis". In: *Proc. of 55th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*. Helsinki, Finland (cit. on pp. 6, 107, 108).

Wierstorf, Ende, and Raake (Mar. 2015). "Klangverfärbung in der Wellenfeldsynthese - Experimente und Modellierung". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 490–493 (cit. on p. 115).

Wierstorf, Raake, and Spors (2017). "Assessing localization accuracy in sound field synthesis". In: *J. Acoust. Soc. Am.* 141.2, pp. 1111–1119. DOI: 10.1121/1.4976061 (cit. on pp. 84, 85, 103).

Wierstorf, Hold, and Raake (2018). "Listener Preference for Wave Field Synthesis, Stereophony, and Different Mixes in Popular Music". In: *J. Aud. Eng. Soc.* 66.5, pp. 385–396 (cit. on p. 123).

Wierstorf, Winter, Rettberg, Erbes, Hahn, Schultz, Hold, and Spors (Mar. 2019). *Sound Field Synthesis Toolbox 2.5.0*. DOI: 10.5281/zenodo.2597212 (cit. on p. 8).

Wightman and Kistler (1992). "The dominant role of low-frequency interaural time differences in sound localization". In: *J. Acoust. Soc. Am.* 91.3, pp. 1648–1661. DOI: 10.1121/1.402445 (cit. on p. 84).

Wilcoxon (1945). "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6, pp. 80–83. DOI: 10.2307/3001968 (cit. on p. 112).

Williams (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press (cit. on pp. 8–16, 18, 25, 32, 53, 73, 121, 131).

Winter, Wierstorf, Hold, Krüger, Raake, and Spors (Oct. 2018a). "Colouration in Local Wave Field Synthesis". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 26.10, pp. 1913–1924. DOI: 10.1109/TASLP.2018.2842435 (cit. on p. 107).

Winter, Schultz, Firtha, and Spors (June 2019a). "A Geometric Model for Prediction of Spatial Aliasing in 2.5D Sound Field Synthesis". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 27.6, pp. 1031–1046. DOI: 10.1109/TASLP.2019.2892895 (cit. on pp. 28, 46, 50).

Winter (2019). *Local Sound Field Synthesis – Supplements 1.0.1*. DOI: 10.5281/zenodo.3543537 (cit. on p. 8).

Winter and Frenkel (2019). *Data of Listening Experiments for Azimuthal Localisation in (Local) Sound Field Synthesis*. DOI: 10.5281/zenodo.3252380 (cit. on p. 99).

Winter and Spors (May 2015). "Physical Properties of Local Wave Field Synthesis using Linear Loudspeaker Arrays". In: *Proc. of 138th Aud. Eng. Soc. Conv.* Warsaw, Poland (cit. on p. 41).

Winter and Spors (Aug. 2016). "On Fractional Delay Interpolation for Local Wave Field Synthesis". In: *Proc. of 24th European Signal Processing Conference (EUSIPCO)*. Budapest, Hungary, pp. 2415–2419. DOI: 10.1109/EUSIPCO.2016.7760682 (cit. on pp. 29–31).

Winter and Spors (Mar. 2017). "Time-Domain Realisations of 2.5-Dimensional Local Sound Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Kiel, Germany, pp. 1134–1137 (cit. on p. 29).

Winter, Ahrens, and Spors (May 2016). "On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.5, pp. 914–926. DOI: 10.1109/TASLP.2016.2531902 (cit. on pp. 22, 25, 37, 46, 47).

Winter, Wierstorf, and Spors (2017a). *Improvement of reporting method for closed-loop human localization experiments – Data*. DOI: 10.5281/zenodo.245826 (cit. on p. 93).

Winter, Wierstorf, and Spors (May 2017b). "Improvement of the reporting method for closed-loop human localization experiments". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany (cit. on pp. 85, 91).

Winter, Hahn, and Spors (Aug. 2017c). "Time-Domain Realisation of Model-Based Rendering for 2.5D Local Wave Field Synthesis Using Spatial Bandwidth-Limitation". In: *Proc. of 25th European Signal Processing Conference (EUSIPCO)*. Kos Island, Greece, pp. 688–692. DOI: 10.23919/EUSIPCO.2017.8081295 (cit. on p. 37).

Winter, Wierstorf, Hold, Krüger, Raake, and Spors (2018b). *Colouration in Local Wave Field Synthesis – Data*. DOI: 10.5281/zenodo.1158027 (cit. on p. 113).

Winter, Schultz, and Spors (Mar. 2019b). "Array Design for Increased Spatial Aliasing Frequency in Wave Field Synthesis Based on a Geometric Model". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Rostock, Germany, pp. 463–466 (cit. on p. 64).

Wittek (2007). "Perceptual differences between wavefield synthesis and stereophony". PhD thesis. University of Surrey (cit. on pp. 6, 108).

Wittek, Rumsey, and Theile (Oct. 2007). "On the Sound Color Properties of Wavefield Synthesis and Stereo". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA (cit. on pp. 6, 107, 108).

Wong (2001). *Asymptotic Approximations of Integrals*. Vol. 34. Classics in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM) (cit. on p. 124).

Wu and Abhayapala (Aug. 2011). "Spatial Multizone Soundfield Reproduction: Theory and Design". In: *IEEE Trans. Audio, Speech, Language Process.* 19.6, pp. 1711–1720. DOI: 10.1109/TASL.2010.2097249 (cit. on p. 23).

Yon, Tanter, and Fink (2003). "Sound focusing in rooms: The time-reversal approach". In: *J. Acoust. Soc. Am.* 113.3. DOI: 10.1121/1.1543587 (cit. on p. 28).

Zahorik, Brungart, and Bronkhorst (2005). "Auditory Distance Perception in Humans: A Summary of Past and Present Research". In: *Acta Acustica united with Acustica* 91.3, pp. 409–420 (cit. on p. 84).

Zielinski, Hardisty, Hummersone, and Rumsey (Oct. 2007). "Potential Biases in MUSHRA Listening Tests". In: *Proc. of 123rd Aud. Eng. Soc. Conv.* New York, USA (cit. on p. 118).

Zotter, Pomberger, and Noisternig (May 2010). "Ambisonic Decoding with and without Mode-Matching: A Case Study Using the Hemisphere". In: *Proc. of 2nd Int. Symp. on Ambisonics and Spherical Acoustics*. Paris, France (cit. on p. 31).

Zotter, Pomberger, and Noisternig (2012). "Energy-Preserving Ambisonic Decoding". In: *Acta Acustica united with Acustica* 98.1. DOI: 10.3813/AAA.918490 (cit. on p. 34).

Zotter and Frank (2019). *Ambisonics*. Springer International Publishing. DOI: 10.1007/978-3-030-17207-7 (cit. on pp. 101, 123).

Zotter and Spors (Sept. 2013). "Is sound field control determined at all frequencies? How is it related to numerical acoustics?" In: *Proc. of 52nd Intl. Aud. Eng. Soc. Conf. on Sound Field Control - Engineering and Perception*. Guildford, UK (cit. on p. 17).

Own Publications

Hahn, Winter, and Spors (June 2016). “Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain”. In: *Proc. of 140th Aud. Eng. Soc. Conv.* Paris, France.

Hahn, Winter, and Spors (Aug. 2017). “Synthesis of a spatially band-limited plane wave in the time-domain using wave field synthesis”. In: *Proc. of 25th European Signal Processing Conference (EUSIPCO)*. Kos Island, Greece, pp. 673–677. DOI: 10.23919/EUSIPCO.2017.8081292.

Hahn, Winter, and Spors (Sept. 2019). “2.5D Local wave field synthesis of a virtual plane wave using a time domain representation of spherical harmonics expansion”. In: *Proc. of 23rd International Congress on Acoustics (ICA)*. Aachen, Germany, pp. 1132–1139.

Schultz, Firtha, Winter, and Spors (Mar. 2019). “On the Connections of High-Frequency Approximated Ambisonics and Wave Field Synthesis”. In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Rostock, Germany, pp. 1446–1449.

Schymura, Winter, Kolossa, and Spors (Sept. 2015). “Binaural Sound Source Localisation and Tracking using a dynamic Spherical Head Model”. In: *Proc. of 16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Dresden, Germany, pp. 165–169.

Winter, Hold, Wierstorf, Raake, and Spors (Oct. 2017a). “Colouration in 2.5D local wave field synthesis using spatial bandwidth-limitation”. In: *Proc. of 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, USA, pp. 160–165. DOI: 10.1109/WASPAA.2017.8170015.

Winter, Ahrens, and Spors (Mar. 2018a). “A Geometric Model for Spatial Aliasing in Wave Field Synthesis”. In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Munich, Germany, pp. 1310–1313.

Winter, Wierstorf, Hold, Krüger, Raake, and Spors (Oct. 2018b). “Colouration in Local Wave Field Synthesis”. In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 26.10, pp. 1913–1924. DOI: 10.1109/TASLP.2018.2842435.

Winter, Schultz, Firtha, and Spors (June 2019a). “A Geometric Model for Prediction of Spatial Aliasing in 2.5D Sound Field Synthesis”. In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 27.6, pp. 1031–1046. DOI: 10.1109/TASLP.2019.2892895.

Winter and Spors (Mar. 2015a). “Parameter Analysis for Range Extrapolation of Head-Related Transfer Functions using Virtual Local Wave Field Synthesis”. In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Nuremberg, Germany, pp. 1131–1134.

Winter and Spors (May 2015b). “Physical Properties of Local Wave Field Synthesis using Circular Loudspeaker Arrays”. In: *Proc. of 10th European Congress and Exposition on Noise Control Engineering (EuroNoise)*. Vol. 31. c. Maastricht, The Netherlands, pp. 263–268.

Winter and Spors (May 2015c). “Physical Properties of Local Wave Field Synthesis using Linear Loudspeaker Arrays”. In: *Proc. of 138th Aud. Eng. Soc. Conv.* Warsaw, Poland.

Winter and Spors (Mar. 2016a). “A Comparison of Sound Field Synthesis Techniques for Non-Smooth Secondary Source Distributions”. In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Aachen, Germany, pp. 1463–1466.

Winter and Spors (Aug. 2016b). "On Fractional Delay Interpolation for Local Wave Field Synthesis". In: *Proc. of 24th European Signal Processing Conference (EUSIPCO)*. Budapest, Hungary, pp. 2415–2419. DOI: 10.1109/EUSIPCO.2016.7760682.

Winter and Spors (Mar. 2017). "Time-Domain Realisations of 2.5-Dimensional Local Sound Field Synthesis". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Kiel, Germany, pp. 1134–1137.

Winter, Schultz, and Spors (Sept. 2014). "Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements". In: *Proc. of Forum Acusticum*. Kraków, Poland.

Winter, Wierstorf, Podlubne, Forgue, Manhès, Herrb, Raake, and Danès (June 2016a). "Database of binaural room impulse responses of an apartment-like environment". In: *Proc. of 140th Aud. Eng. Soc. Conv.* Paris, France, eBrief 252.

Winter, Ahrens, and Spors (May 2016b). "On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.5, pp. 914–926. DOI: 10.1109/TASLP.2016.2531902.

Winter, Wierstorf, and Spors (May 2017b). "Improvement of the reporting method for closed-loop human localization experiments". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany.

Winter, Wierstorf, Raake, and Spors (May 2017c). "The Two!Ears Database". In: *Proc. of 142nd Aud. Eng. Soc. Conv.* Berlin, Germany.

Winter, Hahn, and Spors (Aug. 2017d). "Time-Domain Realisation of Model-Based Rendering for 2.5D Local Wave Field Synthesis Using Spatial Bandwidth-Limitation". In: *Proc. of 25th European Signal Processing Conference (EUSIPCO)*. Kos Island, Greece, pp. 688–692. DOI: 10.23919/EUSIPCO.2017.8081295.

Winter, Schultz, and Spors (Mar. 2019b). "Array Design for Increased Spatial Aliasing Frequency in Wave Field Synthesis Based on a Geometric Model". In: *Proc. of German Annual Conference on Acoustics (DAGA)*. Rostock, Germany, pp. 463–466.

Abstract

This thesis investigates the physical and perceptual properties of selected methods for (Local) Sound Field Synthesis ((L)SFS). First, the mathematical foundations of the approaches are discussed. Special attention is drawn to the implementation of the methods in the discrete-time domain as a consequence of digital signal processing. The influence of their parametrisation on the properties of the synthesised sound field is examined on a qualitative level. A geometric model is developed to predict spatial aliasing artefacts caused by the spatial discretisation of the deployed loudspeaker arrays. In agreement with numerical sound field simulations, the geometric model shows an increase of synthesis accuracy for LSFS compared to conventional SFS approaches. However, the difference in accuracy gets smaller, the closer the listener is located to the active loudspeakers.

With the help of binaural synthesis, the different (L)SFS approaches are assessed within listening experiments targeting their spatial and timbral fidelity. The results show that LSFS performs at least as good as conventional methods for azimuthal sound source localisation. A significant increase of timbral fidelity is achieved with distinct parametrisations of the LSFS approaches.

Zusammenfassung

Die vorliegende Dissertation untersucht die physikalischen und perzeptiven Eigenschaften von ausgewählten Verfahren zur (lokalen) Schallfeldsynthese ((L)SFS). Zunächst werden die mathematischen Grundlagen dieser Methoden diskutiert. Dabei liegt ein besonderes Augenmerk auf der zeitdiskreten Implementierung der Verfahren. Der Einfluss verschiedener Parametrierungen auf die Eigenschaften der synthetisierten Schallfelder wird qualitativ ausgewertet. Ein eigens entwickeltes geometrisches Modell prädiziert räumliches Aliasing als Folge der räumlichen Abtastung durch die eingesetzten Lautsprecheranordnungen. Zusammen mit numerischen Simulationen zeigt das Modell, dass LSFS gegenüber konventioneller SFS zu einer erhöhten Genauigkeit der Synthese führt. Je näher sich der Zuhörer jedoch an den aktiven Lautsprechern befindet, desto kleiner ist der Genauigkeitsunterschied zwischen beiden Ansätzen.

Mit Hilfe der Binauralsynthese werden verschiedene (L)SFS Verfahren in Hörversuchen auf ihre räumliche und klangliche Treue untersucht. Im Bezug auf die horizontale Lokalisierung von Schallquellen erreicht die LSFS eine Genauigkeit, welche mindestens gleich der von konventionellen Methoden ist. Für bestimmte Parametrierungen der LSFS Verfahren wird eine signifikant verbesserte klangliche Treue erreicht.

Selbstständigkeitserklärung

Ich, Fiete Winter, stelle fest und versichere, dass (i) diese der Universität Rostock - erstmalig einer akademischen Institution zur Prüfung - vorgelegte Dissertation mit dem Titel „Local Sound Field Synthesis“ von mir selbständig und ohne fremde Hilfe verfasst wurde, ich (ii) andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt habe und ich (iii) die den benutzten Werken wörtlich oder inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Fiete Winter
Rostock, 25.06.2019 & Rostock, 18.11.2019