Acoustic scene analysis by multiperspective sector-based sound field decomposition

Michael McCrea

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Helsinki 8.8.2021

Supervisor

Prof. Ville Pulkki

Advisor

MSc. Leo McCormack

Aalto University
School of Electrical
Engineering

Aalto University
School of Electrical
Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Michael McCrea

Title Acoustic scene analysis by multiperspective sector-based sound field decomposition

Degree programme Computer, Communication and Information Sciences

Major Acoustics and Audio Technology **Code of major** ELEC3030

Supervisor Prof. Ville Pulkki

Advisor MSc. Leo McCormack

Date 8.8.2021 **Number of pages** 71 **Language** English

Abstract

In this thesis, the performance of high-order sound field analysis by spatial filtering is evaluated in the context of 3D source localization. First, various factors in 'sector-based' processing—including analysis order, the presence of interfering noise, and source direction—are assessed using two key sound field indicators derived from intensimetric analysis: sound field diffuseness estimation and source direction-of-arrival (DoA) estimation. This carries forward to the evaluation of a 3D source localization technique which utilizes concurrent analyses from multiple receivers. The evaluation is carried out by simulation of ideal spherical harmonic receiver signals of sound fields comprised of fundamental components—a mixture of plane wave sound sources and an isotropic diffuse field—which adhere to an underlying plane wave model that has previously been characterized analytically. This model is then challenged with two scenarios: the presence of an interfering sound source, and an anisotropic, partially-correlated reverberant field. The high-order, sector-based approach to source localization is shown to be a consistent improvement upon the (first-order) method without spatial filtering, operating primarily through the principle of increasing the direct-to-diffuse (or signal-to-noise) energy ratio. However, numerous considerations must be taken if robust localization is to be performed over a large spatial extent. These include sector orientation, the arrangement and number of receivers, and estimation filtering and culling. Informed by the system performance under the tested conditions as well as the analytical model describing the sound field under the action of spatial filtering, optimization techniques are proposed, tested, and found to be successful under specified constraints. These include employing the diffuseness metric for weighted-DoA estimation in the localization task, an iterative approach localization using the estimated source distance to weight the contribution of DoA estimates, and a spatial sweep technique which applies a diffuseness constraint for DoA estimation culling which moves toward multisource localization.

Keywords spatial audio, sound field analysis, Ambisonics, HOA, spatial filtering

Preface

This work was carried out in the Department of Signal Processing and Acoustics at Aalto University in Finland.

I would first like to thank Prof. Ville Pulkki for the opportunity to work on this project, and Leo McCormack for his helpful guidance along the way. Thanks also to Rapolas Daugintis for the support in untangling ideas, notation, and contact mic cables. I'm grateful to have found a dear friend, Janis Heldmann, during my time in the program, whose humor, curiosity and evergreen spirit are a constant inspiration. I am eternally thankful for the love and support of my parents throughout the years and across oceans. I also want to express my appreciation to those who were along for the journey, and who made my time in Helsinki shine: Nella for sharing her boundless enthusiasm, her art, and the realest talk, Joonas for bringing the big sound and big support, and of course, Jo and Laura for sharing in all of it more.

I would like to extend special thanks to Dr. Joseph Anderson for his continued guidance, for teaching me how to Think Ambisonically™, and for his spirited insights that always stoke my own curiosity and excitement for spatial sound.

Helsinki, 8.8.2021

Michael T. McCrea

Contents

Symbols and abbreviations

Symbols

a	amplitude density function, e.g. for a plane wave a_{pw}
a_{lm}	ambisonic set component
$\mathbf{a}$	ambisonic set vector
$\mathbf{A}^{\mathrm{i}}$	matrix to generate velocity beamforming weights from sector pattern weights $\mathbf{w}_{\mathrm{p}}$, $\mathbf{A}^{\mathrm{i}} \in \mathbb{R}^{(L+1)^2 \times L^2}$ for $i = \{x, y, z\}$
b	directivity function of an axisymmetric beam pattern
b_l	axisymmetric beamformer weight for spherical harmonic degree l
c	speed of sound
d_x	dipole directivity pattern oriented along the x-axis
$\mathbf{d}$	dipole directivity pattern vector, $\mathbf{d} = [d_x, d_y, d_z]^\top$
f	frequency
L	spherical harmonic order
l	spherical harmonic degree $l \in [0, L]$
m	spherical harmonic index $m \in [-L, L]$
n	Ambisonic Channel Number index $n = l(l+1) + m$, $n \in [0, N]$
p	sound pressure
P_{pw}	power spectral density of a plane wave signal
$\mathbf{I}$	sound intensity vector, with active and reactive parts, $\mathbf{I}_a$ and $\mathbf{I}_r$
k	frequency bin index
S^2	the unit sphere
s	pressure signal, e.g. of an istoropic diffuse field s_{df}
$\mathbf{s}_{\mathrm{pv}}$	cross spectral density of the pressure and velocity signals
S_{pp}	power spectral density of pressure
S_{vv}	power spectral density of velocity
$\mathbf{s}_{\mathrm{w}}$	pressure-velocity signal vector of a sector
t	time
u	one-dimensional sound particle velocity
$\mathbf{u}$	sound particle velocity vector
$\mathbf{v}$	sound particle velocity signal vector
$\mathbf{w}_{\mathrm{p}}$	beamforming coefficient vector to generate a weighted-pressure pattern
$\mathbf{w}_{\mathrm{d,w}}$	beamforming coefficient vector to generate weighted-dipole patterns
w	weight coefficient
w_{lm}	weight coefficients for spherical harmonic components
$\mathbf{w}$	vector of weighting coefficients
$\mathbf{W}$	matrix of weighting coefficients
Y_{lm}	spherical harmonics of degree l, mode index m
$\mathbf{y}$	vector of spherical harmonic coefficients
Z_0	characteristic impedance of air, $Z_0 = c\rho_0$

α	angle-of-arrival with the sector axis
β	direction-of-arrival bias introduced by diffuse field under spatial filtering
$\boldsymbol{\gamma}$	Cartesian unit 3-vector
Γ	direct-to-diffuse ratio
ζ	speed of energy propagation
θ	inclination, $\theta \in [0, \pi]$, zenith $= 0$, nadir $= \pi$
ϑ	user-specified normalization coefficient
ρ_0	equilibrium (ambient) density of air
ϕ	elevation, $\phi \in [-\pi/2, \pi/2]$, 0 at the horizon, zenith $= \pi/2$, nadir $= -\pi/2$, and $\phi = \pi/2 - \theta$
φ	azimuth, $\varphi \in [-\pi, \pi]$
ψ	diffuseness
ω	angular frequency $\omega = 2\pi f$
Ω	spherical direction duple of the azimuth and inclination, $\Omega = (\varphi, \theta)$

Abbreviations

ACN	Ambisonic Channel Number
AoA	angle-of-arrival
CSD	cross-spectral density
DoA	direction-of-arrival
Dof	degrees-of-freedom, e.g. 3DoF, 6Dof
DDR	direct-to-diffuse ratio
HOA	higher-order Ambisonics
HRTF	head-related transfer function
ISHT	inverst spherical harmonic transform
MEE	mean estimation error
PSD	power spectral density
SH	spherical harmonic
SHD	spherical harmonic domain
SHT	spherical harmonic transform
SMA	spherical microphone array
SPL	sound pressure level
SRP	steered-response power

1 Introduction

1.1 Background and motivation

In recent years, digital signal processing has made significant progress in the capture, manipulation, and reproduction of spatial sound for loudspeakers and headphones. In particular, rapid advances in virtual, augmented and synthesized reality research and remote presence applications have brought renewed attention to virtual 3D sound reproduction. This is due to the fact that accurate spatial sound is requisite to create a convincing sense of virtual presence, along with an emerging requirement that a virtual auditory environment be responsive to a listener that is fully mobile. As such, there is increasing demand to accurately capture and reconstruct spatial sound scenes across navigable space. However, a prerequisite to effective auditory scene reconstruction is successful *sound field decomposition*—the ability to infer the spatial organization and dynamic behavior of recorded sound sources—which remains an open challenge.

Spherical microphone arrays (SMAs) have proven to be a central tool in sound field capture and analysis, as they can offer an efficient means of encoding incident sounds from all directions via a truncated spherical harmonic (SH) series [1, 2, 3]. The resultant signals comprise a spatial sound format, known as Higher Order Ambisonics (HOA), representing a *sound scene* from one point in space. HOA signals can be easily transmitted and decoded to a variety of loudspeaker configurations or rendered binaurally to headphones. The SH formulation used to *encode* HOA signals also provides a basis for *decoding* those signals to a listener's head-related transfer functions (HRTFs), allowing for a range of optimizations of improved perceptual accuracy and listener individualization [4, 5, 6, 7]. Furthermore, convenient transformations of HOA signals allow the sound field to be rotated in conjunction with a listener's motion-tracked head rotations. Having three degrees of (rotational) freedom (3DoF) provides increased realism over an otherwise immobile sound scene [8].

Despite the clear utility of HOA, rendering recorded sound fields presents unique challenges in practice. The non-coincident arrangement of the SMA microphone capsules can lead to spectral coloration and frequency-dependent spatial ambiguities [9, §3],[10, 11, 4]. Because the number of capsules is necessarily limited, their signals must be encoded with SH functions of a finite order, which in turn determines an upper limit to the spatial resolution [10, 4]. Furthermore, the listener's auditory perspective is co-located with the SMA and cannot accurately be shifted far beyond the recorded position in space. This is because the accuracy of the sound field expansion degrades only a short distance from the expansion origin, as a function of the wavelength of the reconstructed sound field [3, §2.4][12, §5.1.5] [13] which has detrimental effects for auralization [14, 15]. Accurate perspective translation is further confined to the region between the origin of a single-perspective recording and the nearest sound source [16, 15]. Such limited mobility presents a problem for applications requiring *sound field navigation*, i.e. six degrees of (rotational-plus-translational) freedom (6DoF).

Numerous authors have sought to address the limitations of single-SMA sound

field rendering by *parametric decomposition* of the recorded sound field [17, 18, 19, 20]. In this approach, salient properties of the recorded sound field are analyzed and parameterized to enable selective filtering and reconstruction of the sound scene according to psychoacoustic or other application-specific criteria. Energetic sound field analysis [21] can be efficiently performed on ambisonic signals and offers two useful acoustic quantities of interest for this work. First, the *active sound intensity* indicates a direction of net acoustic energy transport, from which the direction-of-arrival (DoA) of a prominent sound source may be inferred [22]. The second, a *diffuseness* estimate, is a measure describing the probability that sound may arrive from any direction with equal probability [23, 24]. These sound field indicators offer an intuitive description of the sound field as essentially comprised of a directional and a diffuse sound component. As such, they have seen wide use, notably in spatial audio coding [25], spatial impulse response analysis and rendering [26], binaural rendering [27], and spatial upmixing for loudspeaker reproduction [28]. Such methods have been adapted to take advantage of the high-order spatial resolution of HOA signals [29, 30, 31], drawing on the same underlying distinction between direct and diffuse sound field elements. These methods identify multiple directional sound components, and in some in some cases also non-uniform ambient components [32]. Parametric approaches have also improved techniques for sound field navigation, to extend range-limited auditory translation from single-perspective recordings, so-called 3DoF+ systems [33, 34, 35, 36, 37]. Despite great progress in enhanced perceptual rendering and increased mobility within virtual sound fields by parametric methods, navigability of a sound field captured from a single SMA remains inherently confined to the location of recording.

One approach to extending the navigable range of recorded sound fields is the use of multiple SMAs distributed throughout space, i.e. *multiperspective* recordings. The challenge is then to use the plurality of SMA signals to efficiently and plausibly recreate the acoustic scene from a variable listener perspective. One category of solutions involves mixing the ambisonic signals (or spatial impulse responses, in the case of synthetic sound scenes) from multiple SMAs, weighting each ambisonic signal according to interpolation schemes determined by listener and source positions [38, 13, 39, 40]. Such schemes can suffer from erroneously duplicate or spread source images or other spectral coloration on account of differing angles- and times-of-arrival to the SMAs [41]. Some approaches also lack robustness on account of requiring advance knowledge of source locations before rendering. Again, parametric approaches find utility in resolving some of these challenges. For example, the diffuseness metric calculated among the SMAs can inform source localization and interpolative mixing weights of ambisonic signals [42]. In another example, multiperspective energetic DoA estimates are used to improve blind source separation, by which source signals are isolated into "object" streams for directional re-encoding from a new auditory perspective [43].

The success of parametric sound field reconstruction still hinges on sound field parameter estimation, and can easily fail if the assumption of temporal and spectral disjointedness of sound sources is not satisfied. This is especially tenuous when using classical energetic analysis, which offers a single DoA estimate (via the active intensity)

and "global" measure of presumably isotropic diffuseness. A hybrid approach [44] has employed intensity-based DoA estimation to construct a spatial power map by a steered-response power (SRP) beamformer directed around a region local to the initial DoA estimate. Peak-finding is then performed across the power map to refine the estimate, followed by iterative source removal and search for subsequent sources. The results show improvements over either SRP or intensity analysis methods alone and points to the potential of locally-constrained sound field analysis to resolve some of the stringent spatio-temporal source sparsity requirements. Indeed, a method has been developed that uses the compact spatial representation of HOA signals to provide spatially-localized estimates of sound field parameters: *sector-based parametric sound field decomposition* [45, 46, 31]. This technique seeks to minimize the adverse effects of concurrent sound sources and interfering noise by exploiting the full available spatial resolution and beamforming capabilities of HOA signals, while also offering a non-isotropic estimate of the diffuse component of a sound field. Sector-based processing has already shown improvements to parametric sound field analysis and synthesis methods based on single-perspective sound field recordings [30, 29, 32]. However, its performance across multiple SMA signals for sound field analysis in 3D space has not been explored.

1.2 Aim and scope of the thesis

The improvements that sector-based processing offers to single-SMA signal analysis and rendering, together with the compact sound field description from energetic sound field decomposition, shows promise in the domain of multiple-perspective sound field analysis. Therefore, it is the aim of this thesis to evaluate sector-based energetic sound field analysis in the context of 3D source localization across multiple concurrent SH signals. In particular, this work will consider the effect of the SH analysis order, the number and configuration of distributed SH receivers, and the impact of varying acoustic conditions of noise, reverberation, and multiple competing sound sources. The acoustic environment is modeled through simulation of a "shoe-box room" of simple geometry in MATLAB. Each condition—source and receiver count, source power, reverberation time, and isotropic noise power—is parameterized individually for evaluating the system performance. While the preceding Background and Motivation Section places this work in the context—and potentially in service of—sound field navigation, the focus of this work is on the preliminary stage of sound field decomposition and analysis only, which may in turn serve as part of a complete framework for sound field navigation.

The structure of this thesis is as follows: Section 2 will introduce the theory and fundamentals of the sound field model, its representation in spherical harmonic signals, and the analysis of its energetic properties. It goes on to detail spatially localized sound field analysis by sectors. Section 3 describes the implementation of the acoustic simulation, the signal encoding scheme, and the process of analysis for DoA estimation and 3D source localization. The performance of sector-based analysis is evaluated in Section 4. The evaluation begins by isolating influential factors in DoA estimation performance, including both simulation results and an

analysis of analytical formulations of spatial filtering, which carries forward into 3D localization performance. Based on the results, four estimation refinement techniques are proposed in Section 5, including a multisector approach toward multisource localization. Section 6 concludes the thesis and provides an outlook onto future directions.

2 Theory and fundamentals

2.1 Sound energy

Sound energy arises from the movement of an object that acts on the surrounding fluid medium. The force from that object causes displacements of the elements of the medium in the form of disturbances—variations in pressure—that propagate through the medium. The force acting on a particle over the course of its displacement constitute work done through that action, and is considered a *conservative* force because the work is recoverable by the reversal of the work path. Energy lost to friction dissipates as heat and constitutes work that cannot be recovered through a reversal of the process, and is therefore *non-conservative* [47, §4.1].

Disturbances in the medium that result from acoustic energy create pressure fluctuations p, which oscillate around the ambient equilibrium pressure P_0, resulting in a total sound pressure $P = P_0 + p$. The acoustic pressure p is small compared to P_0. To form an intuition about the scale of these fluctuations, consider that the reference pressure for expressing the sound pressure level (SPL)—the faintest audible sound to a person of normal hearing—is $2 \cdot 10^{-5}$ Pa at 1000 Hz. By comparison, the mean (equilibrium) atmospheric pressure is approximately 10^5 Pa [48, §A6.4].

Analogously, the variation of the density of air on account of acoustic energy will be very small compared to the equilibrium density, $(\rho - \rho_0)/\rho_0 \ll 1$. The relationship between pressure and density determines the speed at which sound can propagate; the ratio of the pressure fluctuations to the density fluctuations is equal to the square of the speed of sound [47, Eq. 3.5], expressed as

$$\frac{p}{\rho - \rho_0} = c^2. \tag{1}$$

Sound energy can be further understood by considering the volumetric strain that an elemental mass of fluid (a fluid "element" or "parcel") undergoes when subjected to internal or external force. The resultant change in volume $\mathrm{d}V$ is associated with a change in the *potential energy* E_{pot}, as the internal pressure will 'store' the negative work to be carried out by the restoring force as the work path is reversed. This relationship is expressed as [47, Eq. 4.4]

$$\mathrm{d}E_{\mathrm{pot}} = -p\frac{\mathrm{d}V}{V}. \tag{2}$$

Note that acoustic pressure p is used here instead of the overall sound pressure P because, despite the small proportion of acoustic pressure to overall equilibrium pressure, the action of P_0 contributes little to energy *transport* (rather, it is associated with the relatively macroscopic convection of acoustic energy by the fluid velocity) [47, §4.2]. A change in volume will be inversely proportional to a change in density according to [47, Eq. 3.3]

$$\frac{\mathrm{d}V}{V} = -\frac{\mathrm{d}\rho}{\rho}. \tag{3}$$

So by Eq. 1, which relates the pressure-density fluctuation ratio to the speed of sound, $\mathrm{d}\rho = \mathrm{d}p/c^2$. Further considering the small disturbance approximation of density fluctuation, Eq. 2 can be reformulated and integrated to give the acoustic potential energy as [47, Eqs. 4.5, 4.6]

$$\begin{aligned} \mathrm{d}E_{\mathrm{pot}} = -p\frac{\mathrm{d}V}{V} &= -p\left(-\frac{\mathrm{d}\rho}{\rho}\right) \\ &= p\frac{\mathrm{d}p}{\rho_0 c^2} \end{aligned} \tag{4}$$

$$E_{\mathrm{pot}} = \frac{p^2}{2\rho_0 c^2}. \tag{5}$$

The work done through the displacement of particles in the medium will correspond to a change in the *kinetic energy* of those particles, which, per unit mass, is $\frac{1}{2}u^2$, for a particle velocity u. The kinetic energy per unit volume is therefore

$$E_{\mathrm{kin}} = \frac{1}{2}u^2\rho_0. \tag{6}$$

Taken together, the kinetic and potential energy associated with acoustic disturbance of the medium form the *sound energy density*, which is

$$\begin{aligned} E = \ & E_{\mathrm{pot}} + E_{\mathrm{kin}} \\ &= \frac{p^2}{2\rho_0 c^2} + \frac{1}{2}u^2\rho_0. \end{aligned} \tag{7}$$

For small acoustic disturbances within the linear range of the medium, this total energy will remain constant in accordance with the conservative internal forces involved, unless energy is injected into or extracted from the system by an external sound source.

2.2 Sound intensity and directional sound propagation

The previous subsection introduced sound energy as it relates to sound pressure and particle velocity, but without explicit indication of the directional nature of sound propagation. For this, the sound particle velocity vector $\mathbf{u}$ is introduced, composed of the Cartesian components $[u_x, u_y, u_z]$. In order to estimate the sound energy, it is clear from Eq. 7 that the velocity will be needed. This can be measured directly with a velocity probe, although the more common approach would be to use a microphone array to capture the pressure gradient across space. The velocity can then be calculated from the pressure gradient by the *linearized fluid momentum equation*, which is

$$-\nabla p = \rho_0 \frac{\partial \mathbf{u}}{\partial t}, \tag{8}$$

where ∇ is a vector operator denoting a spatial gradient. The partial derivative term indicates the directional acceleration of particles in the medium (for a derivation of

this equation from Newton's dynamic law, see [23, §2.4.3]). From this it can be seen that, given a spatial sampling of the pressure gradient of the sound field, the particle velocity $\mathbf{u}$ can be attained. This is important in the measurement of directional energy transport, known as the *instantaneous sound intensity*, defined as

$$\mathbf{I}(t) = p\mathbf{u}. \tag{9}$$

In the ideal case of a single monochromatic plane wave, all directional components of $\mathbf{u}(t)$ oscillate phase-synchronously with $p(t)$, wherein zero crossings and peak magnitudes align, with either the same or opposite sign of the pressure depending on the direction of propagation. As the product of the pressure and velocity components, the instantaneous intensity will have a constant sign, indicating that all energy is transported in the direction of $\mathbf{u}(t)$. Put another way, the pressure can be viewed as modulating the vector $\mathbf{u}$, which points in a constant direction in this ideal case. If, however, there is a phase mismatch between the pressure and velocity components, the direction of the intensity vector will fluctuate [23, §2.4.4], indicating a variety of possible complex propagation scenarios. Such scenarios may be a superposition of multiple plane waves propagating in different directions which vary independently in time or differ in frequency, or are otherwise non-planar waves such as spherical waves occurring in the near field of a sound source.

It is common in the literature to inspect properties of acoustic intensity by considering a stationary monochromatic sound field. From this perspective, the pressure and velocity components can be represented as complex phasors oscillating over time [47, 49, 21]. Useful properties of the directional energy propagation are revealed by calculating the instantaneous intensity separately for those components of the complex velocity which are in phase with the sound field pressure and in quadrature with the pressure. Intensity calculated by the *in-phase* (cosinusoidal) components of $\mathbf{u}(t)$ is termed the *active intensity*, which can be represented as [23, Eqs. 2.31-34]

$$\mathbf{I}_a(t) = -\frac{A^2 \cos^2(\omega t + \phi)}{\omega \rho_0} \nabla \phi \tag{10}$$

where A and ϕ denote the space-dependent amplitude and phase, respectively. It can be seen that active intensity will be proportional to the spatial gradient of the phase $\nabla\phi$, which advances in the direction of propagation. Put another way, the phase gradient vanishes in the directions of constant phase which are perpendicular to the direction of propagation. Therefore, $\mathbf{I}_a(t)$ will lengthen with the advance of phase and point in the (negative) direction of propagation. Intensity calculated by the *quadrature* (sinusoidal, 90° out-of-phase from the pressure) components of $\mathbf{u}(t)$ is termed the *reactive intensity*, and is formulated as [23, Eq.2.34]

$$\mathbf{I}_r(t) = -\frac{\sin 2(\omega t + \phi)}{4\omega \rho_0} \nabla A^2 \tag{11}$$

The reactive intensity, in contrast to its active counterpart, is proportional to the spatial gradient of the square of the pressure amplitude. As such, a reactive

component will be observed, for example, in the near field of sound sources where the amplitude gradient is large, but will decrease with distance from the source as the spherical wave front becomes planar. The presence of the sinusoidal oscillation of the reactive intensity, apparent in Eq. 11, can indicate various acoustic conditions such as standing waves or other multiwave interference such as reverberation [50]. It is also clear that this zero-mean oscillation of reactive intensity will vanish when averaged over time, indicating that it does not contribute to the directional transport of energy in a monochromatic or otherwise periodic sound field [47]. The active intensity can therefore be calculated most simply by the time average of the intensity as [50, 21, 23]

$$\mathbf{I}_a = \langle \mathbf{I}(t) \rangle \,, \tag{12}$$

where $\langle \cdot \rangle$ represents the time average. This is applicable to general sound fields and is an efficient means of determining the direction of net flow of energy at a particular location in the field. It should be noted that when intensimetric analysis is performed in the frequency domain, there will necessarily be a time average taking place on account of the time windowing of the Fourier transform, or on account of the filtering operations used in other filterbank approaches to signal decomposition into frequency bands [23].

2.2.1 Intensimetric analysis for general sound fields

The instantaneous measures of intensity in Eqs. 10-12 apply to stationary, monochromatic sound fields. For general sound fields, Schiffer and Stanzial [21] define the velocity vector as $\mathbf{u} = \mathbf{u}_p + \mathbf{u}_q$, where the subscripts denote the in-phase and quadrature components of velocity, respectively, with reference to the phase of the pressure. They are defined as [21, Eq. 18]

$$\mathbf{u}_p = \frac{p \langle p\mathbf{u} \rangle}{\langle p^2 \rangle} \tag{13}$$

$$\mathbf{u}_q = \mathbf{u} - \mathbf{u}_p = \frac{\langle p^2 \rangle \mathbf{u} - p \langle p\mathbf{u} \rangle}{\langle p^2 \rangle}. \tag{14}$$

The instantaneous intensity is then $\mathbf{I}'(t) = p(\mathbf{u}_p + \mathbf{u}_q)$. Note that the time-average window is expected to be set such that the soundfield effectively stationary [51, §II], similar to the assumption of a steady-state signal processed within the frequency transform window in frequency-domain analysis. Transient sounds, for example, may result in deviations from the expected behavior. Decomposing further, the instantaneous active and reactive intensity for general sound fields is then [21, §V]

$$\mathbf{I}'_a(t) = p\mathbf{u}_p = \frac{p^2 \langle p\mathbf{u} \rangle}{\langle p^2 \rangle} \tag{15}$$

$$\mathbf{I}'_r(t) = p\mathbf{u}_q = \frac{\langle p^2 \rangle p\mathbf{u} - p^2 \langle p\mathbf{u} \rangle}{\langle p^2 \rangle}, \tag{16}$$

which have the time-averaged properties [51, Eq. 5]

$$\mathbf{I}'_a = \langle p\mathbf{u} \rangle \quad \text{and} \quad \mathbf{I}'_r = 0. \tag{17}$$

Stanzial and Prodi [52] later designated the active and reactive intensity of Eqs. 15 and 16 as *radiating* and *oscillating* intensity, respectively. These terms are useful in that they are descriptive of the behavior of the acoustic energy they quantify. They also distinguish the above technique from prior methods of intensimetric analysis which regarded the real and imaginary parts of the complex intensity as corresponding to the active and reactive intensity, respectively. Three of these complex analysis methods are outlined in [53], where the terminology is clarified. One of those methods (Heyser [54]) provides a method for calculating the instantaneous active and reactive intensity that is suitable for narrow-band, as opposed to only pure-tone, sound fields. The approach makes use of the Hilbert operator, which imposes a 90° phase shift on the corresponding signals, and is formulated as

$$\mathbf{I}''_a(t) = \frac{1}{2}[p(t)\mathbf{u}(t) + \hat{p}(t)\hat{\mathbf{u}}(t)] \tag{18}$$

$$\mathbf{I}''_r(t) = \frac{1}{2}[\hat{p}(t)\mathbf{u}(t) - p(t)\hat{\mathbf{u}}(t)], \tag{19}$$

where the hat notation here indicates the Hilbert operator. The corresponding time-averaged quantity of the active intensity, $\mathbf{I}''_a = \langle p\mathbf{u} \rangle$, coincides with the results of Eqs. 15 and 17. However the time-averaged reactive intensity, $\mathbf{I}''_r = \langle p\hat{\mathbf{u}} \rangle$, for a general sound field will not vanish an in turn reveal further acoustic quantities of such as the specific acoustic wave impedance and admittance [54, Eqs. 34-35], which will not be considered further here. Heyser's approach is compared to the performance of Eqs. 15 and 16 in [52]. For the present work, directional sound propagation is of primary interest, so the reactive intensity is of limited use. The active intensity will therefore be used exclusively to infer the direction-of-arrival of sound sources, presumed to arrive from the far field and which are in fact simulated by a plane wave sound field model, as will be discussed next.

2.3 The plane wave sound field model

The sound pressure due to a plane wave arriving from the far field will vary according to the frequency, the direction of arrival, and the observation point $\mathbf{r}$ following the exponential function

$$p(\mathbf{r}, \boldsymbol{\gamma}, \omega) = a(\boldsymbol{\gamma}, k)e^{ik\boldsymbol{\gamma}\cdot\mathbf{r}}, \tag{20}$$

where the wavenumber is $k = \omega/c$ at angular frequency $\omega = 2\pi f$, and $i = \sqrt{-1}$ is the imaginary unit. The Cartesian unit vector $\boldsymbol{\gamma}$ points in the direction-of-arrival of the plane wave (opposite the direction of its propagation), with components

$$\boldsymbol{\gamma}(\Omega) = \begin{bmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{bmatrix}. \tag{21}$$

where the direction $\Omega = (\theta, \varphi)$ is comprised of inclination $\theta \in [0, \pi]$ (from the zenith downward) and azimuth $\varphi \in [-\pi, \pi]$ (forward is 0 radians, increasing counterclockwise). Note that each of the three components of $\boldsymbol{\gamma}$ form the directional response of three dipoles oriented along the x, y, and z axes, respectively, as will be used shortly.

The analysis of sound field properties that will be carried out in this work will be from a single point, so it will be more concise to consider the pressure at the coordinate origin by omitting $\mathbf{r}$ which reduces the expression of Eq. 20 to a frequency-dependent complex amplitude density function integrated over the unit sphere. The pressure at the origin resulting from an the amplitude distribution of plane waves is then [45, Eq. 9]

$$p(f) = \int_{\Omega} a(f, \Omega)\, \mathrm{d}\Omega \tag{22}$$

where $\int_{\Omega} \mathrm{d}\Omega = \int_{-\pi}^{\pi} \mathrm{d}\varphi \int_{0}^{\pi} \sin\theta\, \mathrm{d}\theta$ and describes integration over the unit sphere, $\Omega \in S^2$. The dependence on frequency f implies a frequency-domain analysis which decomposes the signal into monochromatic components. The sound field pressure can be captured without any directional sensitivity by an omnidirectional microphone. The same sound field can captured by three orthogonal dipoles oriented along the x, y, and z axes, which will have a directional sensitivity according to the pattern

$$\mathbf{d}(\Omega) = \begin{bmatrix} d_x(\Omega) \\ d_y(\Omega) \\ d_z(\Omega) \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{bmatrix} \tag{23}$$

which has a clear resemblance to Eq. 21. Signals recorded by such a dipole configuration form a signal vector $\mathbf{v}(f) = [v_x, v_y, v_z]^\top$ which correspond to the negative unnormalized Cartesian components of the particle velocity [45, §III A]. These dipole signals relate directly to the acoustic velocity, which, considering a sound field of the same plane wave amplitude distribution as in Eq. 22, is defined as [45, Eq. 10]

$$\mathbf{u}(f) = -\frac{1}{Z_0} \int_{\Omega} \mathbf{d}(\Omega) a(f, \Omega)\, \mathrm{d}\Omega \tag{24}$$

$$= -\frac{1}{Z_0} \mathbf{v}(f), \tag{25}$$

where $Z_0 = c\rho_0$ is the characteristic impedance of air.

Given the above formulations of sound field pressure and velocity, the sound field intensity can be recovered by Eq. 9. As discussed in section 2.2.1, the *active* intensity is of particular interest in this work for its indication of the direction of net energy flow. This can be retrieved for a broadband signal simply by the time-averaging process of Eq. 12. Alternatively, it can be computed in the frequency domain for a frequency-dependent view of energy propagation. So, following a time-frequency signal decomposition—by the Fourier transform or other suitable filter bank approach generating complex time-frequency signals (see [55])—the active intensity can be calculated from the pressure and velocity components (or corresponding dipole

velocity signals of Eq. 25), as

$$\mathbf{I}_a(f) = \frac{1}{2}\,\mathrm{Re}\{p^*(f)\mathbf{u}(f)\} \tag{26}$$

$$= -\frac{1}{2Z_0}\,\mathrm{Re}\{p^*(f)\mathbf{v}(f)\}., \tag{27}$$

where $(\cdot)^*$ denotes the complex conjugate. Note that the condition of a stationary monochromatic sound field is assumed to be satisfied within the duration of time-frequency processing window, so the discussion of Eq. 10 pertains.

The total energy density of the sound field, reflecting the potential and kinetic energy relationship in Eq. 7, is [46, Eq. 12]

$$E(f) = \frac{1}{4Z_0c^2}|p(f)|^2 + \frac{\rho_0}{4}\|\mathbf{u}(f)\|^2 \tag{28}$$

$$= \frac{1}{4Z_0c^2}[|p(f)|^2 + \|\mathbf{v}(f)\|^2]. \tag{29}$$

2.4 Sound field diffuseness

In addition to the primary energetic quantities already introduced, an important derived quantity is the sound field *diffuseness*. As previously discussed, $\mathbf{I}_a$ can be used as an indicator of the direction-of-arrival of a sound source in a free field, describing the direction as well as the magnitude of the *net* energy transport. However, because the oscillatory energy flow captured by the reactive intensity does not propagate, it will not appear in $\mathbf{I}_a = \langle\mathbf{I}\rangle$. And so it is that the ratio between the active intensity and energy density of the sound field, $\varsigma = \|\langle\mathbf{I}_a\rangle\|/\langle E\rangle$, indicates the *speed* of the net propagating energy which is bound by the speed of sound, $\varsigma \in [0, c]$ [21, §VI]. Normalizing this quantity by c then offers a useful scalar metric describing the ratio of transported energy to overall energy at the measurement location. This can be presented as an estimate of the sound field diffuseness by

$$\psi = 1 - \frac{\|\langle\mathbf{I}_a\rangle\|}{c\,\langle E\rangle}. \tag{30}$$

On account of the normalization by c, $\psi \in [0, 1]$. The diffuseness can then be thought of as accounting for energy that is not transported, and indeed Merimaa [23, §2.4.5] defines the diffuseness as the proportion of locally confined energy. When $\psi = 0$ there is no locally oscillating energy, which is interpreted here as exclusively directional energy transport perpendicular to a surface of constant phase, i.e. a plane wave [53] (refer to the deiscussion of Eq. 10). However, a number of acoustic scenarios comprised of just a few plane waves may produce the condition $\psi = 1$. For example, a purely reactive monochromatic field of an ideal standing wave would result in $\langle\|\mathbf{I}_a\|\rangle = 0$. Another condition producing an equivalent result, though largely theoretical, would be two monochromatic plane waves of identical frequency arriving from exactly opposite directions. Diffuseness may also be high if the direct sound of a steady-state sound source interacts strongly with low order room reflections

of appreciable magnitude [50]. Such situations are rare in reality and more useful interpretation of a fully diffuse sound field is that it is equally probable that sound arrives from any direction over a given time period [56, 24].

Diffuse sound fields can be defined from two perspectives: 1) as uniform energy density throughout all points in the sound field, and 2) a uniform probability of sound energy flowing in all directions simultaneously [57],[56, §4.3.2]. The former definition implies that instantaneous diffuseness could be assessed given enough spatially disparate samples to observe the statistical randomness of the energy density throughout the sound field. The latter definition suggests that diffuseness could be measured from fewer local points but requiring more temporal samples to fully realize the property of temporal ergodicity. Such a configuration can be realized with a spherical microphone array (SMA) which can measure energy gradients across its origin, used to indicate directional energy flow and in turn the degree of isotropy at that point in space [58]. Such an approach is used in this work, for which sound fields are modelled by encoding directional plane waves into the SH domain directly, i.e. assuming ideal, coincident SMAs capable of encoding sound perfectly via the spherical harmonic transform (introduced in Section 2.5).

2.4.1 Diffuseness estimators

There are numerous formulations for diffuseness estimation which follow from intensimetric sound field analysis. The first, which follows most directly from the definition of diffuseness in Eq. 30 is the intensity-energy density (IE) ratio estimator [26]

$$\psi_{\mathrm{IE}} = 1 - \frac{\|\mathbb{E}\{\mathbf{I}_a\}\|}{c \cdot \mathbb{E}\{E\}}, \tag{31}$$

Where $\mathbb{E}\{\cdot\}$ is the expectation operator. As discussed in Section 2.4, it is based on the ratio of the magnitude of the active intensity to the overall sound energy density. It derives from the relation that the magnitude of the active intensity of an ideal plane wave in the free field will be equal to the magnitude of the energy density multiplied by the speed of sound. Plane waves incident from many directions will drive the magnitude of the average intensity vector down, thus increasing the diffuseness estimate, as will an excess of oscillating energy, which would be present in the total energy density but will not be captured by the active intensity.

Another estimator which is not dependent on the total energy density but rather the variability of the active intensity vector over time is the temporal variation (TV) diffuseness estimator [59] (also called the "coefficient of variation" estimator in [24]):

$$\psi_{\mathrm{TV}} = \sqrt{1 - \frac{\|\mathbb{E}\{\mathbf{I}_a\}\|}{\mathbb{E}\{\|\mathbf{I}_a\|\}}}. \tag{32}$$

This variant was developed based on the observation that in a diffuse field, instantaneous active intensity vectors vary in their magnitude and direction, while for a sound field consisting of a single prominent plane wave, the intensity vectors are more tightly clustered around the single direction-of-arrival. The estimator is therefore

constructed from the ratio of the magnitude of the averaged intensity vector to the average of the magnitudes those vectors over the window of observations. By excluding the overall sound energy, the estimator will give more weight to prominent directional transient energy within the time window. If the average magnitude of $\mathbf{I}_a$ is high, the direction of each observation will need to be tightly clustered to achieve a low diffuseness. On the other hand, if the average magnitude of the vectors is low and spatially dispersed, highly directional near-instantaneous (transient) values will be emphasized, bringing the overall diffuseness down. The TV estimator also proves useful in the one-dimensional measurement of sound fields (e.g. recorded with an omnidirectional and a single figure-of-eight microphone), and in which a sound source source moves off of that axis [59]. The IE and TV estimators will converge in the case of sound field consisting of a single plane wave or a purely diffuse field, and perform nearly identically over long time averages with steady-state mixtures of the two conditions, assuming a 2D or 3D measure of the sound field, as shown in [24].

Del Galdo et al. present another estimator (AI) that also relies only on the active intensity vector [24]:

$$\psi_{\mathrm{AI}} = 1 - \frac{\|\mathbb{E}\{\mathbf{I}_a\}\|}{\sqrt{2 \cdot \mathbb{E}\{\|\mathbf{I}_a\|^2\} - \|\mathbb{E}\{\mathbf{I}_a\}\|^2}}. \tag{33}$$

The AI method estimates a reference diffuseness very well in both 2D and 3D sound fields and outperforms the IE method in the 1D case. The author of the method points out the drawback that unless there is a sufficiently high spatial sampling to approximate the expectation, the accuracy relies on the plane wave power remaining constant within the time average window. As a result, when using a nearly-coincident sensor arrangement, as with a SMA, the method is not best suited for estimating the diffuseness of a sound field that contains sources with quickly varying energy, like speech.

The spherical variance (SV) diffuseness estimator [60], like the temporal variance estimator, assumes DoA estimates to be uniformly dispersed in a diffuse field and collinear for an ideal plane wave sound field. It is constructed as

$$\psi_{\mathrm{SV}} = 1 - \|\mathbb{E}\{\boldsymbol{\gamma}'(\hat{\Omega}_{\mathrm{DoA}})\}\|, \tag{34}$$

where $\boldsymbol{\gamma}'(\hat{\Omega}_{\mathrm{DoA}})$ are the unit vectors representing DoA estimates. It is analogous to the TV method, in which the $\mathbf{I}_a$ vectors are essentially considered to be DoA estimates with a variable (non-unit-length) magnitude, but the SV estimator discards the magnitude and uses only the directional variance of unit DoA vectors. The method was devised to address erratic DoA estimates that result from sound field energy above the spatial aliasing limit of real microphone arrays. As such, the estimates are not calculated by Eq. 21, but rather by a novel DoA estimator calculated by taking the integral of the unit vector over the sphere, weighted by by the directional magnitude responses of the microphone patterns comprising the SMA [60, Eqs. 5-7]—a method that resembles spatial filtering of analysis results that will be further discussed in Section 2.6. Ultimately the SV method was found to be best suited in

a crossover scheme which combines the TV and SV methods below and above the spatial aliasing frequency, respectively [60].

It should be noted that the above diffuseness estimators are suitable for high-order analysis by use of the sector-based sound field decomposition approach investigated in this work. It is relevant to briefly mention two other diffuseness estimators which can operate on high-order signals directly by spherical harmonic signal covariance methods. The directional power variance method, presented for spherical microphone arrays in [58], uses beamforming to analyze the directivity index and directional dispersion across multiple directions, as well as the dependence of the spatial cross-correlation on direction. Politis adapted the method to operate on SH signals directly [61], using the ratio of the measured directional power to that expected from the maximum directivity beamformer aligned with a plane wave source to form a diffuseness estimate. While the estimator will perform well for uncorrelated sources, it overestimates the diffuseness in the case of multiple coherent sources (as might be found in highly reverberant rooms). The COMEDIE diffuseness estimator [62], on the other hand, will be robust to multiple correlated sources arriving from many directions, up to the number of SH signal components. As this work is focused on spatially-localized sound field quantity estimates (as opposed to global estimates) derived from intensimetric analysis, these signal covariance methods will not be considered further here.

2.5 Spherical harmonic sound field representation

As mentioned in the Introduction, Higher-Order Ambisonics (HOA) is a versatile sound field transmission format for its scalable spatial resolution, its complete description of the sound field from a single point in space, and for a range of mathematical operations useful for manipulating the sound image. Underlying the format is a spherical harmonic (SH) representation of the sound field which encodes various spatial properties therein. As such, SH signals can be analyzed to retrieve spatial information about a sound field which can in turn inform parametric modification and re-synthesis of the sound scene according to application-specific criteria. Such parametric modifications include psychoacoustic enhancement for binaural rendering [63, 31] and adaptive mixing of SH signals for decoding to multichannel loudspeaker layouts [64, 65].

2.5.1 Spherical harmonics

Spherical harmonics are defined in complex form as [1, §1.2]

$$Y_{lm}(\theta, \varphi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\varphi} \tag{35}$$

where $(\cdot)!$ represents the factorial function and P_l^m are associated Legendre functions of *degree* $l \in \mathbb{N}_0$ and mode *index* $m \in \mathbb{R}$. The SH *order* is used to describe the complete set of SH components, for a given truncation of the SH series up to order

L, for degree $l \in [0, L]$ and index $m \in [-l, l]$. The complex conjugate of a spherical harmonic function is

$$Y^*_{lm}(\theta, \varphi) = (-1)^m Y_{l(-m)}(\theta, \varphi). \tag{36}$$

Individual SH functions are visualized in the balloon plots of Figure 1. In general, the complexity of the modal patterns increases with the degree. The azimuthal behavior changes according to the exponential term, which for the real and imaginary parts are $\cos\varphi$ and $\sin\varphi$, respectively. This also leads to the relation $Y_{lm}(\theta, -\varphi) = Y^*_{lm}(\theta, \varphi)$ [1, Eq. 1.17]. The behavior of SH over the inclination are governed by the Legendre functions.

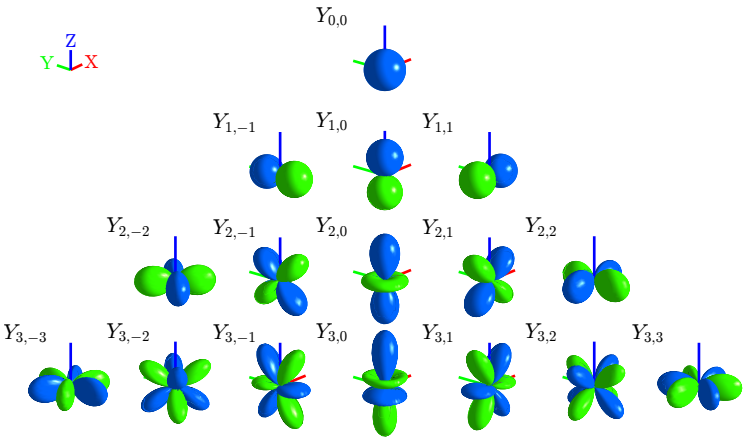


Figure 1: Spherical harmonic functions $Y_{lm}(\Omega)$ up to order $L = 3$. Rows correspond to *degrees* $l = [0, ..., 3]$ and columns correspond to modal *indices* $m = [-l, ..., l]$. The plots can be viewed as the real form of SH functions or, alternatively, as showing the magnitude of the complex functions, without phase, such that the center column holds the real functions $Y_{l,0}(\theta, \varphi)$, the columns left of center hold $\mathrm{Im}\{Y_{lm}(\theta, \varphi)\}$, and columns right of center hold $\mathrm{Re}\{Y_{lm}(\theta, \varphi)\}$ [1, §1.2]. Positive and negative values are blue and green, respectively.

The SHs can also be specified in real form (denoted here by $\hat{Y}$) as [30, Eq.6.3]

$$\hat{Y}_{lm}(\theta, \varphi) = \sqrt{(2l+1)\frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|} \sin\theta y_m(\varphi) \tag{37}$$

where

$$y_m(\varphi) = \begin{cases} \sqrt{2}\sin|m|\varphi & m < 0 \\ 1 & m = 0 \\ \sqrt{2}\cos m\varphi & m > 0 \end{cases}.$$

The spherical harmonic functions are orthogonal, as expressed by [1, Eq. 1.23]

$$\int_\Omega Y^*_{lm}(\Omega) Y_{l'm'}(\Omega)\, \mathrm{d}\Omega = \delta_{ll'}\delta_{mm'} \tag{38}$$

where $\mathrm{d}\Omega = \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\varphi$. For practical purposes SH orthogonality describes the exclusive interaction between SH components of identical degree and index under

various operations such as beamforming, in which modal weights are applied. Further discussion of the modal patterns and properties of SH functions are detailed in [1, §1.2], and various symmetries useful for sound field manipulations are discussed in [66].

2.5.2 The spherical harmonic transform

Any function defined on the sphere $f \in S^2$ can be represented by a weighted combination of spherical harmonics as [67, Eq.6.48]

$$f(\Omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} w_{lm} Y_{lm}(\Omega) \tag{39}$$

where w_{lm} are the direction-independent modal weights of each SH component. The weights fully describe the unique behavior of $f(\Omega)$ and can be derived by the Spherical Fourier Transform, also referred to as the Spherical Harmonic Transform (SHT), defined as [67, Eq.6.49]

$$w_{lm} = \int_{\Omega} f(\Omega) Y_{lm}^{*}(\Omega)\, \mathrm{d}\Omega. \tag{40}$$

Eq. 39 shows that the function value can be recovered for any direction Ω if the weights w_{lm} are known. The SHT may be truncated to a finite order L, representing a continuous spherical domain by a finite number of coefficients [1, §1.4], but with limited spatial resolution according to the truncation order. Therefore, a sound field represented by the continuous spatial distribution function $a(t, \Omega)$ will be transformed into a vector of SH signals which represent these coeffients modulating over time:

$$\mathbf{a}(t) = \mathcal{SHT}\{a(t, \Omega)\} = \int_{\Omega} a(t, \Omega) \mathbf{y}^{*}(\Omega)\, \mathrm{d}\Omega, \tag{41}$$

where $\mathbf{y}(\Omega)$ is a vector containing the spherical harmonic functions evaluated at Ω:

$$\mathbf{y}(\Omega) = [Y_{00}(\Omega), Y_{1(-1)}(\Omega), Y_{10}(\Omega), Y_{11}(\Omega), Y_{2(-2)}(\Omega), ..., Y_{LL}(\Omega)]. \tag{42}$$

The length of the vector $\mathbf{a}$ is $(L+1)^2$, showing that the truncation order determines the number of signals representing the sound field. Note that the conjugate need not be taken if using the real SHs. The time index t will be omitted for brevity. Eq. 41 could also be made a function of frequency, which will not be necessary here.

The double indices (l, m) can be substituted by a single index so that

$$Y_{lm} = Y_n = [\mathbf{y}]_n \qquad a_{lm} = a_n = [\mathbf{a}]_n \tag{43}$$

where $n \in [1, N]$ for $N = (L+1)^2$, and n is known in the ambisonics literature as the Ambisonic Channel Number (ACN). The indices can be exchanged by

$$n = l(l+1) + m, \qquad l = \lfloor \sqrt{n} \rfloor, \qquad m = l^2 - n. \tag{44}$$

The ACN indexing can be useful for compactness, for example when indexing SH signal vectors or more succinctly describing properties of the SH basis function set, such as the orthogonal property of Eq. 38: $\int_\Omega Y_n^*(\Omega) Y_{n'}(\Omega)\,\mathrm{d}\Omega = \delta_{n-n'}$.

The truncation order will determine a minimum number of sampling directions Q required to reconstruct the full spherical domain via the ISHT (Eq. 39). Provided uniform or nearly uniform sampling of the sphere, such as those of platonic solids or a t-design [68], that minimum sampling is $Q \geq N$. For more information on spherical sampling schemes, the reader is referred to [1, §3.4-5]. The discrete form of the ISHT is

$$s_L(\Omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} w_{lm} Y_{lm}(\Omega). \tag{45}$$

Evaluating the order-limited function $s_L(\Omega)$ for Q different directions can be formulated in the compact matrix form as

$$\mathbf{s} = \mathbf{Y}\mathbf{w} \tag{46}$$

where

$$\mathbf{s} = [s(\Omega_1), s(\Omega_2), ..., s(\Omega_Q)]^\top, \tag{47}$$

$$\mathbf{w} = [w_1, w_2, ..., w_N]^\top, \tag{48}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1(\Omega_1) & Y_2(\Omega_1) & ... & Y_N(\Omega_1) \\ Y_1(\Omega_2) & Y_2(\Omega_2) & ... & Y_N(\Omega_2) \\ \vdots & \vdots & ... & \vdots \\ Y_1(\Omega_Q) & Y_2(\Omega_Q) & ... & Y_N(\Omega_Q) \end{bmatrix}. \tag{49}$$

Note the use of ACN indexing. In the special case of (nearly) uniform sampling, where $Q = N$, and $\mathbf{Y}$ is invertible, the corresponding discrete SHT is then

$$\mathbf{w} = \mathbf{Y}^{-1}\mathbf{s}. \tag{50}$$

In the case of general sampling schemes, including the over-sampled case of $Q > N$, the coefficients $\mathbf{w}$ are recovered by

$$\mathbf{w} = \mathbf{Y}^\dagger \mathbf{s}. \tag{51}$$

where the Moore-Penrose pseudo-inverse is $\mathbf{Y}^\dagger = (\mathbf{Y}^\mathsf{H}\mathbf{Y})^{-1}\mathbf{Y}^\mathsf{H}$. Further discussion of non-uniform and under-sampling of spherical functions is outside the scope of this thesis, but can be found in [1, §3.7].

2.5.3 Spherical harmonic beamforming and spatial weighting

Given a spherical harmonic signal set representing a sound field, spatial filters can be applied that impose a directional emphasis on the sound field, or similarly, a single signal can be extracted by a *beamformer* that functions as a "virtual microphone" with a directivity pattern defined by a spatial weighting function $w(\Omega)$. The SHT of

that directivity pattern produces beamforming weights $\mathbf{w}$ which can be used in a weight-and-sum operation on the SH components to give the beamformer output by

$$s(\Omega_0) = \mathbf{w}(\Omega_0)^\top \mathbf{a} \tag{52}$$

where Ω_0 denotes the "look" direction of the beamformer. If the beamformer is axisymmetric, the beam design weights can be separated into two independent terms: *pattern-forming* weights $[b_0, ..., b_L]$ that define the directional pattern $b(\theta)$ which depend only on the SH degree l, and a direction-dependent *steering* term comprised of the SH components evaluated at Ω_0 [30, §6.3.2]:

$$[\mathbf{w}(\Omega_0)]_n = w_{lm}(\Omega_0) = b_l Y_{lm}(\Omega_0). \tag{53}$$

Eq. 53 shows that the degree weights b_l can be designed to form an axisymmetric directivity pattern that suits the application. Common beam patterns are [30]

$$b_l = \frac{1}{(L+1)^2} \qquad \text{hypercardioid,} \tag{54}$$

$$b_l = \frac{L!L!}{(L+l+1)!(L-l)!} \qquad \text{cardioid,} \tag{55}$$

$$b_l = \frac{P_n(\kappa_L)}{\sum_{l=0}^{L}(2l+1)P_n(\kappa_L)} \qquad \text{max-}\mathbf{r}_\text{E}, \tag{56}$$

where $\kappa_L = \cos(2.407/L+1.51)$, as derived in [69] as a suitable approximation of the weight design by Daniel [70].

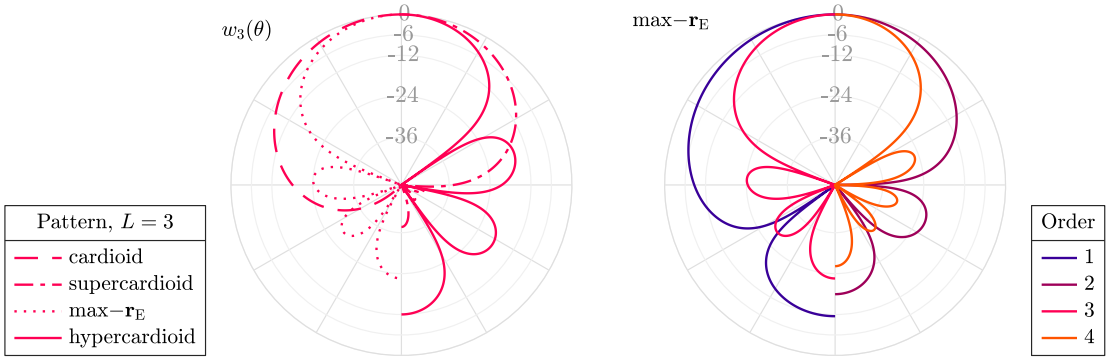


Figure 2: Left: common axisymmetric beam patterns $w_L(\theta)$ of order $L = 3$; Right: the max-$\mathbf{r}_\text{E}$ pattern of orders $L = [1, 2, 3, 4]$.

The beam patterns vary in their ability to maximize on-axis energy or reject off-axis energy. A beam's *directivity factor* is a scalar measure of the ratio between the peak and average magnitude of the squared beam pattern [1, §5.3]:

$$DF = \frac{|w(\Omega_0)|^2}{\frac{1}{4\pi}\int_\Omega |w(\Omega)|^2 \, \mathrm{d}\Omega}. \tag{57}$$

A useful conversion to a log-scale gives the *directivity index*, $DI = 10\log_{10}(DF)$, which is a descriptor of the ability of a beam pattern to provide an increased signal-to-noise ratio (SNR) in its steering direction as compared to an omnidirectional response. Four axisymmetric beam patterns are shown in Figure 2 and their key properties will be outlined next.

The *hypercardioid* pattern (Eq. 54) has been shown to maximize DF by maintaining a unit response on-axis and minimizing the average directional response, and DF increases with increasing order according to $DF_{\max} = (L+1)^2$ [1, Eq. 6.11]. The narrowing of the main lobe width can be approximated by $2\Theta_0 \approx 2\pi/L$ for orders $L \geq 4$ [71, Eq. 22], where Θ_0 is the half-width angle from the beam maximum to the first null. This relates directly to the separability of two incident plane waves in a sound field, down to a minimum angle known as the Rayleigh resolution [1, Eq. 6.14]

$$\Theta_{\text{Rayleigh}} \approx \frac{\pi}{L}. \tag{58}$$

This metric makes explicit the improved ability to discriminate between adjacent plane wave DoAs as the beamforming order increases. Additionally, the hypercardioid pattern is the simplest beam pattern by construction: the beamforming weights are the spherical harmonic coefficients themselves with only a single normalization factor determined by the truncation order. This construction also shows the relationship to a plane wave decomposition of the sound field (Eq. 45), and to a plane wave encoder (Eq. 51), leading to its alternative names as the "plane wave decomposition" [71] or "regular" [72] beamformer.

The *cardioid* pattern has the useful characteristics of a rear-facing null—which minimizes energy in the direction opposite of the steering direction—and an amplitude response with a constant positive sign, although at the expense of a wider main lobe. The *supercardioid* pattern maximizes the ratio of the frontal hemisphere (in the steering direction) to the rear hemisphere [1, §6.5]. Its analytical expression is not presented here but the beamforming coefficients are tabulated up to fourth order in [61].

The *maximum energy* beamformer maximizes the energy concentration along the steering direction, thereby diminishing side lobes compared to the hypercardioid beamformer, at the expense of a slightly wider main lobe. It is also known as the "max-$\mathbf{r}_\text{E}$" beamformer for its ability to maximize the length of the $\mathbf{r}_\text{E}$ vector, which can be thought of as the beam's energetic "barycenter" (a useful depiction is found in [69, Fig. 5]). The max-$\mathbf{r}_\text{E}$ vector is an energetic measure introduced by Gerzon [73] which has seen wide practical use as a psychoacoustic localization metric for the design and evaluation of ambisonic decoders [70, 74, 69] and HOA microphones [75].

Some of the beamformer patterns mentioned above are compared in [76, §III.A], where the authors assess the effectiveness of the patterns to distinguish among multiple spatially distributed sound sources in various SNR conditions. They do so by measuring the output of the beamformer as it is scanned through an exhaustive spatial grid, forming a power map across which energy peak-finding can be performed to form DoA estimates.

2.6 High-order sector-based sound field analysis

Given the clear advantage of using increased spatial resolution to discriminate among different regions of the sound field (see Eq. 58), it is desirable to perform energetic parameter estimation both with a higher degree of spatial resolution and within constrained angular regions of interest of the sound field.

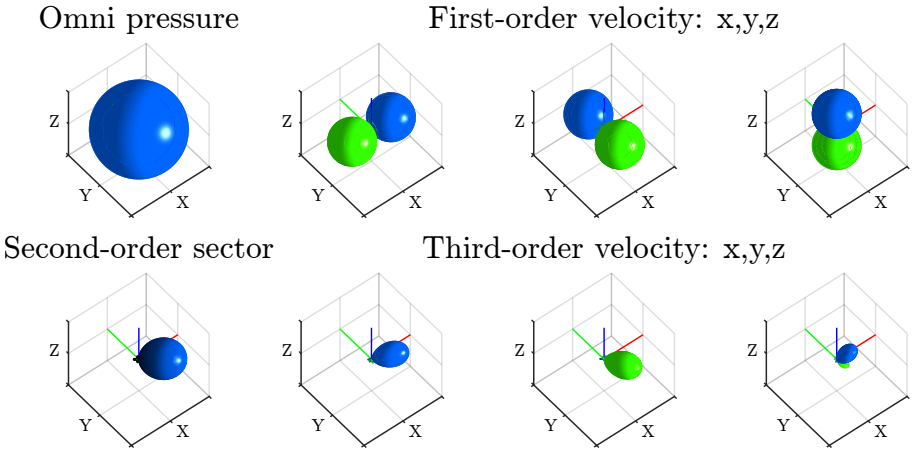


Figure 3: Top row: pressure (omni) and velocity (dipole) patterns used in traditional first-order energetic sound field analysis. Bottom row: pressure p_w and velocity $\mathbf{d}_\mathrm{w}(\Omega)$ beam patterns resulting from a second-order spatial filter $w(\Omega)$ (sector) imposed on the patterns of row one.

It follows from the beamforming relation of Eq. 52 that a pressure signal can be retrieved from a sound field amplitude distribution that has been spatially weighted, giving

$$p_\mathrm{w} = \int_\Omega w(\Omega) s(\Omega)\,\mathrm{d}\Omega = \mathbf{w}_\mathrm{p}^\top \mathbf{a}. \tag{59}$$

where the subscript $(\cdot)_w$ denotes a spatially-weighted quantity, and the subscript $(\cdot)_p$ denotes the SH weights assigned to form the weighted-pressure component of the sound field. The spatial filter will be called the *sector pattern*, as it filters the sound field amplitude distribution down to a region of interest defined by the beamforming pattern $w(\Omega)$. The same sector pattern can be imposed on the sound field velocity components by

$$\mathbf{u}_\mathrm{w} = -\frac{1}{Z_0}\int_\Omega \mathbf{d}_\mathrm{w}(\Omega) s(\Omega)\,\mathrm{d}\Omega = -\frac{1}{Z_0}\mathbf{v}_\mathrm{w} \tag{60}$$

where the spatial filter is

$$\mathbf{d}_\mathrm{w}(\Omega) = w(\Omega)\mathbf{d}(\Omega) = \begin{bmatrix} d_{\mathrm{x,w}}(\Omega) \\ d_{\mathrm{y,w}}(\Omega) \\ d_{\mathrm{z,w}}(\Omega) \end{bmatrix} = \begin{bmatrix} w(\theta,\varphi)\sin\theta\cos\varphi \\ w(\theta,\varphi)\sin\theta\sin\varphi \\ w(\theta,\varphi)\cos\theta \end{bmatrix} \tag{61}$$

(see Figure 3) and the weighted-velocity signals $\mathbf{v}_\mathrm{w}$ are

$$\mathbf{v}_\mathrm{w} = \mathbf{W}_\mathrm{d,w}^\top \mathbf{a} \tag{62}$$

where $\mathbf{W}_{\mathrm{d,w}} = [\mathbf{w}^{\mathrm{x}}_{\mathrm{d,w}}, \mathbf{w}^{\mathrm{y}}_{\mathrm{d,w}}, \mathbf{w}^{\mathrm{z}}_{\mathrm{d,w}}]$. The vector columns of $\mathbf{W}_{\mathrm{d,w}}$ are the velocity-beamforming coefficients and the subscript of $(\cdot)_{\mathrm{d,w}}$ indicates that the SH weights are assigned to form the weighted-dipole signal components. The construction of these coefficients is discussed next.

While the three dipole patterns for a first-order ambisonic signal correspond directly to the signal components $l = 1$, $n = [2, 3, 4]$ (compare the top row of Figure 3 to the top four SH patterns of Figure 1), forming spatially-constrained dipole patterns requires higher-order signal components and the sector beamforming weights of Eq. 59. The final weights are formed by

$$\mathbf{w}^{\mathrm{i}}_{\mathrm{d,w}} = \mathbf{A}^{\mathrm{i}}\mathbf{w}_{\mathrm{p}} \qquad \text{for } i = \{x, y, z\} \tag{63}$$

where $\mathbf{A}^{\mathrm{i}}$ are sparse, deterministic matrices with dimension $(L+1)^2 \times L^2$. Their construction relies only on the SH order and the dipole pattern ($i = \{x, y, z\}$). Because $\mathbf{A}^{\mathrm{i}}$ is independent of the sector pattern, it is compatible with any sector design of order $L - 1$. Its derivation follows from the spherical expansion of the product of two spherical functions (such as Eq. 61) and is detailed in [46].

The operation could be thought of as weighting the sound field with three orthogonal dipole patterns of Eqs. 24 and 25 in parallel, then further weighting each dipole by the same spatial filter pattern of Eq. 59. However, the construction of Eq. 61 shows that the weighted dipole pattern is a product of the first-order dipole patterns and the sector pattern used to generate the weighted pressure signal (Eq. 59). As such, the resultant weighted-dipole coefficients $\mathbf{w}^{\mathrm{i}}_{\mathrm{d,w}}$ will be one order greater than the sector pattern weights $\mathbf{w}_{\mathrm{p}}$. But it is clear from Eqs. 59 and 62 that the sector weights will need to be of the same order L as the input HOA signal $\mathbf{a}$. It follows that the sector pattern will need to be of order $L-1$, indicating also that such sector-based filtering and analysis would not be possible for a first order ambisonic signal.

The resulting pressure and velocity signals after sector-filtering are then

$$\mathbf{s}_{\mathrm{w}} = \begin{bmatrix} p_{\mathrm{w}} \\ \mathbf{v}_{\mathrm{w}} \end{bmatrix} = [\hat{\mathbf{w}}_{\mathrm{p}}, \mathbf{w}^{\mathrm{x}}_{\mathrm{d,w}}, \mathbf{w}^{\mathrm{y}}_{\mathrm{d,w}}, \mathbf{w}^{\mathrm{z}}_{\mathrm{d,w}}]^{\top}\mathbf{a}, \tag{64}$$

where $\hat{\mathbf{w}}_{\mathrm{p}}$ is the sector pattern coefficient vector $\mathbf{w}_{\mathrm{p}}$ which has been zero-padded from length L^2 to the length $(L+1)^2$ to match the dimension of the weighted-velocity matrices. The weighted pressure and velocity signals of Eq. 64 can now be used in the same manner as presented in Section 3.1 by replacing $p, \mathbf{v}$ with $p_{\mathrm{w}}, \mathbf{v}_{\mathrm{w}}$ to generate the sound field intensity and energy density measures of Eqs. 27 and 29 and in turn attain spatially-localized estimates of intensimetric quantities, such as the diffuseness and sound source direction-of-arrival.

3 Implementation

3.1 Statistical sound field modeling

For the purpose of establishing reference field conditions and to concisely frame the signal processing operations for calculating intensimetric quantities, a statistical model of the sound field is used following [46, §5]. A plane wave model (PWM) [24] of sound field is a adopted, in which a sound field may be constructed by a plurality of plane waves. The spatial density and distribution of plane waves are designed to target the full range of the diffuseness boundaries—from a single plane wave ($\psi = 0$), to a statistically isotropic purely diffuse field of many uncorrelated plane waves ($\psi = 1$). A signal represented by the function $a_{\mathrm{pw}}(f)$ constitutes a plane wave and will have a power spectral density (PSD) of

$$P_{\mathrm{pw}}(f) = \mathbb{E}\{|a_{\mathrm{pw}}(f)|^2\} \tag{65}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator. An isotropic diffuse field is modeled with the direction-dependent amplitude density function $a_{\mathrm{df}}(f, \Omega)$ yielding a diffuse field pressure by integration over the sphere as

$$s_{\mathrm{df}}(f) = \int_{\Omega} a_{\mathrm{df}}(f, \Omega)\, \mathrm{d}\Omega \tag{66}$$

with a PSD of $P_{\mathrm{df}}(f) = \mathbb{E}\{|s_{\mathrm{df}}(f)|^2\}$. The constituent plane waves of the diffuse field are mutually uncorrelated such that

$$\mathbb{E}\{{a_{\mathrm{df}}}^*(f, \Omega_1) a_{\mathrm{df}}(f, \Omega_2)\} = \begin{cases} \frac{P_{\mathrm{df}}(f)}{4\pi}, & \text{for } \Omega_1 = \Omega_2 \\ 0, & \text{otherwise.} \end{cases} \tag{67}$$

A plane wave signal of interest is added to the diffuse field forming a *mixed field* and this nominal "source" signal is expected to also be uncorrelated from the diffuse field such that

$$\mathbb{E}\{a^*_{\mathrm{pw}}(f) s_{\mathrm{df}}(f)\} = \mathbb{E}\{a^*_{\mathrm{pw}}(f) a_{\mathrm{df}}(f)\} = 0. \tag{68}$$

Given a sound field modeled from the above properties is captured by the pressure (omnidirectional) signal $p(f)$ and the velocity (dipole) signals $\mathbf{v}(f)$ (Eqs. 22-24), correlative signals then can be defined which are the cross spectral density (CSD) between the pressure and velocity signals

$$\mathbf{s}_{\mathrm{pv}} = \mathbb{E}\{p^*(f)\mathbf{v}(f)\}, \tag{69}$$

as well as the PSDs of each signal, expressed as

$$S_{\mathrm{pp}} = \mathbb{E}\{|p(f)|^2\} \tag{70}$$

$$S_{\mathrm{vv}} = \mathbb{E}\{\mathbf{v}^{\mathsf{H}}(f)\mathbf{v}(f)\} \tag{71}$$

These signals can then used to calculate the *correlative* intensity [21, Eq. 10], energy density and diffuseness from Eqs. 27, 29 and 30, respectively:

$$\mathbf{I}_a(f) = -\frac{1}{2Z_0}\,\mathrm{Re}\{\mathbf{s}_{\mathrm{pv}}(f)\} \tag{72}$$

$$E(f) = \frac{1}{4\rho_0 c^2}[S_{\mathrm{pp}}(f) + S_{\mathrm{vv}}(f)] \tag{73}$$

$$\psi(f) = 1 - \frac{2\|\mathrm{Re}\{\mathbf{s}_{\mathrm{pv}}(f)\}\|}{S_{\mathrm{pp}} + S_{\mathrm{vv}}}. \tag{74}$$

A final measure will be useful in estimating the direction of prominent sound sources relative to the measurement location, the *energetic DoA estimate*, which will point in the opposite direction of the active intensity vector [45, Eq. 17]:

$$\boldsymbol{\gamma}(\Omega_{\mathrm{DoA}}(f)) = \frac{\mathrm{Re}\{\mathbf{s}_{\mathrm{pv}}(f)\}}{\|\mathrm{Re}\{\mathbf{s}_{\mathrm{pv}}(f)\}\|}. \tag{75}$$

3.2 Sound field synthesis

3.2.1 Signal encoding

As discussed in Section 2.5.2, any function defined on the sphere can be described in the SH domain by its modal coefficients. The amplitude density function a_{df} introduced in Eq. 66 is such a spherical function. Therefore its representation in the SH domain is given by the SHT (Eq. 41) as

$$a_{lm} = \int_{\Omega} a_{\mathrm{df}}(\Omega) Y_{lm}^{*}(\Omega)\,\mathrm{d}\Omega. \tag{76}$$

So for the case of an isotropic diffuse sound field composed of Q uncorrelated plane wave signals in the matrix $\mathbf{s}_{\mathrm{df}} \in \mathbb{R}^{Q\times 1}$, the SH signals are recovered by

$$\mathbf{a}_{\mathrm{df}} = \frac{1}{Q}\mathbf{Y}^{\top}\mathbf{s}_{\mathrm{df}} \tag{77}$$

where $1/Q$ normalizes the power of the sound field to unity. $\mathbf{Y}$ have been evaluated at the corresponding directions Ω_q which satisfy the minimum sampling requirement $Q \geq (L+1)^2$, and in practice $Q \gg (L+1)^2$ to ensure isotropic decorrelation, along with satisfying the spatial uniformity condition of Eq. 50. The signals used for the diffuse field synthesis can be generated in the frequency domain by assigning a random amplitude and phase for the phasors comprising each time-frequency slot of the STFT [24, 56]. Alternatively, white noise (or other noise that would be statistically uncorrelated for the intended analysis window size) can be synthesized in the time-domain, as is done for this work. By synthesizing the diffuse field with Gaussian white noise directly, the phase is modulated inherently and the field is made time-ergodic, allowing for observations sufficiently separated in time to be considered as statistically independent [24]. Del Galdo, et al. in their use of isotropic white noise to model microphone self-noise, confirm that the intensimetric estimators

cannot distinguish between the diffuseness estimated for pure tone and Gaussian white noise diffuse fields [24, §IV.3.B].

The plane wave signal modeling is somewhat more straightforward: a source signal is considered the sound field pressure signal is encoded with a simple vector multiplication of the plane wave signal by the SH components evaluated in the plane wave DoA

$$\mathbf{a}_{\mathrm{pw}} = \mathbf{y}(\Omega_{\mathrm{pw}})^{\top} a_{\mathrm{pw}}. \tag{78}$$

A typical approach to diffuseness estimation using the PWM considers monochromatic sound fields in the time-frequency (TF) domain. The total sound pressure of a *mixed* field as a function of time and frequency is then represented by the combined pressures of the diffuse field component p_{df} and plane wave component p_{pw}, as

$$p_{\mathrm{pd}}(k, n) = p_{\mathrm{df}}(k, n) + p_{\mathrm{pw}}(k, n) \tag{79}$$

where k and n are frequency and time indices, respectively. It is also common to perform parametric spatial audio analysis in the TF domain, so the above formulation is applicable assuming substantial time-invariance of the sound field within each TF analysis window. Indeed, TF processing has been advantageous in such tasks as DoA estimation for concurrent sound sources which may be partially overlapping in frequency [44], and in source localization when considering spatially disparate monochromatic sources differing in frequency [77]. Note that the validity of this model relies on the sound field being comprised of *only* a diffuse field and a single deterministic plane wave in various mixtures, which is also an assumption underlying most applications of diffuseness estimators. This assumption would be challenged by the introduction of additional sound sources, which will be discussed in Section 4.

By contrast, this work aims primarily to evaluate the effect of sector processing in discriminating sound sources in *space*. In order to do so, it is important that concurrent sources indeed overlap in time and frequency. Any conclusions regarding spatial discrimination of the studied techniques can then be readily applied to both monochromatic, broadband, or otherwise partially-overlapping frequencies of disparate sound sources. Likewise, advantages of band-wise TF domain processing, such as single-source-zone estimation [78], are applicable also to sector-based analysis techniques. Therefore, intensimetric analysis is performed on broadband noise sources in this work, unless otherwise specified, and the analysis is performed on the (real-valued) correlative signals of Eqs. 69-74, as suggested by Jacobsen [50, §2]. Note also that because the PWM is used and the SH receivers are considered ideal (they have no radial extent), using the real form of spherical harmonics is sufficient.

3.2.2 Ideal diffuseness, direct-to-diffuse ratio

The accuracy of sound field parameter estimation will vary according to the prominence of the source plane wave signal. So the following quantities are useful when measuring and validating the performance of sound field analysis techniques (following [24, §II.A]). Drawing from the diffuse and plane wave pressure components of

the sound field in Eq. 79, an *ideal diffuseness* of a mixed field is calculated by the following ratio of sound powers

$$\psi_{\text{ref}} = \frac{\mathbb{E}\{|p_{\text{df}}|^2\}}{\mathbb{E}\{|p_{\text{df}}|^2\} + \mathbb{E}\{|p_{\text{pw}}|^2\}}, \tag{80}$$

which serves as a target for a diffuseness estimator. The performance of the diffuseness and DoA estimators will vary according to the direct-to-diffuse ratio (DDR)

$$\Gamma = \frac{\mathbb{E}\{|p_{\text{pw}}|^2\}}{\mathbb{E}\{|p_{\text{df}}|^2\}}. \tag{81}$$

If the DDR of the sound field mixture in known in advance, the diffuse and direct sound power can be determined individually by

$$P_{\text{df}} = \frac{1}{\Gamma + 1} P, \qquad P_{\text{pw}} = \frac{\Gamma}{\Gamma + 1} P \tag{82}$$

where the overall soundfield power is the power of the omnidirectional HOA signal component $P = \mathbb{E}\{|p_{\text{pd}}|^2\} = \mathbb{E}\{|[\mathbf{a}]_1|^2\}$, or equivalently $P = \mathbb{E}\{\|\mathbf{a}\|^2\}/(L+1)^2$ (assuming the PWM) [79, §4.1]. The DDR can also be used to reformulate the ideal diffuseness measure as

$$\psi_{\text{ref}} = \frac{1}{1 + \Gamma}. \tag{83}$$

3.2.3 Temporal averaging

On account of the time-ergodicity of the random processes underlying the diffuse field and statistical sound field model, the expectation operation of Eq. 80 can be approximated with a sufficient temporal average, $\mathbb{E}\{\cdot\} \approx \frac{1}{T}\sum_{i=1}^{T}(\cdot)$ [24]. The period of T samples can be tuned for the application, according to the trade-off between accurately approximating the expectation, requiring a longer time window, and quicker responsiveness to the input signal. Depending on the temporal resolution required for the output, the average can be computed for successive blocks of size T, or as a moving average with a hop size that can be adjusted according to the desired down-sampled temporal resolution of the output. For real-time applications, the moving average window might look forward or backward in time, depending on the latency requirements. A weighted average, i.e. $\sum_{i=1}^{T} w_i(\cdot)/\sum_{i=1}^{T} w_i$, could provide further control over the results. For example, non-uniform weights could emphasize sound onsets at the head of a moving average window.

Simulations presented in [24] indicate a time windows of less then approximately 4 ms will systematically underestimate the diffuseness in scenarios with a low a DDR (less than 0 dB). Lengthening the averaging window converges to to ψ_{ref} while also reducing the estimation variance, again, at the expense of lower temporal responsiveness. When considering real microphone arrays with physical extent, a combination of temporal and spatial averaging can be combined to approximate the expectation operator [24]. This has other practical advantages, for example when estimating the diffuseness in the frequency range above the spatial aliasing limit of SMAs, as is shown by the analysis of diffuseness estimation for a tetrahedral SMA in [80].

3.3 Room simulation via the image-source model

To investigate the performance of sound localization by sector-based analysis, it is necessary to model an acoustic scene consisting of multiple sources and variable reverberant conditions, as observed by receivers in various spatial configurations. For this, the room simulation library by Politis [81] was used which produces impulse responses directly in the SH domain. The room simulation is based on the image-source specular reflection model of Allen and Berkley [82], and allows for the specification of rectangular ("shoebox") room dimensions and T_{60} decay times for octave bands. Alternatively, a maximum reflection order can be specified, which can be useful in the case that only the direct sound and earliest reflections are under test. The model will produce appropriate source-to-receiver distance delay and attenuation as well as air absorption, even in anechoic conditions, allowing for the variables of the acoustic scene, such as relative source distance and varying direct-to-diffuse noise ratios, to be isolated when evaluating the estimation of sound field metrics and target source localization.

The shoebox room simulation library was invoked from a larger object-based structure in MATLAB, built for this work, which stores the metadata of the sound scene configuration, such as source and receiver positions, individual source-receiver room impulse responses, time-series buffers of intensimetric analysis results, estimated and true angles-of-arrival between each source and receiver. In this way various configurations can be synthesized, evaluated, and visualized on demand.

3.4 3D source localization via multiple spherical harmonic receivers

Source position estimation was performed by a least squares solution to the intersection of lines [83]. This method minimizes the sum of squared distance between the estimated source position $\hat{\mathbf{p}}$ and the lines projected from the receiver position $\mathbf{r}_j$ in the direction of DoA estimates $\hat{\Omega}_{\mathrm{DoA},j}$. For K receivers and their corresponding unit vector $\boldsymbol{\gamma}_j = \boldsymbol{\gamma}(\hat{\Omega}_{\mathrm{DoA},j})$ pointing toward the DoA estimate for one analysis frame, the estimated source position is found by solving the system of equations [84]

$$\mathbf{R}\mathbf{p} = \mathbf{q}, \quad \text{where} \quad \mathbf{R} = \sum_{j=1}^{K} (\mathbf{I} - \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^\top), \quad \mathbf{q} = \sum_{j=1}^{K} (\mathbf{I} - \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^\top)\mathbf{r}_j. \tag{84}$$

The solution is found by solving the system of equations directly, as done in the chosen MATLAB implementation [85], or through the use of the Moore-Penrose pseudoinverse

$$\hat{\mathbf{p}} = \mathbf{R}^\dagger \mathbf{q}. \tag{85}$$

4 Evaluation

The success of source localization by intensimetric analysis relies foremost on the accuracy of the direction-of-arrival (DoA) estimation. Based on the IE estimator of Eq. 31 it is clear that this depends on the ratio of the propagating energy, as quantified by the active intensity vector, and the overall sound field energy. This ratio can be visualized by the negated intensity vector, wherein its direction indicates the DoA estimate. Its length indicates the degree to which the sound field is diffuse or directive for values that tend toward 0 or 1, respectively. Many of the following visualizations use this representation to demonstrate the response of the DoA estimator to convey the impact of the direct-to-diffuse ratio (DDR), sector pattern directivity, source angle-of-arrival (AoA), interferer AoA, and temporal averaging. For clarity, in the (negated) active intensity vector depiction will be referred to as the "DoA vector". Note that the visualization of the DoA vectors is a top-down projection of 3D vectors, for sources and receivers positioned on the same X-Y plane. When order is specified in the plots, it refers to the order of the analyzed receiver signals, which is one order greater than the sector (pressure) pattern used to filter the sound field (refer to Section 2.6).

For the results that follow, the simulated room condition is anechoic and isotropic diffuse Gaussian white noise is added to the receiver signal to attain a DDR according to Eq. 81. The source is white noise and broadband analysis is performed on one second of steady state signal (avoiding the onset of the source after propagation delay to the receiver) and, unless otherwise specified, the temporal averaging window is 132 samples (3 ms) at a sample rate of 44100 Hz. This temporal average window size is chosen to more clearly illustrate the differences between the conditions being compared and to be close to a frame size typically chosen for parametric spatial audio processing in the time-frequency (TF) domain. In practice, the time averaging window would be tuned to the application, and results from processing performed in the TF domain would also be further averaged over multiple frames for estimation stability [45].

4.1 Sector-based DoA estimation

4.1.1 Varying direct-to-diffuse ratio

The direct-to-diffuse ratio at the receiver is central to DoA estimation. Figure 4a shows the impact of the DDR on DoA estimation for a source that is *on-axis* with sectors of various orders. For very low DDR, its apparent that a first-order analysis produces essentially a diffuse field response, in terms of both the DoA vector length and the spatial spread of the estimates. Sector-based analysis, on the other hand, already begins to concentrate DoA estimates around the peak of the filter's main lobe, which *in this case* coincides with the direction of the source. So while the sector profile clearly rejects competing off-axis noise, as would be expected from an axisymmetric spatial filter, a simple averaging of the DoA estimates would not be sufficient to make an accurate prediction if the source were to move off-axis to the sector. However, considering the length and directional variance (spread) of the DoA

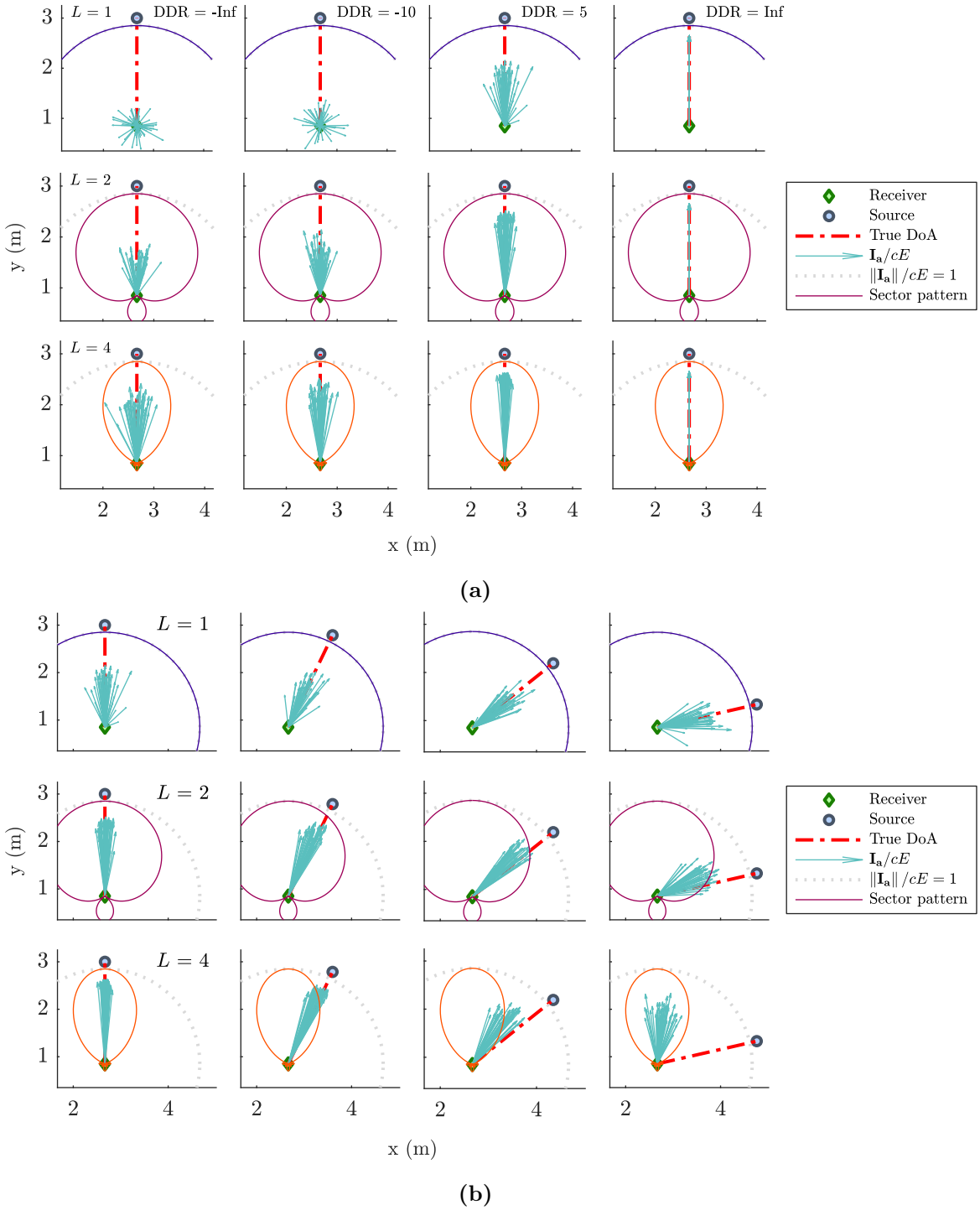


Figure 4: a) DoA estimation with varying DDR for an on-axis source. **b)** DoA estimation with varying source angle-of-arrival, $\alpha = [0°, 25.7°, 51.4°, 77.1°]$, with a constant DDR of 5 dB.

vectors would provide an indication of the presence of a sound source, as can be inferred from the Figure 4a as the DDR condition increases from left to right. This directional variance underlies the TV and SV diffuseness measures (Eqs. 32, 34). The behavior of the DoA vector length shift with DDR is addressed in Section 4.2.

4.1.2 Source angle-of-arrival

It can be seen that the accuracy of first-order analysis in the single-source case can only be improved by an increase in the DDR, as its omnidirectional "field of view" is insensitive to the direction of the source. This can be seen in the spread and length of the DoA estimates in the first rows of both sub-figures of Figure 4, and the estimation error plots of Figure 5, which remain constant with the source direction. On the other hand, as the source moves off-axis it is attenuated on account of the sector's directional gain response. As a result, the DoA vectors shorten and the angular variance of the estimates increase, eventually converging with the diffuse-field response once the source AoA, denoted α, coincides with the sector null. This behavior is visible in the cases of $L = [2, 4]$ in Figure 4b.

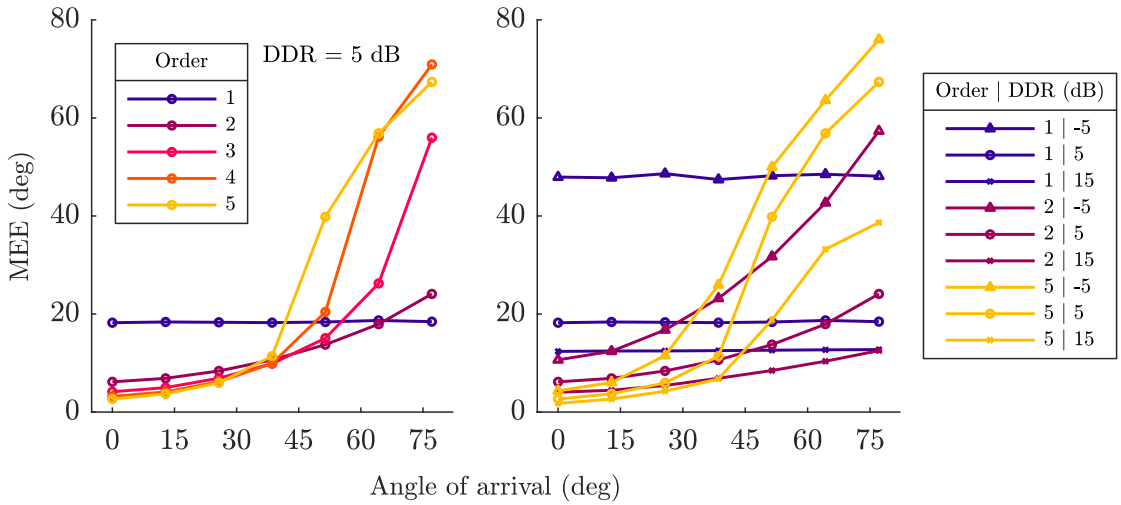


Figure 5: Mean estimation error (MEE) of source DoA in mixed field conditions of a single source in a diffuse field at varying angles-of-arrival α. Performance is compared across orders $L = [1, ..., 5]$ for a constant DDR of 5 dB (left), and for orders $L = [1, 2, 5]$ at varying DDRs (right). $\alpha = 0°$ can be considered best-case performance for a given DDR. The advantage of higher-order processing deteriorates beyond a certain threshold of α, which decreases with increasing order, contingent upon the DDR.

The source position does not affect the isotropic diffuse noise, so as the source moves off-axis it is attenuated by the sector pattern, effectively *reducing* the DDR of the mixed field, denoted as Γ. Note the ambiguity in the analysis between the conditions of $L = 4$, $\Gamma = -\infty$, $\alpha = 0°$ (bottom leftmost panel, Fig. 4a) and $L = 4$, $\Gamma = 5$ dB, $\alpha = 77.1°$ (bottom rightmost panel, Fig. 4b). This ambiguity is not present at $L = 2$ (middle row, Fig. 4b) where the off-axis source still lies substantially within the main beam, which attenuates the source by approximately -5.8 dB, and the

average DoA estimate still largely tracks the true DoA. This suggests the possibility of disambiguating the source position and diffuse field bias by comparing analyses across orders.

The DoA estimation performance is quantified by the mean estimation error (MEE) over the full set of observations in Figure 5. The left panel shows that for a single DDR of 5 dB, sector processing is clearly beneficial for on-axis sources. The improvement deteriorates as the source moves off axis and as the DDR is forced down by the spatial selectivity of the sector. This trend occurs more rapidly with a higher sector order, indicating the trade-off between robustness of estimation across a larger field of view with a broader, low-order sector, with the higher accuracy of a narrower higher-order sector.

4.1.3 Impact of temporal averaging

The impacts of temporal averaging, discussed in Section 3.2.3, are visualized in Figure 6 for a DDR of 5 dB. While the longer averaging window will produce a DoA estimates which are more stable in both direction and length, observations which are more directional but transient could be lost. The shortest time average in the first order analysis (top left panel), for example, shows that those DoA vectors with greater magnitude tend toward to the source direction, while erroneous directional estimates tend to have smaller magnitudes, indicating a higher diffuseness (by the IE diffuseness estimator). This begins to motivate the concept of estimation filtering based on the accompanying diffuseness metric, which will be discussed in Section 5.4. Sector-processed estimates for $L = 4$ (bottom row) show an improvement in directional estimation for sources within the main lobe of the spatial filter due to the rejection of off-axis noise by the spatial filter, potentially allowing for the use of shorter time average windows.

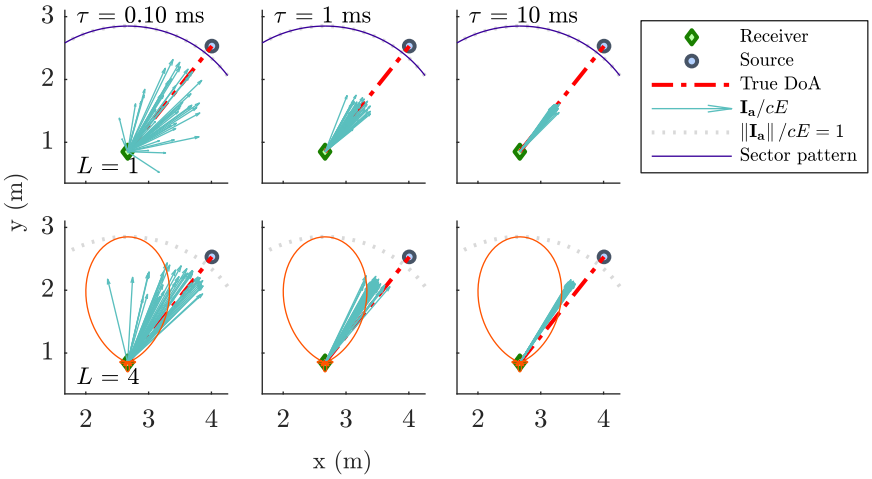


Figure 6: DoA estimation vectors resulting from different moving average window sizes $\tau = [0.1, 1, 10]$ ms, for analysis orders $L = [1, 4]$ and a DDR of 5 dB.

4.1.4 Interfering sources

The ability of a spatial filter to reject off-axis noise is advantageous for improving the DDR for a source of interest, provided the source lies substantially within the main lobe of the analysis sector. The same principle applies to the rejection of interfering sound sources that are separated from the target source by a sufficient angular distance. Figure 7a illustrates the behavior of the DoA estimation given a sound scene free of diffuse noise with two equal-power uncorrelated sound sources at an equal distance from the receiver. The on-axis source is considered the target to be distinguished from the interferer to its right. The MEE over the full set of observations is shown in Figure 7b, for orders $L = [1, 2, 5]$.

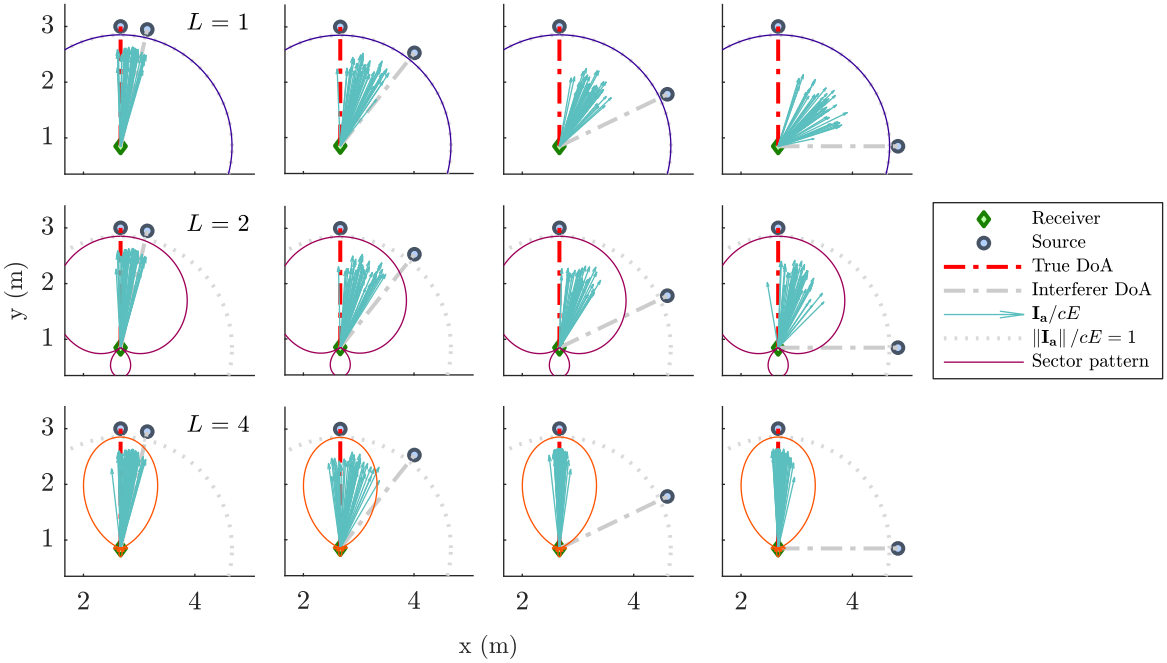
First-order analysis shows a linear increase in estimation error as the interferer moves away from target source on account of the total propagated sound energy being equally divided between the two sources, but averaged equally across their directions of incidence to the receiver. This spatial averaging is also observed in the length of the DoA vectors which shorten on as the proportion of total sound energy, which remains constant, is divided across a wider spatial region. The line plot of Figure 7b shows the clear advantage of sector-based analysis in DoA estimation to reject off-axis interferers. The level of the interferer is reduced simply as a function of the sector's directivity, with maximum rejection occurring at the null of the sector pattern. The left panel shows estimation error for a sound scene without interfering noise, while the right panel shows the error with diffuse noise present at varying levels. Note that first order estimation degrades, somewhat predictably, as the noise level increases. However, this characteristic is only consistent for sector-based analysis for which the interferer is near to the target (see the first data point for $L = 2$), and as the interferer moves off-axis, the performance appears to *improve* with increasing noise levels. This is explained by the fact that with a higher DDR, the interferer comprises a larger percentage of the overall sound field energy, thereby causing an increasing directional estimation distortion as its DoA moves away from the target.

4.1.5 Interferer distance

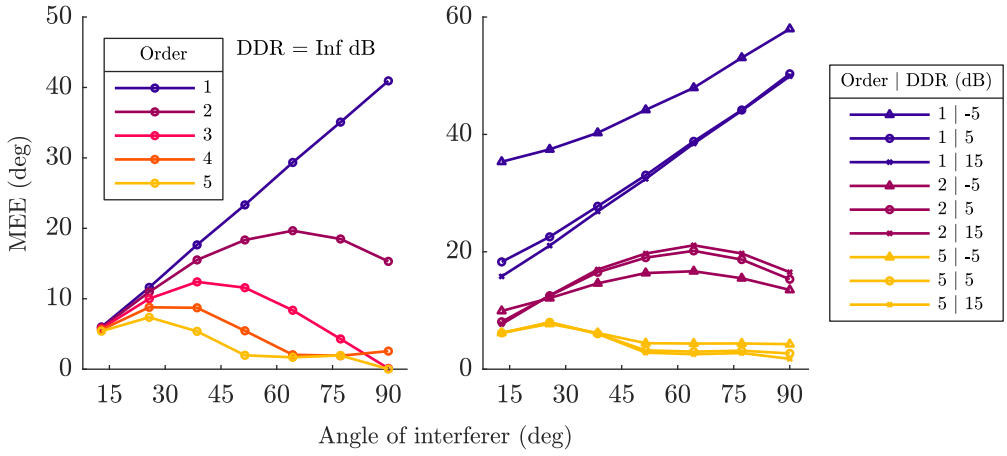
In addition to the angular distance of an interferer from the target source, the relative distance of the interferer will also impact estimation accuracy. Figure 8 shows how estimation improves as an equal-power interferer recedes from the receiver at a constant AoA. The DoA vectors shift toward the target source as the source increasingly dominates the energy ratio between them. First order analysis shows the error scales approximately with the inverse distance law governing the sound energy contribution of the interferer as it recedes. The spatial filter used in the sector-based analysis adds a further attenuation factor to the interferer, which in this idealized, anechoic case, shows predictable improvement over first-order analysis.

4.2 Spatially-localized diffuseness estimation

It has already been observed that the spatial filter imposed on a sound field comprised of a purely diffuse field or a mixed field will impact the active intensity measure,

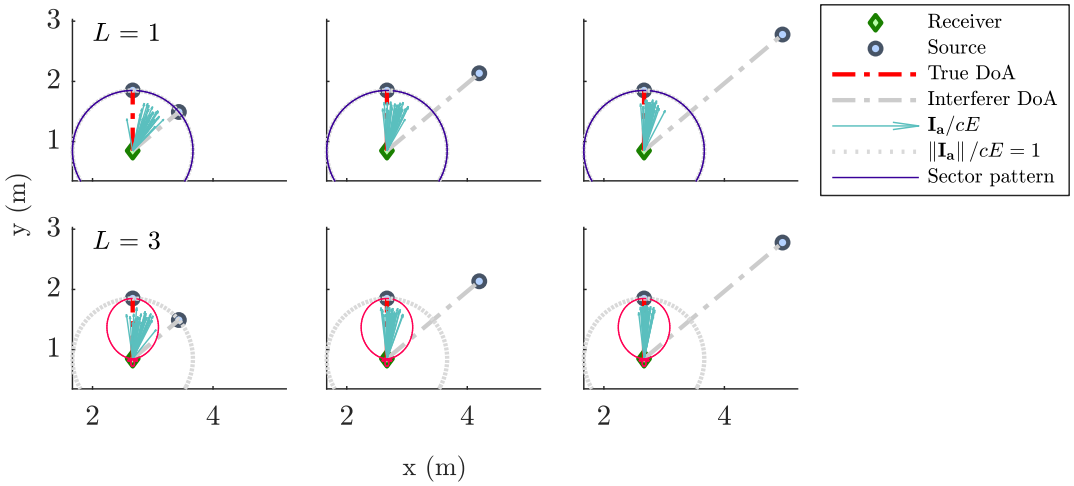


(a)

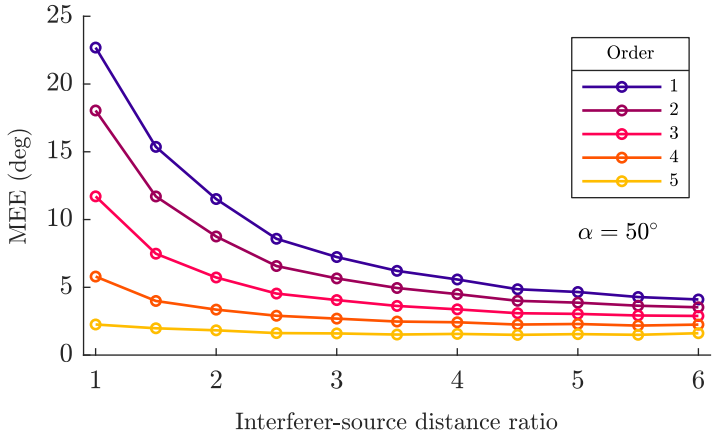


(b)

Figure 7: a) DoA estimation vectors of a sound scene with an equal power interferer at varying AoAs: $\alpha = [0°, 25.7°, 51.4°, 77.1°]$, $\Gamma = \infty$ dB. **b)** Corresponding mean estimation errors (MEE) for source and interferer only (left), and with diffuse noise present at varying DDRs (right). Note the different scales of the y-axes.



(a)



(b)

Figure 8: a) DoA estimation vectors of a sound scene with an equal-power interferer at $\alpha = 50°$, varying source-interferer distance ratios of $[1, 1.8, 2.6]$, for analysis orders $L = [1, 3]$. **b)** Corresponding mean estimation errors for analysis orders $L = \{1, ..., 5\}$ and source-interferer distance ratios between 1 and 6.

which necessarily impacts the diffuseness estimate and in turn the DoA vector length (calculated by the IE estimator). The effect of the spatial filter on energetic analysis has been defined analytically in [46, §5.1-3], for three reference field conditions: a single plane wave, a purely diffuse field, and a mixture of the two. The directivity factor (see also Eq. 57) for a normalized sector pattern can be formulated in terms of its sector coefficients by

$$Q = \frac{4\pi}{\int_\Omega w^2(\Omega)\,\mathrm{d}\Omega} = \frac{4\pi}{\mathbf{w}_\mathrm{p}^\top \mathbf{w}_\mathrm{p}}. \tag{86}$$

For a completely diffuse field weighted by the sector pattern, the vector representing the average DoA over the sphere will be the product of the squared sector pattern and the uniformly distributed DoAs (Eq. 23), as

$$\mathbf{k} = \int_\Omega w^2(\Omega)\mathbf{d}(\Omega)\,\mathrm{d}\Omega = \hat{\mathbf{w}}_\mathrm{p}^\top \mathbf{W}_\mathrm{d,w} \tag{87}$$

where $\mathbf{W}_\mathrm{d,w} = [\mathbf{w}^\mathrm{x}_\mathrm{d,w}, \mathbf{w}^\mathrm{y}_\mathrm{d,w}, \mathbf{w}^\mathrm{z}_\mathrm{d,w}]$. For an axisymmetric sector pattern, $\mathbf{k}$ will point toward the pattern maximum and therefore is equivalent to magnitude $K = \|\mathbf{k}\|$ scaling the unit vector representing the look direction of the beam, $\mathbf{k} = K\mathbf{d}(\Omega_0)$. Calculating the magnitude K can then be simplified by considering the one-dimensional sector directivity pattern $c(\theta)$ (refer to Section 2.5.3) and $c_\mathrm{z}(\theta) = c(\theta)d_\mathrm{z}(\theta)$. Using the beamforming weights of each pattern, $\mathbf{c}$ and $\mathbf{c}_\mathrm{z}$, respectively, K can be defined as [46, Eq. 39]

$$K = \hat{\mathbf{c}}^\top \mathbf{c}_\mathrm{z} \tag{88}$$

where the coefficient vector $\hat{\mathbf{c}}$ is $\mathbf{c}$ zero-padded from length L to $L+1$ to match the length of $\mathbf{c}_\mathrm{z}$, similar to the construction of Eq. 64. Using Eqs. 86 and 88, the diffuseness of a purely diffuse sound field that has been spatially filtered, following [46, Eq. 38], is

$$\psi_\mathrm{df,w} = 1 - \frac{QK}{4\pi}. \tag{89}$$

This value represents the theoretical diffuseness in a *purely diffuse* field that has undergone sector processing, and decreases with increasing order. Figure 9 shows these diffuseness thresholds (dashed lines) along with values realized in simulation, for which the bar lines represent one standard deviation for the observations. The underestimation of diffuseness, for $L = 1$ in particular, results from the short time-average window—a longer time average converges on the theoretical value but with little variance to display.

The addition of a single plane wave source to the sound field mixture will alter the diffuseness estimate as a function of the DDR and source angle, Γ and α, respectively, according to [46, Eq. 48]

$$\psi_\mathrm{dp,w}(\Gamma, \alpha) = 1 - \frac{\Delta(\Gamma, \alpha)}{\Gamma c^2(\alpha) + 1/Q}, \tag{90}$$

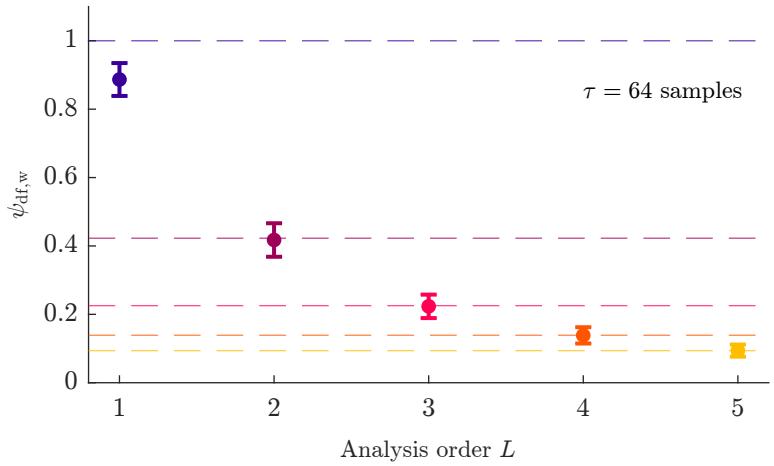


Figure 9: Ideal diffuseness for a spatially-filtered sound field. The dashed lines are the reference level according to Eq. 89, with mean diffuseness values calculated through simulation indicated by the markers. The sector pattern is max$-\mathbf{r}_{\mathrm{E}}$ and the moving average window size is $\tau = 64$ samples, with one standard deviation indicated by the bar lines.

where $\Delta(\Gamma, \alpha) = \sqrt{\Gamma^2 c^4(\alpha) + (K/4\pi)^2 + 2\Gamma(K/4\pi)c^2(\alpha)\cos\alpha}$. If the source DoA coincides with the sector heading, $\alpha = 0$, the resulting diffuseness simplifies to the diffuse field response of Eq. 89, scaled by the sector pattern directivity factor Q and the DDR, according to [46, Eq. 50]

$$\psi_{\mathrm{dp,w}}(\Gamma, 0) = 1 - \frac{|\Gamma + (K/4\pi)|}{\Gamma + 1/Q} = \frac{\psi_{\mathrm{df,w}}}{Q\Gamma + 1}. \tag{91}$$

This relation in turn forms a lower bound of the diffuseness range for the mixed field case. The effect of Eqs. 90 and 91 on defining theoretical diffuseness boundaries can be observed in Figure 10. The order-limited upper bound is shown by the case of $\Gamma = -\infty$ dB (dotted line), and the lower bound for the plotted DDRs is marked by the dashed line, which occurs only for an on-axis source. Comparing the diffuseness range between the left and right panels for $L = 2$ and $L = 4$, respectively, shows that the range of diffuseness when α lies within the sector's main lobe will become increasingly compressed with increased order. This needs to be accounted for when performing source localization tasks that use the diffuseness estimation as a variable for estimation refinement (see Section 5.4).

Complex behavior is observed when α is near the sector pattern null or in the range of the side or rear lobes, where the diffuseness will increase *beyond* $\psi_{\mathrm{df,w}}$. This indicates the source energy draws the active intensity away from $\mathbf{k}$, thereby driving its magnitude below K, so its proportion to the overall sound energy diminishes, resulting in a *higher* diffuseness than would be predicted by Eq. 89 for a purely diffuse field. If the single-source condition of the model is met, this could be a rough indicator of the presence of an off-axis source.

Lastly, in the single-source mixed field case, if the source does not align with the sector heading, sector-based analysis will introduce bias in the DoA estimation

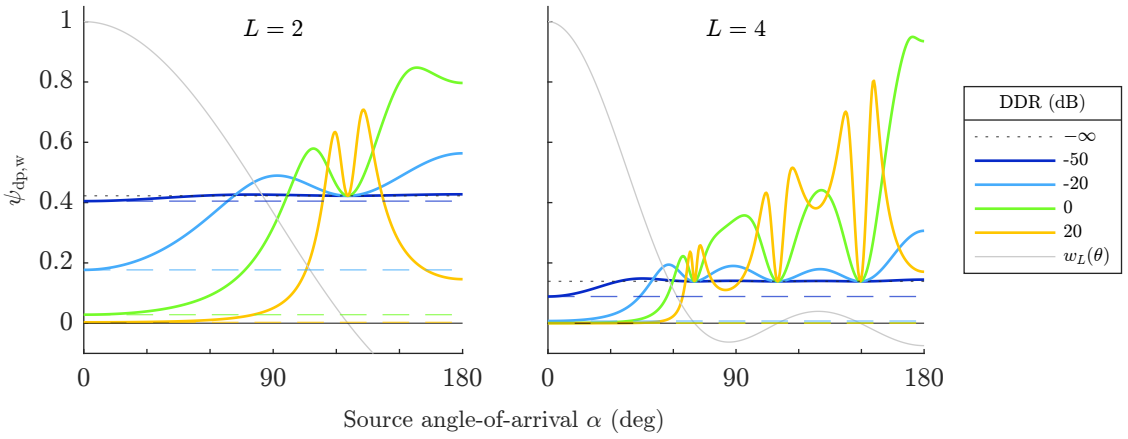


Figure 10: Theoretical diffuseness $\psi_{\text{dp,w}}$ observed when sector processing of a mixed sound field—a plane wave and isotropic diffuse noise. The diffuseness varies according to the source AoA α and DDR. The lower bound is governed by Eq. 91, corresponding to $\alpha = 0$. The upper diffuseness range anchors around $\psi_{\text{df,w}}$ as the source moves off-axis and observed precisely when α coincides with the sector pattern nulls. Note the diffuseness overshoots this boundary near the pattern nulls and beyond the pattern's main lobe.

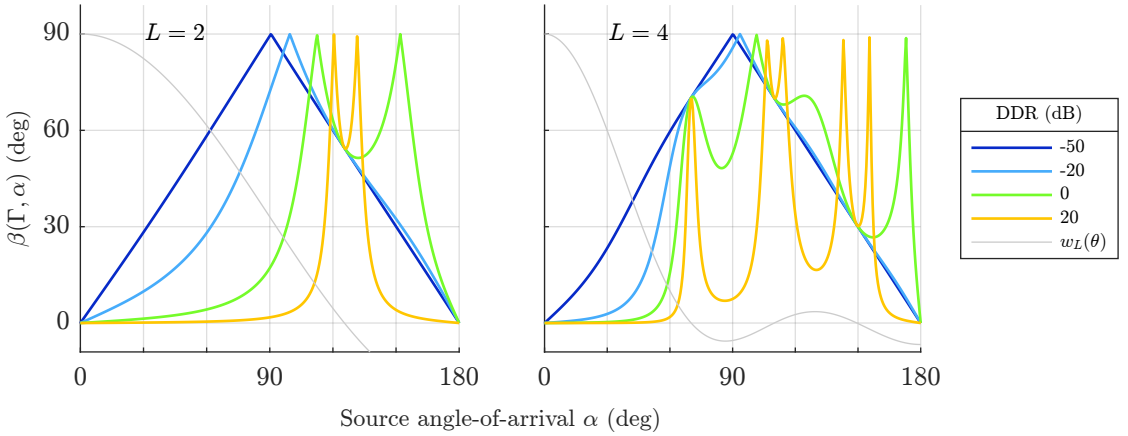


Figure 11: DoA estimation bias due to isotropic diffuse noise present with a target sound source, as a function of the source angle of arrival to the sector axis and the DDR. The sector pattern $w_L(\theta)$ is plotted in gray, normalized to the maximum angular bias, for visual reference as to the directional gain and null position.

toward Ω_0, as discussed in Section 4.1 (see Figure 4, lower panel). This bias can also be analytically defined as a function of the DDR and AoA as [46, Eq. 51]

$$\beta(\Gamma, \alpha) = \arcsin\left(\frac{K \sin\alpha}{4\pi\Delta(\Gamma, \alpha)}\right). \tag{92}$$

As can be seen in Figure 11, for a very low DDR the bias approaches a linear function with respect to the AoA, while the bias diminishes with increasing DDR, arriving at a perfect estimate when $\Gamma = \infty$. This transfer function presents the possibility of bias compensation, which is presented in Section 5.3. Note, however, the method requires

that the DDR is known or estimated and that a only a single source is present. With mixed field diffuseness estimation, complex behavior emerges for source arrival angles outside of the sector's main lobe, which further complicates bias compensation.

4.3 Sector-based 3D localization

The improvements offered by high-order sector-based processing over first-order analysis, carry forward into the task of three-dimensional source *position estimation*, here called source *localization*, which is performed with analyses across multiple SH receivers. However, the stipulations of sector-based DoA and diffuseness estimation—such as the assumed sound field model, effects of the DDR, source and interferer distance and proximity, sector heading with regard to source directions—need to be considered as well. In the evaluation that follows it will be seen that sector processing can serve to improve source position estimation, but require the insights from the evaluations in the previous sections about single-receiver analysis to understand the limitations of the method and approaches to improving performance.

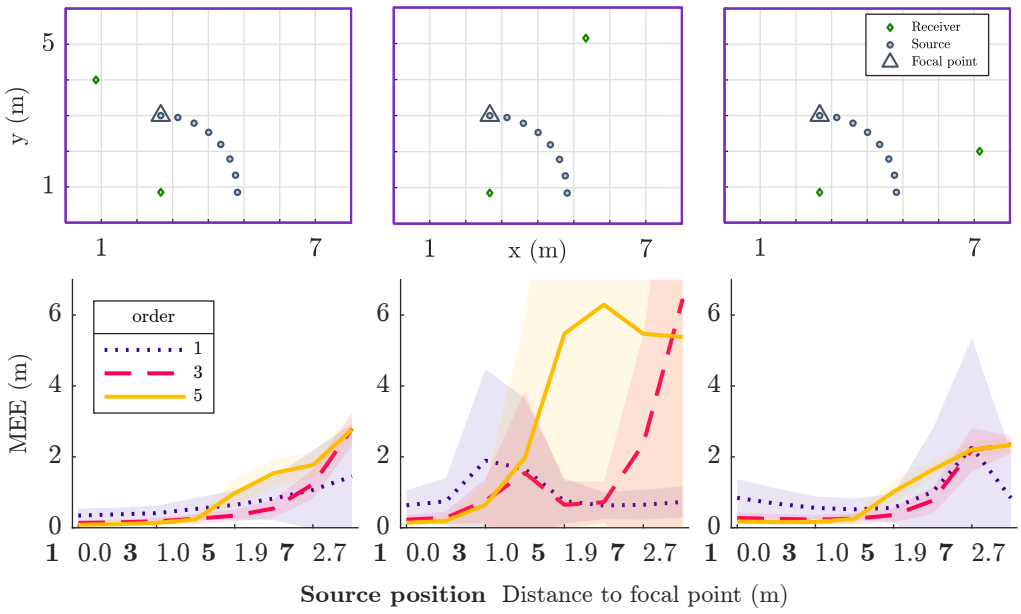
4.3.1 Single-source localization

To evaluate the accuracy of localization by multireceiver analyses, a sound scene is constructed with a single source located at a "focal position". A single spatial filter (sector) from each receiver is then directed toward this position so that the source signal is maximized when in the focal position. A specified DDR is achieved by injecting isotropic diffuse noise into each receiver signal equally (refer to Section 3.2.1), with reference to an omnidirectional source at the focal position and a reference receiver position (the bottom receiver position in the scene diagram plots that follow). In doing so, the effects of source distance to each receiver are properly accounted for in the energetic analysis.

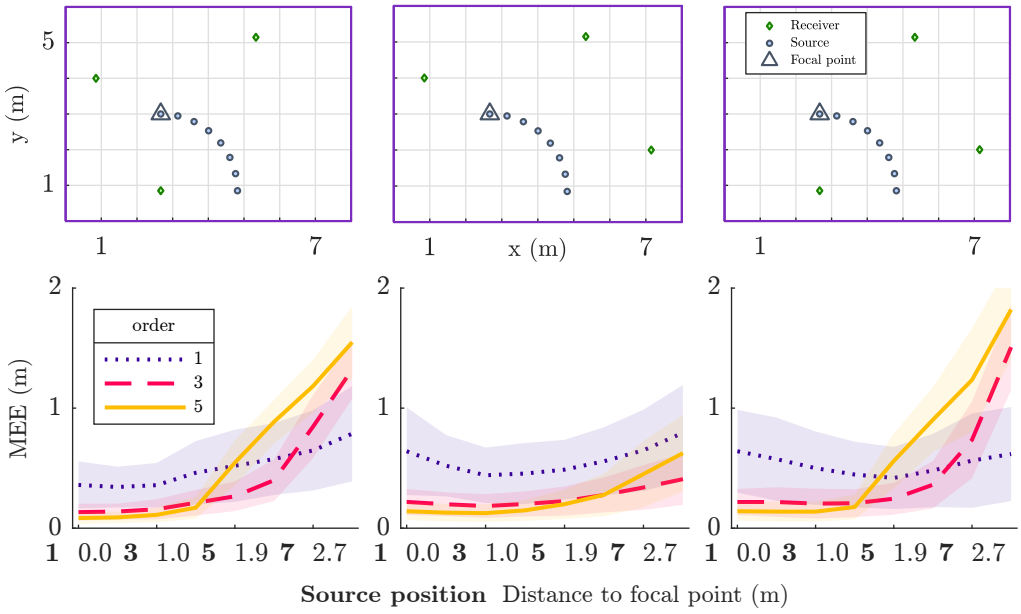
Figure 12 illustrates the MEE of source position estimation, for two- and three-receiver configurations, evaluated for eight different source positions. Source position 1 is the focal position and the others draw a 90° arc sequentially clockwise. In this way the DDR at the reference receiver will be constant while for the other receivers, the DDR will vary with source position.

In the two-receiver configuration, it is clear that the position estimation error is larger, with larger variance, when first-order processing (i.e. no spatial filtering) is used, as might be predicted from a similar trend in DoA estimation (refer to Section 4.1). With higher-order processing, the error is consistently lower for the first few source positions but increases as the source moves away from the focal position, off-axis with receivers' analysis sectors. In the leftmost configuration, for example, the DDR will decrease in the upper left receiver as the source recedes from the first to the last position, degrading its DoA estimates as a result. A corresponding degradation of DoA estimation will occur in the lower receiver as the source moves off-axis compounding the error for source positions farthest from the focal point. Refer to Figure 4 for a visualization of both effects.

The two-receiver configuration in the middle panel of Figure 12a illustrates a



(a) Two-receiver configurations



(b) Three-receiver configurations

Figure 12: Mean estimation errors (MEE) of single-source position estimation, resulting from varying 2- and 3-receiver configurations in panel 12a and 12b, respectively, with source-receiver scene configurations shown above each error plot. Diffuse noise level is set equally across all receiver, with reference to a DDR $\Gamma = 10$ dB between the source positioned at the focal position ($\triangle$) and the lower receiver. Estimation error is performed with a time-averaging window $\tau = 3$ ms for analysis orders $L = [1, 3, 5]$, with one standard deviation shaded around the MEE. Labels on the x-axis display the interferer position index and the distance from the focal position in bold and regular type weight, respectively. Note the difference of scale on the y-axis between the two- and three-receiver error plots.

particular shortcoming of the intersecting-line localization approach (Section 3.4): when a source moves directly between two receivers, DoA estimation vectors which may individually have a small DoA error can result in intersection points very far from the target source as the vectors approach a parallel or collinear orientation. This is observed in the very high MEE and variance of the center panel and, to a lesser extent, for source positions 6 to 8 in the right panel. Notably, under this unstable condition the MEE peaks at different source positions for different orders, and has the highest variance (i.e. least reliable estimates) for the highest-order sector processing. The problem is visualized in Figure 13. One approach to mitigating the issue is to use multiple time-series DoA estimates, rather than only two time-averaged vectors per frame, but care must be taken on account of each DoA projection intersecting back at each receiver. The detrimental effect of off-axis source directions is visible across all receiver configurations, which in some cases allows first-order estimation to outperform sector-basted estimation.

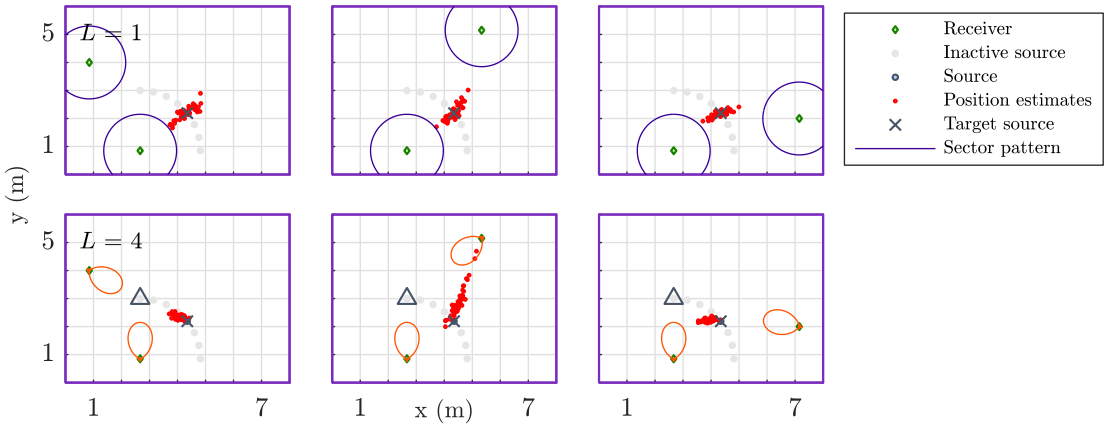


Figure 13: Source position estimates with two-receiver configurations analysized at orders $L = [1, 4]$. Error increases as the source moves between receivers, drawing DoA estimates nearer to parallel thereby making the intersection point calculation highly variable and error-prone. Scene and analysis conditions match those of Figure 12.

The variability in position estimation is improved considerably with the addition of a third receiver, as seen in the error plots of Figure 12b (note the difference in the MEE scale compared to Figure 12a). In addition to improved MEE overall, the error variance is made largely consistent across source positions. As with the two-receiver cases, estimation accuracy degrades with increasing order for those source positions which lie substantially off-axis from one or more receivers. This trend reflects the contour of the DoA error observed in Figure 5 and discussed in Section 4.1.2.

While Figures 12 and 13 highlight particular features of position-dependent localization error, the performance of the technique is extrapolated across broad spatial regions enclosed by the receivers in Figure 14. It can be seen that the region of highest localization accuracy anchors to the focal position of the sectors and draws toward the barycenter of the receiver arrangement. Close inspection of the error contours shows regions of highest accuracy form at positions in which rays projected towards the receivers would be maximally orthogonal.

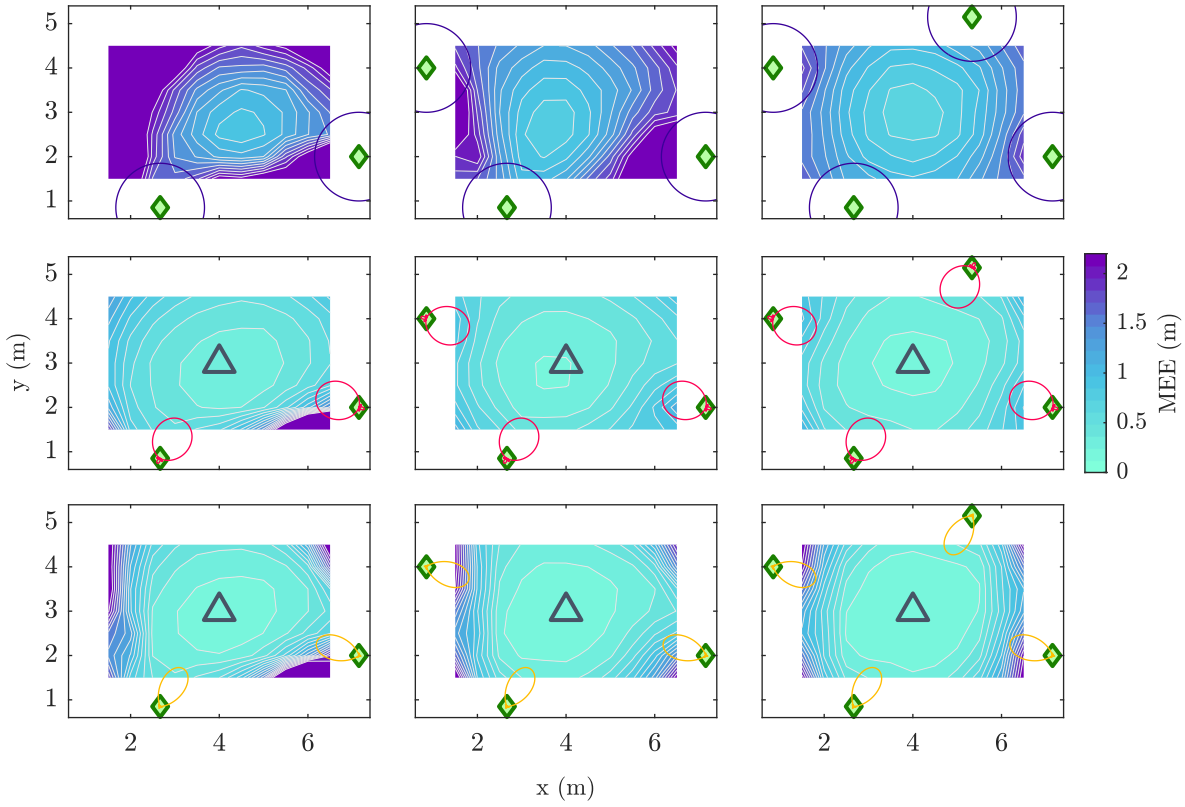
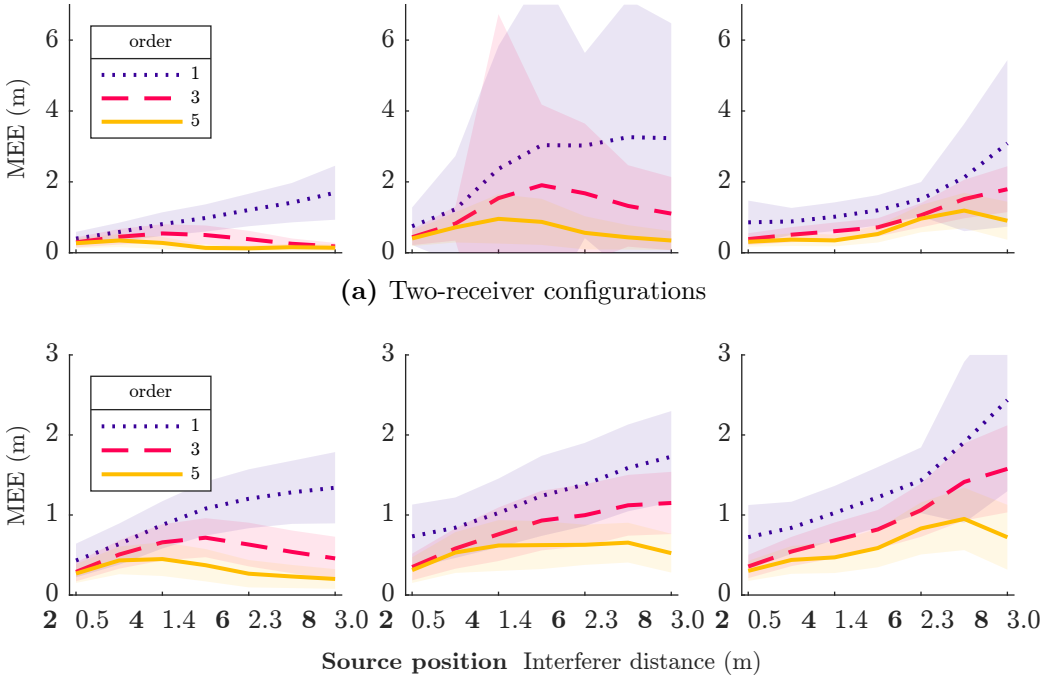


Figure 14: Mean estimation error (MEE) for single-source position estimation by two-, three- and four-receiver arrangements. Receivers are oriented toward the focal point (△), while the source moves throughout the plotted region. The scene DDR is $\Gamma = 5$ dB, rows correspond to analysis orders $L = [1, 3, 5]$.

4.3.2 Source localization with an interferer

In order to quantify the impact that an interfering sound source has on localizing a target source, two sound scenes were evaluated. Similar to the previous single-source analysis, the first scenario considers the influence of the *number* and *configuration* of receivers with the target source located at the focal position with an equal-power uncorrelated interferer at varying proximity. The interferer positions match source positions 2-8 of Figure 12, i.e. the target-to-interferer power ratio is constant for the bottom receiver, and varies for all other receivers as the interferer position changes. The results are shown in Figure 15 for analysis orders $L = [1, 3, 5]$.

The results differ from the single-source case in that the estimation performance is consistently improved by sector-processing, and further by increasing analysis order. Note that the conditions for DoA estimation performed with each individual receiver can be considered a best-case scenario; the target source lies directly on-axis with the sector heading, $\alpha = 0$. As such, the discussion of the impact of interferers on DoA error from Section 4.1.4 can be readily applied to this scenario (see Figures 7 and 8). In the case an interferer is very close to the target will result in low localization error, but this should be viewed only as advantageous in terms of source *detection*



(a) Two-receiver configurations

(b) Three-receiver configurations

Figure 15: Mean estimation error (MEE) of source localization with an equal-power interferer at varying distances from the target source. The source and receiver configurations correspond to those of Figure 12, as do the analysis settings ($\Gamma = 10$ dB, $\tau = 3$ ms, $L = [1, 3, 5]$). The target source is at the focal position while the interferer positions follow positions 2 to 8. One standard deviation is shaded around the MEE. Labels on the x-axis display the interferer position index and the distance from the focal position in bold and regular type weight, respectively. Note the different scale on the y-axis of **(a)** and **(b)**.

at that location, but not source *separation* (the target and interfering source are indistinguishable without further processing). As such, the performance for the case of an *off-axis* interferer, i.e. interferer rejection, is of greater interest.

Localization error can be expected to rise initially as the interferer moves away from the source, as the DoA vectors at each receiver are drawn to the energetic spatial average between the source and interferer. But the error begins to fall as the interferer moves away from the focal position, indicating source separation by the spatial filter's rejection of interfering off-axis energy. This can be observed in all fifth-order analysis cases (orange solid line), and to a lesser extent in the third-order case (red dashed line). First-order estimation (blue dotted line), in contrast, progressively worsens as the interferer moves away from the target, as the DoA vectors for the receivers are drawn to the energetic spatial average between the source and interferer. Interestingly, for two out of three of the chosen receiver configurations the performance is similar between two- and three-receiver cases (left and right columns of Figure 15, note the difference in y-axis scale between rows). However, the center configuration indicates the same detrimental effects of poor DoA-intersection accuracy arising from collinear DoA vectors, as observed in the single-source case (see Figure 12a). This especially impacts the two-receiver configurations and is largely resolved by the addition of a third receiver.

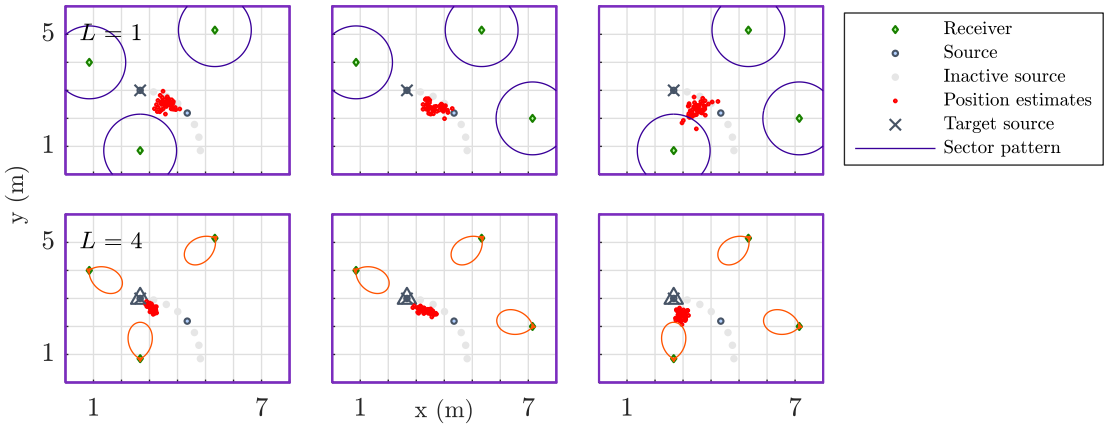


Figure 16: Source localization estimates for a target source (×) with an equal-power interferer present, for various three-receiver configurations (matching those in Figures 12b and 15b), produced from analysis at orders $L = [1, 4]$. Error increases as the interferer draws DoA estimates away from the target, and depends on the relative proximity and AoA of both the target and interfering source. Scene and analysis conditions match those of Figure 12.

The localization performance for the three-receiver configurations, corresponding to Figures 12b and 15b, are visualized in Figure 16 for orders $L = [1, 4]$. It can be seen that the centroid and variance of the position estimates are similar in all first-order configurations, whereas the centroid of fifth-order position estimates shifts according to the receiver configuration, which is not otherwise obvious from the MEE metric alone. This effect becomes important in applications such as sound source signal extraction, in which the centroid of the estimates may be used for

choosing which receiver to produce the beam-formed signal (by source proximity or path of least interference), or choosing mixing weights and temporal alignment of beam-formed signals from multiple receivers used to reconstruct the source signal.

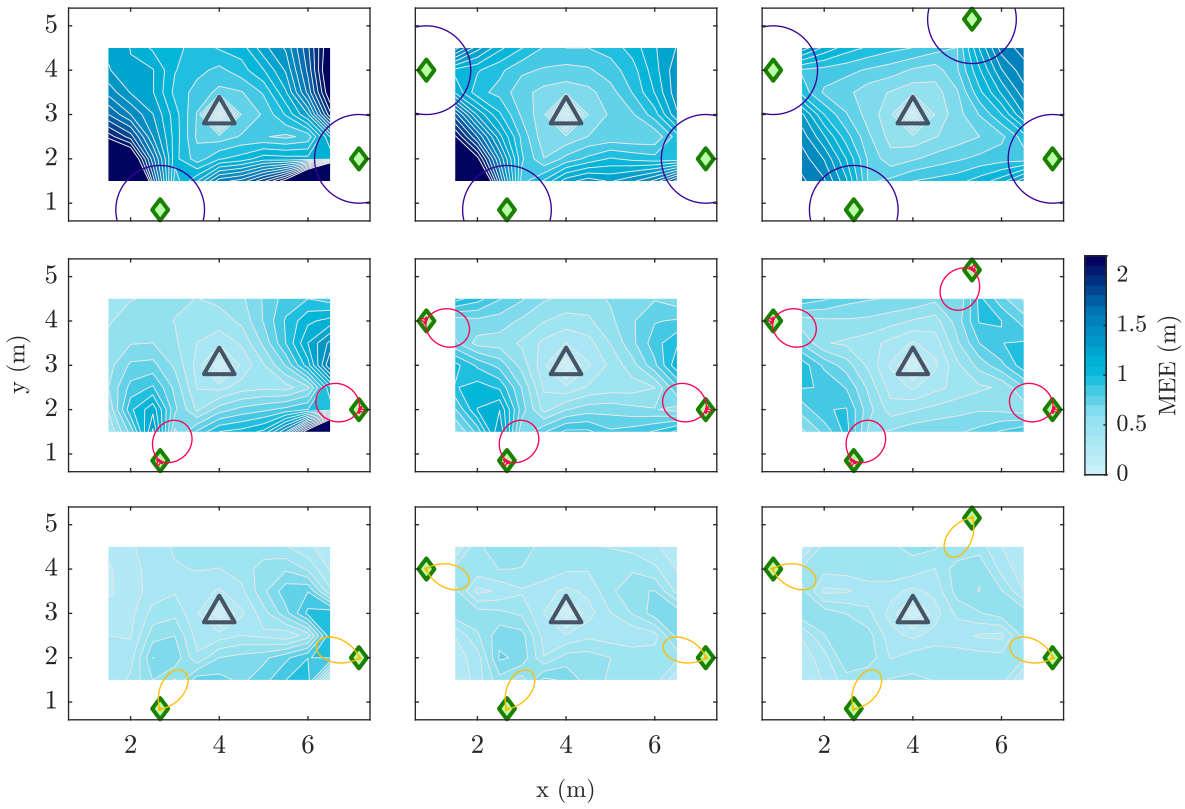


Figure 17: Mean estimation error (MEE) of source localization in the presence of an equal-power interferer by two-, three- and four-receiver arrangements (left to right). The target source remains in the focal point (△), while the interferer moves throughout the plotted region. The scene DDR is $\Gamma = 5$ dB, rows correspond to analysis orders $L = [1, 3, 5]$.

The localization error was evaluated for a second sound scene in which the target source remains fixed in the focal position as an interferer traverses a spatial region enclosed by multiple receivers. Figure 17 should be read as the error of localizing a source at the focal position for an interferer located at each plotted location (differentiated from the interpretation of Figure 14 by a difference in the colormap). A dominant trend in this plot is the increase in apparent source localization accuracy as the interferer moves in between any of the receivers and the target, as indicated by the lower error contours "stretching" toward the receivers. As with the analysis of the previous sound scene, it can be more informative to consider the areas in which the localization deteriorates. Higher error is observed in regions where the interferer approaches a receiver, but from a DoA that is misaligned with the target DoA. In such a case the energetic DDR for the nearby receiver will be dominated by the proximate interfering source, drawing the DoA estimates toward it, away from the target. This is especially apparent in the two-receiver configurations (left column), where DoA bias more strongly distorts the resulting estimate. It can also

be seen that the higher-order estimates mitigate the error by rejecting off-axis energy, indicated by the more shallow error contours with increasing order.

4.4 Estimation performance in reverberant conditions

Thus far, the sound field model used for analysis has been comprised of highly uncorrelated sources and diffuse field components. This approach is useful to capture idealized performance statistics, but does not account for important aspects of real sound fields, such as temporal dynamics and varying degrees of correlated interference. Reverberation is an important source of correlated interference in real spaces and presents various challenges for intensimetric sound field analysis. For example, a receiver placed near a reflecting surface will record the direct sound with a highly correlated reflection, slightly offset in time, from another direction. In the case that the paths traveled by the direct and reflected sound are similar in length, multiple sound wave arrivals may fall within the same analysis window causing a spatial and temporal averaging of that introduces error into the DoA estimation. Furthermore, the directional pattern of reflections will be anisotropic, challenging the diffuse sound field model assumed in the previous evaluations.

Early reverberant reflections will be spatially discrete and, when analyzed at a short enough time scale, can be localized as numerous individual sound sources before the mixing time of a room. This principle has been put to use processing room impulse responses (RIRs) to localize reflections [86], and even localize a source with a single SMA (DoA combined with range estimation) through tracing multipath propagation over known surface geometry [87], although practical applicability is yet to be evaluated. RIRs are, however, sparse by nature and complications arise with analysing continuous signals. Numerous TF-domain approaches have been developed to address the problem of DoA estimation in reverberant environments in which reflections are spectrally and temporally correlated with the emitting source. Well performing techniques center on performing estimation on a subset of frequency bins determined to be dominated by energy from the direct arrival path of a sound source, either through eigendecomposition of the spatial spectrum matrix or by directivity of the measured sound field. A useful overview of techniques can be found in [88]. As mentioned previously, the present work aims to evaluate broadband localization performance of sector-based approaches in order that it may inform the adaptation of such TF approaches and other estimation refinement techniques.

The preceding analyses have used isotropic noise as a proxy for the diffuse field. In what follows, and DoA estimation and localization will be evaluated for a reverberant sound field, which will have anisotropic and partially correlated diffuse characteristics, thereby violating the model in Eq. 68. The sound scene is modeled by a single broadband noise source in a "shoebox" room with dimensions 8 x 6 x 4 meters with a four-band reverberation time of $[400, 283, 200, 141]$ ms corresponding to three octave bands centered at $[125, 250, 500]$ Hz and fourth high-pass band above 1000 Hz.

Figure 18 shows time-series DoA estimates for a receiver located in the lower position of all the schematic figures, with a source moving along positions that are equidistant from the receiver and, in the case of order $L = 3$ analysis (lower row),

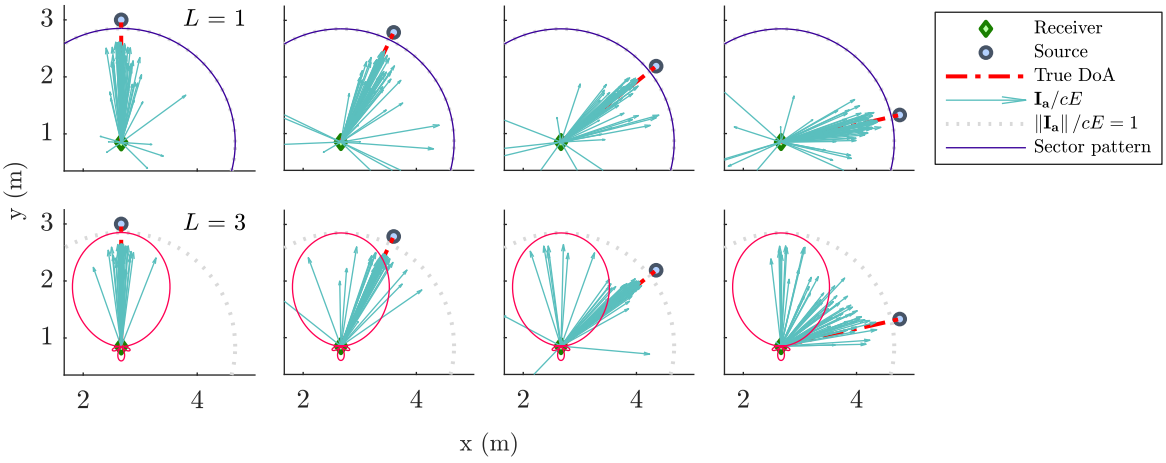


Figure 18: Time-series DoA estimates in reverberant conditions, performed at orders $L = [1, 3]$. The reverberant sound field is modeled by a single broadband noise source in a "shoebox" room with dimensions 8 x 6 x 4 meters and a four-band reverberation time of $[400, 283, 200, 141]$ ms corresponding to three octave bands centered at $[125, 250, 500]$ Hz and fourth high-pass band above 1000 Hz.

are increasingly off-axis from the sector. The presence of coherent reflections in the reverberant diffuse field can be seen in the errant DoA estimates which deviate from the target source (compare to the analysis of a source amid isotropic noise in Figure 4b). The magnitudes of these estimates can be quite large compared to those of the more accurate DoA estimates, indicating erroneous frame estimates.

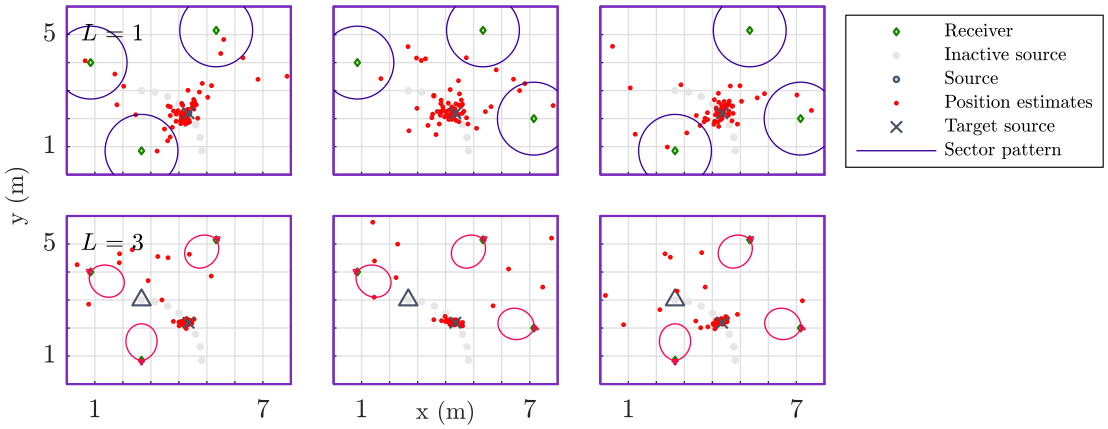


Figure 19: Time-series source position estimates in a reverberant sound field, performed at orders $L = [1, 3]$. The purple outline delineates the room boundaries. The reverberant conditions are the same as those of Figure 18.

Spatial filtering shows a clear advantage for on-axis source targets on account of the spatial filter rejecting lateral energy from reflections. However, this proves increasingly detrimental as the source moves off-axis and the direct sound is attenuated, while reflections for which α lies largely in the main lobe are relatively prominent. This results in a DoA spread that resembles low DDR conditions of Figure 4b or the

presence of an interfering source in Figure 7a. The erroneous DoA estimates then carry forward into the localization task, as is seen in Figure 19, where many DoA outliers lead to a high variance in the overall performance, and therefore require further post processing to smooth the results.

As with previous tests, source localization performance is evaluated for the full spatial region enclosed by the receivers, of varying count and analysis order. The results, shown in Figure 20, follow the general error contours of the isotropic noise case (compare to Figure 14), but with error boundaries that are more erratic. The regions of highest accuracy are considerably narrowed and do not consistently align with the focal position. However, despite the interference caused by the complex reflection pattern, anticipating best accuracy at the focal position (with an additional bias toward the centroid of the receiver arrangement) would still appear to be the best guiding design principle.

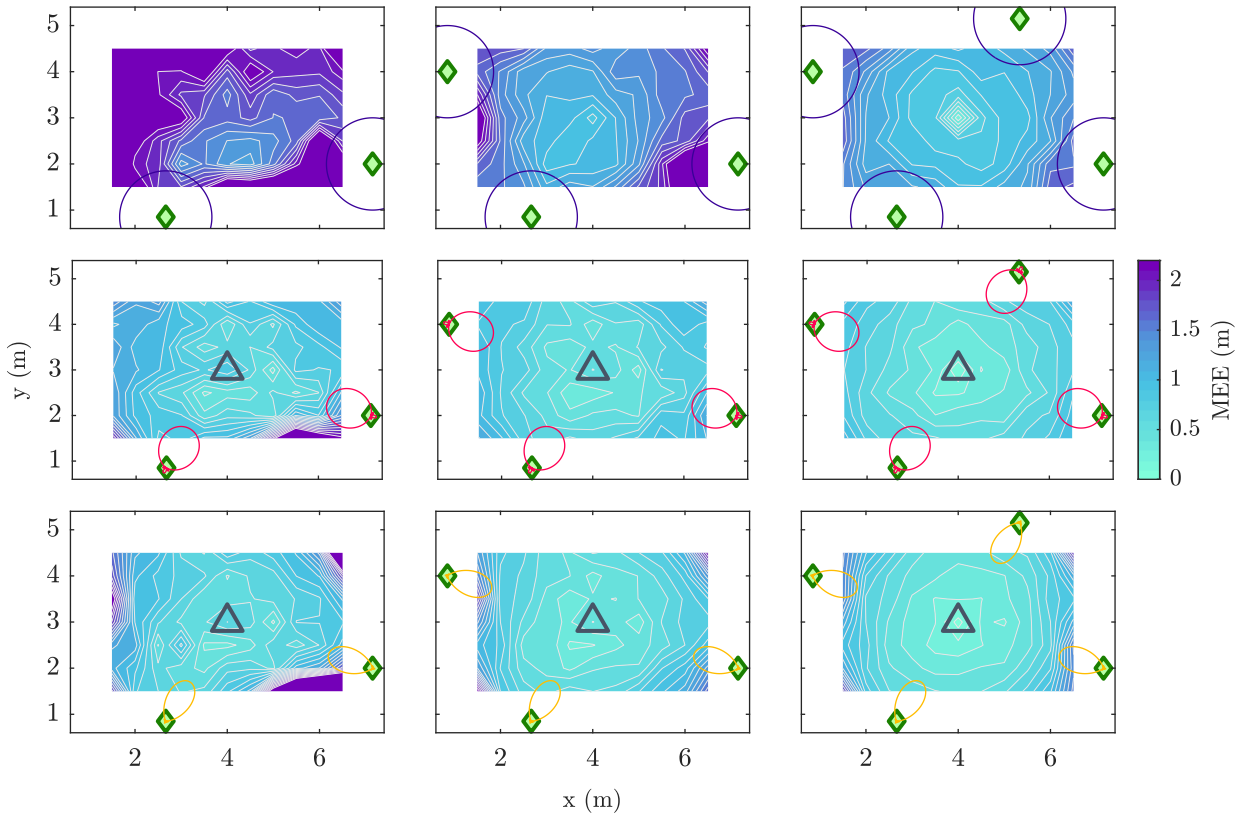


Figure 20: Mean estimation error (MEE) for single-source position estimation in reverberant conditions by two-, three- and four-receiver arrangements. Receivers are oriented toward the focal point (△), while the source moves throughout the plotted region. The reverberant conditions are the same as those of Figure 18.

5 Estimation refinement

Up to this point, the source localizing tests have made use of DoA estimates from two or more receivers per time frame, each contributing equally to the final source position estimation. However, numerous optimizations are possible, considering the analytical models describing the behavior of DoA estimation under specific sound field conditions (see Section 4.2) and considering the observations made in the preceding tests about the impact of the receiver configurations and source positions. In what follows, four optimizations are explored which make use of either estimated sound field quantities or initial position estimates that in turn inform a second (or more) iteration of the localization algorithm.

5.1 Refinement by distance weighting

It has be discussed previously that DoA estimates, and subsequently localization accuracy, improves as a sound source draws nearer to a receiver on account a higher DDR. It follows that as a source traverses a region of space monitored by multiple receivers, some of those receivers will have a better DoA estimate than others. This leads to an approach to modifying the localization algorithm to accepted weights which modify the contribution of each DoA to the resultant estimate. One such weighting scheme introduces "confidence" weights w_j for each DoA estimate into Eq. 84 as [84, §3.1]

$$\mathbf{Rp} = \mathbf{q}, \quad \text{where} \quad \mathbf{R} = \sum_{j=1}^{K} w_j(\mathbf{I} - \boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top), \quad \mathbf{q} = \sum_{j=1}^{K} w_j(\mathbf{I} - \boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top)\mathbf{r}_j. \tag{93}$$

The solution is found as in Eq. 84. The weighting scheme chosen for demonstration is

$$w_j = 1 - \frac{d_j}{d_{\max}}(1 - \vartheta) \quad \text{for } j \in [1, K],\ \vartheta \in [0, 1] \tag{94}$$

where d_j is the Euclidean distance between the estimated source position and the j-th receiver, and $d_{\max}$ is the maximum of the K distances. In this case ϑ represents a user-defined compression of the normalized range of weights applied to the corresponding DoA estimate, i.e. $\vartheta = 0$ would entirely diminish the contribute of the DoA estimate by the farthest receiver from the estimated source position, and $\vartheta = 1$ would apply equal weighting to all DoAs, regardless of the estimated source distance.

Results for localization with distance weighting for a single-source sound scene are shown in Figure 21. Comparing the left and right columns shows that the accuracy is increased for a larger region, particularly extending into areas between adjacent receivers, as the estimates from those nearer receivers are given a higher confidence weight. Note that this result would not hold, and in fact could be made worse, in the presence of an interferer because the initial source position estimate will be drawn toward the interferer, thereby erroneously applying higher weights to receivers farther from the target. That is to say, this method would be most useful for scenarios (or individual time windows) dominated by a single source.

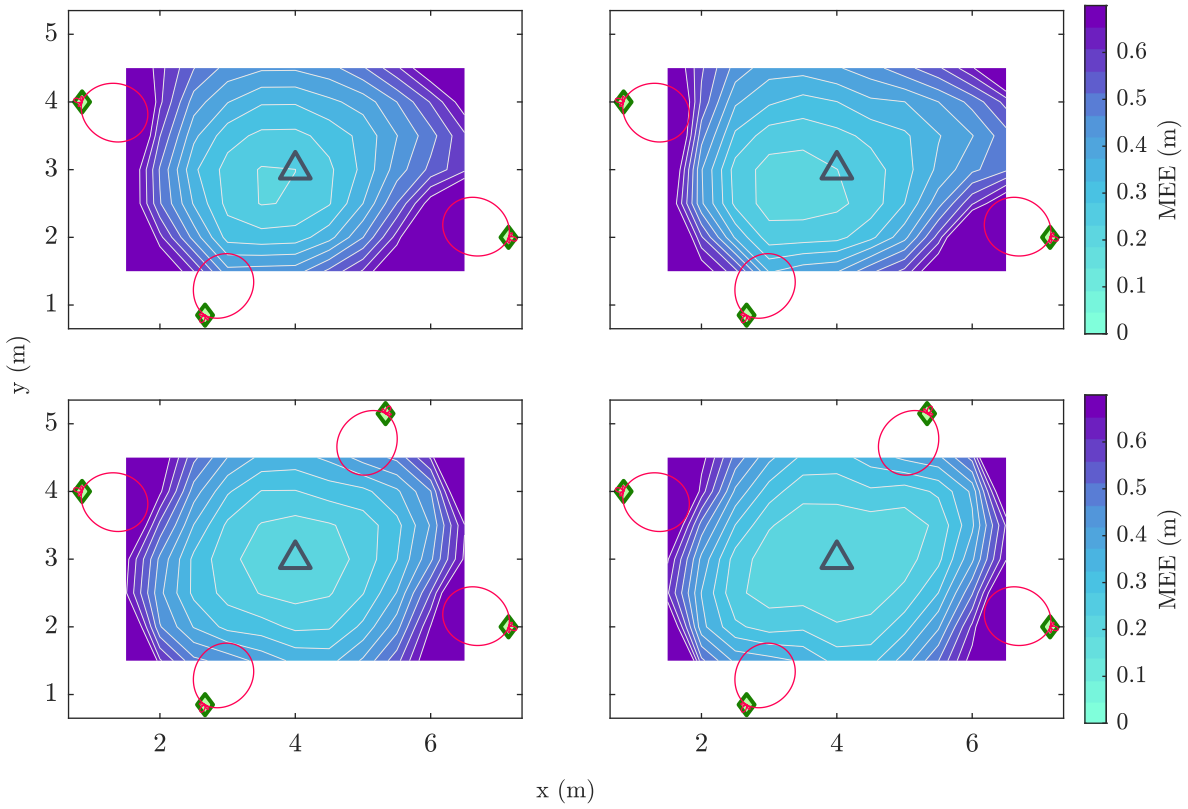


Figure 21: Mean estimation error (MEE) for single-source localization with estimation refinement by distance weighting. Results are shown for three- and four-receiver arrangements (top and bottom rows, respectively) at analysis order $L = 3$. Left panels show no weighting and the right shows the weighted estimates with the weight-scaling parameter $\vartheta_{\min} = 0.2$. Receivers are oriented toward the focal point ($\triangle$), while the source moves throughout the plotted region. The scene DDR is $\Gamma = 5$ dB.

5.2 Refinement by diffuseness

For scenarios in which it is known that an interfering source or high DDR is distorting localization of a target source, the diffuseness estimation can offer improved localization accuracy. Using the same "confidence" weighting scheme introduced previously, diffuseness can be used to penalize those DoA estimates which have high corresponding diffuseness, thereby emphasizing estimates with a higher degree of directional energy transport. A simple weighting scheme is chosen here as

$$w_j = \max(\psi_j - \vartheta_{\min}, \epsilon) \tag{95}$$

where $\vartheta_{\min}$ is a minimum expected diffuseness and should be set accounting for the changing diffuseness bounds with sector order governed by Eq. 89 (refer to Figure 9), though would ultimately be tuned for the application. The $\max(\cdot)$ operation ensures the weighted contribution remains positive non-zero clipped at the value $\epsilon > 0$, which can also serve as a minimum contribution bound. Further normalization can be performed by the maximum of all estimate weights within one frame or a running maximum across multiple time frames, depending on the nature of the source signal.

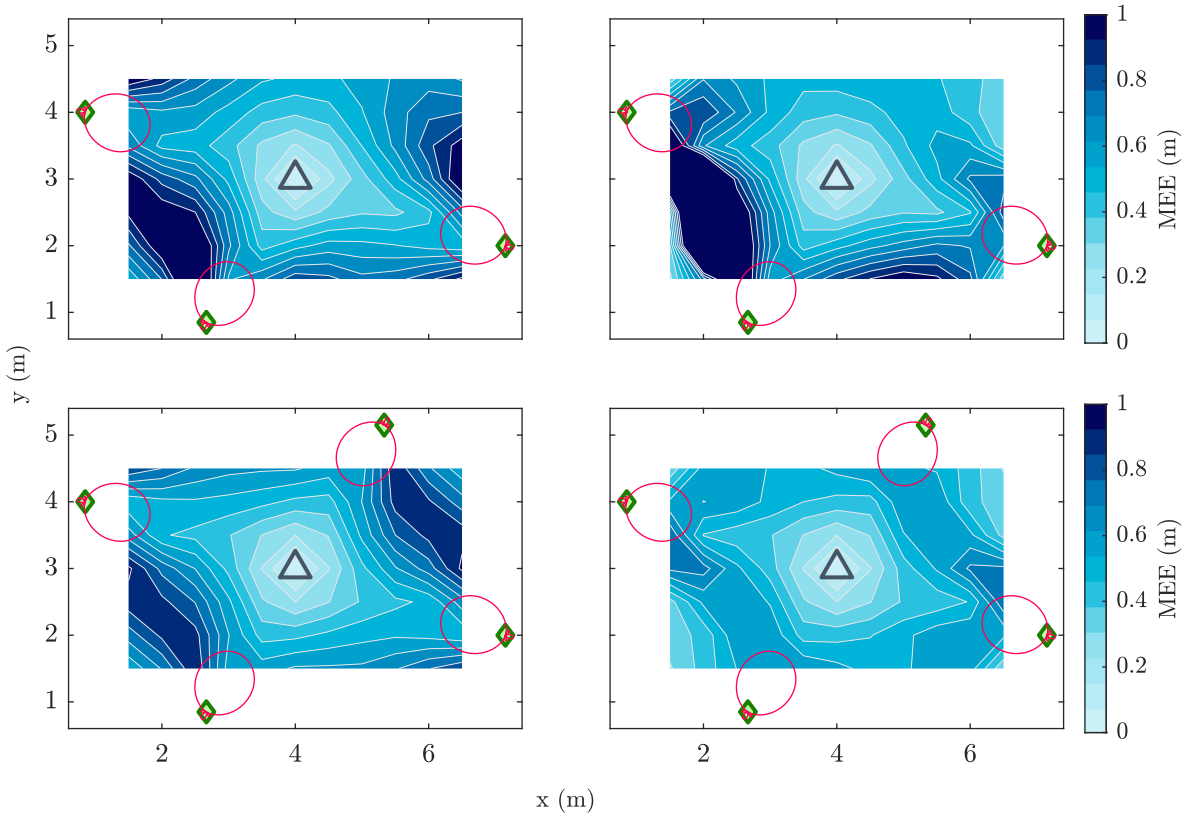


Figure 22: Mean estimation error (MEE) for source localization in the presence of an interferer with estimation refinement by diffuseness weighting. Results are shown for three- and four-receiver arrangements (top and bottom rows, respectively) at analysis order $L = 3$. Left panels show no weighting and the right shows the weighted estimates with the weight-scaling parameter $\vartheta_{\text{min}} = 0.2$. Receivers are oriented toward the focal point ($\triangle$), while the source moves throughout the plotted region. The scene DDR is $\Gamma = 10$ dB.

Similar to the test presented in Section 4.3.2, Figure 22 shows the localization accuracy for a target source at the focal position as an interferer traverses the monitored region. The right panels show the results after diffuseness weighting has been applied, showing improved performance particularly in regions which suffered most from the presence of the interfering source. Note, for example, regions between adjacent receivers and where the interferer is proximate but off-axes to a receiver. As with prior interferer accuracy plots, the peak accuracy at the focal position is naturally high on account if the interferer coinciding with the target source, and should not be interpreted as a measure of source separation.

5.3 Refinement by DoA bias transfer function

As discussed in Section 4.2 a bias is introduced into the DoA estimation as a result of spatial filtering amid a diffuse field, as a function of the sector order, DDR and source AoA. Estimation refinement can be performed by compensating for this bias in the case that the conditions of the underlying model are satisfied; namely that there is

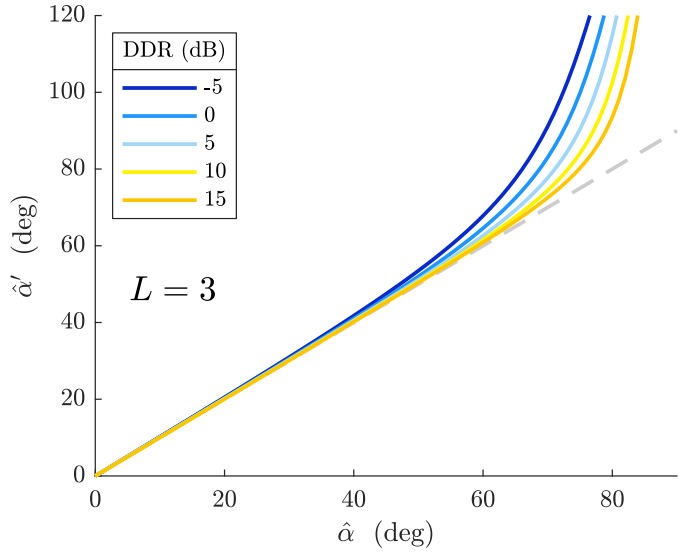


Figure 23: Diffuse field bias compensation transfer function, derived from Eq. 92 for analysis order $L = 3$. The angle-of-arrival estimate $\hat{\alpha}$ is shown on the x-axis, and the bias-compensated estimate $\hat{\alpha}'$, which is subsequently supplied for a second iteration to the localization algorithm, is on the y-axis.

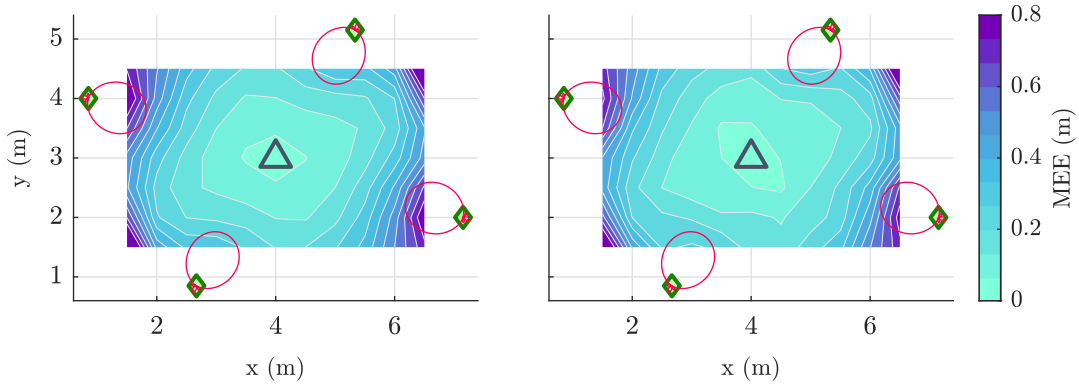


Figure 24: Mean estimation error (MEE) for single-source localization with estimation refinement by diffuse field bias compensation. Results are shown for a four-receiver arrangement without bias compensation in the left panel and with compensation applied on the right. The analysis order is $L = 3$ and the DDR is $\Gamma = 5$ dB. Receivers are oriented toward the focal point ($\triangle$), while the source moves throughout the plotted region.

a relatively isotropic diffuse field, a single sound source and a known or estimated DDR. The bias compensation corresponds to a transfer function derived from Eq. 92, shown in Figure 23 for analysis order order $L = 3$ and varying DDRs. The plot shows that for reasonably high DDR values, little compensation is needed until the estimated source AoA, denoted $\hat{\alpha}$, veers far from the sector pattern maximum. The plot is trimmed both for visibility and because in practice a source estimated to be very far off-axis from the sector heading would reasonably be discarded and deferred to a neighboring sector that aligns more closely with its heading axis.

The bias compensation function is applied in simulation for a sound scene containing a single source and the result is shown in Figure 24. As the monitored region falls largely within a shallow AoA to the analysis sectors, the magnitude of improvement is modest, though the improvement spreads appreciably over the area of coverage.

5.4 Toward multisource localization by spatial sweep with diffuseness constraint

Diffuseness is a primary derived feature of intensimetric sound field analysis. It has been used up to this point to assess the ability of a sector pattern to reduce off-axis interfering sound, while a target source is presumed to be nearly on-axis. The diffuseness estimate has been shown to be sensitive to the DDR, to the target source AoA, and to the AoA and proximity of interfering sources. As such, it can be a useful tool for source *detection*.

In order to detect one or more sources, multiple analysis sectors, each with a different orientation and producing a unique DoA estimate, may be used within an analysis time frame. This will be here referred to as a "spatial sweep", as the approach can be thought of conceptually as sweeping a sector through a region of interest. The method is flexible in that any spherical region (cap, segment, path, etc.) can be analyzed, with a sampling density appropriate to the sound scene composition and analysis order (such as spherical coverings of appropriate density or paths based on the Rayleigh criteria of Eq. 58). In the present case, sources will be assumed to be confined to the horizontal plane, so a spatial sweep will be conducted from each receiver to cover a 180° range oriented away from the nearby wall. The spatial sampling interval will be over-sampled for illustrative purposes in 3° steps.

Once analyses are produced for each sector in the sweep, the diffuseness measure can be used to cull the corresponding DoA estimates. Recalling that a spatial filter applied to the pressure and velocity components of a sound field will alter the diffuseness estimate (see Section 4.2), order-dependent bounds need to be established. Equation 91 defines the expected diffuseness for a sound field with a given DDR and a source arriving on-axis with the sector. Assuming the DDR is known or estimated, this can serve as the minimum diffuseness, i.e. the best-case value that can be expected for detecting a sound source. Equation 89, on the other hand, defines the diffuseness that can be expected for a sector peering into a purely diffuse field, in the absence of a sound source (refer to the shifting maximum diffuseness by sector order of Figure 9). These two values then serve as reasonable bounds within which DoA estimates can be culled according to a diffuseness threshold

$$\psi_< = \vartheta(\psi_{\mathrm{df,w}} - \psi_{\mathrm{dp,w}}) + \psi_{\mathrm{dp,w}} \tag{96}$$

where $\vartheta \in [0, 1]$ is a user-defined, normalized threshold below which DoA estimates are kept. Given a set of DoA estimates $D = \{\Omega_r\}_R$ for R observations in the spatial sweep, the subset selected for subsequent source localization is

$$D_< = \{\Omega_r \mid \Omega_r \in D,\ \hat{\psi}_r < \psi_<\} \tag{97}$$

where $\hat{\psi}_r$ is the diffuseness estimate for the r-th observation.

Figure 25 illustrates the DoA culling process for a scene comprised of three sound sources, their DoAs indicated by dashed lines of different color, with the threshold $\vartheta = 0.25$. Note the differing diffuseness scales of the panels in the top row on account of the diffuseness range being compressed with increasing order. This illustrates the utility of formulating a normalized diffuseness threshold value for consistency across analysis orders.

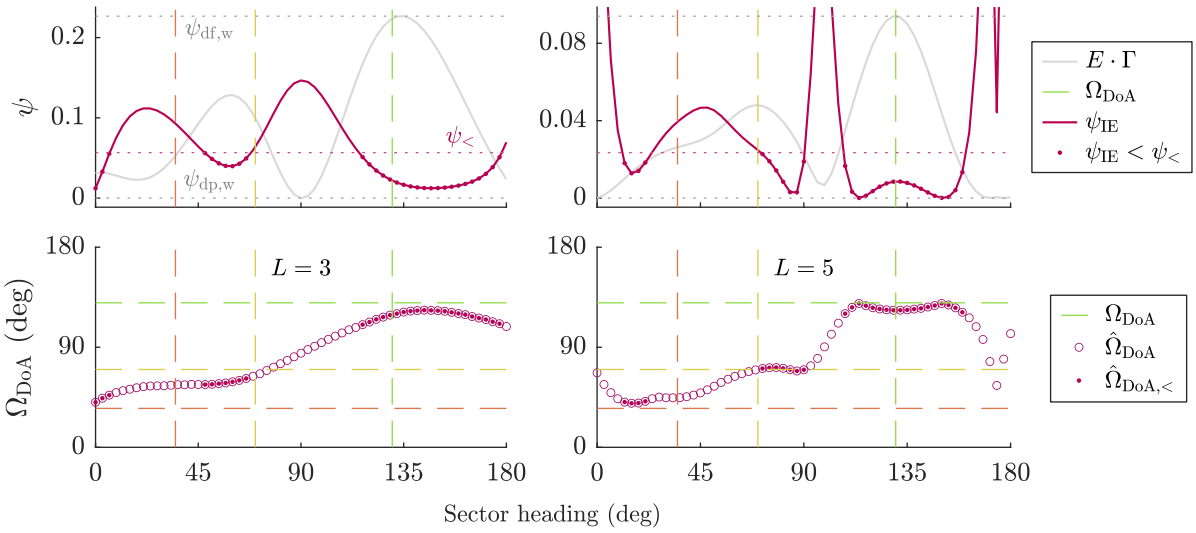


Figure 25: DoA estimation culling by diffuseness constraint for observations made over a spatial sweep. An analysis sector is "swept" across a 180° range on the azimuthal plane in steps of 3°. Left and right columns correspond to analysis orders $L = [3, 5]$, respectively. **Top** panels show the diffuseness estimate contours across the sweep (maroon line). Dotted lines ($\cdots$) indicate the diffuseness bounds within which the diffuseness threshold is selected according to Eq. 96, and $\vartheta = 0.25$ (note the different y-axis scales). Grey lines indicate the sector energy contour, normalized across all observation in D such that $E \in [\psi_{\mathrm{df,w}}, \psi_{\mathrm{dp,w}}]$ for display purposes. **Bottom** panels show the corresponding DoA estimates, where x- and y-axes are azimuthal directions of the sector heading and source DoA, respectively. True source DoAs are marked with dashed lines ($--$), individuated by color. DoA estimates are retained for which the diffuseness measure $\hat{\psi} < \psi_<$, highlighted by points ($\bullet$) in all panels.

The diffuseness metric proves to be a useful culling strategy in this example, as the selected DoA estimates are observed to be reasonably accurate, and the accuracy improves with analysis order. Furthermore, the estimates of highest accuracy *do not* correspond to those sectors which are oriented directly toward the source. This indicates that for a sound field composed of multiple sources, the complex interactions among the sources and the beam pattern (side lobes in particular) are not easily anticipated. The diffuseness constraint therefore offers improved DoA estimation over an energy constraint—note the grey line of the sound field directional energy contour—the peaks of which either obscure some sources or are misaligned with the source DoAs.

6 Conclusion and outlook

This thesis has presented an evaluation of high-order sector-based analysis of sound fields with distributed spherical harmonic receivers, in the context of 3D source localization. After a review of intensimetric sound field analysis using first-order spherical harmonic (SH) signals, including intensity-based diffuseness estimators, spatially-localized analysis of high-order SH signals by sector-based sound field decomposition was revisited. Through simulation of acoustic scenes comprised of fundamental sound field components, performance characteristics were evaluated for source direction-of-arrival (DoA) estimation using ideal SH receivers. Specific conditions were examined in detail: the impacts of varying direct-to-diffuse energy ratios (DDR), temporal averaging of intensity measures, and source angle-of-arrival (AoA). The influence of an interfering sound source at varying proximity to both the target and the receiver was also assessed. Reverberant conditions were simulated to consider the performance of intensimetric analysis in non-ideal diffuse circumstances. The latter two conditions were of interest for challenging the underlying simple sound field model of an isotropic diffuse field and plane wave mixture. These assessments established a basis for evaluating a 3D source localization scheme which employed DoA estimates derived from analyses across multiple distributed SH receivers. The advantages and caveats of sector-based sound field analysis were detailed for each acoustic condition.

Across all evaluations, it is clear that a primary advantage of sector-based analysis in source localization is the improvement of the DDR for a target source signal (or similarly, the direct-to-interferer ratio), an effect which compounds when using multiple receivers. Through the rejection of diffuse or otherwise interfering energy that is off-axis to the sector's spatial filter, the proportion of energy from the target source contributing to the intensimetric analysis—from which the DoA and diffuseness estimates derive—is maximized. However, in the single-source case the improvements in DoA estimation by higher-order processing diminishes above a threshold angle-of-arrival which decreases with increasing order, contingent upon the DDR. Source localization performance across multiple receivers reflected many of the constraints of successful DoA estimation with regard to the sector pattern's interactions with the DDR and source AoA. The region of highest localization accuracy consistently anchors to the focal position of the sectors and further draws toward the centroid of the receiver arrangement where projected DoAs would be maximally orthogonal. In cases cases where DoA estimates across receivers are collinear, estimation becomes erratic but is largely resolved by an additional receiver. Finally, source proximity has a considerable impact on 3D localization. In particular, *interferer* proximity can highly impact the DoA estimate of the nearby receiver, skewing subsequent localization results. In such a case, a higher receiver count is advantageous to minimize the impact of the erroneous contribution to the ensemble of estimates.

The evaluation results, along with consideration of analytical expressions describing the behavior of the diffuseness metric under a spatially-filtered sound field, led to the proposal of various refinement strategies. Recognizing that not all receivers will be optimally positioned, their contribution to the localization estimate can be weighted

by their corresponding diffuseness measure, or by their distance from or orientation to an initial source position estimate. Such techniques show improvements but are sensitive to sound field conditions. Finally, a multisector "spatial sweep" technique is used in combination with a diffuseness constraint used to cull DoA estimates that can serve multiple-source localization. The spatial sweep also revealed that, in a complex acoustic scene, the DoA estimates of highest accuracy *do not* necessarily correspond to those sectors which are oriented directly toward the source, suggesting that redundant spatial coverage by multiple sectors offers increased accuracy.

Sector-based intensimetric analysis shows clear advantages over traditional first-order methods in complex acoustic scenarios. The method shows promise in a variety of applications, especially those with multiple concurrent sources or in high-noise environments. Opportunities remain to incorporate optimization schemes, such as iterative estimation weight refinement and source subtraction, integration with subspace methods, and cross-sector and cross-order correlation metrics. Other intensimetric analysis, such as numerous reactive field indicators, can be incorporated into the technique to assist in sound field characterization. Receiver and sector arrangements should be further explored, along with other beamforming and nulling techniques, to provide robust source localization and signal extraction. Further progress in these areas would inform parametric sound field reproduction, creative or restorative sound scene processing, and object-based transmission and reconstruction, all of which serve the emerging paradigm of sound field synthesis for fully mobile 6DoF navigation.

References

[1] B. Rafaely, *Fundamentals of Spherical Array Processing*, 2nd ed., ser. Springer Topics in Signal Processing. Springer International Publishing, 2019, vol. 16. [Online]. Available: http://link.springer.com/10.1007/978-3-319-99561-8

[2] J. Meyer and G. W. Elko, "Audio Signal Processing for Next-Generation Multimedia Communication Systems," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Boston, MA, USA: Kluwer Academic Publishers, 2004, ch. 3, pp. 67–90. [Online]. Available: http://link.springer.com/10.1007/b117685

[3] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.

[4] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *Journal of the Audio Engineering Society*, vol. 68, no. 6, pp. 428–440, Jul. 2020. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20859

[5] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. J. Tashev, "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, United Kingdom: IEEE, May 2019, pp. 261–265. [Online]. Available: https://ieeexplore.ieee.org/document/8683751/

[6] T. McKenzie, D. Murphy, and G. Kearney, "Diffuse-Field Equalisation of Binaural Ambisonic Rendering," *Applied Sciences*, vol. 8, no. 10, Oct. 2018. [Online]. Available: http://www.mdpi.com/2076-3417/8/10/1956

[7] A. Politis, M. R. P. Thomas, H. Gamper, and I. J. Tashev, "Applications of 3D spherical transforms to personalization of head-related transfer functions," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai: IEEE, Mar. 2016, pp. 306–310. [Online]. Available: http://ieeexplore.ieee.org/document/7471686/

[8] D. R. Begault, A. S. Lee, E. M. Wenzel, and M. R. Anderson, "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," in *Proc. Audio Engineering Society 108th Conv.*, Paris, France, Feb. 2000.

[9] B. Bernschütz, "Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording," Ph.D. dissertation, Technische Universität Berlin, Berlin, Germany, 2016.

[10] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, Jan. 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1369318/

[11] J. Ahrens and C. Andersson, "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," *Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2783–2794, Apr. 2019. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.5096164

[12] F. Zotter, "Analysis and Synthesis of Sound-Radiation with Spherical Arrays," Ph.D. dissertation, University of Music and Performing Arts, Austria, Graz, Austria, Sep. 2009.

[13] F. Schultz and S. Spors, "Data-Based Binaural Synthesis Including Rotational and Translatory Head-Movements," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, Sep. 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16894

[14] F. Winter, F. Schultz, and S. Spors, "Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements," in *7th EAA Forum Acusticum*, Kraków, Poland, Sep. 2014.

[15] J. G. Tylka and E. Choueiri, "Comparison of Techniques for Binaural Navigation of Higher-Order Ambisonic Soundfields," in *Audio Engineering Society Convention 139*, Oct. 2015. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17977

[16] E. Fernandez-Grande and T. Walton, "Reconstruction of sound fields with a spherical microphone array," in *Inter-noise 2014*, Melbourne, Australia, Nov. 2014.

[17] V. Pulkki, S. Delikaris-Manias, and A. Politis, *Parametric time-frequency domain spatial audio*, 1st ed. Hoboken, New Jersey, USA: John Wiley & Sons, Ltd., 2018.

[18] K. Kowalczyk, O. Thiergart, M. Taseska, G. Del Galdo, V. Pulkki, and E. A. P. Habets, "Parametric Spatial Sound Processing: A flexible and efficient solution to sound scene acquisition, modification, and reproduction," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 31–42, Mar. 2015. [Online]. Available: https://ieeexplore.ieee.org/document/7038281

[19] A. Politis, "Microphone array processing for parametric spatial audio techniques," Ph.D. dissertation, Aalto University, Helsinki, Finland, 2016.

[20] S. Delikaris-Manias, "Parametric spatial audio processing utilising compact microphone arrays," Ph.D. dissertation, Aalto University, Espoo, Finland, 2017.

[21] G. Schiffrer and D. Stanzial, "Energetic properties of acoustic fields," *Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3645–3653, Dec. 1994. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.410585

[22] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *18th European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010.

[23] J. Merimaa, "Analysis, Synthesis, and Perception of Spatial Sound - Binaural Localization Modeling and Multichannel Loudspeaker Reproduction," Ph.D. dissertation, Helsinki University of Technology, Espoo, Finland, 2006.

[24] G. Del Galdo, M. Taseska, O. Thiergart, J. Ahonen, and V. Pulkki, "The diffuse sound field in energetic analysis," *Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2141–2151, Mar. 2012. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.3682064

[25] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, Jun. 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14170

[26] J. Merimaa and V. Pulkki, "Spatial impulse response rendering i: Analysis and synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 12, pp. 1115–1127, 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13401

[27] M.-V. Laitinen, "Binaural Reproduction for Directional Audio Coding," Master's thesis, Helsinki University of Technology, Helsinki, Finland, 2008.

[28] S. Berge and N. Barrett, "High Angular Resolution Planewave Expansion," in *2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, France, May 2010.

[29] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall, "Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution," *Journal of the Audio Engineering Society*, vol. 68, no. 5, pp. 338–354, Jun. 2020. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20852

[30] A. Politis and V. Pulkki, "Higher-Order Directional Audio Coding," in *Parametric Time-Frequency Domain Spatial Audio*. John Wiley & Sons, Ltd, 2018, ch. 6, pp. 141–159. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119252634.ch6

[31] A. Politis, L. McCormack, and V. Pulkki, "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, NY, USA: IEEE, Oct. 2017, pp. 379–383. [Online]. Available: http://ieeexplore.ieee.org/document/8170059/

[32] A. Politis, S. Tervo, and V. Pulkki, "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, AB: IEEE, Apr. 2018, pp. 6802–6806. [Online]. Available: https://ieeexplore.ieee.org/document/8462608/

[33] T. Pihlajamäki and V. Pulkki, "Projecting simulated or recorded spatial sound onto 3d-surfaces," in *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio*, Mar 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16198

[34] A. Politis, T. Pihlajamäki, and V. Pulkki, "Parametric spatial audio effects," in *15th International Conference on Digital Audio Effects (DAFx-12)*, York, United Kingdom, Sep. 2012.

[35] A. Plinge, S. J. Schlecht, O. Thiergart, T. Robotham, O. Rummukainen, and E. A. P. Habets, "Six-Degrees-of-Freedom Binaural Audio Reproduction of First-Order Ambisonics with Distance Information," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Redmond, WA, USA, Aug. 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19684

[36] M. Kentgens, A. Behler, and P. Jax, "Translation of a Higher Order Ambisonics Sound Scene Based on Parametric Decomposition," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, May 2020, pp. 151–155. [Online]. Available: https://ieeexplore.ieee.org/document/9054414/

[37] L. McCormack, A. Politis, and V. Pulkki, "Parametric spatial audio effects based on the multi-directional decomposition of Ambisonic sound scenes," in *23rd International Conference on Digital Audio Effects (DAFx2020)*, Vienna, Austria, Sep. 2021.

[38] A. Southern, J. Wells, and D. Murphy, "Rendering walk-through auralisations using wave-based acoustical models," in *17th European Signal Processing Conference*, Glasgow, Scotland, Aug. 2009, pp. 715–719.

[39] J. G. Tylka and E. Y. Choueiri, "Soundfield Navigation using an Array of Higher-Order Ambisonics Microphones," in *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*, Los Angeles, CA, USA, Sep. 2016.

[40] E. Stein and M. M. Goodwin, "Ambisonics Depth Extensions for Six Degrees of Freedom," in *Audio Engineering Society Conference: 2019 AES International Conference on Headphone Technology*, Aug. 2019. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20514

[41] N. Hahn and S. Spors, "Physical properties of modal beamforming in the context of data-based sound reproduction," in *Audio Engineering Society Convention 139*, Oct 2015. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18024

[42] C. Schörkhuber, R. Holdrich, and F. Zotter, "Triplet-based variable-perspective (6DoF) audio rendering from simultaneous surround recordings taken at multiple perspectives," in *Fortschritte der Akustik (DAGA)*, Hanover, Germany, Mar. 2020.

[43] X. Zheng, "Soundfield navigation: Separation, compression and transmission," Ph.D. dissertation, University of Wollongong, 2013.

[44] D. Pavlidi, S. Delikaris-Manias, V. Pulkki, and A. Mouchtaris, "3D DOA estimation of multiple sound sources based on spatially constrained beamforming driven by intensity vectors," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai, China: IEEE, Mar. 2016, pp. 96–100. [Online]. Available: http://ieeexplore.ieee.org/document/7471644/

[45] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, Aug. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7065255/

[46] A. Politis and V. Pulkki, "Acoustic intensity, energy-density and diffuseness estimation in a directionally-constrained region," arXiv:1609.03409, Tech. Rep., Sep. 2016. [Online]. Available: https://arxiv.org/abs/1609.03409v2

[47] F. Fahy, *Sound Intensity*, 2nd ed. London, England: E & FN Spon, 1995.

[48] ——, *Foundations of Engineering Acoustics*. San Diego, CA, USA: Academic, 2001.

[49] J. A. Mann, J. Tichy, and A. J. Romano, "Instantaneous and time-averaged energy transfer in acoustic fields," *Journal of the Acoustical Society of America*, vol. 82, no. 1, pp. 17–30, Jul. 1987. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.395562

[50] F. Jacobsen, "Active and reactive sound intensity in a reverberant sound field," *Journal of Sound and Vibration*, vol. 143, no. 2, pp. 231–240, Dec. 1990. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0022460X9090952V

[51] D. Stanzial, N. Prodi, and G. Schiffrer, "Reactive acoustic intensity for general fields and energy polarization," *Journal of the Acoustical Society of America*, vol. 99, no. 4, pp. 1868–1876, Apr. 1996. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.415369

[52] D. Stanzial and N. Prodi, "Measurements of newly defined intensimetric quantities and their physical interpretation," *Journal of the Acoustical Society*

of America, vol. 102, no. 4, pp. 2033–2039, Oct. 1997. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.419693

[53] F. Jacobsen, "A note on instantaneous and time-averaged active and reactive sound intensity," *Journal of Sound and Vibration*, vol. 147, no. 3, pp. 489–496, Jun. 1991. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0022460X91904967

[54] R. C. Heyser, "Instantaneous intensity," *Journal of the Acoustical Society of America*, vol. 80, no. S1, pp. S103–S103, Dec. 1986. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.2023514

[55] J. Vilkamo and T. Bäckström, "Time–frequency processing: Methods and tools," in *Parametric time-frequency domain spatial audio*, 1st ed., V. Pulkki, S. Delikaris-Manias, and A. Politis, Eds. John Wiley & Sons, Ltd., 2018, ch. 1, pp. 3–24.

[56] D. Romblom, "Diffuse Field Modeling: The Physical and Perceptual Properties of Spatialized Reverberation," Ph.D. dissertation, McGill University, Montréal, Quebec, Aug. 2016.

[57] F. Jacobsen, "The diffuse sound field," The Acoustic Laboratory, Technical University of Denmark, Ph.D. dissertation Report number 27, 1979.

[58] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," *Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 1980–1991, 2002.

[59] J. Ahonen and V. Pulkki, "Diffuseness estimation using temporal variation of intensity vectors," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, NY, USA: IEEE, Oct. 2009, pp. 285–288. [Online]. Available: http://ieeexplore.ieee.org/document/5346496/

[60] A. Politis, S. Delikaris-Manias, and V. Pulkki, "Direction-of-arrival and diffuseness estimation above spatial aliasing for symmetrical directional microphone arrays," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. South Brisbane, Queensland, Australia: IEEE, Apr. 2015, pp. 6–10. [Online]. Available: http://ieeexplore.ieee.org/document/7177921/

[61] A. Politis, "Spherical array processing MATLAB library," 2016, MATLAB library. [Online]. Available: https://github.com/polarch/Spherical-Array-Processing

[62] N. Epain and C. T. Jin, "Spherical Harmonic Signal Covariance and Sound Field Diffuseness," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1796–1807, Oct. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7501622/

[63] S. Delikaris-Manias, J. Vilkamo, and V. Pulkki, "Parametric binaural rendering utilizing compact microphone arrays," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. South Brisbane, Queensland, Australia: IEEE, Apr. 2015, pp. 629–633. [Online]. Available: http://ieeexplore.ieee.org/document/7178045/

[64] J. Vilkamo, "Optimized Covariance Domain Framework for Time-Frequency Processing of Spatial Audio," *Journal of the Audio Engineering Society*, vol. 61, no. 6, 2013.

[65] J. Vilkamo and S. Delikaris-Manias, "Perceptual Reproduction of Spatial Sound Using Loudspeaker-Signal-Domain Parametrization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1660–1669, Oct. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7122278/

[66] M. Chapman, "Symmetries of Spherical Harmonics: applications to ambisonics," in *Symmetries of Spherical Harmonics: applications to ambisonics*, Graz, Austria, Jun. 2009.

[67] E. G. E. G. Williams, *Fourier acoustics sound radiation and nearfield acoustical holography*, 1st ed. San Diego, CA, USA: Academic Press, 1999.

[68] R. H. Hardin and N. J. A. Sloane, "McLaren's improved snub cube and other new spherical designs in three dimensions," *Discrete & Computational Geometry*, vol. 15, no. 4, pp. 429–441, Apr. 1996. [Online]. Available: http://link.springer.com/10.1007/BF02711518

[69] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 807–820, 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16554

[70] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," in *Audio Engineering Society Convention 105*, Sep. 1998. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8385

[71] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.1792643

[72] Z. Li and R. Duraiswami, "Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 2, pp. 702–714, Feb. 2007. [Online]. Available: http://ieeexplore.ieee.org/document/4067056/

[73] M. A. Gerzon, "General Metatheory of Auditory Localisation," in *Audio Engineering Society Convention 92*, Mar. 1992. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=6827

[74] A. Heller, E. Benjamin, and R. Lee, "Design of Ambisonic Decoders for Irregular Arrays of Loudspeakers by Non-Linear Optimization," in *Preceedings of the 129th Convention of the Audio Engineering Society*, San Francisco, CA, USA, Nov. 2010.

[75] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone," in *Audio Engineering Society Convention 120*, 2006.

[76] S. Delikaris-Manias, D. Pavlidi, V. Pulkki, and A. Mouchtaris, "3D localization of multiple audio sources utilizing 2D DOA histograms," in *2016 24th European Signal Processing Conference (EUSIPCO)*. Budapest, Hungary: IEEE, Aug. 2016, pp. 1473–1477. [Online]. Available: http://ieeexplore.ieee.org/document/7760493/

[77] C. Schörkhuber, P. Hack, M. Zaunschirm, F. Zotter, and A. Sontacchi, "Localization of Multiple Acoustic Sources with a Distributed Array of Unsynchronized First-Order Ambisonics Microphones," in *Proceedings of the 6th Congress of Alps-Adria Acoustics Assosiation*, Graz, Austria, Oct. 2014.

[78] D. Pavlidi, S. Delikaris-Manias, V. Pulkki, and A. Mouchtaris, "3D localization of multiple sound sources with intensity vector estimates in single source zones," in *23rd European Signal Processing Conference (EUSIPCO)*. Nice, France: IEEE, Aug. 2015, pp. 1556–1560. [Online]. Available: http://ieeexplore.ieee.org/document/7362645/

[79] A. Politis, S. Tervo, T. Lokki, and V. Pulkki, "Parametric Multidirectional Decomposition of Microphone Recordings for Broadband High-Order Ambisonic Encoding," in *Audio Engineering Society Convention 144*, May 2018. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19457

[80] A. Politis and V. Pulkki, "Broadband Analysis and Synthesis for Directional Audio Coding Using A-Format Input Signals," in *Audio Engineering Society Convention 131*, Oct. 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16051

[81] A. Politis, "Shoebox roomsim MATLAB library," 2015, MATLAB library. [Online]. Available: https://github.com/polarch/shoebox-roomsim

[82] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.382599

[83] Wikipedia contributors, "Line–line intersection — Wikipedia, the free encyclopedia," 2021, [Online; accessed 24-June-2021]. [Online]. Available: https://en.wikipedia.org/wiki/Line%E2%80%93line_intersection#General_derivation

[84] J. Traa, "Least-Squares Intersection of Lines," 2013, [Online; accessed 15-January-2021]. [Online]. Available: https://silo.tips/download/least-squares-intersection-of-lines

[85] A. Eikenes, "lineintersect3d," 2012, MATLAB function. [Online]. Available: https://se.mathworks.com/MATLABcentral/fileexchange/37192-intersection-point-of-lines-in-3d-space

[86] S. Tervo and A. Politis, "Direction of Arrival Estimation of Reflections from Room Impulse Responses Using a Spherical Microphone Array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1539–1551, Oct. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7115080/

[87] J. Daniel and S. Kitic, "Time Domain Velocity Vector for Retracing the Multipath Propagation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, May 2020, pp. 421–425. [Online]. Available: https://ieeexplore.ieee.org/document/9054561/

[88] L. Madmoni and B. Rafaely, "Direction of Arrival Estimation for Reverberant Speech Based on Enhanced Decomposition of the Direct Sound," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 131–142, Mar. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8570815/