Analysis, Treatment, and Manipulation Methods for Spatial Room Impulse Responses Measured with Spherical Microphone Arrays

Pierre Massé

Analysis, Treatment, and Manipulation Methods for Spatial Room Impulse Responses Measured with Spherical Microphone Arrays

Thesis Dissertation
Doctorate in Acoustics and Signal Processing

—

École doctorale Informatique, Télécommunications, Électronique (EDITE) de Paris
Sorbonne Université

Pierre Massé

—

Acoustic and Cognitive Spaces team, STMS
Ircam–Sorbonne Université–CNRS–Ministère de la Culture

Supervisors: Gérard Assayag and Markus Noisternig

—

Manuscript reviewers: Michael Vorländer and Ville Pulkki
Defence committee: Roland Badeau, Efren Fernandez-Grande, and Rozenn Nicol



Defended February 14th, 2022
Final manuscript version: March 21st, 2022

Abstract

The use of spatial room impulse responses (SRIR) for the reproduction of three-dimensional reverberation effects through multi-channel convolution over immersive surround-sound loudspeaker systems has become commonplace within the last few years, thanks in large part to the commercial availability of various spherical microphone arrays (SMA) as well as a constant increase in computing power. This use has in turn created a demand for analysis and treatment techniques not only capable of ensuring the faithful reproduction of the measured reverberation effect, but which could also be used to control various modifications of the SRIR in a more "creative" approach, as is often encountered in the production of immersive musical performances and installations.

Within this context, the principal objective of the current thesis is the definition of a complete space-time-frequency framework for the analysis, treatment, and manipulation of SRIRs. The analysis tools should lead to an in-depth model allowing for measurements to first be treated with respect to their inherent limitations (measurement conditions, background noise, etc.), as well as offering the ability to modify different characteristics of the final reverberation effect described by the SRIR. These characteristics can be either completely objective, even physical, or otherwise informed by knowledge of human auditory perception with regard to room acoustics.

The theoretical work in this research project is therefore presented in two main parts. First, the underlying SRIR signal model is described, heavily inspired by the historical approaches from the fields of artificial reverberation synthesis and SMA signal processing, while at the same time (incrementally) extending both. The signal model is then used to define the analysis methods that form the core of the final framework; these focus particularly on (a) identifying the "mixing time" that defines the moment of transition between the early reflection and late reverberation regimes, (b) obtaining a space-time cartography of the early reflections, and (c) estimating the frequency- and direction-dependent properties of the late reverberation's exponential energy decay envelope. In order to account for the directional dependence of these properties, a procedure for generating directional SRIR representations (i.e. directional room impulse responses, DRIR) that guarantee the preservation of certain fundamental reverberation properties must also be defined.

In the second part, the model parameters made explicit by the analysis methods are exploited in order to either treat (i.e. attempt to correct some of the inevitable limitations inherent to the SMA measurement process) or more creatively manipulate and modify the SRIR. Two treatment methods in particular are developed in this thesis: (1) a pre-analysis procedure acting directly on repeated exponential sweep method (ESM) SMA measurement signals in an attempt to simultaneously increase the resulting SRIR's signal-to-noise ratio (SNR) while reducing its vulnerability to non-stationary noise events, and (2) a post-analysis denoising technique based on replacing the SRIR's background noise floor with a resynthesized extrapolation of the late reverberation tail.

The theoretical descriptions thus complete, the main analysis methods as well as the DRIR generation and the denoising treatment procedures are then subjected to a series of validation tests, wherein simulated SRIRs (or parts thereof) are used to evaluate the performance, discuss the limitations, and parameterize the implementation of the different techniques. These sub-studies allow each method to be individually verified, resulting in a comprehensive investigation into the inner workings of the analysis toolbox (as well as the denoising process).

Finally, to provide a concluding overview of the complete analysis-treatment-manipulation framework, similar studies are carried out using examples of real-world SMA SRIR measurements. One such measurement is also used here to validate the pre-analysis treatment procedure. In closing, illustrative "proof of concept" examples of the various manipulation methods are presented in order to demonstrate the potential capabilities of the framework; these also serve to open the discussion as to the many directions the work completed in this thesis could subsequently be extended.

Résumé

L'utilisation de réponses impulsionnelles spatiales de salles (*spatial room impulse response*, SRIR) dans la reproduction d'effets de réverbération de salle tri-dimensionnels connaît aujourd'hui une réelle démocratisation grâce à la commercialisation répandue d'antennes sphériques de microphones (*spherical microphone array*, SMA) et à une capacité de calcul numérique en croissance continue. Ces SRIR peuvent reproduire des effets de réverbération spatialisés sur des dispositifs immersifs ("*surround-sound*") à travers des convolutions multicanal de plus en plus performantes. De cette utilisation découle naturellement une demande pour des techniques d'analyse et de traitement non seulement capables d'assurer une reproduction fidèle, mais qui pourraient éventuellement aussi servir à contrôler différentes modifications de la SRIR de façon plus créative que réaliste.

Dans ce contexte, l'objectif principal de cette thèse est de développer un environnement complet d'analyse, de traitement, et de manipulation temps-fréquence-espace de SRIR. Les outils d'analyse doivent mener à une modélisation approfondie permettant ensuite un traitement de la mesure vis-à-vis de ses limitations intrinsèques (conditions de mesure, accumulation de bruit de fond, etc.) ainsi qu'une capacité à faire évoluer certaines caractéristiques de l'effet de réverbération décrit par la SRIR. Ces caractéristiques peuvent être tout à fait objectives, c'est-à-dire explicitement reliées à différents paramètres du modèle, ou alors plutôt informées par une connaissance de la perception humaine de l'acoustique des salles.

Les aspects théoriques de ce projet de recherche sont présentés en deux parties principales. Dans un premier temps, le modèle de signal de SRIR sous-jacent est décrit en s'inspirant directement des approches historiques dans les domaines de la réverbération artificielle et le traitement de SMA, tout en y proposant plusieurs extensions. Le modèle de signal est alors exploité afin de définir les méthodes d'analyse qui forment le noyau du cadre de traitement-manipulation final. Ces méthodes se focalisent particulièrement sur (a) l'identification du "temps de mélange" décrivant le moment de transition entre les premières réflexions et la réverbération tardive, (b) la génération d'une cartographie temps-espace des premières réflexions, et (c) l'estimation des paramètres régissant la décroissance exponentielle de l'enveloppe d'énergie de la réverbération tardive, à la fois en fréquence et en direction. La définition d'une procédure de génération de représentations directionnelles de SRIR (*directional room impulse response*, DRIR) est aussi nécessaire pour pouvoir prendre en compte la dépendance directionnelle de ces propriétés.

En seconde partie, les paramètres de modélisation explicités par les méthodes d'analyse sont exploités à des fins soit de traitement (c'est-à-dire tenter de corriger certaines des limitations inhérentes au processus de mesure par SMA), soit de manipulation ou de modification plus créative de la SRIR. Deux méthodes de traitement sont développées en particulier : (1) une procedure d'atténuation de bruits non stationnaires agissant directement sur les signaux de mesure par balayages de fréquence exponentiels (*exponential sweep method*, ESM) répétés, et (2) une technique de débruitage basée sur une extrapolation et une resynthèse de la queue de réverbération tardive.

Les descriptions théoriques ainsi complétées, les principales méthodes d'analyse ainsi que la génération de DRIR et le débruitage sont sujets à une série de tests de validation au cours desquels des SRIR simulées sont utilisées afin d'évaluer la performance, les limitations, et la paramétrisation des différentes techniques. Ces sous-études permettent à chaque méthode d'être vérifiée individuellement, et donnent un aperçu détaillé du fonctionnement interne des outils d'analyse.

Enfin, une vue d'ensemble de l'environnement d'analyse-traitement-manipulation est obtenue en effectuant des études similaires sur des exemples de mesures réelles de SRIR. En guise de conclusion générale, des exemples purement démonstratifs des différentes méthodes de manipulation sont présentés afin d'illustrer leurs potentielles capacités dans d'éventuelles applications créatives.

To Rod, et à Annick.

Acknowledgments

The tangled web of happenstance that has led to the completion of this Ph.D. thesis involves so many amazing human beings that it would be impossible for me to list every single one in these customary *remerciements*. So let me simply begin by extending the widest-reaching thanks I can muster to everyone responsible, from near or afar, for the existence of this dissertation.

This work is the culmination of an intertwined and lifelong love of music, science, and technology, and no one was more pivotal in enlightening me to the myriad ways in which those fields can intersect than Seth Bernstein. Many years later, I found myself taking the first steps towards this doctorate by joining the EAC team at Ircam for my Master's internship/research project; my endless recognition goes out to Markus Noisternig for having been convinced enough during that time to subsequently take me on in the pursuit of a Ph.D., and for having been an inspired guide ever since. An equal amount of thanks to Thibaut Carpentier, for the countless forms and generous quantities of help offered when needed, from fixing code snippets to re-reading manuscripts and everything else in between. The very same to EAC's fearless team leader, Olivier Warusfel, and in particular for having provided the opportunity to go measure SRIRs in some very cool places (on top of all it takes to keep the good ship EAC sailing straight). Many thanks as well to Gérard Assayag for having signed off on this thesis project and having thus provided the final piece of the puzzle that was needed to make it happen.

Tons of love to my fellow *thésards*, at Ircam and Jussieu and beyond: Judy, Victor, Adrien, Carlos, Nadège, Quentin, Marie, J.-C., Franck, Vincent, Antoine, Aidan... Whether we shared *apéros*, lunches, pints, concerts, shop talk, music talk, or (much) more, you all know how much you helped me get through this thing. Special thanks as well to: Benoît Alary, for the super fun and super fruitful collaboration; David Poirier-Quinot for the IS simulation data and the many interesting technical conversations; Nadine Schütz for giving me the opportunity to participate in some amazing interdisciplinary work; Augustin Muller for the incredible database of SRIR measurements (with Pedro Garcia-Velasquez) and the discussions on their use in spatial audio productions; Wolfgang Kreuzer for the enriching work sessions on spatial decompositions and the use of the Herglotz wavefunction; and Éric Raynaud for setting up SRIR measurements in a couple of really fun spaces and taking lots of pictures throughout.

Finally, I can only ever understate the abundance of support I've ceaselessly received from the Massé contingent in Toronto: my parents, Xavier and Sheila, who have always known to use the right dose of both encouragement and pressure; my grandmother, Joan a.k.a. Nanny, whose boundless curiosity on even the most technical of subjects has made fulfilling Einstein's idiom an absolute pleasure; and my brothers, Antoine and Samuel, whose Schrödinger-esque view of my work as simultaneously cool and desperately nerdy has been ever refreshing and a great source of laughter. A wink also to Tess, whose wine selection has probably been responsible for more than a few of my family dinner ramblings on the finer points of spatialized reverberation.

Glossary

IR	Impulse response.
FDN	Feedback delay network.
RIR	Room impulse response.
FIR	Finite impulse response.
SRIR	Spatial room impulse response.
DRIR	Directional room impulse response.
t_{mix}	Mixing time.
T_{60}	60 dB (Sabine) reverberation time.
E{.}	Mathematical expectation.
Δf_{max}	Frequency spacing between room transfer function maxima.
f_{Sch}	Schroeder frequency.
EDC	Energy decay cuve.
EDR	Energy decay relief.
IS	Image source.
P_0	Initial reverberation power.
SH	Spherical harmonic(s).
HOA	Higher-order Ambisonics.
$\mathbf{\Omega}$	Direction vector in spherical coordinates.
$\mathcal{S}^2$	Spherical surface for a given radius r.
θ	Azimuthal angle, 0 at x-axis, $-\frac{\pi}{2}$ at positive y-axis, $\frac{\pi}{2}$ at negative y-axis.
φ	Elevational angle, 0 at (x, y)-plane, $\frac{\pi}{2}$ at zenith, $-\frac{\pi}{2}$ at nadir.
$Y_{l,m}$	Spherical harmonic of order l and degree m.
$\mathbb{N}_0$	The set of natural numbers (integers) including zero.
$.^*$	Complex conjugation.
SMA	Spherical microphone array.
SLA	Spherical loudspeaker array.
SSLD	Surround-sound loudspeaker dome.
ESM	Exponential sweep method.
ASW	Apparent source width.
LEV	Listener envelopment.

LF	Lateral fraction.
LF_E	Early lateral fraction.
LF_L	Late lateral fraction.
IACC	Interaural cross-correlation.
IACC_E	Early interaural cross-correlation.
IACC_L	Late interaural cross-correlation.
G	Strength of an RIR.
G_L	Late strength of an RIR.
C_{80}	80 ms clarity index.
FBR	Front-to-back energy ratio.
SBT	Spatially-balanced center time.
α	Quadrature weights for discretized integration points on the sphere.
Q	Number of spatial sampling points/transducers on an SMA.
L	Maximum SH order.
K	Number of non-vanishing singular values of the SH encoding matrix $\mathbf{Y}$.
PW	Plane wave.
PWD	Plane wave decomposition.
$j_l(x)$	Spherical Bessel function of the first kind of order l.
$h_l(x)$	Spherical Hankel function of the first kind of order l.
$b_l(kr)$	SMA mode strength or holographic function of order l.
k	Plane wave wavenumber [m^{-1}].
r_s	SMA radius [m].
ω	Angular frequency [rad/s].
c	Speed of sound in air [m/s].
$\delta_{n,m}$	Kronecker delta symbol, $= 1$ iff $n = m$, 0 otherwise (n and m are discrete integers).
$\delta(x - x_0)$	Dirac delta distribution, $= 1$ iff $x = x_0$, 0 otherwise (x is a continuous variable).
$a_l(f)$	Frequency response of HOA SMA encoding filters.
$c_l(f)$	Combined frequency response of $a_l(f)$ and $b_l(f)$ [$b_l(kr)$ for a set $r = r_s$].
$w(\mathbf{\Omega})$	Directivity function.
$\mathcal{P}_l(x)$	Legendre polynomial of order l.
Θ	Central angle/great circle distance between two points $\mathbf{\Omega}$ and $\mathbf{\Omega}'$ on the sphere.
$\tilde{d}_{l,m}(\mathbf{\Omega}_d)$	SH beamforming coefficients for a given look direction $\mathbf{\Omega}_d$.

d_l	Order-dependent axisymmetric SH beamformer coefficients.
DI	Directivity index.
WNG	White noise gain.
SLSHT	Spatially localized SH transform.
f_{alias}	Spatial aliasing frequency.
DoA	Direction of arrival.
SRP	Steered response power.
PIV	Pseudo-intensity vector.
Re{.}	Real part of a complex number.
||.||	Euclidean/2-norm of a vector.
EB-MUSIC	Eigenbeam multiple signal classification.
EB-ESPRIT	Eigenbeam estimation of signal parameters via rotational invariance techniques.
$\lfloor . \rfloor$	Floor operator.
$\lceil . \rceil$	Ceiling operator.
STFT	Short-term Fourier transform.
ToA	Time of arrival.
TDoA	Time difference of arrival.
GCC	Generalized cross-correlation.
SLF	Spatial localization function.
ψ	Diffuseness or incoherence measure.
ϕ	Phase angle in radians, $\in [0, 2\pi]$.
$\mathcal{U}[a, b]$	Uniform probability distribution bounded by values a and b.
Φ	Coherence function.
SDR	Signal-to-diffuse ratio.
CoMEDiE	Covariance matrix eigenvalue diffuseness estimation.
$\mathcal{I}$	Sound field isotropy measure.
$\lvert . \rvert$	Magnitude of a complex number.
SIRR	Spatial impulse response rendering (not to be confused with SRIR).
DirAC	Directional audio coding (not to be confused with Paul Dirac, the physicist).
SDM	Spatial decomposition method.
DEDC	Directional energy decay curve.
γ	Exponential decay rate/damping coefficient.

L^2	Space of square integrable functions (i.e. Lebesgue L^p space with $p = 2$).
$\mathcal{C}^1$	First-order continuity (i.e. functions with continuous first derivatives).
$.^{\mathrm{H}}$	Hermitian or conjugate transpose of a matrix.
PSD	Power spectral density.
SNR	Signal-to-noise ratio.
RMS	Root-mean-square.
κ	Incoherence profile segment score for mixing time estimation.
$\mathbf{b}_c$	Spatio-temporal IS barycentre in Cartesian coordinates, $\mathbf{b}_c = [B_x, B_y, B_z]$.
η	Early reflection distribution "spread" measure/indicator.
ρ	Echo/early reflection density measure.
EDD	Energy decay deviation.
$\xi(f,t)$	Repeated ESM measurement deviation factor.
$\mu(f,t)$	Mean repeated ESM measurement magnitude spectrogram.
λ	Generic control parameter.
$\sigma(f,t)$	Repeated ESM measurement magnitude spectrogram standard deviation.
$\mathrm{sgn}(x) = \frac{x}{\lvert x\rvert}$	Signum function, $= 1$ iff $x > 0$, $= 0$ iff $x = 0$, and $= -1$ iff $x < 0$.
GCV	Generalized cross-validation.
WDI	Weighted directivity index.
MWDI	Maximum weighted directivity index (beamformer).
χ	Under-determined Herglotz inversion regularization parameter.
DEED	Directional echo energy density.
ε	Generic error measurement.
$f_{\mathrm{dir},1}$	First-order directivity limit.
ζ	Weighting function.
DRR	Direct-to-reverberant ratio.
MFBR	Maximum front-to-back ratio (beamformer).
$\mathcal{G}_n^k$	Set of all size k sequences without repetition from within $\{1, 2, \ldots, n\}$.
ATE	Artefact-to-total-energy ratio.
JND	Just-noticeable difference.

Contents

List of Figures

List of Tables

Introduction

"It's not the notes you play, it's the notes you don't play."
— Miles Davis
"Music is the space between the notes."
— Claude Debussy

What exists in the space a musician chooses to leave between two notes? Had Miles and Claude gone a touch further in their one-liners, perhaps they might have mentioned how you can begin to hear the room, studio, or concert hall breathe its own sonic signature back into the music. The nature of this breath inevitably affects the very way that music is played in a given acoustic space, as David Byrne explains in the opening chapter to *How Music Works* (McSweeney's, 2012): there is a reason Gregorian chants were made for (or with?) stone cathedrals and not the Serengeti Plain (where, on the other hand, rhythmically complex and highly precise percussive music was developed).

More specifically, after the natural decay or release of the instrument in question (the *source*), a listener will continue to *receive* echoes of the initial sound as it reflects off the different surfaces in the space (floor, walls, ceiling, furniture, etc.). The specific way these reflections accumulate to form a dense *reverberation* "tail" describes the space's unique acoustic stamp, which can be mathematically formalized as an *impulse response* (IR). Recreating a given IR (or at least a "good enough" approximation of it) is therefore akin to recreating the acoustical properties of the space.

As studio production techniques rapidly developed throughout the second half of the 20^{th} century, controlling the amount and character of the reverberation excited by any instrument in the "mix" became an increasingly important technical and creative challenge. George Martin and The Beatles, for instance, famously pioneered the use of controlled reverberation chambers at Abbey Road Studios, along with completely artificial methods using damped metal springs or plates.

The related idea of artificially recreating a generic facsimile of an acoustic space's IR had been around since the 1950s and the foundational work of Manfred R. Schroeder. This approach led to the development of reverberators based on looping feedback delay networks (FDN), which enabled the very first experiments with a *spatialized* output over multiple loudspeakers. Recording an actual room's impulse response (RIR) and using it to recreate the exact sonic imprint of the space by convolving it with an input sound (in the manner of a linear causal finite impulse response [FIR] filter), on the other hand, would have to wait for the advent of modern recording technology and digital signal processing.

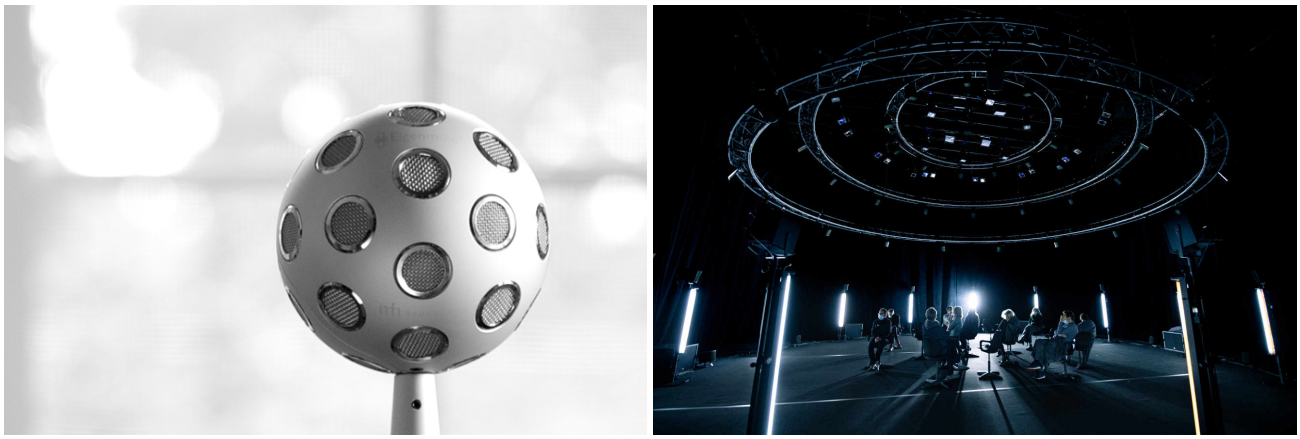
Mono- and stereophonic RIR convolution has since become an integral part of the artificial reverberation landscape, with many well-established commercial options available and a significant history of algorithmic optimization. Many of these implementations also include methods for modifying the character of the reverberation effect by manipulating different parts of the loaded RIR, in the vein of the "one knob" control paradigm readily offered by fully algorithmic approaches (e.g. FDNs or various digital emulations of spring and plate reverbs). In this thesis, we are interested in the extension of these ideas to the context of "spatial" or "immersive" audio (also known as "surround sound"), terms describing the field of research and technology generally dealing with the capture, synthesis, and reproduction of three-dimensional representations of sound.

Context and Motivation

The term *spatial room impulse response* (SRIR) refers, in general, to any formal representation or encoding of a multi-channel RIR measured with a spherical microphone array (SMA). As such, the choice of representation under any given circumstance must be either be clearly specified or otherwise implied; for example, under higher-order Ambisonics (HOA) encoding, this might be made explicit using a prefix (e.g. HOA-SRIR). In this thesis it can be safely assumed that, unless noted otherwise, SRIR refers to such an HOA-encoded SMA RIR measurement (though it may, on occasion, refer to the "raw" SMA RIR signals themselves).

As a spatialized extension of the mono (or stereo) RIR, the SRIR can similarly recreate room reverberation effects through convolution with a mono input sound. The spatialized reverberation effect can then be reproduced in a three-dimensional manner using immersive surround-sound loudspeaker systems. These are sometimes referred to as "spherical loudspeaker arrays" (SLA), which is, unfortunately, an identical term to that used for compact directional sound sources. To avoid the confusion, the term *surround-sound loudspeaker dome* (SSLD) will instead be used in this thesis.

Thanks in large part to the commercial availability of various spherical microphone array (SMA) models, the use of SRIRs in the reproduction of three-dimensional reverberation effects on SSLDs has become commonplace within the last few years. An ever-increasing access to greater computing power, coupled with efficient multi-channel convolution algorithms, has additionally rendered such applications viable even in live stage productions (though [at least partly] pre-rendered immersive audiovisual experiences remain the most common use case).



(a) A commercially available spherical microphone array (SMA), the mh acoustics Eigenmike®.
Photo: mh acoustics.

(b) An immersive audio experience making use of a surround-sound loudspeaker dome (SSLD).
Photo: Hervé Véronèse.

Figure I: Examples of a commercially available SMA (a, left) and an immersive audio/surround-sound experience using an SSLD (b, right).

Figure I gives examples of the type of SMA (a, left) and SSLD-based immersive audio production (b, right) that have become popular in the world of experimental and avant-garde music. The mh acoustics Eigenmike® is a commercially available 32-capsule rigid SMA with a radius of 4.2 cm, and is one of the most widely used solutions for capturing HOA recordings. At Ircam, the Eigenmike has been used to measure SRIRs in a wide variety of acoustic spaces, either for use in immersive audio productions, such as the *Musiques-Fictions* pictured in Fig. Ib (part of Ircam's ManiFeste-2021 festival), or in the interest of more fundamental research. As of the end of this doctorate, a database of over 450 Eigenmike SRIR measurements covering 28 unique venues (many of which include several separate acoustic spaces) has thus been accumulated.

The diversity in the acoustical properties of the measured spaces, especially those displaying less

"traditional" reverberation characteristics (such as coupled volumes or semi-open spaces, for example), has served to highlight the specific advantages offered by the context of spatial audio. The use of multi-channel SRIR convolution to recreate the three-dimensional acoustic signature of these spaces has subsequently opened a host of creative possibilities when producing immersive sonic content. However, this method remains rigid compared to the parametric controls available in algorithmic or mono/stereo convolution reverberators: there is, as of yet, no "one knob" approach to manipulating the properties of the spatialized reverberation effects described by SMA-measured SRIRs.

This research project is therefore fundamentally motivated by a demand for high-level control over the perceptually predominant characteristics of the spatialized reverberation effect generated by an SRIR convolution. The term "high-level" is used here in the sense that the final control parameters should represent intuitive, perceptually relevant descriptors of the reverberation effect; in other words, they should exist several layers of abstraction above the fundamental or "low-level" signal properties[1].

Research Project

Bridging the levels of abstraction between an SRIR's most fundamental signal properties, i.e. its *directly measurable* or *observable* quantities, and any given high-level concepts (e.g. "reverberance" or "envelopment"), requires knowledge of how the measurables interact amongst each other and how these interactions then impact the SRIR's behaviour: that is, we first need to *analyze* the SRIR with respect to a hypothesized signal model. Before attempting to *manipulate* the different properties brought to light by the SRIR analysis, we must also ensure that any limitations or inaccuracies inherent to the measurement process are either corrected or compensated for. Most commonly, these take the form of different types of noise, against which the SRIR measurement must therefore be *treated*.

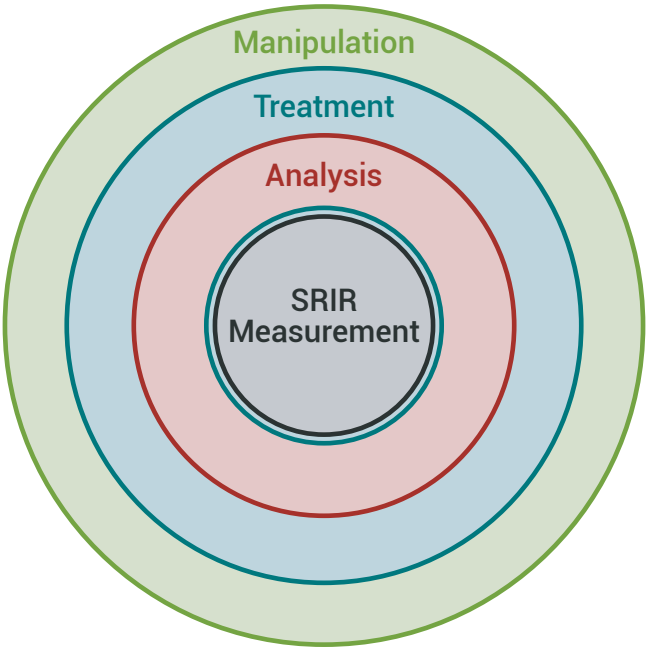


Figure II: Schematic presentation of the analysis-treatment-manipulation framework.

The resulting analysis-treatment-manipulation framework is represented schematically in Fig. II. Its concentric structure illustrates how each outer level relies on the results of the layers within it: as described above, a measurement must first be analyzed before it can be treated and eventually manipulated. (The thin blue layer surrounding the measurement "core" represents a pre-analysis

[1]The concept of audio "volume" is a good example of such a descriptor. Volume, usually formalized as "loudness", is in fact a higher-level perceptual abstraction based on the low-level signal properties of amplitude, energy, and power.

treatment procedure that operates somewhat "out of turn" with respect to the framework's general structure – the exception that proves the rule, in a sense.)

The main goal of this thesis project is therefore the definition and construction of such a self-contained, top-down framework for the analysis, treatment, and manipulation of SRIRs measured with SMAs. In order to offer as complete a description as possible, the framework must consider SRIRs along the dimensions of space, time, and frequency simultaneously. Indeed, part of the motivation behind this work is also the desire to be able to apply the framework to the various acoustically complex spaces that are present in the Ircam SMA SRIR measurement database (see above). As we will see, this requires the creation of new methods that directly exploit the space-time-frequency dependence of different SRIR model parameters.

To give an example, the main treatment method that will be addressed in this work involves compensating for the non-decaying background noise floor inevitably present in SRIR measurements. The approach taken here seeks to replace the noise floor with an extrapolated reconstruction of the SRIR's late reverberation tail. The late reverberation tail's energy envelope is described by a frequency-dependent exponential decay (in general, higher frequencies decay faster than lower ones). In some of the particularly complex spaces measured, the properties of the exponential decay also vary significantly with direction (from the point of view of the SMA, i.e. the receiver). In these cases, the framework must operate on a direction-dependent view of the SRIR, a representation referred to in this dissertation as a *directional* room impulse response (DRIR).

Finally, as noted above, high-level control over the SRIR manipulations would ideally be based on perceptual descriptors of the reverberation effect. Unfortunately, due to time constraints, this extra step has fallen outside the scope of the current research project. A natural extension, however, would be the addition of such a perceptive control layer capable of "translating" or abstracting the lower-level SRIR model properties into intuitive parameters. To help lay the groundwork to this end, a few perceptual considerations are nonetheless presented throughout this thesis, and the extension to a full control layer is further discussed in the Conclusion.

Dissertation Structure

This dissertation is organized into five principal chapters, the latter four of which actually present the work accomplished during this doctorate. Chapter 1 first provides an overview of the theoretical background required in order to define the space-time-frequency framework's underlying signal models, from the historical approaches to room acoustics to the state of the art in SMA processing.

Chapter 2 sets the foundations for the analysis layer of the framework, beginning with a global view of the SRIR and then developing specific signal models for each of the individual acoustical regimes identified within. Chapter 3 follows by introducing the different treatment and manipulation methods made possible by the analysis and modelling of Ch. 2, and in addition, further clarifies the conceptual difference between *treatment* and *manipulation* methods.

Chapter 4 presents the wide range of validation tests carried out in order to verify the different analysis methods from Ch. 2 (including the DRIR representation itself) as well as the noise compensation treatment from Ch. 3. Finally, Chapter 5 concludes the main content of the thesis with a comprehensive demonstration of the complete space-time-frequency analysis-treatment-manipulation framework's capabilities through its application to "real-world" SMA SRIR measurements, i.e. examples taken from the Ircam Eigenmike SRIR database described above. A brief Conclusion closes the dissertation with a summary, a short discussion, and an overview of potential topics for future study.

1 / Theoretical Background

This chapter begins with a brief run through various aspects of the history of room acoustics and reverberation modelling. Particular attention is brought to the various physical assumptions that have subsequently guided the ways in which the reproduction of artificial reverberation effects has been approached in modern digital signal processing, including the ever-expanding field of spatial audio (Sec. 1.1). This historical review culminates in a summary of contemporary research on the measurement and processing of spatial room impulse responses (SRIR) measured with spherical microphone arrays (SMA; Sec. 1.2). Together, these constitute the fundamental knowledge necessary for the proposed SRIR signal model that underpins the work carried out in this thesis and is presented in the following chapter (Ch. 2). As such, expert readers may jump directly to Ch. 2 should the thought of yet another historical review of the field prove to be too much to bear.

1.1 / A Brief History of (Spatial) Reverberation Modelling

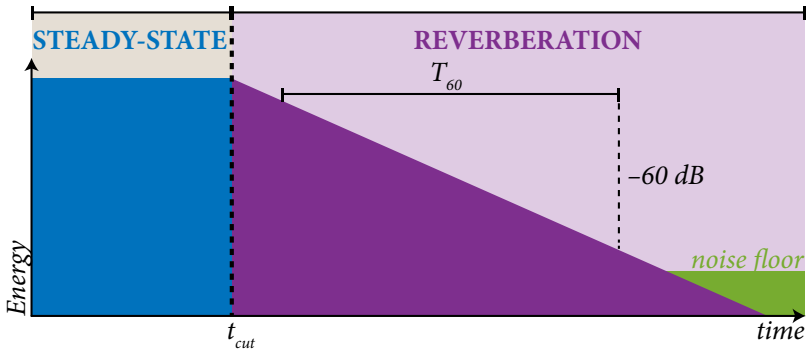
1.1.1 / Geometrical Room Acoustics and Statistical Analysis Methods

The investigation of room acoustics through statistical analysis methods can be traced back to Wallace Clement Sabine's late 19$^{\text{th}}$-century derivation of an exponential energy decay and corresponding "reverberation time", defined as the time necessary for the acoustical energy to decay 60 dB following the end of a "steady-state" white noise excitation (see Fig. 1.1a). Sabine's derivations and their subsequent applications, e.g. for the measurement of the acoustical properties of enclosed spaces [1] or the generation of artificial reverberation effects [2], rely on several important hypotheses relating to the distribution of the acoustical energy in the room.

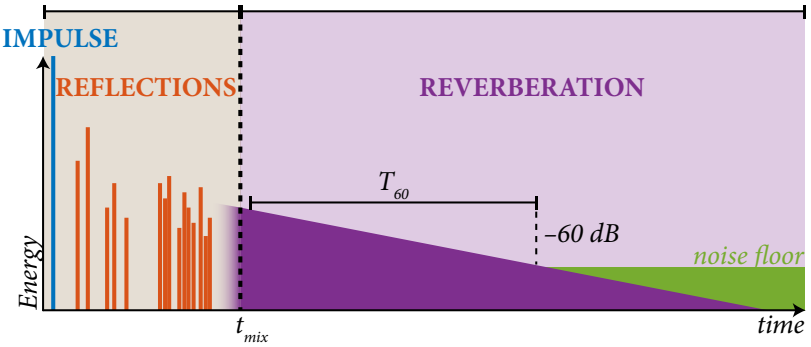
Following Sabine, room acoustics were usually measured and analyzed through steady-state white noise excitations [3] that would rapidly create "diffuse field" conditions throughout the room [4]. This diffuse field is described as consisting of a near-infinite number of uncorrelated, equal-power plane waves evenly distributed throughout the space, resulting in an incoherent, isotropic, and homogeneous acoustic field [4] [5]. Throughout this dissertation, the term "diffuse" will thus refer to this particular definition, and in particular the triple-headed assumption of incoherence, isotropy, and homogeneity.

The artificial reproduction of reverberation effects, however, involves the simulation of a space's response to an impulsive excitation rather than a steady-state one [1]. Exciting a space with an impulse generates discrete reflections or echoes of the original (the so-called "direct sound") that accumulate as each one hits a new reflective or partially absorptive surface. The resulting *impulse response* (IR) therefore gives a complete description of how sound will travel from a given source to a given receiver in the excited space (and, theoretically at least, at any frequency, since an ideal impulse has a completely white spectrum). As later shown more rigorously by Polack [6], this accumulation eventually leads to an exponentially decaying diffuse sound field identical to the one described by Schroeder [3] following a steady-state excitation (see Fig. 1.1b).

In his original design for an artificial reverberation engine, Schroeder [2] was already aware of the need to achieve a minimum echo density in order to satisfy the conditions for a stochastic signal



(a) Schematic depiction of reverberation analysis through steady-state excitation. The acoustic space is excited with a constant-power white noise source long enough for steady-state diffuse field conditions to be achieved, at which point (t_{cut}) the source is switched off and the diffuse field decays exponentially (linearly in dB). The reverberation time T_{60} measures how long it takes the decaying power envelope to lose 60 dB.



(b) Schematic depiction of a room impulse response (RIR). The acoustic space is excited with an impulse, generating reflections of itself that accumulate to the point (beyond the mixing time t_{mix}) of recreating the exponentially decaying late reverberation tail seen in steady-state excitations.

Figure 1.1: Schematic comparison of steady-state versus impulsive excitations of acoustic spaces.

and thereby produce a "natural-sounding" effect. Using billiard theory and a geometric ray-tracing approach, Polack [6] formalized the "mixing" process of reflection accumulation, eventually determining that a density of 10 overlapping echoes within 24 ms [7] could be used as a condition for considering reflections to be fully uncorrelated (the 24 ms figure is a "characteristic time" of the human auditory system that essentially acts as the minimum resolvable time difference between discrete echoes) [5].

Thus Polack [6] crucially makes the link between mixing echoes and the diffuse field properties described above: beyond a critical time following the arrival of the direct sound, the "late reverberation tail" can subsequently be properly described in the context of an IR as a diffuse field subjected to a power envelope that follows an exponential decay. This critical time has (sometimes controversially) been termed the "mixing time" (t_{mix}), though some may sometimes prefer the use of the term "transition time". For the sake of continuity with the majority of the existing literature on the subject, this thesis will use the former.

Schroeder and Kuttruff [8] were also aware that a stochastic approach could not be used to model the deterministic and discrete (in the frequency domain) damped room modes that dominate the low frequency region of a room's response, be it steady-state or impulsive. However, they showed that above a certain critical frequency, the stochastic properties of any "large" room's response will converge regardless of its specific geometric or absorptive properties [1] [9]. For example, it was demonstrated that the frequency response would reach an "average frequency spacing between adjacent maxima" of $\mathrm{E}\{\Delta f_{\max}\} \simeq 3.91/T_{60} \pm 0.04$, where T_{60} is the room's characteristic (broadband) 60 dB Sabine reverberation time and $\mathrm{E}\{.\}$ denotes mathematical expectation. The critical frequency above which this holds, rather appropriately known today as the "Schroeder frequency" (f_{Sch}), was then empirically

related to the measured room's volume and T_{60} reverberation time [8]:

$$f_{\text{Sch}} = 2000\sqrt{\frac{T_{60}}{V}}. \tag{1.1}$$

Rooms whose acoustic response satisfies these conditions are commonly known as "mixing rooms", in reference to Polack [7]. To summarize the historical body of work reviewed above, the "classic" models for the diffuse late reverberation of these spaces can be seen to rely on the following set of assumptions:

1. *Incoherence*: the diffuse field is comprised of a near-infinite number of uncorrelated plane waves such that, at any given point in the space, the observed field is completely incoherent (none of the plane waves "passing through" the point share any information with each other);
2. *Isotropy*: the diffuse field presents the same conditions regardless of the direction it is observed in, or, equivalently, regardless of its orientation;
3. *Homogeneity*: the diffuse field presents the same conditions no matter where it is observed in the space, i.e. all observation points are equivalent.

In order for these assumptions to be valid in the context of an IR, two conditions must be met:

1. *Minimum echo density*, achieved at the mixing time (t_{mix}) and beyond: the number of echoes arriving at a given observation point per unit time becomes so great that individual reflections cannot be resolved deterministically;
2. *Minimum modal density*, achieved at the Schroeder frequency (f_{Sch}) and above: the number of room modes present per unit frequency in a space's transfer function becomes so great that individual eigenfrequencies cannot be resolved deterministically.

This classic view of late reverberation and its relation to IR measurements quickly led to the development of artificial reverberation processors that could reproduce the effects observed in a mixing room. As noted above, Schroeder [2] again was at the forefront of this work, proposing an all-pass reverberator composed of simple delay lines and decaying feedback loops. Such processors would later become known as "feedback delay networks" (FDN), following the work notably of Gerzon [10] [11], Stautner and Puckette [12], and finally Jot and Chaigne [13].

FDNs have since become a popular choice for the reproduction of artificial reverberation in real-time, as they have proven to be highly manipulable whilst remaining very light on computational cost, even in three-dimensional spatial audio systems. However, as was already pointed out by Schroeder [2], they suffer by nature from a lack of faithful reproduction in the early stages of the impulse response, which is dominated by discrete and deterministic echoes (early reflections). This has lead to the development of hybrid systems, e.g. by Moorer [14] using an image source (IS) approach, or more recently by Carpentier et al. [15], wherein convolution with a measured IR is used to recreate the early reflections before "handing off" to an FDN to save on computing power.

In an analysis-synthesis approach more aligned with the context of the current thesis, Jot et al. [16] then showed that the late reverberation tail could be digitally synthesized using a zero-mean Gaussian white noise signal subjected to an exponentially decaying time-frequency power envelope. In order to match the decay properties of a measured IR, Jot et al. [16] proposed extracting the frequency-dependent $T_{60}(f)$ reverberation time and initial power spectrum $P_0(f)$ from the energy decay relief (EDR), a time-frequency extension of the broadband energy decay curve (EDC) first formalized by Schroeder [1]. These two parameters completely describe the exponentially decaying energy envelope

and therefore allow the IR's late reverberation tail to be fully reconstructed using a zero-mean Gaussian noise. As originally pointed out by Jot et al. [16], a practical application of this idea is in the treatment of real-world IR measurements, which are always somewhat corrupted by a non-decaying background noise floor (see Fig. 1.1b). If the late reverberation tail can be re-synthesized as described, the noise floor can be replaced with a prolongation of the IR's actual decaying tail.

1.1.2/ Towards Spatialized Reverberation

Due to its inherent role in the description of acoustic space, the reproduction of reverberation effects has long been tied to advances in "spatialized" or "surround-sound" systems. Indeed, the use of "spatial audio" to create immersive scenes occurring in virtual acoustic spaces has proven to be a natural application for artificial reverberation techniques that themselves seek to mimic real acoustic spaces. Schroeder [2] ended his original proto-FDN paper with a proposal for a three-dimensional "ambiophonic" system in which the mixing matrix is used to send uncorrelated reverberation tails to a surround-sound loudspeaker setup, an idea later expanded upon by Stautner and Puckette [12] and Jot and Chaigne [13] among others. And in parallel to his work on artificial reverberation, Gerzon [17] laid the foundations for what would later become known as "Ambisonics" by using the spherical harmonics as a basis to describe the sound field measured with a tetraphonic microphone arrangement (even going so far as to show the theoretical possibility of higher-order systems).

The spherical harmonics (SH) have since become a widespread tool in the description and manipulation of spatialized audio. Following Gerzon's work, Daniel [18] formalized "higher-order" Ambisonics (HOA) in a complete encoding-decoding framework (i.e. from microphone array recording to loudspeaker array playback) based on an SH representation of the sound field. For a continuous function $X(\mathbf{\Omega})$ defined on the spherical surface S^2, the ideal SH decomposition is written as follows [19]:

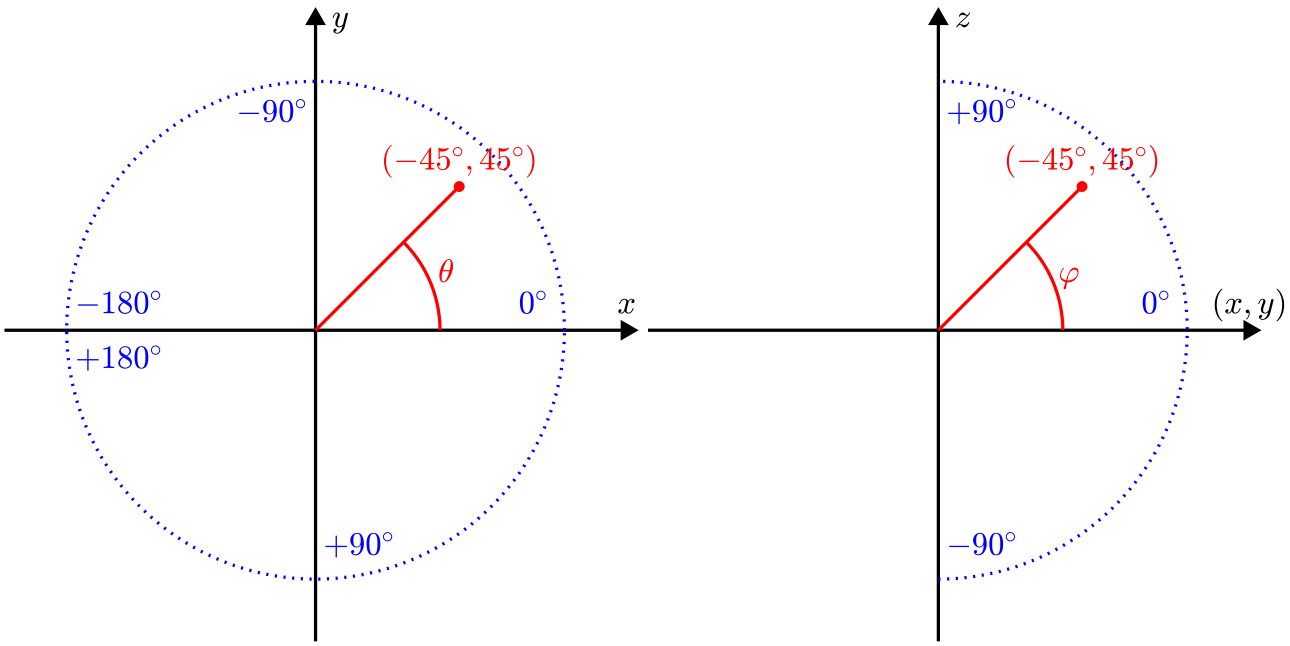
$$X_{l,m} = \int_{\mathbf{\Omega}\in S^2} X(\mathbf{\Omega}) Y^*_{l,m}(\mathbf{\Omega}) d\mathbf{\Omega}, \tag{1.2}$$

where $\mathbf{\Omega} \equiv (\theta, \varphi) \in S^2$ in spherical coordinates with azimuthal angle θ measured from the x-axis and elevation φ measured from the (x, y)-plane (see Fig. 1.2), $Y_{l,m}(\mathbf{\Omega})$ are the spherical harmonics of order $l \in \mathbb{N}_0$ and degree $m \in [-l, l]$, and $.^*$ denotes complex conjugation.

By their very nature, the spherical harmonics favour the use of spherical microphone and loudspeaker arrays; the literature on the optimization of both recording and reproduction systems is extensive and mostly lies beyond the scope of this thesis (not to mention this section). Nevertheless, an overview of contemporary SMA processing will be useful and will be presented in Sec. 1.2 below.

As a consequence of the aforementioned draw of recreating spatialized reverberation effects in three-dimensional surround-sound systems, SMAs quickly became a tool of choice for the measurement of SRIRs. Using techniques well established for monophonic IR measurement such as the exponential sweep method (ESM) [20], a multi-channel IR measured with an SMA can be encoded into HOA, convolved with a monophonic sound source, and the resulting effect (of the monophonic sound source being played at the source position used in the measured space) can then be reproduced on a surround-sound loudspeaker dome (SSLD).

This convolution approach, well-known in the monophonic case, has the advantage of exactly reproducing the acoustic "signature" of the measured space. In multi-channel systems such as HOA, however, the computational cost of many parallel convolutions can rapidly become prohibitive and is another reason why hybrid processors [15] are a subject of interest. Preceding this Ph.D. research project, in fact, the author's Master's thesis [21] focused on extending such a hybrid reverberation engine to SRIRs.



(a) Azimuthal view. (b) Elevational view.

Figure 1.2: Spherical "navigation" coordinate system, displaying the point $\mathbf{\Omega} = (-45°, 45°)$ on the unit sphere. The Cartesian unit x-axis vector $\mathbf{u}_x = [1, 0, 0]$ corresponds to $\mathbf{\Omega} = (0°, 0°)$.

1.1.3/ Considering Perceptual Aspects of Reverberation

The interest in recreating spatial reverberation effects also lies in developing some capacity for creative manipulation of the original acoustic response, in a similar or at least analogous manner to the myriad other ways a sound designer or mixer is capable of shaping an audio signal. The decisions involved in such an approach are overwhelmingly based on our human auditory perception of the result (of the effect applied to a source signal) rather than the direct influence of technical parameters. In order to offer this kind of manipulation, whether it be through a purely convolutional or a hybrid approach, it is therefore necessary to look towards an analysis-treatment framework that is capable of providing a link between low-level signal processing and higher-level perceptual control. This is, indeed, the view suggested by Carpentier et al. [15] for hybrid processors and that has long served to control algorithmic (e.g. FDN) reverberation engines such as the one at the heart of Ircam's Spat~ suite of spatial audio tools.

A full bibliographical review on the history of perceptual evaluations of room acoustics is wildly beyond the scope of this research project, but a brief summary of some of the more widespread measures in use today will allow us to close out this thesis work with an evaluation of how these could be manipulated by the low-level treatments presented beforehand (and therefore vice-versa, as well). Two common indicators to which significant attention has historically been dedicated are the notions of apparent source width (ASW) and listener envelopment (LEV). As their names suggest, ASW aims to describe how "wide" the input signal is made to appear while LEV assesses how "enveloping" the room acoustics tend to feel.

ASW is usually linked to characteristics of the "early" part of RIRs, and has been shown to correlate with the early lateral energy fraction (LF_E) [22] as well as the early interaural cross-correlation (IACC_E) [23]. The lateral energy fraction measures how much of an RIR's energy is contained in its "lateral" portions, i.e. as recorded with a side-to-side bidirectional or "figure-8" microphone. The IACC is a binaural measure of the short-term correlation between the left and right ear signals. It should however be noted that in these definitions "early" refers to the first 80 ms of an RIR following the arrival of the direct sound, in contrast to the geometrical acoustic view presented above using the mixing time to

separate the early and late sections of an IR.

Conversely, LEV has traditionally been tied to the "late" part of RIRs, i.e. after the first 80 ms and through to the end of the tail. This has led to the definition of predictors based on the late lateral energy fraction (LF_L) [24] or the late "strength" (G_L) and late IACC (IACC_L) [25]. The global strength (G) of an RIR can be seen as a measure of its "amplification" of the direct sound; it is defined as the ratio of the total omnidirectional RIR energy to that of a reference anechoic source placed 10 m away from the receiver (which can be simulated by renormalizing the measured RIR's direct sound with respect to the known source-receiver measurement distance). The late strength is then defined by calculating the omnidirectional energy from the 80 ms mark onwards instead of the overall total.

Beranek [25] then defines the late strength using the global strength and the 80 ms "clarity index" (C_{80}), another common room acoustics descriptor that aims to quantify the perceived clarity of the sound source using the ratio of total omnidirectional early to late energies (under the assumption that discrete early reflections contribute positively to clarity by reinforcing the direct sound while late echoes and reverberation have a muddling or "smearing" effect). Although more spatially informed indicators have also been proposed, such as the front-to-back energy ratio (FBR) [26] or spatially balanced centre times (SBTs) [27], recent research suggests that the level-based measures described above correlate better with a subjective evaluation of LEV, even in an SMA measurement/SLA reproduction context [28] (see Sec. 1.2.4 below).

To close out this section, it will be worth mentioning the "DirE" measure, which combines the relative levels from a four-part temporal sectioning of an IR (direct sound, first early reflections, secondary early reflections or "cluster", and late reverberation) [29] and has been shown to provide a good indication of the source's "presence" [30]. It has since been successfully implemented in Ircam's Spat∼ reverberation engine [15] and empirically confirmed through its use in a multitude of spatialized musical audio productions.

1.2/ State of the Art

Following on from the historical review above, this section provides a summary of the contemporary literature upon which the work carried out in this research project is more or less directly based.

1.2.1/ Spherical Microphone Array Processing

Much of the theory on modern SMA design and processing stems from the work of Meyer and Elko [31], Abhayapala and Ward [32], and Gover et al. [33], as well as Rafaely [34] [35] and Zotter [36] (among many others). Though the physical fundamentals can be traced back to the study of inverse problems in acoustical holography [37] (as well as Gerzon's [17] aforementioned approach to spatialized microphone setups), this work from the early 2000s provides the mathematical details for practical implementations from an engineering perspective, both in terms of array design and signal processing. Crucially, it makes the link between the SMA's spatial sampling configuration (i.e. the layout of the transducer positions on the sphere), the discretization of the SH transform, and the subsequent limitations of the captured sound field.

The SH transform (Eq. 1.2) can be seen as a Fourier transform on the 2-sphere using the SH functions as an orthonormal basis [19]. Many of the considerations involved in the discretization of Fourier time series therefore have their counterparts in SMA processing. Foremost of these are the consequences of truncating such transforms due to the use of a finite set of discrete sampling points.

In the case of the SH transform, Eq. 1.2 thus becomes:

$$\widetilde{X}_{l,m} = \sum_{q=1}^{Q} \alpha_q X(\mathbf{\Omega}_q) Y^*_{l,m}(\mathbf{\Omega}_q), \tag{1.3}$$

where α_q are quadrature weights that depend on the layout of the Q spatial sampling points and are optimized such that $\widetilde{X}_{l,m} \to X_{l,m}$ (e.g. through least-squares minimization [38]). Since Q must be a finite integer, Eq. 1.3 can be re-written as a linear system in matrix form:

$$\mathbf{x}_{\text{SH}} = \mathbf{Y}\mathbf{x}, \tag{1.4}$$

where $\mathbf{x}_{\text{SH}}$ is an $(L+1)^2 \times 1$ vector containing the SH coefficients X_n (with the monotonically increasing index $n = l^2 + l + m \in [0, (L+1)^2[$), $\mathbf{Y}$ is an $(L+1)^2 \times Q$ matrix often referred to as the *encoding matrix* and containing the terms $\alpha_q Y^*_n(\mathbf{\Omega}_q)$, and finally $\mathbf{x}$ is a $Q \times 1$ vector containing the signal measured at each sampling point.

Whereas for Fourier time series such discretization and truncation leads to a maximum frequency limit depending on the number of sampling points per unit time (through the Shannon-Nyquist theorem), the system described by Eq. 1.4 imposes a maximum SH order depending on the number of spatial sampling points and their layout on the sphere. In effect, the $(L+1)^2 \times Q$ encoding matrix $\mathbf{Y}$ must present $K = (L+1)^2 \leq Q$ non-vanishing singular values [32] [39].

In order to simplify the following mathematical descriptions of the sound field measured by an SMA, we will now make the common assumption of a *plane wave decomposition* (PWD). The PWD simply assumes that the measured sound field is composed solely of progressive plane waves; as we have seen, this also happens to be a fitting assumption in the case of room reverberation (under the further assumption that the SMA is a large enough distance away from the source and any reflective surfaces).

A single incident plane wave arriving from a given direction $\mathbf{\Omega}_d$ with unit amplitude can be fully described in the SH basis by [37, p. 227]:

$$X_{\text{in}}(kr, \mathbf{\Omega}) = 4\pi \sum_{l=0}^{\infty} i^l j_l(kr) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y^*_{l,m}(\mathbf{\Omega}_d), \tag{1.5}$$

where $i = \sqrt{-1}$ is the imaginary unit, $k = \omega/c$ is the plane wave's wavenumber (with ω the angular frequency and c the speed of sound), and $j_l(kr)$ are the spherical Bessel functions of the first kind.

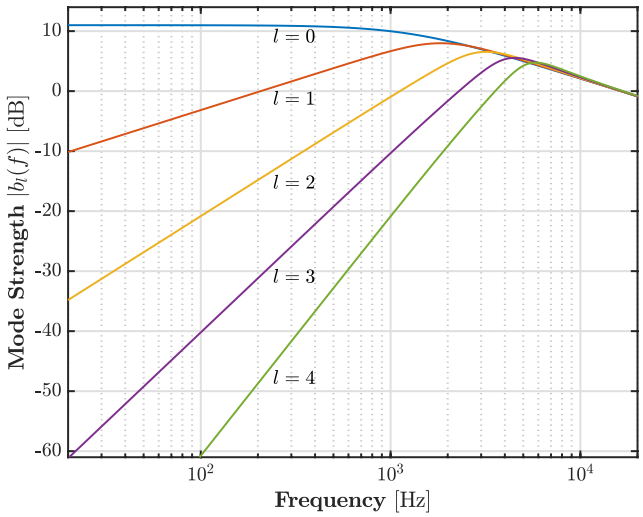
In the case of a rigid or "closed-sphere" SMA of radius $r = r_s$, which represents the majority of commercially available options today (e.g. the mh acoustics Eigenmike® mentioned in the Introduction), the scattered field must also be considered [37, p. 228]:

$$X_{\text{sc}}(kr, \mathbf{\Omega}) = -4\pi \sum_{l=0}^{\infty} i^l \frac{h'_l(kr_s)}{j'_l(kr_s)} h_l(kr) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y^*_{l,m}(\mathbf{\Omega}_d), \tag{1.6}$$

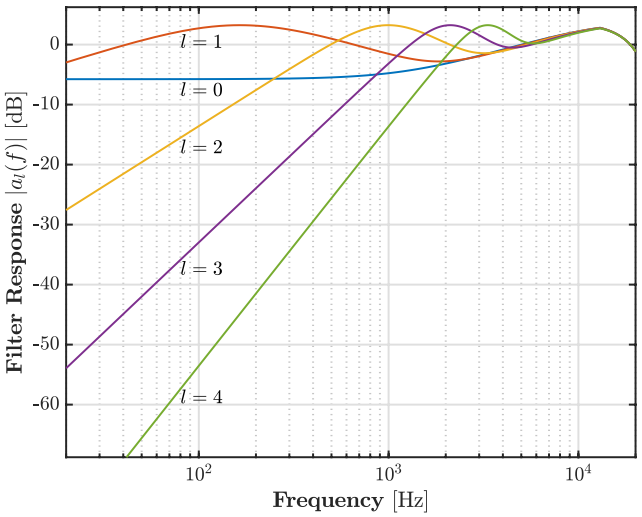
where $h_l(kr)$ are the spherical Hankel functions of the first kind and $.'$ represents differentiation with respect to the argument kr. By restricting our description to the surface of the SMA ($r = r_s$), we can then replace the dependence on the product kr by a dependence on the temporal frequency $f = \omega/2\pi$. The total sound field generated by a plane wave incident on the surface of a rigid SMA is thus:

$$\begin{aligned} X_{\text{PW}}(f, \mathbf{\Omega}) &= X_{\text{in}}(f, \mathbf{\Omega}) + X_{\text{sc}}(f, \mathbf{\Omega}) \\ &= \sum_{l=0}^{\infty} b_l(f) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y^*_{l,m}(\mathbf{\Omega}_d), \end{aligned} \tag{1.7}$$

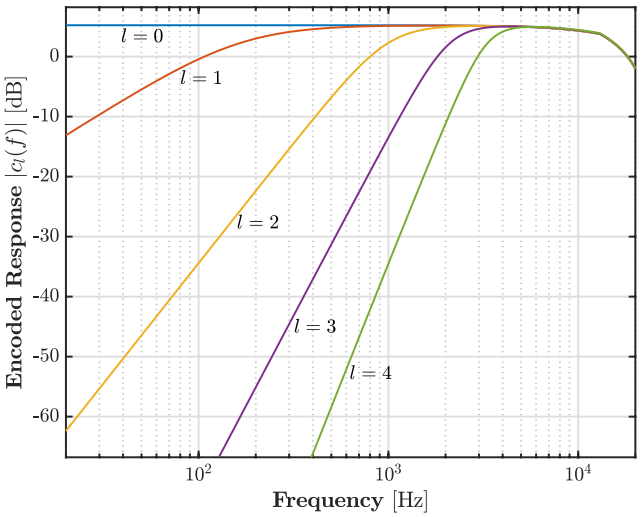
where $b_l(kr) = 4\pi i^l \left[j_l(kr) - \frac{h'_l(kr_s)}{j'_l(kr_s)} h_l(kr) \right]$ is known as the mode strength or holographic function for a



(a) Mode strength magnitude frequency response curves $|b_l(f)|$ for $l \in [0, 4]$ and a rigid SMA radius of 4.2 cm.



(b) Magnitude frequency response $|a_l(f)|$ of high-passed 4$^{\text{th}}$-order HOA encoding filters, with a maximum gain regularization of 18 dB.



(c) The resulting magnitude frequency response $|c_l(f)| = |a_l(f)b_l(f)|$.

Figure 1.3: Magnitude frequency response curves for the 4$^{\text{th}}$-order HOA encoding of a plane wave field using a rigid SMA with a 4.2 cm radius.

rigid sphere (of radius $r = r_s$). The magnitude frequency response $|b_l(f)|$ for $r_s = 4.2$ cm and $l \in [0, 4]$ is shown in Fig. 1.3a.

Taking the ideal SH transform of such a field (Eq. 1.2), laying aside the discretization and truncation considerations discussed above for the moment, gives:

$$\begin{aligned} X_{l,m}^{\text{PW}}(f) &= \sum_{l'=0}^{\infty} b_{l'}(f) \sum_{m'=-l'}^{l'} Y_{l',m'}^{*}(\mathbf{\Omega}_d) \int_{S_{r_s}^2} Y_{l',m'}(\mathbf{\Omega}) Y_{l,m}^{*}(\mathbf{\Omega}) d\mathbf{\Omega} \\ &= b_l(f) Y_{l,m}^{*}(\mathbf{\Omega}_d), \end{aligned} \tag{1.8}$$

by exploiting the orthogonality of the SH basis:

$$\int_{S^2} Y_{l,m}(\mathbf{\Omega}) Y_{l',m'}^{*}(\mathbf{\Omega}) d\mathbf{\Omega} = \delta_{l,l'} \delta_{m,m'}, \tag{1.9}$$

where $\delta_{n,n'}$ is the Kronecker delta. (Indeed, one of the challenges in designing spatial sampling

configurations is ensuring that this orthogonality holds under the discretization and truncation described above with the lowest possible error [35].)

Furthermore, since the SH functions form a complete orthonormal basis, the following closure relation also holds [37, p. 191]:

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega})Y_{l,m}^{*}(\mathbf{\Omega}_d) = \delta(\theta-\theta_d)\delta(\sin\varphi-\sin\varphi_d), \tag{1.10}$$

where $\delta(x-x_0)$ is the Dirac delta distribution (note the use of the sine of the elevation φ due to its measurement with respect to the $[x,y]$-plane, see Fig. 1.2). As a result, an ideal incident plane wave at $\mathbf{\Omega}_d$ can be simply described in the SH domain as the (infinite) set of all $Y_{l,m}^{*}(\mathbf{\Omega}_d)$ values, combined with a radial dependence carried by spherical Bessel functions as in Eq. 1.5, and then reconstructed in the spatial domain by taking its inverse SH transform on a given spherical surface S^2. This is, in fact, one of the fundamental ideas behind the higher-order Ambisonics (HOA) encoding-decoding framework [18].

Equation 1.8 is therefore all but a single $b_l(f)$ factor away from exactly describing such an ideal point source. If the holographic function could be corrected for, i.e. Eq. 1.8 multiplied by $1/b_l(f)$, the SMA measurement of an incident free-field plane wave would therefore be exactly equivalent to its theoretical representation in the SH domain.

Unfortunately, due to the nature of the Bessel functions, the inversion $1/b_l(f)$ is highly unstable. As shown in Fig. 1.3a, the $|b_l(f)|$ magnitude frequency curves present increasing attenuation at low frequencies as the SH order l increases. A direct $1/b_l(f)$ inversion would therefore result in an uncontrolled "bass boost" effect at audible frequencies. As such, the design of stable "radial" or "encoding" filters (e.g. with causal finite impulse response [FIR] implementations) presents a significant challenge.

In answer to this problem, Moreau and Daniel [40] [41] provide a filter design strategy that seeks to optimize the encoding filters with respect to the subsequent sound field reconstruction, i.e. the HOA decoding step. Their approach aims to limit the bass boost effect by applying high-pass filters with increasing cutoff frequencies by SH order to the theoretical $1/b_l(f)$ curves, and is usually parameterized by the maximum allowable bass boost. An example of the encoding filters obtained by such a method is shown in Fig. 1.3b; note that these also include a hard high-cut at around 13 kHz ensuring the regularization is also applied at very high audible frequencies.

Considering once again the discretization and truncation of the SH transform, and assuming that orthogonality holds sufficiently well under the sampling configuration $\mathbf{\Omega}_q$, the HOA encoding of a plane wave measured by an SMA can be therefore written as:

$$\begin{aligned}
\widetilde{X}_{l,m}^{\mathrm{PW}}(f) &= a_l(f)\sum_{q=1}^{Q}\alpha_q X_{\mathrm{PW}}(f,\mathbf{\Omega}_q)Y_{l,m}^{*}(\mathbf{\Omega}_q) \\
&= a_l(f)\sum_{q=1}^{Q}\alpha_q\sum_{l'=0}^{\infty}b_l(f)\sum_{m'=-l'}^{l'}Y_{l',m'}(\mathbf{\Omega}_q)Y_{l',m'}^{*}(\mathbf{\Omega}_d)Y_{l,m}^{*}(\mathbf{\Omega}_q) \\
&\approx c_l(f)Y_{l,m}^{*}(\mathbf{\Omega}_d),
\end{aligned} \tag{1.11}$$

where $a_l(f)$ is the frequency response of the HOA encoding filters and $c_l(f)=a_l(f)b_l(f)$ is the resulting frequency response of the HOA-encoded plane wave, shown in Fig. 1.3c.

One of the main advantages of the HOA framework is thus its capacity to (theoretically) represent spatial sound fields independently from either the SMA used for measurement or the destination reproduction system. This is achievable in practice by using the methods detailed above to design encoding filters that ensure the resulting frequency response $c_l(f)$ is independent of the SMA characteristics carried by the holographic function $b_l(f)$.

A further advantage, of particular interest in the context of this thesis, is the capacity to generate a purely directional view of the sound field by taking the inverse SH transform of the encoded HOA signal, as suggested by the closure relation (Eq. 1.10). In the discretized and truncated HOA framework, of course, the relation is not fully verified; much like in Fourier time series discretization, the spatial Dirac delta is "spread" out into an axisymmetric directivity function w_L around $\mathbf{\Omega}_d$ [42]:

$$\begin{aligned} w_L(\mathbf{\Omega}) &= \sum_{l=0}^{L} \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d) \\ &= \sum_{l=0}^{L} \frac{2l+1}{4\pi} \mathcal{P}_l(\cos\Theta) \\ &= w_L(\Theta), \end{aligned} \tag{1.12}$$

where Θ is the central angle between $\mathbf{\Omega}$ and $\mathbf{\Omega}_d$, $\mathcal{P}_l(\zeta)$ are the Legendre polynomials, and the SH addition theorem has been used to obtain the second line [43]. The directivity function $w_L(\Theta)$ is shown in Fig. 1.4 for three different maximum truncation orders ($L = \{1, 4, 7\}$), demonstrating how $w_L(\Theta)$ approaches an ideal spatial Dirac delta as L increases.

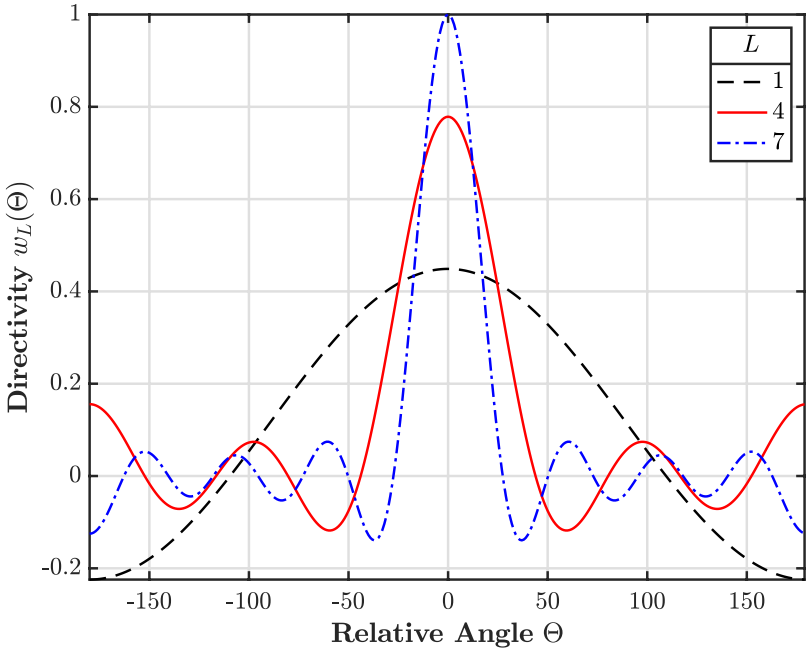


Figure 1.4: PWD directivity $w_L(\Theta)$ as a function of the axisymmetric relative angle Θ for three maximum orders $L = \{1, 4, 7\}$, normalized to equal total energy.

This "spreading" of the Dirac delta illustrates how truncating the SH transform affects the spatial resolution of the encoded signal. This can be characterized by the main lobe half-width Θ_{PN} of the directivity function, i.e. the first zero of $w_L(\Theta)$. Rafaely [42] provides an empirical approximation of $\Theta_{\text{PN}} \approx \pi/L$ as a function of the maximum truncation order that is valid to within 2° for $L \in [4, 40]$.

The rise of a directivity function with a main lobe and several side lobes also highlights the beamforming view of the inverse SH transform under a PWD. In this sense, the theoretical incident plane wave at $\mathbf{\Omega}_d$ represents a test or probe input often referred to as a "steering" or "look" direction. Generally, an SH beamforming interpretation describes an output signal y_d (a "beam") corresponding to the look direction $\mathbf{\Omega}_d$, using beamforming coefficients $\tilde{d}_{l,m}(\mathbf{\Omega}_d)$ that can be determined through various beamformer design strategies [35]:

$$y_d = \sum_{l=0}^{L} \sum_{m=-l}^{l} \tilde{d}^*_{l,m}(\mathbf{\Omega}_d)\widetilde{X}_{l,m}. \tag{1.13}$$

If the beamforming coefficients are now chosen as $\tilde{d}_{l,m}(\mathbf{\Omega}_d) = d_l Y_{l,m}(\mathbf{\Omega}_d)$, with $d_l = 1 \; \forall \; l \in [0, L]$, and the beamformer's directivity is evaluated over $\mathbf{\Omega}$ using $\widetilde{X}_{l,m} = Y_{l,m}(\mathbf{\Omega})$, the equivalence between the beamforming view and the PWD becomes apparent (Eq. 1.13 becomes the first line of Eq. 1.12). As such, this specific choice of $\tilde{d}_{l,m}(\mathbf{\Omega}_d)$ and d_l coefficients is known as the "natural" or PWD beamformer.

A measure commonly associated with beamformer directivity functions is the *directivity index* (DI), defined in dB as [44]:

$$\mathrm{DI} = 10 \log_{10} \left(\frac{2\pi |w_L(0)|^2}{\int_0^{2\pi} |w_L(\Theta)|^2 d\Theta} \right), \tag{1.14}$$

i.e. the ratio between the square magnitude of the beamformer's response at the steering direction $\mathbf{\Omega}_d$ and the spatial average of the directivity function's square magnitude.

The DI is often used in beamforming applications either as a design target or as a performance evaluator. In fact, it can be shown that the PWD beamformer is the solution to the beamformer design optimization problem that maximizes the DI [45]. Other beamformer designs seek to optimize with respect to different measures such as the white noise gain (WNG, a measure of the beamformer's sensitivity to error noise, including sensor self-noise and positioning errors) [44] [46], or a side lobe level to main lobe width compromise (the Dolph-Chebyshev beamformer) [47], among others.

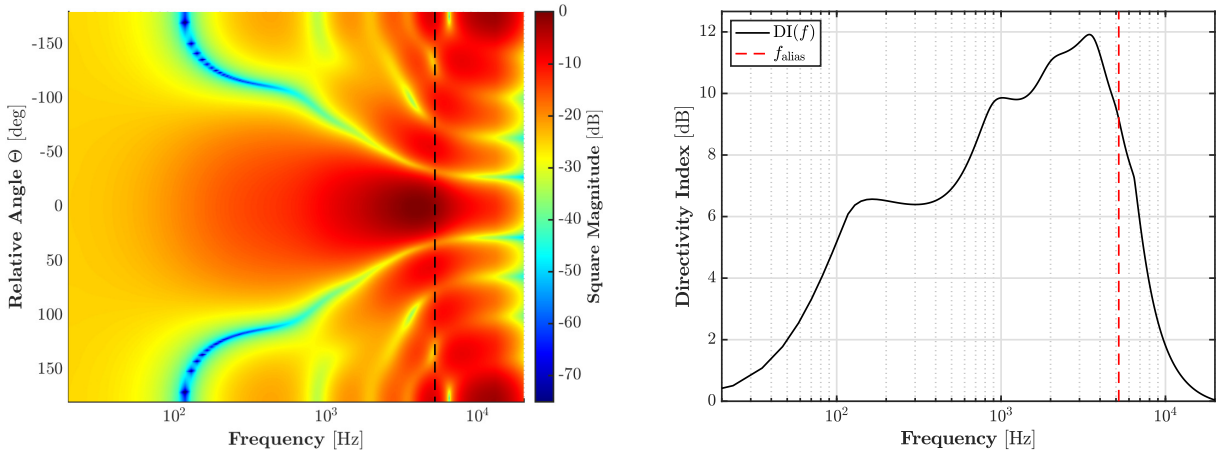
By defining a set of D look directions spread out over the sphere, SH beamforming can therefore generate a purely directional view of the HOA-encoded sound field. The spatial resolution of this representation is ultimately determined by the main lobe width of the chosen beamformer design, while the side lobe characteristics influence how the directivities of the different look directions overlap on the sphere to form the overall "coverage" of the directional decomposition.

In this sense, the generation of such a directional representation of the sound field is simply a special case of a more general family of spatial decomposition or "windowing" techniques. Indeed, the challenge of placing D beamforming look directions on the sphere while taking into account their individual directivities and overlap is an equivalent problem to the design of a spatial filter-bank on the sphere [48]. Mathematically, these can even be further generalized as special cases of the function discretization approach known as "frames" applied to the 2-sphere [49]. Although a thorough exploration of frames or the spatial filter-bank approach is out of the scope of this research project, the use of the spatially localized SH transform (SLSHT) [50] will be briefly evaluated as an alternative to the PWD beamformer.

Regardless of the chosen mathematical framework, it should finally be noted that any spatial decomposition of an HOA-encoded SMA measurement will be affected by the frequency dependence $c_l(f)$ as well as the upper measurement resolution limit $kr \leq L$ [42]. The former, as evidenced by Fig. 1.3c, limits the achievable directivity at lower frequencies due to the order-dependent high-pass filtering applied through the encoding filters $a_l(f)$. The latter is a consequence of the mode strengths $b_l(kr)$ rapidly diminishing for $kr > L$ [42]; in other words, higher-frequency plane waves require higher order SH functions to be resolved on a sphere of given radius r. This is therefore a *spatial aliasing* limit, and for a fixed SMA radius $r = r_s$ it leads directly to an aliasing frequency f_{alias}:

$$f_{\text{alias}} = \frac{cL}{2\pi r_s}. \tag{1.15}$$

Figure 1.5a shows the frequency dependence of the directivity function $w_L(\Theta, f)$ obtained by using $\tilde{d}_{l,m}(\mathbf{\Omega}_d) = d_l Y_{l,m}(\mathbf{\Omega}_d)$ with $d_l = 1 \; \forall \; l \in [0, L]$ and $\widetilde{X}_{l,m} = c_l(f) Y_{l,m}(\mathbf{\Omega})$ in Eq. 1.13. The resulting



(a) Frequency dependence of the directivity function $w_L(\Theta, f)$ for a PWD beamformer applied to a 4^{th}-order HOA-encoded SMA measurement.

(b) Frequency dependence of the DI measure for a PWD beamformer applied to a 4^{th}-order HOA-encoded SMA measurement.

Figure 1.5: Frequency dependence of the PWD beamformer's directivity when applied to a 4^{th}-order HOA-encoded SMA measurement, including the aliasing frequency $f_{\text{alias}} \simeq 5.2$ kHz for the simulated rigid SMA with a radius of 4.2 cm (vertical dashed lines).

frequency-dependent DI is shown in Fig. 1.5b. The aliasing frequency $f_{\text{alias}} \simeq 5.2$ kHz for the given 4.2 cm radius SMA is also displayed on both figures (vertical dashed lines).

1.2.2/ Spatial Sound Field Analysis

The processing techniques described in the previous section have subsequently been applied to the analysis of spatial sound fields as measured by an SMA [51]. Several important properties of a field can be evaluated in this manner; this section will focus on the estimation of the number and direction of arrival (DoA) of any source(s) present, as well as measures of the sound field's spatial coherence, isotropy, and "diffuseness".

Direction of Arrival Estimation

The source localization or DoA estimation problem is one of the most widely researched SMA applications and as such has given rise to many different approaches. One of the most straightforward methods consists of generating a steered response power (SRP) map over the sphere by beamforming the measured signal and then estimating its power (e.g. by averaging over short time frames or narrow frequency bands). The spatial peak of the power map is then usually taken as the DoA estimate.

SRP methods, while relatively robust, suffer from the beamformer directivity and resolution limitations discussed in the previous section, and can rapidly become computationally expensive depending on the number of beams formed. Furthermore, they make no use of any assumptions or characterizations on the underlying signal model for the sources, and simply taking the peak of the SRP limits the method to a single DoA per frame. Recent work by McCormack et al. [52] aims to remedy some of these drawbacks by sharpening the SRP using a directional re-assignment algorithm.

HOA-encoded SMA signals can also be used to construct an analogue of the acoustical intensity vector often referred to as a pseudo-intensity vector (PIV) [53]. The PIV is especially interesting as it carries both energetic and directional information. Indeed, the zeroth-order SH component $\widetilde{X}_{0,0}$ can be seen as an estimate of the sound pressure measured by a reference omnidirectional microphone at the centre of the sphere, and the first-order components $\widetilde{X}_{1,m}$ can be used to construct axial dipole (figure-8) measurements along the x-, y-, and z-axes respectively [53]:

$$\widetilde{X}_{\{x,y,z\}} = \sum_{m=-1}^{1} Y_{1,m}(\mathbf{\Omega}_{\{x,y,z\}})\widetilde{X}_{1,m}, \tag{1.16}$$

where $\mathbf{\Omega}_{\{x,y,z\}} = \{(0,0),(-\pi/2,0),(0,\pi/2)\}$ following the spherical coordinate system in Fig. 1.2. Using short-term Fourier transform (STFT) time-frequency estimates of the signal components, the PIV can then be defined as:

$$\mathbf{I}(f,t) = \frac{1}{2}\mathrm{Re}\left\{\widetilde{X}_{0,0}(f,t)\begin{bmatrix}\widetilde{X}_x^*(f,t)\\ \widetilde{X}_y^*(f,t)\\ \widetilde{X}_z^*(f,t)\end{bmatrix}\right\}, \tag{1.17}$$

where Re{.} represents the real part of a complex number. The unit vector giving the DoA is then:

$$\hat{\mathbf{u}}(f,t) = -\frac{\mathbf{I}(f,t)}{||\mathbf{I}(f,t)||}, \tag{1.18}$$

where ||.|| denotes the Euclidean or 2-norm of a vector.

The PIV has the advantage of being a relatively lightweight method, as well as being a quantity with multiple applications (see the PIV diffuseness measure presented below). However, it can only estimate a single DoA per time-frequency window, and furthermore fails to exploit the localization information made available by the higher-order ($L > 1$) SH components. These concerns have recently been addressed by Politis et al. [54] through a sector-based "localized" extension of the PIV (in which higher-order SH components are essentially used to beamform the PIV over different look directions on the sphere).

"Subspace" methods such as (eigenbeam) multiple signal classification (EB-MUSIC) and the estimation of signal parameters via rotational invariance techniques (EB-ESPRIT) exploit the decomposition of the SH components into a "signal" and a "noise" subspace by eigenvalue decomposition of the SH covariance matrix [51, p. 72]. Determining the number of "largest" eigenvalues in this decomposition can provide an estimate of the number of individual sources present, i.e. the size of the signal subspace, if this information is not available *a priori*. Recent work on EB-ESPRIT by Jo et al. [55] has increased the maximum number of simultaneously detectable sources to $L^2 + L + \lfloor L/3 \rfloor$ ($\lfloor . \rfloor$ is the floor operator), while Herzog and Habets [56] have addressed the angular ambiguity issues encountered by previous implementations.

Both of these subspace methods, however, assume the sources are at most very weakly mutually correlated, and ideally completely mutually incoherent. Jo and Choi [57] have also proposed SH smoothing techniques to remedy this limitation at the cost of reducing the maximum useable order L (and consequently the maximum number of detectable sources). Additionally, estimating the covariance matrix requires time averaging over several STFT frames, which can greatly reduce the final temporal resolution of the method.

Finally, time of arrival (ToA) and time difference of arrival (TDoA) methods based on the generalized cross-correlation (GCC) have long been used in various microphone array applications, e.g. in acoustic imaging [58], and have been adapted to SRIR analysis by Tervo et al. [59] (see also Sec. 1.2.3 below). Combined ToA/TDoA methods have the potential to provide highly precise estimates in both time and space, but can also be very sensitive to background and sensor noise. The spatial localization function (SLF) generated by these methods also requires the use of a peak search algorithm in order to detect the DoA (and ToA) estimates. Depending on the chosen resolution for the spatial scanning grid, the peak search can quickly become computationally expensive.

Spatial Coherence, Isotropy, and Diffuseness Measures

Characterizing the "diffuseness" of a measured sound field is another fundamental problem in spatial audio. Indeed, estimating the degree to which the field tends from being composed of a single incident plane wave to the fully diffuse conditions described in Sec. 1.1.1 provides powerful insight into its acoustical nature. Various approaches have been proposed in order to quantify this property using different estimates and combinations of the two underlying conditions of a diffuse field: its spatial coherence (or lack thereof) and its isotropy. Note that these are the two conditions that can always be tested for with a single SMA measurement; evaluating the field's homogeneity would require the comparison of multiple measurements in different positions.

As previously noted, the PIV can also be exploited to produce a diffuseness measure [60]. Since the acoustic intensity vector theoretically represents the active sound energy over a given surface, the PIV contains information as to the manner in which the incident signal energy is distributed over the sphere. The idea is then to use the temporal variation of the PIV to estimate the diffuseness ψ_{PIV}:

$$\psi_{\text{PIV}}(f,t) = \sqrt{1 - \frac{||\,\text{E}\,\{\mathbf{I}(f,t)\}\,||}{\text{E}\,\{||\mathbf{I}(f,t)||\}}}. \tag{1.19}$$

In a fully diffuse field, the expectation value of the PIV's components averages out to zero, while in a purely directional field the PIV is stable and thus $||\,\text{E}\,\{\mathbf{I}(f,t)\}\,|| = \text{E}\,\{||\mathbf{I}(f,t)||\}$. The diffuseness $\psi_{\text{PIV}}(f,t)$ is therefore bound between 0 (a purely directional field) and 1 (a fully diffuse field).

The fact that the components of the PIV average out to zero in a fully diffuse field is a consequence of its complete spatial incoherence. A basic signal model for such a field can be written as:

$$X_{\text{dif}}(f,\mathbf{\Omega}) = \sqrt{P(f)}e^{i\hat{\phi}(f,\mathbf{\Omega})}, \tag{1.20}$$

where $P(f)$ is an isotropic power spectrum and $\hat{\phi}(f,\mathbf{\Omega})$ is a stochastic plane wave phase, uniformly distributed over $[0,2\pi]$ and independent over both frequency and direction (the hat is used to specify that this is a random variable). The spatial incoherence of this field can be demonstrated:

$$\Phi_{\mathbf{\Omega},\mathbf{\Omega}'}(f) = \frac{\text{E}\,\{X_{\text{dif}}(f,\mathbf{\Omega})X^*_{\text{dif}}(f,\mathbf{\Omega}')\}}{\sqrt{\text{E}\,\{X_{\text{dif}}(f,\mathbf{\Omega})X^*_{\text{dif}}(f,\mathbf{\Omega})\}\,\text{E}\,\{X_{\text{dif}}(f,\mathbf{\Omega}')X^*_{\text{dif}}(f,\mathbf{\Omega}')\}}} = \delta_{\mathbf{\Omega},\mathbf{\Omega}'}, \tag{1.21}$$

since

$$\text{E}\,\{X_{\text{dif}}(f,\mathbf{\Omega})X^*_{\text{dif}}(f,\mathbf{\Omega}')\} = P(f)\text{E}\left\{e^{i\hat{\phi}(f,\mathbf{\Omega})}e^{-i\hat{\phi}(f,\mathbf{\Omega}')}\right\} = P(f)\delta_{\mathbf{\Omega},\mathbf{\Omega}'}. \tag{1.22}$$

In the case of an isotropic power distribution, the result furthermore holds in the SH domain (using the ideal SH transform for clarity):

$$X^{\text{dif}}_{l,m}(f) = \sqrt{P(f)}\int_{S^2} e^{i\hat{\phi}(f,\mathbf{\Omega})}Y^*_{l,m}(\mathbf{\Omega})d\mathbf{\Omega}, \tag{1.23}$$

such that

$$\Phi_{n,n'}(f) = \frac{\text{E}\left\{X^{\text{dif}}_{l,m}(f)X^{*\text{dif}}_{l',m'}(f)\right\}}{\sqrt{\text{E}\left\{X^{\text{dif}}_{l,m}(f)X^{*\text{dif}}_{l,m}(f)\right\}\text{E}\left\{X^{\text{dif}}_{l',m'}(f)X^{*\text{dif}}_{l',m'}(f)\right\}}} = \delta_{n,n'}, \tag{1.24}$$

since

$$\begin{aligned}\mathrm{E}\left\{X^{\mathrm{dif}}_{l,m}(f)X^{*\mathrm{dif}}_{l',m'}(f)\right\} &= P(f)\int_{S^2}\int_{S^2}\mathrm{E}\left\{e^{i\hat{\phi}(f,\mathbf{\Omega})}e^{-i\hat{\phi}(f,\mathbf{\Omega}')}\right\}Y^*_{l,m}(\mathbf{\Omega})Y_{l',m'}(\mathbf{\Omega}')d\mathbf{\Omega}d\mathbf{\Omega}' \\ &= P(f)\int_{S^2}Y^*_{l,m}(\mathbf{\Omega})Y_{l',m'}(\mathbf{\Omega})d\mathbf{\Omega} \\ &= P(f)\delta_{n,n'},\end{aligned} \tag{1.25}$$

following Eq. 1.22 (first two lines) and Eq. 1.9 (second two lines), with once again $n = l^2 + l + m$.

Both Jarrett et al. [61] and Epain and Jin [62] have used this result to define diffuseness measures directly in the SH domain. Jarrett et al. [61] define a signal-to-diffuse ratio (SDR) by performing a weighted average over all possible $\Phi_{n,n'}$ pairs, with the weights parameterized by a DoA estimate of the predominant coherent signal component. Epain and Jin [62] observe that the SH component incoherence means that the SH domain covariance matrix will be diagonal and therefore present a flat eigenspectrum; their diffuseness measure (CoMEDiE) is thus defined according to the average absolute difference between the eigenvalues of the SH covariance matrix and their mean.

It is important to note that these three measures of diffuseness (PIV, SDR, and CoMEDiE) all assume a perfectly isotropic power distribution, following the historical definition of a diffuse field as presented in Sec. 1.1.1. Until recently, however, very little work had been available with respect to the analysis of this property, i.e. the definition of a measure that quantifies the isotropy of a measured field. Nolan et al. [63] have since proposed the use of a "wavenumber" approach that is essentially equivalent to the PWD formalism presented above but where only the direction-dependent magnitude $\sqrt{P(f,\mathbf{\Omega})}$ (strictly speaking the directional amplitude density) is transformed into the SH domain. The magnitude of the corresponding SH components, $B_{l,m}(f) = \int_{S^2}\sqrt{P(f,\mathbf{\Omega})}Y^*_{l,m}(\mathbf{\Omega})d\mathbf{\Omega}$, can then be used to determine the sound field's isotropy:

$$\mathcal{I}_{\mathrm{PW}}(f) = \frac{|B_{0,0}(f)|}{\sum_{l=0}^{\infty}\sum_{m=-l}^{l}|B_{l,m}(f)|}, \tag{1.26}$$

under the observation that in a completely isotropic field, the omnidirectional component $|B_{0,0}(f)|$ will contain all the present energy (and thus $\mathcal{I}_{\mathrm{PW}} = 1$) while any degree of anisotropy will "transfer" energy into different higher-order harmonics.

1.2.3/ Spatial Room Impulse Response Analysis and Modelling

Several of the SMA processing techniques described in the previous section have subsequently been used in the interest of reproducing spatial reverberation effects more efficiently than straightforward SRIR convolution. The general idea behind both spatial impulse response rendering (SIRR) [64], its generalization as directional audio coding (DirAC) [65], and the more recent spatial decomposition method (SDM) [66] is to analyze a measured SRIR and encode only the most perceptually relevant information, i.e. only those properties necessary for the reproduction of a perceptually convincing reverberation effect.

SIRR was originally developed to work with first-order, so-called "B-format" microphone arrays, and thus makes use of the PIV for which higher SH orders are unnecessary. Time-frequency representations of the individual component channels are obtained as STFTs, allowing a time-frequency PIV $\mathbf{I}(f,t)$ to be constructed (Eq. 1.17). Then a single DoA estimate $\widehat{\mathbf{\Omega}}^{\mathrm{PIV}}_i(f,t)$ and diffuseness value $\psi_{\mathrm{PIV}}(f,t)$ can be calculated per time-frequency window.

The reproduction or synthesis step in SIRR exploits the aforementioned fact that the omnidirectional component is proportional to the total sound pressure as would be measured from a reference microphone at the centre of the sphere. SIRR therefore splits the omnidirectional signal into diffuse and non-diffuse

parts: diffuse energy $\psi_{\text{PIV}}(f,t)|\widetilde{X}_{0,0}(f,t)|^2$ and a non-diffuse signal $\sqrt{1-\psi_{\text{PIV}}(f,t)}\widetilde{X}_{0,0}(f,t)$ associated with the DoA $\widehat{\mathbf{\Omega}}_i^{\text{PIV}}(f,t)$. These can then be reproduced on an SLA, e.g. by decorrelation techniques for the diffuse part and spatial panning/virtual source synthesis techniques for the non-diffuse signal.

Consequently, SIRR only needs to store three pieces of information per time-frequency window: the omnidirectional signal $\widetilde{X}_{0,0}(f,t)$, the DoA $\widehat{\mathbf{\Omega}}_i^{\text{PIV}}(f,t)$, and the diffuseness $\psi_{\text{PIV}}(f,t)$. This approach was later generalized as DirAC in order to treat any arbitrary continuous spatial signal [65].

In a sense, SDM can itself be seen as a generalization of SIRR, or at least as an extension to allow for the use of higher-order SMA measurements (in theory SDM actually allows for the use of any type of compact microphone array). The basic idea of a windowed analysis for which each time step encodes an omnidirectional pressure value and a DoA remains the same, but whereas SIRR/DirAC relies on STFT time-frequency analysis, SDM remains solely in the time domain.

For each time window, SDM performs a source localization procedure. Tervo et al. [66] use a straightforward TDoA method, but in theory the choice of specific algorithm is open. As in SIRR/DirAC, this assigns a DoA to the corresponding window of the reference omnidirectional microphone signal. Once again in the case of an SMA this omnidirectional signal can be taken as the zeroth-order SH component $\widetilde{X}_{0,0}(t)$.

Note that SDM does not aim to estimate the sound field's diffuseness at any point, nor does it separate the signal into diffuse and non-diffuse components. The rationale behind this is that in the context of an SRIR and with small enough time windows, SDM will naturally recreate diffuse late reverberation conditions since the localization step will begin to return random DoAs once the late field is reached [66]. It is thus assumed that the rapid succession of windows with random DoAs will create the perception of a diffuse late reverberation tail.

Furthermore, both SIRR/DirAC and SDM assume that the DoA returned by the localization step is the "correct perceptual location" [66] and constitutes an appropriate mapping for the omnidirectional signal values in that window. In SIRR/DirAC, the synthesis of an artificial diffuse signal component restricts the method to the historical assumption of isotropic late reverberation (see Sec. 1.1.1), though recent work by McCormack et al. [67] extending the SIRR method to higher orders has the potential to reproduce anisotropic late fields. To the author's knowledge, there does not currently seem to exist a rigorous evaluation of SDM's performance with respect to anisotropic late reverberation.

Some recent research, however, has explicitly interested itself in the analysis of anisotropic late reverberation. Berzborn and Vorländer [68] have, for example, proposed the use of directional energy decay curves (DEDC) for investigating deviations from isotropy in the decay envelope. They define the DEDC by first beamforming a measured SRIR over a set of look directions covering the sphere, thereby creating a directional room impulse response (DRIR), and then calculating the EDC (see again Sec. 1.1.1) on each directional signal. In order to focus solely on the spatial variations of the late reverberation tail, they finally propose to study the directional deviations to the spatial mean over all DEDCs at each time step.

1.2.4/ An Overview of Related Contemporary Research

In closing to this chapter, a few ongoing research projects exploring topics closely related to the current thesis will be worth mentioning. First and foremost, as this particular subject gave rise to collaborative work that will be presented throughout the following chapters, Alary et al. [69] have formalized a spatialized FDN implementation allowing for the reproduction of anisotropic late reverberation. This work has provided further motivation for the analysis of corresponding phenomena in measured SRIRs.

Meanwhile Badeau [70] has proposed a unified stochastic framework for modelling the complete space-time-frequency characteristics of reverberation, including both the early and late regimes, using an underlying uniform Poisson IS distribution. Beyond the elegance of describing the different aspects

of reverberation under a common mathematical framework, this approach has the further advantage of allowing for both isotropic and anisotropic late reverberation, as well as the transition from discrete early reflections to the incoherent late field.

Meacham et al. [71], on the other hand, have taken an energy-stress tensor approach to modelling diffuse late reverberation, with the potential to eventually consider spatial decay variations as well as arbitrary source models. Though both these research topics are more oriented towards the pure simulation of reverberation effects, their evolution has proven a good reference point for the work carried out in this thesis.

Finally, with respect to the perceptual aspects of spatialized reverberation, Dick and Vigeant [28] have used SMA SRIR processing techniques to investigate LEV evaluation under SLA reproduction conditions. Their research suggests that, contrary to the classic early/late ASW/LEV separation presented in Sec. 1.1.3 but in accordance with other earlier studies [72] [73], LEV can be highly influenced by the early characteristics of an SRIR. Furthermore, they indicate that the overall relative levels between the direct sound, early reflections, and late reverberation have a greater effect on LEV than the finer spatial structure of the SRIR (e.g. the spatiotemporal distribution of the early reflections or anisotropic variations in the late decay).

Current research trends are therefore clearly anchored in an effort to move beyond the historically accepted tenets of room reverberation modelling and perception. Part of the motivation behind this evolution, as posed in the Introduction, is a desire to analyze, treat, and manipulate more acoustically complex spaces than the nicely behaving "large mixing rooms" so often assumed in the historical descriptions (see Sec. 1.1.1). The research project presented in this dissertation is fully inscribed in this context, and as such seeks to define an analysis and treatment framework allowing for as many potential complexities as possible (e.g. slow-mixing reflections, anisotropic late reverberation, multiple-slope decays, etc.). The general signal model and analysis methods implemented towards this goal are thus presented in the following chapter.

---

2 / Analysis Methods

With the essential theoretical knowledge fundamental to the work carried out in this thesis having been presented in Ch. 1, this second chapter aims to describe the SRIR analysis methods upon which the treatment procedures and manipulation strategies of Ch. 3 are reliant (see also Fig. II). The analysis methods themselves, however, require the definition of a general SRIR signal model – this is therefore the subject of the opening section, Sec. 2.1.

More precisely, the general SRIR model is segmented into several different time-frequency regimes (Sec. 2.1.1); of these, the temporal separation between the early reflections (Sec. 2.1.2) and the late reverberation tail (Sec. 2.1.3) will be of particular interest in this work. As we will see, the overarching objective of allowing each and every property of the SRIR to evolve in space, and therefore in incident direction from the point of view of an SMA, requires a directional representation of the SRIR. The spatial decomposition methods used to generate this *directional room impulse response* (DRIR) are thus presented in Sec. 2.2.

The first analysis method to be described is the estimation of the "mixing time", i.e. the moment of transition between the early reflection and late reverberation regimes. A strategy based on the use of a time-dependent scalar measure of the sound field's spatial incoherence is given in Sec. 2.3. Finally, the specific analysis methods relating to the modelling of the early reflections (Sec. 2.4) and the late reverberation tail (Sec. 2.5) can be presented. The early reflection analysis begins with a procedure for detecting the "direct sound" (i.e. the first impulse to reach the receiver, Sec. 2.4.1), before the general strategy for obtaining a complete time of arrival (ToA)-direction of arrival (DoA)-energy echo cartography is described (Sec. 2.4.2). A measure of "directional echo energy density" (DEED) is also defined (Sec. 2.4.3) in a first attempt to abstract some of the low-level model parameters (following the objectives outlined in the Introduction). In terms of late reverberation, the main analysis method is the modelling of the exponentially decaying energy envelope described by the EDR (Sec. 2.5.1). Finally, a space-time-frequency approach to analyzing the isotropy of the late reverberation tail using the "energy decay deviation" (EDD) is also presented.

2.1 / Spatial Room Impulse Response Signal Model

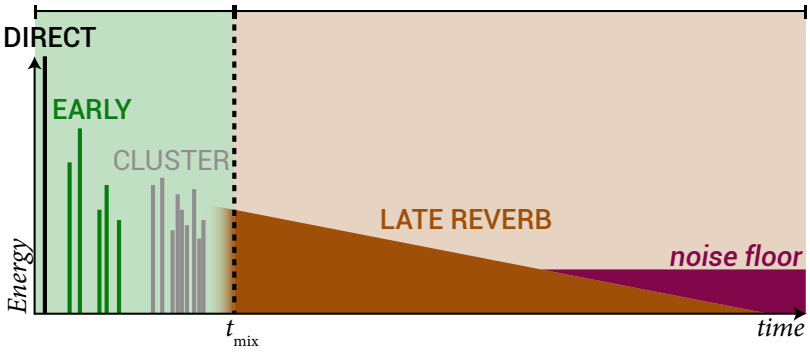
The signal model proposed in this section follows naturally from many of the theoretical considerations presented in Ch. 1. Within the research context outlined in the Introduction, it attempts to overcome several of the limitations inherent to the historical approaches. It has been conceived to be as versatile as possible, offering the combined space-time-frequency framework that will serve as the basis for the various investigations carried out in this thesis work.

As such, it must be flexible enough to describe the different aspects of the SRIR while providing enough detail within each application to enable thorough analysis and treatment. The model therefore begins by sectioning the SRIR into six distinct time-frequency regimes, following the statistical geometrical acoustics approach presented in Sec. 1.1.1. Each SRIR section then has its own underlying mathematical description. The scope of this research project has been limited to the direct sound, early reflections (Sec. 2.1.2), late reverberation (Sec. 2.1.3), and noise floor; a brief note will be made

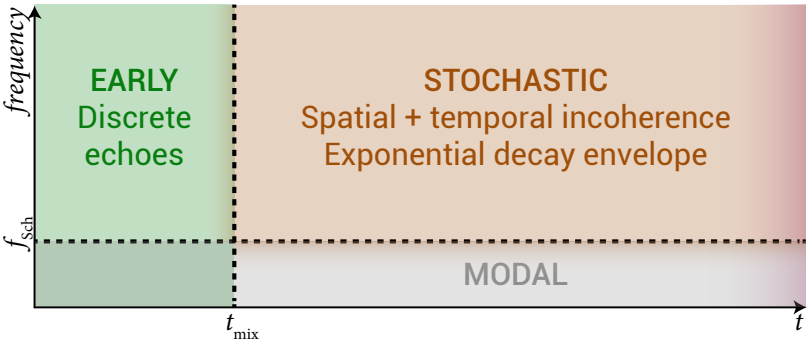
concerning possible approaches to the low-frequency modal domain and the semi-stochastic transition "cluster", but detailed investigations of these are left to future work.

2.1.1/ Time-Frequency Regime Sectioning

As mentioned in the historical overview presented in Sec. 1.1, the stochastic late reverberation models developed around the work of Schroeder [1] describe the steady-state or diffuse field response of an acoustic space, i.e. once a sufficient number of uncorrelated plane waves have accumulated. In terms of an IR, this implies that a certain number of reflections of the direct sound need to have been generated, a process that occurs over a period of time following the arrival of the direct sound known as the "mixing time" (t_{mix}). This condition, described by Schroeder [2] as a minimum echo density condition, leads us to our first main time-domain regime separation, as illustrated in Fig. 2.1a. Before the mixing time, the SRIR is considered to be dominated by the arrival of spatially coherent early reflections, highly correlated with the direct sound impulse, while the late reverberation tail that occurs after the mixing time obeys the stochastic properties of the diffuse field response, subject to an exponentially decaying power envelope. Note that in Sec. 2.1.3 we will see how the diffuse field condition can be relaxed while retaining the necessary stochastic plane wave properties; however, as seen in Sec. 1.1 [4], the historical models on which the concept of the mixing time is based all rely on this strict assumption.



(a) Time-sectioned IR model.



(b) Time-frequency IR model domains.

Figure 2.1: Schematic time (a, above) and time-frequency (b, below) representations of the different regimes in the general RIR model. Note that the "early" label in (b, below) includes both the deterministic first reflections and the semi-stochastic "cluster", shown separately in (a, above).

In addition to this crucial temporal separation, Schroeder and Kuttruff [8] showed that a modal density condition is also required in order for the stochastic late reverberation model to be valid: as seen once again in Sec. 1.1, this describes how the number of room modes increases with frequency until the transfer function peaks overlap so much as to be deterministically indistinguishable. Schroeder and Kuttruff [8] relate this condition to an average frequency spacing of transfer function maxima $\mathrm{E}\{\Delta f_{\text{max}}\} \simeq 3.9/T_{60}$, which is achieved above a frequency limit now appropriately known as the

Schroeder frequency (f_{Sch}). Combined with the mixing time, this results in a general time-frequency sectioning of the SRIR model and defines the validity domain for the stochastic late reverberation approach (see Fig. 2.1b).

Following Kahle [29] and Jot [74], the early section is then further temporally refined by separately considering the direct sound, the "true" early reflections, and a semi-stochastic transition regime termed the "cluster". This view is shown schematically in Fig. 2.1a. Though a thorough investigation of the cluster is out of the scope of this thesis, a brief description of its properties and some potential modelling approaches will be given at the end of Sec. 2.1.2.

Finally, since we are interested in modelling and analyzing real-world SRIR measurements, the presence of background noise must evidently be taken into account. In this work, a single term will be used to encompass all noise sources, defined as any signal that is not an integral part of the SRIR. As such, this term includes both sensor-self noise from the SMA transducers as well as external acoustic noises. Though no explicit assumption will need to be made on the actual signal content of the noise, it will nevertheless be assumed to be stationary with a constant power spectrum.

In its most general form, the directional space-time-frequency representation of an SRIR can therefore be written as the sum of these six sections:

$$\begin{aligned} H(f, \boldsymbol{\Omega}, t) = H_{\text{dir}}(f, \boldsymbol{\Omega}_0, t_0) + H_{\text{early}}(f, \boldsymbol{\Omega}, t) + H_{\text{clust}}(f, \boldsymbol{\Omega}, t) \\ + H_{\text{late}}(f, \boldsymbol{\Omega}, t) + H_{\text{modal}}(f, \boldsymbol{\Omega}, t) + N(f, \boldsymbol{\Omega}, t), \end{aligned} \tag{2.1}$$

where $\boldsymbol{\Omega}_0$ and t_0 represent the DoA and ToA of the direct sound, respectively, and $N(f, \boldsymbol{\Omega}, t)$ is the generic stationary background noise term.

A Note on the Low-Frequency Modal Domain

As mentioned in the opening to this section, detailed treatment of the low-frequency modal region illustrated in Fig. 2.1b will be left to future work. In theory, below f_{Sch}, individual amplitude peaks corresponding to discrete room modes should be resolvable on the omnidirectional RIR's frequency response, i.e. the spectrum of the zeroth-order component $H_{0,0}(f)$. It would therefore be tempting to model this region as the sum of M damped sinusoidal modes:

$$h_{\text{modal}}(t) = \sum_{k=1}^{M} A_k e^{-\gamma_k t} e^{2i\pi f_k t} \Leftrightarrow H_{\text{modal}}(f) = \sum_{k=1}^{M} \frac{A_k}{2i\pi(f - f_k) + \gamma_k}, \ \forall\ f < f_{\text{Sch}}, \tag{2.2}$$

where A_k, γ_k, and f_k are the modal amplitudes, damping coefficients and resonant frequencies, respectively.

Such an approach would enable the definition of a parametric fitting problem for estimating the $3M$ total parameter values. However, any resolution strategy would inevitably have to rely on two pieces of *a priori* information: the number of room modes M to fit and the maximum frequency up to which to search, i.e. the Schroeder frequency (f_{Sch}). Though methods for estimating f_{Sch} from the room transfer function (RTF) are available in the literature [75], determining a value for M without external knowledge of the space's geometry and absorption properties is in itself a complex problem that would require a dedicated research effort.

Furthermore, the large majority of measured spaces present T_{60} vs. volume relationships that result in very low f_{Sch} values with respect to human audition, according to Eq. 1.1. In effect, the discrete room modes that could be modelled are usually confined to a frequency range within which the human auditory system is incapable of resolving such fine frequential details. Considering the additional masking effects from the stochastic mid-frequency reverberation, it will be assumed that a full modelling of the low-frequency modal domain will not be necessary for the present work.

$H_{\text{modal}}(f, \mathbf{\Omega}, t)$ will thus be treated in terms of its energy envelope only. In other words, the actual signal content will not be reconstructed, but the noise floor and decay envelope analysis and treatment procedures (Sec. 2.5) will nonetheless be applied to ensure continuity with the other domains.

2.1.2/ Early Reflections

The fundamental underlying approach for modelling the discrete early reflections is that of the classic image-source (IS) description [14]. In this view, the individual specular echoes are highly correlated with the direct sound as well as with each other. Furthermore, following Polack [6], it is assumed that a limited number of reflections are simultaneously incident in any given short time frame; in other words, the early sound field is also highly coherent.

In order to tie the discrete IS view to the analysis of an SMA-measured signal, we make use of the Herglotz wavefunction formalism [76, p. 60]. The Herglotz approach enables the description of a field containing any number of incident plane waves over a given surface, and whose respective presence at any given point on the surface is given by a probability function called the *Herglotz kernel*.

The Herglotz kernel is a nice continuous (at least $\mathcal{C}^1$) and square integrable (i.e. L^2) function over the given surface, so the problem of detecting the positions of the incident plane waves on the surface becomes a peak detection problem on the kernel function. In the case of a spherical surface S^2, the Herglotz wavefunction $v(k, \mathbf{r})$ can be written as:

$$v(k, \mathbf{r}) = \int_{S^2} g(\mathbf{\Omega}_d) e^{ik\mathbf{r}\cdot\mathbf{d}} d\mathbf{\Omega}_d, \tag{2.3}$$

where $\mathbf{r}$ is the position vector on S^2, $g(\mathbf{\Omega}_d)$ is the Herglotz kernel function, and $\mathbf{d}$ is the unitary vector pointing towards the incident direction $\mathbf{\Omega}_d$.

The term $e^{ik\mathbf{r}\cdot\mathbf{d}}$ simply describes a plane wave, so the formalism initially presented in Sec. 1.2.1 can be reprised to write[1]:

$$X_{\text{early}}(k, \mathbf{\Omega}) = \sum_{l=0}^{\infty} b_l(kr_s) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) \int_{S^2_{r_s}} g(\mathbf{\Omega}_d) Y^*_{l,m}(\mathbf{\Omega}_d) d\mathbf{\Omega}_d, \tag{2.4}$$

i.e. the field generated by a Herglotz wavefunction on the surface of a rigid SMA of radius $r = r_s$, decomposed over the SH domain. Note the use of the generic field variable X in Eq. 2.4: this is not yet a description of the early SRIR region $H_{\text{early}}(f, \mathbf{\Omega}, t)$ as we would like to obtain.

First, as in Ch. 1, the dependence on the wavenumber k can be replaced by a dependence on the frequency f since the field is being described at a fixed $r = r_s$ (i.e. on the surface of the SMA). Then Eq. 2.4 must be extended to include the time-frequency representation of the measured field. This can be done by assigning a total real amplitude $A(f, t)$ and average phase $\phi_e(f, t)$ to the Herglotz wavefunction over the incident direction sphere $\mathbf{\Omega}_d$ [i.e. separate from the SMA's self-scattering effects contained in $b_l(f)$]. As such, any arbitrary measured field can be considered: its directional properties are carried solely by the Herglotz kernel, and the detection of any incident reflections present is indeed a peak detection problem on $g(\mathbf{\Omega}_d, t)$.

The resulting model for $H_{\text{early}}(f, \mathbf{\Omega}, t)$ can therefore be written:

$$H_{\text{early}}(f, \mathbf{\Omega}, t) = A(f, t) e^{i\phi_e(f,t)} \sum_{l=0}^{\infty} b_l(kr_s) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) \int_{S^2_{r_s}} g(\mathbf{\Omega}_d, t) Y^*_{l,m}(\mathbf{\Omega}_d) d\mathbf{\Omega}_d, \quad \forall\ t < t_{\text{mix}}. \tag{2.5}$$

[1] A more detailed description of the Herglotz wavefunction formalism as used in this work is given in Appx. A.1.

In the context of SMA processing, as described in Sec. 1.2.1, the continuous integral over $S^2_{r_s}$ must then be discretized using a weighted sum, and the infinite SH series must be truncated to a maximum order L. However, if the orthogonality properties of the SH basis are sufficiently verified by the SMA layout, Eq. 2.5 can be simplified by taking its discrete SH transform:

$$\begin{aligned}\widetilde{H}^{\text{early}}_{l,m}(f,t) &= A(f,t)e^{i\phi_e(f,t)} \sum_{l'=0}^{L} b_{l'}(f) \sum_{m'=-l'}^{l'} \sum_{q=1}^{Q} \beta_q Y_{l',m'}(\mathbf{\Omega}_q) Y^*_{l,m}(\mathbf{\Omega}_q) \sum_{d=1}^{D} \alpha_d g(\mathbf{\Omega}_d, t) Y^*_{l',m'}(\mathbf{\Omega}_d) \\ &\approx A(f,t)e^{i\phi_e(f,t)} \sum_{l'=0}^{L} b_{l'}(f) \sum_{m'=-l'}^{l'} \delta_{l,l'}\delta_{m,m'} \sum_{d=1}^{D} \alpha_d g(\mathbf{\Omega}_d, t) Y^*_{l',m'}(\mathbf{\Omega}_d) \\ &= A(f,t)e^{i\phi_e(f,t)} b_l(f) \sum_{d=1}^{D} \alpha_d g(\mathbf{\Omega}_d, t) Y^*_{l,m}(\mathbf{\Omega}_d),\end{aligned} \tag{2.6}$$

where β_q and α_d are the quadrature weights for the directional layouts of the Q SMA transducers and the D incident look directions, respectively.

This discretized system of linear equations can now be written in matrix form:

$$\mathbf{h}^{\text{early}}_{\text{SH}}(f,t) = A(f,t)e^{i\phi_e(f,t)}\mathbf{D}_{\text{SH}}(f)\mathbf{g}(t), \tag{2.7}$$

where $\mathbf{h}^{\text{early}}_{\text{SH}}$ is an $(L+1)^2 \times 1$ column vector containing the early SRIR signal in the SH domain, $\mathbf{D}_{\text{SH}}$ is an $(L+1)^2 \times D$ matrix containing the terms $\alpha_d b_l(f) Y^*_{l,m}(\mathbf{\Omega}_d)$, and $\mathbf{g}$ is a $D \times 1$ column vector containing the Herglotz probabilities over the D look directions. Note that an equivalent "HOA" formulation can also be obtained by applying the encoding filters $a_l(f)$, in which case the matrix $\mathbf{D}_{\text{HOA}}$ would contain elements $\alpha_d c_l(f) Y^*_{l,m}(\mathbf{\Omega}_d)$, with $c_l(f) = a_l(f)b_l(f)$ the resulting frequency response, and the signal vector would then be $\mathbf{h}^{\text{early}}_{\text{HOA}}(f,t)$ with the encoding filters applied as well.

This formulation naturally leads to an inverse problem in which $\tilde{\mathbf{g}}$ is estimated by inverting $\mathbf{D}_{\text{SH}}$ (or equivalently $\mathbf{D}_{\text{HOA}}$):

$$\tilde{\mathbf{g}}(t) = \frac{\mathbf{D}^{-1}_{\text{SH}}(f)\mathbf{h}^{\text{early}}_{\text{SH}}(f,t)}{A(f,t)e^{i\phi_e(f,t)}} = \frac{\mathbf{D}^{-1}_{\text{SH}}(f)\mathbf{h}^{\text{early}}_{\text{SH}}(f,t)}{H_{0,0}(f,t)}, \tag{2.8}$$

where $H_{0,0}(f,t)$ is the omnidirectional zeroth-order component of $\mathbf{h}^{\text{early}}_{\text{SH}}$, which corresponds precisely to the total magnitude and average phase over the sphere described by $A(f,t)e^{i\phi_e(f,t)}$ above.

This approach is further detailed in Sec. 2.4 in comparison to the state of the art DoA estimation methods presented in Ch. 1, and will crucially explore the spatial resolution limitations on $g(\mathbf{\Omega}_d)$ and, in consequence, the maximum detectable number of simultaneous reflections.

A Note on the Semi-Stochastic "Cluster" Transition Domain

In a majority of real-world measurements, the above assumption of a "limited number" of simultaneous early reflections is rapidly invalidated. Nonetheless, high echo density does not always coincide with the spatial incoherence condition used to define the mixing time in this work (a point that will be further explored in Sec. 2.3 and followed up on in Ch. 4). As such, high echo density can often be encountered during the early segment of an RIR without being considered late reverberation.

Under these conditions, the Herglotz model can only serve to indicate the predominant echoes in any given time frame. The remaining signal is hence considered the "cluster", i.e. dense yet still relatively coherent reflections that can neither be individually resolved nor entirely treated as late reverberation. Acoustically, these arise from non-specular diffusive and diffractive reflection phenomena in the space (and are, in fact, what eventually lead to the emergence of the incoherent late field).

In IS-based RIR simulation methods, these effects are often addressed by including an acoustical

radiosity model, which is essentially a boundary element method (BEM) for calculating the radiation density of the room's surfaces [77]. To the author's knowledge, little to no work has been done on this topic from a purely analytical view such as the one taken in this research project, i.e. using no *a priori* information on a space's geometry.

As it would merit a complete thorough investigation, modelling the cluster is therefore left to future work. In this thesis, the presence of the cluster will be taken into account by (a) recognizing that any detected echoes represent only the most predominant specular reflections and (b) ensuring that any manipulation of the early SRIR section preserves its underlying space-time-frequency characteristics (i.e. no signal information is discarded).

A Note on the Overlap Between Early Reflections and the Modal Domain

It is worth mentioning a singular region of the time-frequency sectioning shown in Fig. 2.1b, namely the lower-left corner in which the modal and early reflection domains appear to coexist. Indeed, in a geometrical acoustics approach, the direct sound and its specular reflections are considered to be ideal impulses that, in theory, cover all frequencies.

In reality, however, the geometrical acoustics description is only valid for frequencies with wavelengths much smaller than the characteristic dimensions of the room [4]. In other words, depending on the size of the room, there is a frequency limit below which the wavelengths of the traveling sound are of the same order of magnitude as the surfaces they encounter and their reflections can thus no longer be considered entirely specular.

In large rooms, this frequency limit can be so low as to be entirely inconsequential, but in spaces with complex geometries, small obstacles, or apertures to coupled volumes, frequency-dependent reflection phenomena may become prevalent. For IS-based simulators, this has often meant augmenting the geometrical results with direct low-frequency solutions of the wave equation, e.g. using finite-difference time-domain (FDTD) methods. Once again, to the author's knowledge, little to no work has been undertaken to characterize this limit from a "blind" analytical standpoint.

Finally, it should be noted that the lower frequency limit for the geometrical acoustics approach is an entirely separate phenomenon from the previously discussed Schroeder frequency limit of the stochastic late reverberation model. Indeed, repeating orders of a given specular reflection in a geometrical description would coincide with a particular damped room mode (and vice-versa). For the early segment, a geometrical echo detection approach and the damped sine model of Eq. 2.2 would thus be equivalent in the range between this "specular limit" and f_{Sch} (assuming the former is lower).

A practical approach would therefore be to make use of the damped sine model all the way up to f_{Sch} throughout the SRIR. Since the modal description has been left to future work and all $f < f_{\text{Sch}}$ are being treated energetically according to their decay envelope, the geometrical echo detection model detailed above will simply be restricted to $f > f_{\text{Sch}}$. As we will see in Sec. 2.4, there is an additional interest in limiting echo detection to higher frequencies in SMA processing due to the directivity properties described in Ch. 1.

2.1.3/ Late Reverberation

As stated in Sec. 2.1.1 and following from the discussion in Ch. 1, the late reverberation tail is characterized by an exponentially decaying incoherent field composed of an "infinite" number of uncorrelated plane waves. Traditionally, this field has been further assumed to be fully diffuse, i.e. energetically isotropic and homogeneous on top of being incoherent. In complex spaces and under the directional representation offered by SMA measurements, the condition of isotropy can easily be invalidated (and, by definition, so too that of homogeneity, though it is less of a factor in the analysis of single SRIRs).

The underlying acoustical causes for anisotropic late reverberation are wide and varied but usually stem from particular combinations of uneven geometries and inhomogeneous reflection/absorption properties. Though these will be briefly discussed in an intuitive sense in Ch. 5, it is not the aim of this work to specifically characterize them, as has recently been done by Nolan et al. [63] and Berzborn and Vorländer [68]. Rather, the observation of their effects are an additional motivation for considering a model that allows for the analysis and treatment of anisotropic late reverberation.

As such, and reprising the SH PWD formalism presented in Sec. 1.2.1, the late reverberation tail measured by an SMA is described here in its most general form as:

$$H_{\text{late}}(f,\mathbf{\Omega},t)=\sum_{l=0}^{\infty}b_l(kr_s)\sum_{m=-l}^{l}Y_{l,m}(\mathbf{\Omega})\int_{S^2_{r_s}}\sqrt{P(f,\mathbf{\Omega}_d,t)}e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)}Y^*_{l,m}(\mathbf{\Omega}_d)d\mathbf{\Omega}_d, \quad \forall\ f>f_{\text{Sch}},\ t>t_{\text{mix}}, \tag{2.9}$$

where $P(f,\mathbf{\Omega}_d,t)$ is the field's directional time-frequency power envelope and $\hat{\phi}(f,\mathbf{\Omega}_d,t)\sim\mathcal{U}[0,2\pi)$ is a stochastic plane wave phase uniformly distributed over $[0,2\pi)$ and independent with respect to time, frequency, and incident direction. The random nature of $\hat{\phi}(f,\mathbf{\Omega}_d,t)$ translates the late field's spatial incoherence (due to the uncorrelated nature of the plane waves), as well as its density (each and every space-time-frequency point is assumed to contain an independent incident plane wave).

The power envelope is assumed to be exponentially decaying in time, with potential smooth variations in both frequency and direction. The discretized set of directions $\mathbf{\Omega}_d$ thus represents an idealized spatial sampling of the underlying smooth power distribution $P(f,\mathbf{\Omega},t)$. Specifically, $P(f,\mathbf{\Omega},t)\in L^2$ is considered to be at least $\mathcal{C}^1$ smooth over $\mathbf{\Omega}$.

We can furthermore consider the coupling of C exponential decays in order to account for the multi-slope envelopes found in certain configurations of coupled volumes:

$$P(f,\mathbf{\Omega}_d,t)=\sum_{j=1}^{C}P_{0,j}(f,\mathbf{\Omega}_d)e^{-2\gamma_j(f,\mathbf{\Omega}_d)t}, \tag{2.10}$$

where $P_{0,j}(f,\mathbf{\Omega}_d)$ is the direction-dependent initial power spectrum of each slope and $\gamma_j(f,\mathbf{\Omega}_d)$ are their decay rates, related to the Sabine 60 dB reverberation time by:

$$T_{60}=\frac{3\ln(10)}{\gamma}. \tag{2.11}$$

In the same manner as Eq. 2.6 for the early segment, Eq. 2.9 can be simplified under the (discretized and truncated) SH transform:

$$\widetilde{H}^{\text{late}}_{l,m}(f,t)=b_l(f)\sum_{d=1}^{D}\alpha_d\sqrt{P(f,\mathbf{\Omega}_d,t)}e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)}Y^*_{l,m}(\mathbf{\Omega}_d), \tag{2.12}$$

or in matrix form:

$$\mathbf{h}^{\text{late}}_{\text{SH}}(f,t)=\mathbf{D}_{\text{SH}}(f)\mathbf{p}(f,t), \tag{2.13}$$

where $\mathbf{D}_{\text{SH}}$ is the same as in Eq. 2.7 and $\mathbf{p}$ is a $D\times 1$ column vector containing the field model elements $\sqrt{P(f,\mathbf{\Omega}_d,t)}e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)}$. From this view, the spatial incoherence of the model can be demonstrated. Taking the square modulus of Eq. 2.13 yields:

$$\mathbf{S}(f,t)=\mathbf{D}(f)\mathbf{p}(f,t)\mathbf{p}^{\text{H}}(f,t)\mathbf{D}^{\text{H}}(f)=\mathbf{D}(f)\mathbf{P}(f,t)\mathbf{D}^{\text{H}}(f), \tag{2.14}$$

where $.^{\text{H}}$ denotes the Hermitian or conjugate transpose. The square $(L+1)^2\times(L+1)^2$ matrix $\mathbf{S}(f,t)$

thus contains all the cross- and auto-spectra in the SH domain, while the central term $\mathbf{P}(f,t)$ (a $D \times D$ square matrix) carries the directional equivalents:

$$\mathcal{P}_{d,j}(f,t) = \sqrt{P(f,\mathbf{\Omega}_d,t)P(f,\mathbf{\Omega}_j,t)}e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)}e^{-i\hat{\phi}(f,\mathbf{\Omega}_j,t)}. \tag{2.15}$$

Finally, taking the expectation value of Eq. 2.14 to obtain the SH-domain covariance matrix $\mathbf{R}_{\text{SH}}(f,t) = \text{E}\{\mathbf{S}(f,t)\}$ shows that the directional covariance matrix $\mathbf{R}_{\text{dir}}(f,t) = \text{E}\{\mathbf{P}(f,t)\}$ reduces to a diagonal matrix:

$$\begin{aligned}\text{E}\{\mathcal{P}_{d,j}(f,t)\} &= \sqrt{P(f,\mathbf{\Omega}_d,t)P(f,\mathbf{\Omega}_j,t)}\text{E}\left\{e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)}e^{-i\hat{\phi}(f,\mathbf{\Omega}_j,t)}\right\} \\ &= P(f,\mathbf{\Omega}_d,t)\delta_{d,j},\end{aligned} \tag{2.16}$$

since the random variable $\hat{\phi}$ is independent and identically distributed (i.i.d.) over the three variables of direction, time, and frequency. Normalizing $\mathbf{R}_{\text{dir}}(f,t)$ by the power spectral densities (PSD) $\sqrt{\text{E}\{P_d(f,t)\}\,\text{E}\{P_j(f,t)\}}$ would therefore yield the identity matrix, demonstrating that the field signal model contained in $\mathbf{p}(f,t)$ is indeed spatially incoherent.

It is also worth noting here that in the case of an isotropic late reverberation field (i.e. both P_0 and T_{60} independent of direction), ideal diffuse conditions are automatically recovered. In such a case, the power decay term $P(f,t)$ no longer plays a role in the spatial discretization and the SH-domain covariance matrix $\mathbf{R}_{\text{SH}}(f,t)$ is also diagonal, as shown by Epain and Jin [62].

This concludes the essence of the underlying signal model for this research project. The remaining sections in this chapter will turn to describing the various analysis methods that have been developed in order to estimate several of its properties and parameters. These methods form the "toolbox" with which the experimental treatment and manipulation methods of Ch. 3 will be implemented (and later evaluated in Ch. 4).

2.2 / Spatial Decomposition Methods

As we will see throughout the following sections, as well as in the next chapter, the definition of a directional SRIR representation (i.e. a directional RIR or DRIR) is a fundamental tool for the estimation and eventual modification of many (if not all) of the signal model properties described above. A DRIR should itself be seen as an attempt to obtain an estimate of a spatial sampling of the measured SRIR over the sphere: in other words, we seek to approximate a directional discretization of the general signal model in Eq. 2.1 over a set of S look or steering directions. This directional representation must therefore preserve as many of the space-time-frequency characteristics of the measured SRIR as possible, at least within the inherent limitations of a compact rigid sphere SMA measurement. This implies several important considerations.

First, the set of S look directions should be arranged so as to form a comprehensive and regular sampling scheme over the entire sphere. Assuming that the SMA's mode strengths have been properly compensated for (i.e. through HOA encoding), the spatial decomposition should also be defined in a broadband manner so as to preserve the SRIR's spectral characteristics. Finally, since we would like to be able to transform back and forth between the HOA-SRIR and DRIR representations without loss of generality, the decomposition (and more specifically its matrix form) should be invertible.

We will therefore favour a steered beamforming approach as our chosen spatial decomposition method. For the reasons hinted at above, we will restrict our attention to broadband axisymmetric beamformer designs (i.e. with constant, order-dependent weights). Furthermore, to begin the discussion in the simplest terms, we will first make use of the "natural" PWD beamformer described in Sec. 1.2.1 before discussing alternative beamformer designs towards to end of the section.

Reprising the beamforming view given in Eq. 1.13 with the natural PWD weights $d_l = 1$, the directional decomposition of the signal model described above can now be written (with HOA encoding/SMA mode strength correcting filters applied):

$$\begin{aligned}\mathbf{h}_{\text{dir}}(f,t) &= \mathbf{B}\mathbf{h}_{\text{HOA}}(f,t) \\ &= \mathbf{B}\mathbf{D}_{\text{HOA}}(f)\mathbf{h}(f,t),\end{aligned} \tag{2.17}$$

where $\mathbf{B}$ is an $S \times (L+1)^2$ matrix with elements $\alpha_s Y_{l,m}(\mathbf{\Omega}_s)$, $\mathbf{\Omega}_s$ is a grid of S steering directions on the sphere (assumed to be different to the set $\mathbf{\Omega}_d$ used to define $\mathbf{D}_{\text{SH}}$ and $\mathbf{D}_{\text{HOA}}$), and $\mathbf{h}$ is a $D \times 1$ column vector containing a spatially discretized version of the general SRIR signal model from Eq. 2.1.

As shown above (as well as in Massé et al. [78]), one of the most important signal properties that must be preserved under this decomposition is the spatial incoherence of the late sound field: it is indeed a necessary condition for the faithful representation (and therefore reproduction) of the late reverberation tail. However, it also places an important limitation on the beamforming matrix $\mathbf{B}$.

Equation 2.17 highlights the fact that such a directional decomposition is inherently limited by the maximum SH order L and the resulting number of components $(L+1)^2$. Assuming a theoretical SRIR $\mathbf{h}(f,t)$ defined over a large number $D \gg (L+1)^2$ of incident directions, its truncated SH transform $\mathbf{h}_{\text{HOA}}(f,t)$ following its measurement by an SMA would nonetheless be limited to the $(L+1)^2$ linearly independent basis functions available [32] [39] (see Sec. 1.2.1 once again).

Although any phase incoherence is preserved through the initial truncated SH transform by the summing of independent stochastic variables, there are now only $(L+1)^2$ independent components from which to perform the directional decomposition. In order to ensure that the set of beamformed signals $\mathbf{h}_{\text{dir}}(f,t)$ are also independent (thereby preserving spatial incoherence), the beamforming matrix $\mathbf{B}$ must therefore not only be full-rank but also limited to $S \leq (L+1)^2$ look directions.

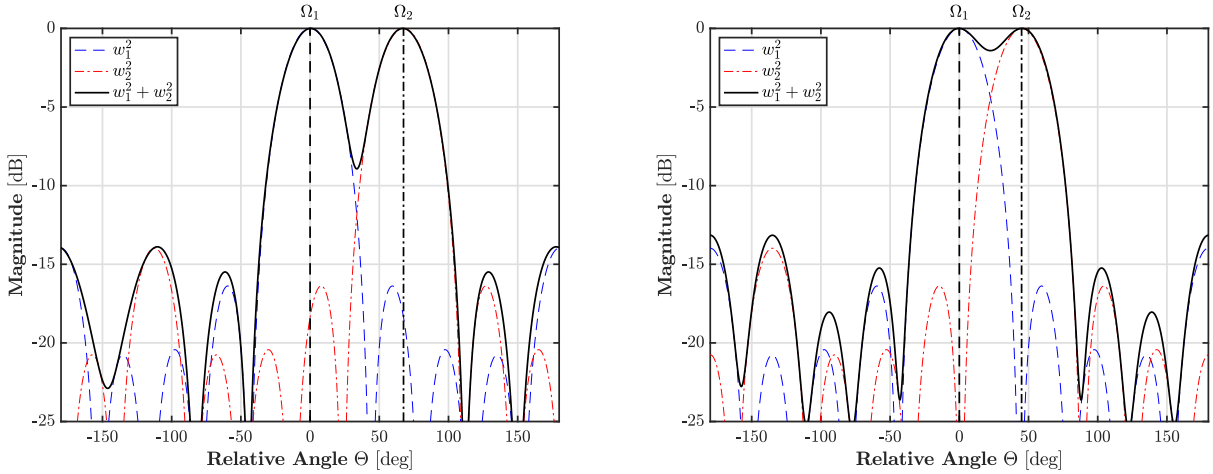
As mentioned above, we furthermore require that the DRIR representation must be able to be transformed back to the SH domain after directional analysis/treatment/manipulation for posterior reproduction in HOA-based spatialization systems. This implies that we must choose the maximum $S = (L+1)^2$ such that the re-encoding step preserves linear independence in the same manner as the beamforming (i.e. $\mathbf{B}$ is a full-rank square matrix that can simply be inverted for re-encoding).

However, successfully preserving the late field's spatial incoherence does not only depend on ensuring linear independence in the directional decomposition. Indeed, as hinted at in Sec. 1.2.1, the beamforming view of the PWD highlights the issue of lobe overlap when generating a set of directional signals in an attempt to cover the sphere. To illustrate this, Fig. 2.2 shows the interaction between the directivity functions of two natural PWD look directions for three different separation angles $\Delta\mathbf{\Omega}$.

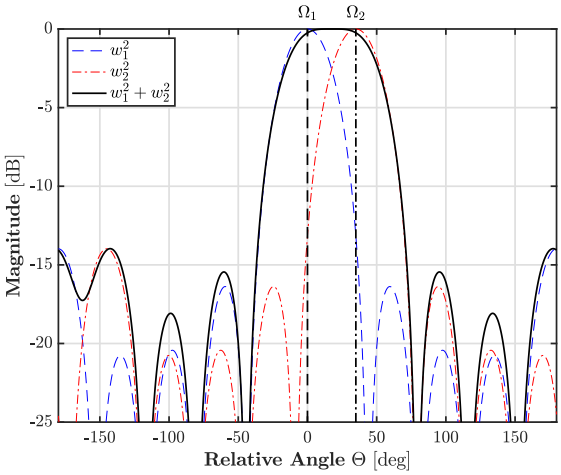
The first (Fig. 2.2a) is an arbitrarily chosen "large" separation angle with minimal overlap between the two main lobes. At this distance, the look directions are well separated and their combined power response presents a marked gap between them. When the separation angle $\Delta\mathbf{\Omega} = \Theta_{\text{PN}}$ (Fig. 2.2b), i.e. the main lobe peak-to-null half-width (see once more Sec. 1.2.1), the gap is significantly reduced at the cost of increased main lobe overlap. Finally, as the angular separation reaches a "point of equal energy", denoted Θ_{eqE} (Fig. 2.2c), the gap disappears and the combined power response flattens, but with greater overlap once again. This equal energy point is defined as the largest angular separation for which the maximum of the combined response is at the midpoint between the two look directions.

More importantly, however, any amount of lobe overlap inevitably reduces independence between the beamformed signals, since they end up partly describing shared areas on the sphere. Extended to $S = (L+1)^2$ look directions laid out over the sphere (and thus all overlapping with each other), the effect on the resulting DRIR's ability to describe a spatially incoherent field, at least in terms of one defined as above using $\mathbf{p}(f,t)$ with $D \gg (L+1)^2$, can become drastic.

This ability is additionally affected by the frequency dependence of the PWD directivity (as shown



(a) $\Delta\boldsymbol{\Omega} = 67.5°$. (b) $\Delta\boldsymbol{\Omega} = \Theta_{\text{PN}} = 45°$.



(c) $\Delta\boldsymbol{\Omega} = \Theta_{\text{eqE}} = 35°$.

Figure 2.2: Combined 4[th]-order broadband directivity patterns for two natural PWD look directions with an angular separation of $\Delta\boldsymbol{\Omega} = 67.5°$ (a, top left), 45° (b, top right), and 35° (c, bottom), in terms of the central angle Θ along the great circle joining them on the unit sphere. Θ_{PN} represents the main lobe peak-to-null half-width (see Sec. 1.2.1) and Θ_{eqE} represents the "point of equal energy", defined as the largest angular separation for which the maximum of the combined response is at the midpoint between the two look directions.

in Fig. 1.5). Though Fig. 2.2 shows the ideal broadband directivity functions (equivalent to those shown in Fig. 1.4 but here in square-magnitude dB), these are once again never fully achievable due to the combined frequency response $c_l(f)$ of an HOA-encoded SMA-measured plane wave (as also described in Sec. 1.2.1). Incoherence preservation is thus further limited by increased overlap both at lower frequencies where higher SH orders are not present (see Fig. 1.3), and higher frequencies above the spatial aliasing limit where side lobes become increasingly present (see once more Fig. 1.5).

In order to quantify the complete overlap over the entire sphere, we propose the following definition of a directivity energy ratio $W_{\text{ER}}(\boldsymbol{\Omega}_s, \boldsymbol{\Omega})$, using the ideal broadband directivities $w_L(\boldsymbol{\Omega}_s, \boldsymbol{\Omega})$ for each one of the S look directions:

$$W_{\text{ER}}(\boldsymbol{\Omega}_s, \boldsymbol{\Omega}) = \frac{w_L^2(\boldsymbol{\Omega}_s, \boldsymbol{\Omega})}{\sum_{s'=1}^{S}(1 - \delta_{s,s'})w_L^2(\boldsymbol{\Omega}_{s'}, \boldsymbol{\Omega})}. \tag{2.18}$$

In other words, $W_{\text{ER}}(\boldsymbol{\Omega}_s, \boldsymbol{\Omega})$ quantifies the energy ratio between the directivity function for the look direction $\boldsymbol{\Omega}_s$ and the total energy response from all the other steering points on the current layout grid. As a result, any point on the sphere where $W_{\text{ER}}(\boldsymbol{\Omega}_s, \boldsymbol{\Omega}) > 1$ can be considered to "belong" to the

look direction $\mathbf{\Omega}_s$ in the sense that its directivity has a stronger response there than all the others combined; that point can thus be seen as being uniquely represented by the $\mathbf{\Omega}_s$ beam.

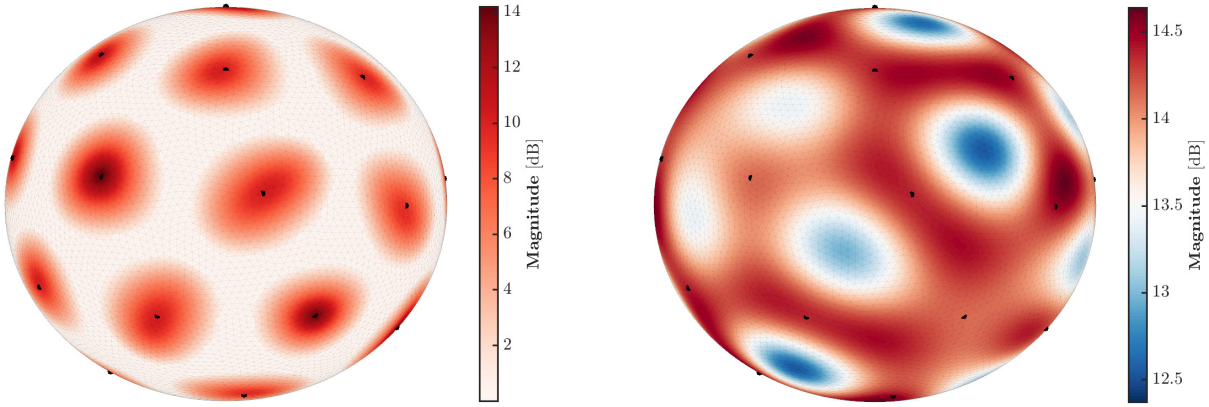
Ideally, therefore, all points on the sphere would be uniquely "assigned" to one of the look directions on the layout grid. Of course, this may in general not be the case, but we can at least quantify the total surface of the sphere that is thus uniquely described by defining a "unique coverage" factor:

$$\mathrm{UC} = \frac{1}{4\pi}\int_{\mathbf{\Omega}} \mathcal{D}_{\mathrm{UC}}(\mathbf{\Omega})d\mathbf{\Omega}, \text{ with } \mathcal{D}_{\mathrm{UC}}(\mathbf{\Omega}) = \begin{cases} 1 & \text{if } \overline{W}_{\mathrm{UC}}(\mathbf{\Omega}) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.19}$$

where $\overline{W}_{\mathrm{UC}}(\mathbf{\Omega}) = \sum_{s=1}^{S} W_{\mathrm{UC}}(\mathbf{\Omega}_s, \mathbf{\Omega})$ and

$$W_{\mathrm{UC}}(\mathbf{\Omega}_s, \mathbf{\Omega}) = \begin{cases} W_{\mathrm{ER}}(\mathbf{\Omega}_s, \mathbf{\Omega}) & \text{if } W_{\mathrm{ER}}(\mathbf{\Omega}_s, \mathbf{\Omega}) > 1. \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

The global average directivity energy ratio $\overline{W}_{\mathrm{ER}} = \frac{1}{4\pi S}\int_{\mathbf{\Omega}} \sum_{s=1}^{S} W_{\mathrm{ER}}(\mathbf{\Omega}_s, \mathbf{\Omega})d\mathbf{\Omega}$ will also be a useful quantity to consider: the larger it is, the more each look direction is uniquely describing a region of the sphere and therefore the less it is overlapping with the other beam responses. An illustration of the unique coverage $\overline{W}_{\mathrm{UC}}(\mathbf{\Omega})$ is given in Fig. 2.3a for the 4^{th}-order natural PWD beamformer applied over a 25-point Fliege-Maier [79] look direction layout grid. The dB-scale visualization of Fig. 2.3a is cut off at 0 dB, and all points on the sphere with a positive dB-scale $\overline{W}_{\mathrm{UC}}$ magnitude can be considered "uniquely assigned" to a given look direction. Ideally, of course, this would be the case for every single point on the sphere (i.e. there would be no white regions between look directions in Fig. 2.3a).



(a) Unique coverage, $\overline{W}_{\mathrm{UC}}(\mathbf{\Omega})$, UC = 68.9%. (b) Total directivity coverage, $W(\mathbf{\Omega})$, $\sigma_W = 0.400$ dB.

Figure 2.3: Illustrations of (a, left) the unique coverage measure $\overline{W}_{\mathrm{UC}}(\mathbf{\Omega}) = \sum_{s=1}^{S} W_{\mathrm{UC}}(\mathbf{\Omega}_s, \mathbf{\Omega})$, where $W_{\mathrm{UC}}(\mathbf{\Omega}_s, \mathbf{\Omega})$ is given by Eq. 2.20, and (b, right) the total directivity coverage $W(\mathbf{\Omega})$ given by Eq. 2.21. This example makes use of the natural PWD beamformer on a 25-point Fliege-Maier [79] look direction layout grid, which is represented by the black points on the sphere.

On top of ensuring that the individual directivities overlap with each other as little as possible, the total coverage over the sphere should also be as "flat" as possible. Indeed, since the late field is composed of an infinity of plane waves under an arbitrary spatial power distribution, the spatial decomposition must not privilege any particular directions or subspaces on the sphere. To this end, a "total directivity coverage" function can be defined for the S look directions over the sphere:

$$W(\mathbf{\Omega}) = \sum_{s=1}^{S} w_L^2(\mathbf{\Omega}_s, \mathbf{\Omega}), \tag{2.21}$$

and its standard deviation σ_W (in dB) provides a quantitative measure of total coverage "flatness". The total directivity coverage is illustrated in Fig. 2.3b, once again using the 4^{th}-order natural PWD beamformer on a 25-point Fliege-Maier grid.

Note that flat coverage can be achieved exactly using t-designs with $S > (L+1)^2$ [54], but this would break the linear independence conditions detailed above. It should also be pointed out that the choice made to define this total coverage using the ideal broadband directivities is not only simpler than taking into account the true frequency dependence, it also implies any optimization will inherently improve performance across the spectrum (albeit globally).

The larger question of whether a frequency-dependent optimization of the spatial decomposition and DRIR generation can be used to address some of the limitations of the current method will be discussed in the final Conclusion and is thus mostly left to future work. There is, however, one line of optimization readily available here: throughout this section (beginning at Eq. 2.17), we have simply assumed the use of the natural PWD beamformer, i.e. with weights $d_l = 1$.

Although the natural beamformer can be shown to provide the maximum theoretical directivity index (DI) [45], a closer look at its beampattern (see Fig. 2.2) reveals a suboptimal characteristic: the most important secondary lobe is located at the "rear" (i.e. $\Theta = \pm 180°$). That is, besides the main lobe, the next most represented spherical point in a given beam is the one diametrically opposed to the look direction $\mathbf{\Omega}_s$. There is a relatively simple reason for this: from a beamforming view, the natural beampattern can be seen as an axisymmetric design maximizing the DI with respect to a single incident plane wave (see Sec. 1.2.1 once again). But with no additional constraints on the directional properties of the beampattern, the rear lobe is allowed to be larger than the sides.

This feature goes directly against many of the considerations discussed above, for which one of the main objectives is a directionally independent sectioning of the sphere. As such, it may be useful to consider the use of alternative beamformer designs in order to address this issue.

By definition, the maximum front-to-back ratio (FBR) or super-cardioid beamformer is the most optimal with respect to the rear lobe issue, as it maximizes the directivity's energy in the hemisphere towards the look direction relative to that opposite it. However, its solution achieves this at the expense of a greatly reduced DI and a widely enlarged main lobe. Dolph-Chebyshev designs, on the other hand, produce beampatterns with equal side lobe levels [47]. These are parameterized either in terms of main lobe width, in which case the solution minimizes the side lobe level, or conversely using a set side lobe level, for which a minimum main lobe width is then found.

Finally, we propose a "maximum weighted DI" (maxWDI) design, which is detailed mathematically in Appx. A.2. As its name suggests, the maxWDI beamformer aims to maximize a weighted version of the DI measure: specifically, the weighting function is proportional to the angular distance from the look direction. In other words, energy further from the look direction contributes more to lowering the WDI value, and the optimal solution therefore attempts to maximize the WDI by concentrating the beampattern's energy around the look direction[2].

The specific choice of spherical layout grid for the S look directions as well as the particular directivity properties of the chosen beamformer are thus absolutely critical for the generation of a DRIR that preserves the space-time-frequency characteristics of the measured field. As such, a thorough optimization protocol (similar to that given in Massé et al. [78]) will be used in Sec. 4.1 to evaluate several candidate look direction layout grids and beamformer choices with respect to the properties mentioned throughout this section.

[2]Expert readers may have noticed that this definition is essentially a reformulation of the "max-rE" HOA-SSLD reproduction formalism [80], but from a purely beamforming perspective.

2.3 / Mixing Time Estimation Using a Measure of Spatial Incoherence

In the case of monophonic IRs, various windowed statistical measures have been used in order to estimate the moment the sound field verifies the stochastic properties of the late reverberation field. The echo density profile (EDP) [81] and cumulative kurtosis [82] are notable approaches in this context. As we have seen, however, the use of SMA-measured SRIRs allows for a much richer description of the sound field, allowing more refined measures to be exploited.

The use of diffuseness measures to estimate the mixing time can be traced in the literature to Götz et al. [83], though the idea seems to have been independently percolating amongst various research groups around the same time. Whereas Götz et al. [83] use only the PIV diffuseness measure (in comparison to the statistical methods mentioned above as well as some perceptual estimators), the author's Master's thesis [21] offers a comparison between the PIV and SDR measures. Finally, the CoMEDiE measure was chosen in Massé et al. [84] for its robustness, ability to exploit higher-order SH components (unlike the PIV, which is limited to the first order), and independence from other estimated signal properties (unlike the SDR, which requires a reference signal DoA).

Regardless of the specific choice of diffuseness measure, the common underlying idea is to generate a *diffuseness profile* by calculating the diffuseness over short time frames throughout the SRIR. Such a profile is shown in Fig. 2.4 using the CoMEDiE measure on a simulated 4^{th}-order SH-encoded SRIR with a perfectly diffuse late reverberation tail[3]. Estimating the mixing time then corresponds to detecting the moment the diffuseness profile reaches a maximum value that is maintained throughout the late reverberation tail.

In practice, however, a choice must be made at this point as to the exact definition of the mixing time. As described above in discussing the transitional cluster regime (Sec. 2.1.2), there may be some overlap between the emergence of the stochastic late field and some "late arriving" specular reflections. The question of "placing" the mixing time is therefore open to some interpretation, and may vary with respect to its expected use in any final applications.

On one hand, the mixing time can be seen as the first instant at which maximum diffuseness is reached, i.e. a t_{mix} that refers to the earliest time the stochastic model is valid. Conversely, it could also refer to the final, most stable segment of maximum diffuseness, i.e. a t_{mix} after which the stochastic model can be assumed to describe the entirety of the SRIR's tail. Of course, a compromise between the two views can also be considered, and indeed this is the approach that will be taken in this work.

Additionally, in keeping with this thesis' objective of allowing for the analysis of SRIRs with direction-dependent decay properties, a measure of spatial incoherence will be used instead of the diffuseness measures mentioned above. As presented in Massé et al. [78], this follows directly from the late reverberation model described in Sec. 2.1.3 and the fact that the SH domain cannot properly represent the spatial incoherence of a stochastic plane wave field under an anisotropic power distribution. Indeed, though we have demonstrated in Eqs. 2.12 to 2.16 that our stochastic late reverberation field model is directionally incoherent [and that therefore $\mathbf{R}_{\text{dir}}(f,t) = \text{E}\{\mathbf{P}(f,t)\}$ is diagonal], this does not hold under the SH transform. As opposed to Eq. 1.25 for the ideal isotropic diffuse case, the elements of $\mathbf{R}_{\text{SH}}(f,t) = \text{E}\{\mathbf{S}(f,t)\}$ are now given by:

[3]This SRIR has been simulated based on an arbitrary IS model performed by the EVERTims auralization engine, to which a late reverberation tail has been matched and added directly in the SH domain. The decay rate of the late envelope is chosen to match the decay of the early reflections by analyzing the IS-only SRIR in the SH domain using the method described in Sec. 2.5.1. The late tail is synthesized following Jot et al. [16], as further detailed in Sec. 3.3. An arbitrary target mixing time is defined as halfway through the decay analyzed on the IS SRIR. The synthesized tail and IS SRIR are then matched energetically through their EDRs at this chosen t_{mix}. To further ensure a smooth transition, a constant diffuse field with the same power spectrum as the late tail is added to the early reflections (i.e. the IS SRIR). Finally, the IS SRIR is cut off at t_{mix} and the synthesized reverberation tail is added in.

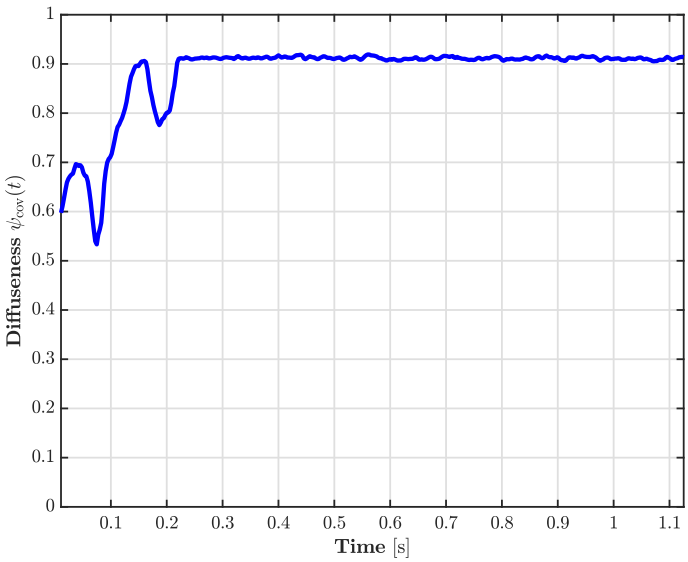


Figure 2.4: CoMEDiE diffuseness profile $\psi_{\mathrm{cov}}(t)$ calculated on a simulated 4$^{\mathrm{th}}$-order SH-encoded IS-based SRIR to which a perfectly diffuse late reverberation tail has been added (but without fully simulating a "realistic" SMA measurement, for simplicity). This SRIR simulation is also described in a few more details in the surrounding text.

$$
\begin{aligned}
R^{\mathrm{SH}}_{n,n'}(f) &= \mathrm{E}\left\{\widetilde{H}^{\mathrm{late}}_{l,m}(f)\widetilde{H}^{\mathrm{late}*}_{l',m'}(f)\right\} \\
&= b_l(f)b^*_{l'}(f)\sum_{d=1}^{D}\sum_{d'=1}^{D}\alpha_d\alpha_{d'}\sqrt{P(f,\boldsymbol{\Omega}_d)P(f,\boldsymbol{\Omega}_{d'})}\,\mathrm{E}\left\{e^{i\hat{\phi}(f,\boldsymbol{\Omega}_d)}e^{-i\hat{\phi}(f,\boldsymbol{\Omega}_{d'})}\right\}Y_{l,m}(\boldsymbol{\Omega}_d)Y^*_{l',m'}(\boldsymbol{\Omega}_{d'}) \\
&= b_l(f)b^*_{l'}(f)\sum_{d=1}^{D}\alpha_d P(f,\boldsymbol{\Omega}_d)Y_{l,m}(\boldsymbol{\Omega}_d)Y^*_{l',m'}(\boldsymbol{\Omega}_d),
\end{aligned}
\tag{2.22}
$$

where the time dependence has been dropped for simplicity.

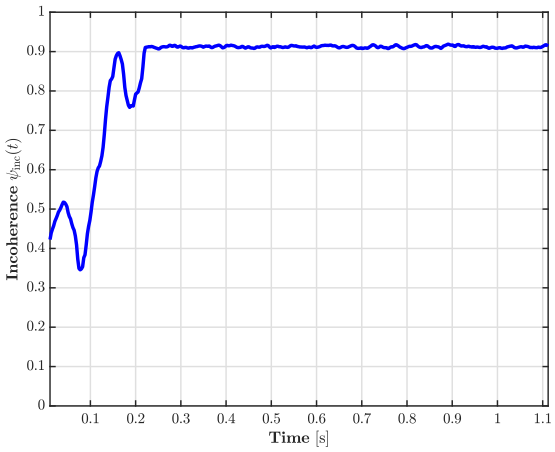
In other words, $\mathbf{R}_{\mathrm{SH}}(f,t) = \mathrm{E}\{\mathbf{S}(f,t)\}$ is not, in general, diagonal under anisotropic incoherent field conditions; in fact, it does not take on any single characteristic shape and depends solely on the power distribution $P(f,\boldsymbol{\Omega}_d,t)$. For these reasons, the use of the SH-domain CoMEDiE diffuseness measure (or, for that matter, any other diffuseness measure based on SH-domain covariance or coherence, such as the SDR or PIV measures) may not be appropriate when taking direction-dependent and potentially anisotropic power distributions into account.

However, if an estimate of the directional sound field can be obtained through a DRIR representation (as described in Sec. 2.2 above) such that $\mathbf{h}_{\mathrm{dir}}(f,t) \to \tilde{\mathbf{p}}(f,t)$ as $t \to t_{\mathrm{mix}}$, then its covariance matrix $\widetilde{\mathbf{R}}_{\mathrm{dir}}(f,t)$ should tend toward the diagonal form obtained in Eq. 2.16 as it reaches the late reverberation tail. If the covariance matrix is further normalized by the estimated PSDs of $\tilde{\mathbf{p}}(f,t)$, the resulting $\widetilde{\mathbf{R}}^{\mathrm{norm}}_{\mathrm{dir}}(f,t)$ will tend toward the identity matrix and the same approach as for the CoMEDiE measure [62] can be applied in order to obtain a scalar-valued incoherence ψ_{dir}.

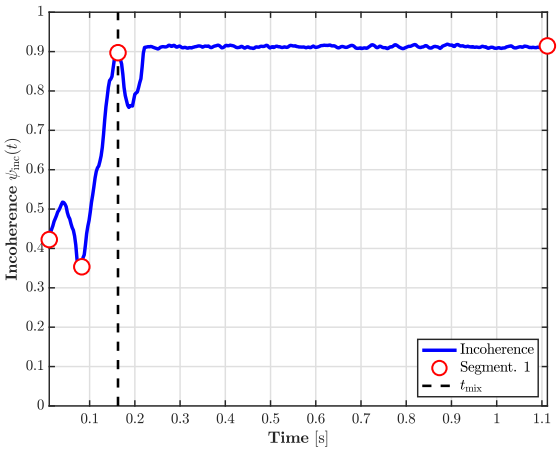
Note that the CoMEDiE measure is defined solely in the time domain [62]; the final calculation can either be adapted to our general space-time-frequency view by summing the normalized time-frequency covariance matrix over its discrete frequency bins to obtain $\underline{\widetilde{\mathbf{R}}}^{\mathrm{norm}}_{\mathrm{dir}}(t) = \sum_k \widetilde{\mathbf{R}}^{\mathrm{norm}}_{\mathrm{dir}}(f_k,t)$, or a time-domain signal representation of $\tilde{\mathbf{p}}$ can be used to define $\widetilde{\mathbf{R}}^{\mathrm{norm}}_{\mathrm{dir}}(t)$ directly (with appropriately adjusted time windows for an equivalent power estimation, which would be an RMS measure in this case). The

incoherence profile $\psi_{\text{dir}}(t)$ can then be calculated through the eigen-decomposition of $\underline{\widetilde{\mathbf{R}}}_{\text{dir}}^{\text{norm}}(t)$ as in Epain and Jin [62] and presented in the same way as Fig. 2.4.

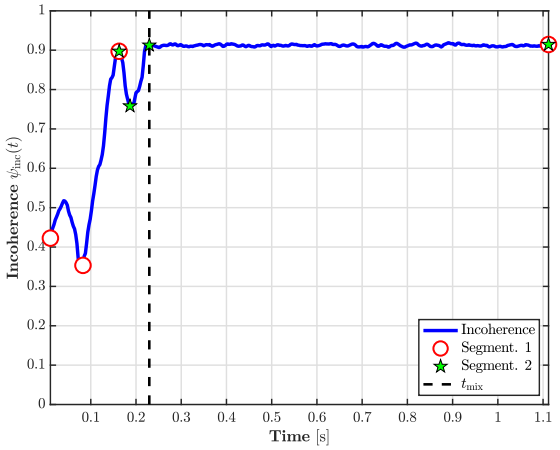
Figure 2.5a thus shows $\psi_{\text{dir}}(t)$ for a simulated SRIR similar to that in Fig. 2.4, but whose late reverberation tail has been rendered anisotropic by making $T_{60}(\mathbf{\Omega}_d)$ vary in direction according to a cardioid distribution pattern[4]. Furthermore, the energy matching at t_{mix} of the IS SRIR and synthesized reverberation tail is performed direction-wise (see also Sec. 4.2.2 for additional details, though once again the SMA measurement is not simulated in the preliminary examples shown here).



(a) Spatial incoherence profile.



(b) First segmentation and initial t_{mix} estimate.



(c) Second segmentation and final t_{mix} estimate.

Figure 2.5: Spatial incoherence profile $\psi_{\text{dir}}(t)$ and mixing time estimation for a simulated 4^{th}-order SH-encoded SRIR with an anisotropic late reverberation tail (once again without applying an SMA measurement simulation). The spatial incoherence is measured by covariance matrix eigen-decomposition on a directional SRIR representation (DRIR) obtained using a natural PWD beamformer steered around a 25-point Fliege-Maier [79] look direction layout grid.

Considering the discussion above with respect to the choice of mixing time definitions, we now propose a t_{mix} estimation algorithm that aims for a compromise between the least restrictive condition (i.e. set at the earliest maximum incoherence point) and the most restrictive one (i.e. find the most stable maximum incoherence segment). The algorithm is outlined in pseudo-code in Fig. 2.6 and detailed step by step below. Illustrations of the main steps are shown in Fig. 2.5 for the simulated anisotropic SRIR described above.

[4] In this case, the reverberation time is made to vary from half the reference omnidirectional value analyzed on the IS SRIR, at the minimum of the cardioid pattern, to 1.5 times the reference T_{60} at the cardioid's maximum.

Data: A list of spatial incoherence profile values `spatInc` $= [\psi(t_0), \ldots, \psi(t_n), \ldots, \psi(t_N)]$, a corresponding list of time values `t` $= [t_0, \ldots, t_n, \ldots, t_N]$, and a re-segmentation parameter `reSegParam` $= \lambda_{\text{reSeg}}$.

Result: The mixing time `tMix` $= t_{\text{mix}}$ and the average late reverberation incoherence `avgLateInc` $= \overline{\psi}_{\text{late}}$.

```
initSeg ← RDP(spatInc);      /* Adaptive Ramer-Douglas-Peucker segmentation. */
initSegReg ← LinReg(initSeg,spatInc);    /* Segment-wise linear regressions. */
initSegScores ← Score(initSegReg);
i ← argmax(initSegScores);
j ← initSeg[i][1];                 /* Start point of segment with highest score. */
tMix ← t[j];
avgLateInc ← mean(spatInc[j : end]);
if ValidTest(tMixInitEst,avgLateIncInitEst) returns true then
    if ReSegTest(spatInc,initSegReg[i],reSegParam) returns true then
        reSeg ← RDP(spatInc[j : end]);
        reSegReg ← LinReg(reSeg,spatInc[j : end]);
        reSegScores ← Score(reSegReg);
        scoreThresh ← mean([mean(reSegScores),median(reSegScores)]);
        k ← min(reSeg[:][1]) such that reSegScores[k] ≥ scoreThresh;
        tMix ← t[j + k];
        avgLateInc ← mean(spatInc[j + k : end]);
    end
else
    tMix ← NaN;
    avgLateInc ← NaN;
end
```

Figure 2.6: Pseudo-code for the mixing time estimation algorithm.

1. *Initial segmentation*: The list `spatInc` containing the discrete spatial incoherence values $\psi(t_n)$ is segmented using an adaptive Ramer-Douglas-Peucker (RDP) algorithm [85].

2. *Segment-wise linear regressions*: A fit is performed over each detected segment. The average incoherence of each segment is also stored.

3. *Segment scoring*: Each segment is then given a score

$$\kappa_i = \frac{N_i - N_{\min}}{N_{\max}} + 1 - \frac{|m_i| - |m_{\min}|}{|m_{\max}|} + \frac{\overline{\psi}_i - \overline{\psi}_{\min}}{\overline{\psi}_{\max}},$$

   where N_i is the segment's length, m_i is its slope, $\overline{\psi}_i$ is its average incoherence, and the min/max indices indicate the min/max values over all segments.

4. *Initial t_{mix} estimate*: The onset of the segment with the highest score is taken as a first guess for the mixing time. A corresponding initial $\overline{\psi}_{\text{late}}$ value is also calculated.

5. *Estimate validity check*: The validity of the initial guess is verified by ensuring it meets two conditions:

   (a) $\overline{\psi}_{\text{late}} > 0.5$,
   (b) $\overline{\psi}_{\text{late}} > \{\min[\psi(t_n)] + \max[\psi(t_n)]\}/2$.

6. *Re-segmentation check*: If the initial estimate was valid, a second verification is run to determine whether the chosen segment should itself be re-segmented. This check seeks to determine if there are any important fluctuations in the incoherence measure throughout the segment, potentially due to "late arriving" early reflections. The verification condition is

$$\max\left\{\left|\psi(t_n) - \hat{\psi}(t_n)\right|\right\} > \lambda_{\text{reSeg}}\{\max[\psi(t_n)] - \min[\psi(t_n)]\},$$

   where $\hat{\psi}(t_n)$ are the linear regression values for the segment, t_n is constrained to within the segment's time bounds, and λ_{reSeg} is a tuning parameter allowing for the re-segmentation condition to be adjusted with respect to the parameterization of the incoherence profile itself.

7. *Re-segmentation*: If re-segmentation is deemed necessary, the adaptive RDP algorithm [85] is applied to the initially selected segment and the resulting refined segments are once again scored in the same manner as step 3.

8. *t_{mix} estimate adjustment*: This time, a score threshold κ_{thresh} is calculated as the mean between the mean and the median of the new segment scores. The onset of the first refined segment whose score is greater than or equal to κ_{thresh} is then used to adjust the mixing time estimate. The average late incoherence value $\overline{\psi}_{\text{late}}$ is also updated accordingly.

Note that an "early" estimate of the mixing time can also be obtained by simply taking the result of the initial segmentation and ignoring steps 6-8. Conversely, the estimate can be further constrained by adjusting the verification condition in step 6 and taking the maximum of the re-segmentation scores (instead of the first above the threshold κ_{thresh} as in step 8). This "safe" estimate is useful in the context of hybrid reverberators [15], for example, where prominent early reflections are at risk of being cut off if the t_{mix} is too short. It is also the version used for the validation test in Sec. 4.2.2.

Once the mixing time has been estimated, thereby sectioning the analyzed SRIR into its early reflection and late reverberation segments, these two regimes can subsequently be analyzed according to their respective signal models (see Sec. 2.1). We begin chronologically (from an IR standpoint) with the early reflections in Sec. 2.4 below, before closing the chapter with the late reverberation in Sec. 2.5.

2.4/ Detecting Early Reflections

As stated in the Introduction, one of the main objectives of this research project is the extraction of low-level properties with a view to their subsequent use in higher-level treatment and manipulation procedures. This approach is markedly different from analysis/synthesis methods such as SDM or SIRR/DirAC that aim to encode a "complete" (albeit dimensionally reduced) description of an SRIR in order to re-create it efficiently (see Sec. 1.2.3). In this work, we are further interested in obtaining a "qualitative" description that can be used to inform (potentially perceptual) manipulation models.

In further contrast to SDM/SIRR, this objective implies we cannot simply analyze an SRIR frame-by-frame and assign space-frequency properties to each time step. Although this has been shown to work well in re-creating the SRIR's original sonic signature in a perceptually transparent fashion [66], the lack of discrimination and underlying acoustical models in the analysis means the properties obtained have limited uses in informing subsequent manipulations of the SRIR's characteristics.

In the context of analyzing the discrete early reflections, this entails a strategy that searches for the predominant echoes while attempting to limit the number of "false positives", e.g. doubly detected reflections, noise, or semi-stochastic signals more appropriately considered as part of the cluster. On top of estimating DoAs, therefore, a characterization of the underlying signal (i.e. an impulsive reflection) must also be used to perform this discrimination.

The most basic property that can be exploited to this end is that of the early reflections' high spatial coherence, which can be translated as low incoherence [i.e. $\Phi(t) = 1 - \psi(t)$]. Similarly to the t_{mix} estimation above, a refined "early incoherence" profile $\psi_{\text{early}}(t)$ can be calculated[5]. A threshold value $\psi_{\text{early}}^{\text{thresh}}$ can then be derived from this incoherence profile, with windows with $\psi_{\text{early}}(t) < \psi_{\text{early}}^{\text{thresh}}$ then identified as "coherent". The constituent frames of each window can then be considered for echo detection (in general, an incoherence estimation window will be made up of several STFT frames, a point further elaborated on in Sec. 4.2.1). In this work, we define $\psi_{\text{early}}^{\text{thresh}}$ as:

$$\psi_{\text{early}}^{\text{thresh}} = \overline{\psi}_{\text{late}} - \lambda_{\text{coh}}\sigma_{\psi}^{\text{late}}, \tag{2.23}$$

where $\overline{\psi}_{\text{late}}$ and $\sigma_{\psi}^{\text{late}}$ are the mean and standard deviation of the incoherence measure taken over the length of the late reverberation tail, and $\lambda_{\text{coh}} > 0$ is a control parameter that can govern the overall sensitivity of the echo detection procedure.

Within each identified coherent frame, several different DoA (and ToA) estimation methods can be used. In this work, two of these will be compared in particular (and subsequently evaluated Ch. 4): a classic beamforming SRP approach, as presented in Sec. 1.2, and a novel approach (to the author's knowledge) consisting of inverting the Herglotz framework described in Sec. 2.1.2, as in Eq. 2.8[6].

Of the additional literature methods also mentioned in Sec. 1.2, the extended EB-ESPRIT method has been discarded [55] due to its assumption that any simultaneous signals should not be strongly correlated, a condition clearly violated when searching for the predominant reflections of an impulsive excitation, while the PIV method [60] is left aside due to its limitation to a single DoA estimate at each time frame. Furthermore, both of these methods rely on the estimation of expectation values through frame averaging, which severely limits their temporal resolution (i.e. it would not be better than the coherence estimation used in the first step).

Finally, GCC-based ToA/TDoA methods function directly on the SMA transducer signals. In

[5]The term "refined" refers here to the fact that this "early incoherence profile" is calculated using an STFT parameterization specific to the echo detection procedure, which can be expected to involve shorter time windows than the general incoherence profile $\psi(t)$ used to estimate t_{mix}. See Sec. 4.2.1 for a more complete discussion on this point.

[6]Though this exact formalism does not seem to be currently present in the literature, similar inversion approaches for reconstructing the sound field measured by an SMA have been presented, e.g. by Fernandez-Grande [86], and have indeed been used as an inspiration for this approach.

the case of compact rigid-sphere SMAs, they are thus subject to two main issues: the very small distances between transducer positions, and the self-scattering effects of the SMA's body. For this reason, ToA/TDoA methods are usually reserved for larger, open-configuration microphone arrays (as in SDM, for example) [59].

The beamformed SRP and Herglotz inversion approaches are both subject to the combined order-dependent frequency response shown in Fig. 1.3c. In particular, and as presented in Sec. 1.2, the accuracy of the SRP is conditioned by the resulting frequency-dependent directivity illustrated in Fig. 1.5. For these methods, it is therefore optimal to limit the DoA estimation to frequencies around the maximum directivity index (see Fig. 1.5b). The influence of the exact choice of bandwidth will be investigated in Ch. 4, and will also depend on the time-frequency resolution compromises inherent to STFT analysis (which are to be evaluated for all the analysis methods in this chapter).

This frequential bandlimiting is additionally important for the inversion of the matrix $\mathbf{D}_{\mathrm{SH}}(f)$ in the Herglotz framework described in Sec. 2.1.2. Indeed, since the elements $\alpha_d b_l(f) Y^*_{l,m}(\mathbf{\Omega}_d)$ depend on the mode strength $b_l(f)$ [and on the combined response $c_l(f)$ in the case of the HOA equivalent $\mathbf{D}_{\mathrm{HOA}}(f)$], as well as the layout of the directional sampling points $\mathbf{\Omega}_d$, the inversion may be rank-deficient at low frequencies. Though in general the linear independence of the SH basis guarantees that $\mathbf{D}_{\mathrm{SH}}$ and $\mathbf{D}_{\mathrm{HOA}}$ will be full-rank for any number of sampling points $D \leq (L+1)^2$, the extreme attenuation of higher orders at low frequencies due to the mode strengths $b_l(f)$ can further limit this.

Working solely at frequencies around the maximum directivity therefore ensures that the matrix has maximum rank $(L+1)^2$ and that the complete SH spectrum is exploited for localization. In general, we would then like to obtain the highest possible spatial resolution by decomposing over a larger number of sampling points, i.e. $D_{\mathrm{Herg}} > (L+1)^2$. However, this inevitably results in an under-determined inversion of $\mathbf{D}_{\mathrm{SH}}$ or $\mathbf{D}_{\mathrm{HOA}}$, and a regularization scheme must be applied. For example, using Tikhonov regularization, Eq. 2.8 can be re-written:

$$\tilde{\mathbf{g}}(t) = \frac{\left[\mathbf{D}^{\top}_{\mathrm{SH}}(f)\mathbf{D}_{\mathrm{SH}}(f) + \mathbf{T}^{\top}\mathbf{T}\right]^{-1}\mathbf{D}^{\top}_{\mathrm{SH}}(f)\mathbf{h}^{\mathrm{early}}_{\mathrm{SH}}(f,t)}{H_{0,0}(f,t)}, \tag{2.24}$$

where $\mathbf{T} = \chi\mathbf{I}_D$ is an L^2 regularization Tikhonov matrix defined using a regularization parameter χ and the $D_{\mathrm{Herg}} \times D_{\mathrm{Herg}}$ identity matrix $\mathbf{I}_D$. The regularization parameter χ itself can be estimated using the generalized cross-validation (GCV) method, or otherwise optimized with respect to expected performance (i.e. through a form of expectation-maximization); this is further explored in Sec. 4.2.1 in the context of evaluating the overall performance of the echo detection methods.

The main advantage of the beamformed SRP method is that, since we are simply interested in generating a directional power map, an arbitrarily high number of directional beams can be generated to completely cover the sphere with high precision and low variance (though the actual underlying resolution remains order-limited by the width of the directivity's main lobe, as described in Ch. 1 and further discussed in Sec. 2.2 below)[7].

In the case of the Herglotz inversion, the maximum number of sampling points D_{Herg} may be limited by the stability of the regularization scheme (see further Sec. 4.2.1). To obtain an arbitrary point precision equivalent to the SRP, a cubic Hermite spherical interpolation scheme [87] can be subsequently applied to the estimated Herglotz kernel $\tilde{\mathbf{g}}$. As with the SRP, this does not increase the underlying resolution, which in this case is limited by the maximum D_{Herg}; however, by using local gradient estimations at each point on the sphere, such an approach may end up generating a "sharper" localization map than the SRP.

[7] The main advantage in generating a power map with high precision even though the underlying resolution remains limited is the fact that it will tend to be smooth and quasi-continuous, which helps regularize the performance of the surface peak detection algorithm.

Finally, the actual echo localization (whether on the SRP map or the interpolated Herglotz kernel) is performed by means of a spherical surface peak detection algorithm specifically developed for this purpose and detailed in Appx. A.4.

We now turn to specific applications of these methods, first in the detection of the direct sound and then in the general context of generating a "cartography" of the early reflections, before briefly discussing the definition of some indicative measures describing the spatiotemporal properties and layout of the detected echoes.

2.4.1/ Direct Sound Detection

Detecting the direct sound impulse on an SRIR measured using an multi-channel ESM [20] can be rendered non-trivial by the presence of harmonic distortion in the excitation signal or non-stationary elements in the background noise. Harmonic distortion is often a result of having to overdrive the source loudspeaker in order to obtain an acceptable signal-to-noise ratio (SNR)[8]. Upon convolution with the inverse sweep (generally but technically erroneously referred to as "deconvolution"), this distortion can generate preceding copies of the IR [20].

Careful parameterization of the ESM measurement can greatly reduce the impact of these artefacts (e.g. using a long enough sweep and zero-padding to ensure that any harmonic copies are generated beyond the true $t = 0$ of the measurement). Non-stationary background noise can be mitigated by combining several repeated sweeps (see the pre-treatment processing method presented in Sec. 3.1). Nonetheless, artefacts resulting from these phenomena may still pollute the time preceding the arrival of the direct sound (i.e. the direct sound's ToA, sometimes referred to as the "pre-delay").

A robust and accurate detection of the direct sound is necessary due to the fact that most of the analysis methods presented within this chapter use a time scale defined relative to the direct sound's ToA ($t = t_0$). Furthermore, precise characterization of the direct sound signal provides an important tool for eventual perceptually informed manipulations of the SRIR, as the properties of the direct sound are known to play a crucial role in reverberation perception [88], especially through the direct-to-reverberant ratio (DRR) [89].

The main lines of the direct sound detection algorithm used in this work are thus presented below. Note that for descriptive simplicity in the following, a generic time variable t is used indiscriminately regardless of the actual time resolutions involved (samples, STFT frames).

1. *Define maximum search range*: Estimate a maximum time up to which to search for the direct sound as $t_{\max}^{\text{DSdet}} = 2\,\text{argmax}\left[|h_{0,0}(t)|^2\right]$, where $h_{0,0}(t)$ is the time-domain omnidirectional component of the HOA-encoded SRIR. This is a completely arbitrary choice made to ensure that the direct sound is contained within the search range, under the assumption that max $\left[|h_{0,0}(t)|^2\right]$ is either the direct sound itself (as is often the case) or a very early reflection.

2. *Calculate spatial coherence measure*: Since we are working with the HOA-encoded SRIR, the SH-domain CoMEDiE measure is exploited to define $\Phi_{\text{cov}}(t) = 1 - \psi_{\text{cov}}(t)\ \forall\ t \in \left[0, t_{\max}^{\text{DSdet}}\right]$.

3. *Spatial coherence peak detection*: Find temporal peaks in the spatial coherence measure that are above a certain threshold $\Phi_{\text{pre}}^{\text{thresh}} = 0.5\left\{\min\left[\Phi_{\text{pre}}(t)\right] + \max\left[\Phi_{\text{pre}}(t)\right]\right\}$, and choose the first of these as the window within which to search for the direct sound. If no peaks are above the threshold, fall back on the maximum coherence value (usually a sign that it is the first frame and as such doesn't technically constitute a "peak").

[8] The notion of "acceptable SNR" will be further discussed in context of background noise detection (Sec. 2.5.1) and denoising (Sec. 3.3.1).

4. *Signal smoothing*: Within the signal window defined by the chosen peak coherence frame (itself consisting of multiple STFT frames and therefore covering a relatively large sample window), smooth out high-frequency noise from the $h_{0,0}^{\text{pre}}(t)$ component using a Gaussian kernel smoothing function (revealing the overall shape of the direct sound impulse).

5. *Energy peak detection*: From the smoothed signal $\tilde{h}_{0,0}^{\text{pre}}(t)$, detect peaks in its instantaneous energy $\widetilde{E}_{\text{pre}}(t) = \left|\tilde{h}_{0,0}^{\text{pre}}(t)\right|^2$. Choose the first energy peak whose value is above the threshold $\widetilde{E}_{\text{pre}}^{\text{thresh}} = \text{mean}\left[\widetilde{E}_{\text{pre}}(t)\right] + \text{std}\left[\widetilde{E}_{\text{pre}}(t)\right]$ as the energy peak of the direct sound impulse.

6. *Impulse signal detection*: Assuming the archetypal shape of a measured impulse, search for the extent of the direct sound signal by looking for zero-crossings in the smoothed signal $\tilde{h}_{0,0}^{\text{pre}}(t)$ on either side of the detected energy peak. This can be made more robust by considering "noise-crossings" instead of zero-crossings, using an estimate of the noise root-mean-square (RMS) power on $h_{0,0}(t)$ from $t = 0$ to the start of the current analysis window.

7. *DoA estimation*: The DoA of the direct sound is estimated using the SRP and/or Herglotz inversion methods. By design, the generated localization map (for both methods) should contain only a single peak corresponding to the direct sound: the DoA can thus be taken directly as the maximum of the localization map.

Note that the ToA of the direct sound can be taken to be either the start of the detected impulse signal as defined in Step 6 or the maximum energy peak chosen in Step 5. Unless specified otherwise, the former will be used throughout this work in order to avoid cutting off non-zero signal values (for example when removing the pre-delay before analysis).

2.4.2/ Strategies for Echo DoA, Energy, and ToA Estimation

Once the direct sound has been detected, a cartography of the remaining (predominant) early reflections can be obtained by analyzing the SRIR over the time range $[t_0, t_{\text{mix}}]$. As mentioned above, a spatial incoherence measure $\psi_{\text{early}}(t)$ is exploited in order to select time windows with low incoherence (high coherence) within which to detect echoes. Again, this strategy seeks to eliminate highly incoherent early segment frames such as the one taken as an initial t_{mix} estimate in Fig. 2.5b, which must predominantly contain either background noise or semi-stochastic cluster plane waves.

To obtain this cartography of early reflections, three parameters must therefore be estimated for each detected echo: its DoA, its energy, and its ToA.

DoA Estimation

Direction of arrival estimation is performed on the constituent STFT frames of any spatial incoherence window with a value below the threshold defined above. The sample length of the STFT frames is therefore the temporal resolution limit of the echo detection procedure (though a more precise ToA estimation can subsequently be obtained as described below). As such, the parameterization of the STFT as well as the subsequent expectation value estimation necessary for the (in)coherence calculation is a crucial factor that will be thoroughly assessed in Sec. 4.2.1.

Within each frame, the SRP and/or Herglotz inversion methods are used to generate a localization map on which the spherical surface peak detection algorithm detailed in Appx. A.4 is applied in order to extract the DoAs of any reflections present. Note that the peak detection algorithm normalizes the localization maps between 0 and 1 and operates on the assumption that at least one reflection is present (hence the need to ensure that the frames are not within a fully incoherent window). Of course, there is still a chance that a frame with no echoes could exist within a highly coherent window; in

practice, however, echo density increases rapidly enough that this will rarely be the case. Additionally, the noise floor verification step described below can help eliminate any reflections "detected" in a frame that should not have contained any.

The bandlimiting described previously, which ensures maximum directivity and well-conditioned inversion, restricts the generation of these localization maps to a limited number of frequency bins, depending on the chosen STFT time-frequency resolution and the exact bandwidth defined around the maximum directivity frequency. In general, however, more than one bin will be present in this range and a sharpened localization map can be obtained by taking their product over the available bins. This procedure is directly inspired by similar approaches commonly used when combining cross-correlation pairs in GCC and TDoA-based methods [58] [59]. To avoid generating values so small as to risk being numerically unstable, the localization maps should first be normalized to their maximum value over these frequencies (since, as noted above, the maps will then get renormalized between 0 and 1 by the peak detection algorithm, it is good practice to avoid potentially boosting floating-point quantization noise that could be generated when taking the product of very small values).

Energy Estimation

Even without the various renormalizations, the localization maps described above cannot provide accurate estimations of reflection energies. In the case of the SRP method, this is due not only to the fact that the localization map is defined over a limited subset of frequency bins, but also to the total disregard for directivity overlap effects and flat spherical coverage during the beamforming process (in contrast to the strict approach taken when defining the DRIR representation in Sec. 2.2 above).

For the Herglotz inversion method, the estimation of the kernel $\tilde{\mathbf{g}}(t)$ is deliberately rendered amplitude- (and thus energy-) independent by dividing out the omnidirectional SH component (Eq. 2.8 and Eq. 2.24). In both cases, therefore, a complementary procedure must be used in order to obtain an accurate energy estimate. To do so, we will in fact exploit the DRIR definition from Sec. 2.2.

At each frame of the echo detection procedure, the directional energy spectra provided by taking the STFT of the DRIR can be interpolated to the arbitrary precision of the DoA estimation localization maps, i.e. $D_{\text{int}} \gg (L+1)^2$, through the same cubic Hermite spherical interpolation scheme as described above. Thus the resulting interpolated energy spectra present at the DoAs estimated through the spherical surface peak detection algorithm can be taken as estimates of the reflections' energy spectra, thereby providing an estimate of their total energy as well.

Obtaining echo energy estimates also enables the definition and application of an additional verification step. We will see in Sec. 2.5.1 how analyzing the late reverberation tail also provides an estimate of the SRIR's measurement noise floor, and more importantly its power, P_{noise}. This estimate can then be used as an energy threshold for the detected early reflections: any echo with an estimated energy at or below the noise floor may thus be discarded from the detection results.

ToA Estimation

Though a straightforward estimate of the ToA can be provided by taking the midpoint of the current STFT frame, such a definition is inherently limited by the temporal resolution of the STFT (i.e. the chosen window length). However, a refined estimation of the ToA is possible through the fact that an ideal impulse signal [a temporal Dirac delta $\delta(t - t_j)$ at a given ToA t_j] has a linear phase over frequency, i.e. $\phi_{e,j}(f, t_n) = -2\pi f(t_j - t_n)$, for a given reflection with a ToA t_j in an STFT window centred on t_n. A linear regression on the "unwrapped" phase angle spectrum of a detected echo can

therefore provide a ToA estimate as:

$$\tilde{t}_j = \frac{-\hat{\phi}_{e,j}(f_k, t_n)}{2\pi f_k} + t_n \;\forall\; f_k > 0, \tag{2.25}$$

where $\hat{\phi}_{e,j}(f_k, t_n)$ is the linear regression model for the unwrapped phase spectrum.

Note that a linear regression is used to obtain $\tilde{t}_j$ instead of averaging over direct estimates at each frequency as it is less sensitive to error fluctuations in certain bins (e.g. from noise or interpolation errors). In a sense, the linear regression itself provides an "average" estimation.

In order to ensure the best possible fit for the linear regression, it is further limited to a frequency range $[f_{k_s}, f_{k_e}]$ such that $f_{k_s} \geq f_{\text{dir},1}$ and $f_{k_e} \leq f_{\text{alias}}$, where $f_{\text{dir},1}$ is the "first-order directivity limit" and f_{alias} is the spatial aliasing frequency presented in Sec. 1.2.1. The first-order directivity limit corresponds to the lowest frequency at which the first order of the combined frequency response of an HOA-encoded SMA-measured plane wave (see Fig. 1.3c) comes within 6 dB of the omnidirectional (zeroth order) response, resulting in a "useful" DI > 6 dB (see Fig. 1.5b).

Since frequencies $f < f_{\text{dir},1}$ are essentially omnidirectional and $f > f_{\text{alias}}$ are exposed to spatial aliasing, the phase spectrum is highly prone to non-linear behaviour in these ranges. Limiting the linear regression to the range $[f_{\text{dir},1}, f_{\text{alias}}]$ is therefore crucial in safely assuming that the unwrapped phase spectrum will indeed be linear.

The (complex) DRIR spectra used to interpolate and estimate the echo energies in the previous section can thus be re-used in order to similarly interpolate and extract the phase spectra corresponding to the detected DoAs. The linear regressions and $\tilde{t}_j$ estimations are then performed on the extracted phase spectra as described above (Eq. 2.25).

2.4.3/ Directional Echo Energy Density

A traditional measure of (average) echo density can be defined using the length of the early segment as $\rho = N_{\text{ech}}/t_{\text{mix}}$. Through the spatio-temporal analysis described in this section, however, it is further possible to describe a *directional* echo *energy* density. Instead of only counting the total number of echoes detected in the early segment, their relative contribution in a region surrounding a particular look direction can be calculated for a full set of such directions covering the sphere.

Following the PWD resolution discussion from Sec. 1.2.1, the set of look directions around which to estimate the echo density should be limited to $D_{\text{dens}} = (L+1)^2$: in other words, the detected reflections are "gathered" by spatial sectors that more or less correspond to the given maximum SH order's resolution limit. Thus the reflection DoAs estimated at the arbitrarily higher precision used for the SRP and/or interpolated Herglotz inversion, i.e. $\boldsymbol{\Omega}_j$ where $j \in \{1, 2, \ldots, D_{\text{int}}\}$ with $D_{\text{int}} \gg (L+1)^2$, are assigned to the nearest look direction $\boldsymbol{\Omega}_d$ ($d \in \{1, 2, \ldots, D_{\text{dens}}\}$), and their corresponding total number and energy then contribute to the echo density for that look direction:

$$\rho_{\text{NE}}(\boldsymbol{\Omega}_d) = \frac{\widetilde{N}_{\text{ech}}(\boldsymbol{\Omega}_d) \sum_{j=1}^{\widetilde{N}_{\text{ech}}(\boldsymbol{\Omega}_d)} E_j}{\sum_{d=1}^{D_{\text{dens}}} \widetilde{N}_{\text{ech}}(\boldsymbol{\Omega}_d) \sum_{j=1}^{\widetilde{N}_{\text{ech}}(\boldsymbol{\Omega}_d)} E_j}, \tag{2.26}$$

where $\widetilde{N}_{\text{ech}}(\boldsymbol{\Omega}_d)$ is the total number of echoes assigned to the look direction $\boldsymbol{\Omega}_d$, and E_j are the broadband detected echo energies as above.

The echo density values for the set of look directions $\boldsymbol{\Omega}_d$ can then be interpolated "back" to the arbitrary $D_{\text{int}} \gg (L+1)^2$ DoA estimation grid $\boldsymbol{\Omega}_j$, once again through the aforementioned cubic Hermite spherical interpolation scheme [87]. This results in a full-precision directional echo density map, a preliminary application of which is presented in Sec. 3.2.3.

2.5 / Late Reverberation Analysis

The final analysis methods to be presented in this chapter are those relating to the modelling of the late reverberation tail and the estimation of its properties as described by the theory in Sec. 2.1.3. The first step in this analysis is to obtain a direction-dependent view of the late field that preserves the underlying stochastic signal properties. This is crucial because, as demonstrated in Massé et al. [78], anisotropic power distributions cannot be properly analyzed and reconstructed directly in the SH domain. This stems from the fact that power distributions are proportional to the expectation value of the square modulus (i.e. instantaneous energy) of a directional signal (see Appx. A.3 for a more detailed mathematical explanation).

Once again, this implies working with a DRIR representation, from which the energy decay envelope parameters can be estimated by analyzing the time-frequency energy decay relief (EDR) in each look direction. This chapter then closes with a discussion on the evaluation of the late tail's isotropy and how a choice can be made between an analysis-synthesis approach directly in the SH domain for highly isotropic cases (i.e. close to diffuse, where the anisotropic power distribution issue above is no longer a problem), and the use of the generalized direction-dependent framework when anisotropic characteristics become significant.

2.5.1 / EDR Analysis

Following the generation of a DRIR through the spatial decomposition approach detailed in Sec. 2.2, the direction-dependent energy decay envelope parameters $P_0(f, \mathbf{\Omega}_s)$ (the initial power spectrum) and $T_{60}(f, \mathbf{\Omega}_s)$ (the 60 dB reverberation time) can be estimated by analyzing the EDR of each beamformed DRIR signal in turn. Combining Schroeder [3], Polack [7], and Jot et al. [16], we can write:

$$\begin{aligned} \mathrm{EDR}(f, \mathbf{\Omega}_s, t) &= \int_t^\infty \left| \widetilde{H}_{\mathrm{late}}(f, \mathbf{\Omega}_s, \tau) \right|^2 d\tau \\ &\approx \int_t^\infty \mathrm{E}\left\{ \left| \widetilde{H}_{\mathrm{late}}(f, \mathbf{\Omega}_s, \tau) \right|^2 \right\} d\tau \\ &\approx \int_t^\infty P(f, \mathbf{\Omega}_s, \tau) d\tau, \end{aligned} \tag{2.27}$$

where $P(f, \mathbf{\Omega}_s, t)$ is given by Eq. 2.10. In a sense, the resulting estimated space-frequency values $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$ can thus be considered estimates of the "true" underlying $P_0(f, \mathbf{\Omega}_d)$ and $T_{60}(f, \mathbf{\Omega}_d)$ that parameterize $P(f, \mathbf{\Omega}_d, t)$ in the signal model from Sec. 2.1.3 (to within the limitations of the spatial decomposition, as discussed above and further detailed in Ch. 4).

Note that in order to simplify the various explanations, this section will first present the basic approach to fitting an exponential decay model to an EDR with a single-slope decay, and the extension to multiple-slope decays as given in Eq. 2.10 will be discussed afterwards. For reference, the majority of the analysis procedure presented in this section was originally published in Massé et al. [84].

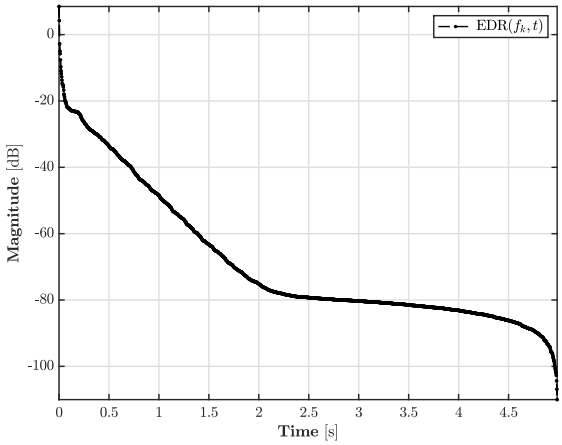
In addition to estimating the decay envelope parameters, analysis of the late reverberation on an SRIR measured in real-world conditions also requires a characterization of the non-decaying background noise floor. Generally, a single frequency bin of a noisy single-slope, dB-scale EDR can be illustrated schematically as in Fig. 2.7a. Since the EDR is a Schroeder-type reverse-integrated curve [3], the (discrete-time) constant power noise floor takes on the characteristic shape:

$$N(f, \mathbf{\Omega}_s, t_n) = \sum_{j=n}^{N_t} P_{\mathrm{noise}}(f, \mathbf{\Omega}_s), \tag{2.28}$$

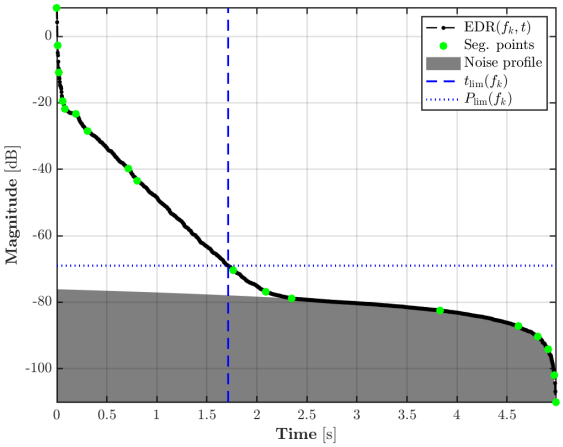
where N_t is the total number of time frames in the EDR (such that $t_{N_t} = T_{\mathrm{IR}}$, the duration of the

measured SRIR) and $P_{\text{noise}}(f, \mathbf{\Omega}_s)$ is the direction-dependent noise power spectrum. The dB-scale noise profile $10\log_{10}[N(f, \mathbf{\Omega}_s, t_n)]$ is illustrated by the grey shaded region in Fig. 2.7b.

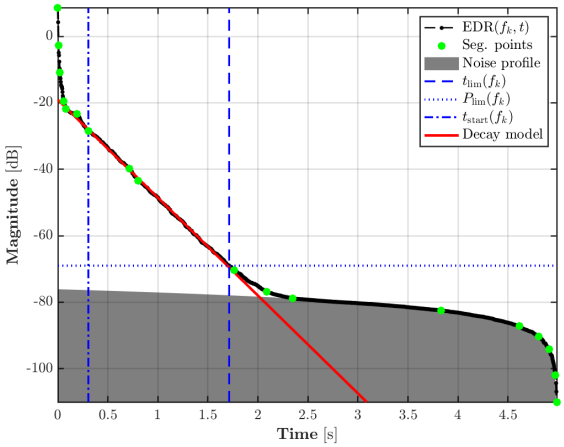
As a preliminary step in order to aid identification of the various sections of interest in the decay curve, the dB-scale EDR bin is segmented using an adaptive RDP algorithm (in the same manner as the incoherence profile in Sec. 2.3 for the mixing time estimation algorithm). The result of this segmentation is shown by the green points on Fig. 2.7b.



(a) Single frequency bin of a single-slope EDR.



(b) RDP segmentation and noise floor detection.



(c) Exponential decay model fitting.

Figure 2.7: EDR analysis schematic for a single frequency bin of a single RIR signal (either a monophonic RIR or a single component of a multi-channel SRIR representation), showing the three principal steps described in the text.

Following RDP segmentation, the noise profile is fitted to the late portion of the dB-scale EDR curve (EDR_{dB}). This is done by successively matching a unit-power [i.e. $P_{\text{noise}}(f, \mathbf{\Omega}_s) = 1$] noise profile to the first point of each segment, and at each j^{th} matched segment:

1. Calculating the mean squared error (MSE) ε_j^N between the matched noise profile $\widehat{N}_j^{\text{dB}}(f, \mathbf{\Omega}_s, t_n)$ and the actual decay curve $\text{EDR}_{\text{dB}}(f, \mathbf{\Omega}_s, t_n)$ over their full length (i.e. t_0 to T_{IR}),

2. Defining an initial guess for the noise floor time limit t_{lim} as the first point verifying:

$$\left|\widehat{N}_j^{\text{dB}}(f, \mathbf{\Omega}_s, t_n) - \text{EDR}_{\text{dB}}(f, \mathbf{\Omega}_s, t_n)\right| \leq \varepsilon_j^N. \tag{2.29}$$

3. Calculating the "late MSE" ($\varepsilon_{\text{late}}^N$), i.e. from $t_{\text{lim},j}$ to T_{IR}.

The segment having resulted in the lowest $\varepsilon^N_{\text{late}}$ is then considered to provide the best matching point for fitting the noise profile. To ensure that the noise floor limiting point $\{P_{\text{lim}}, t_{\text{lim}}\}$ belongs to the "true" exponential decay section of the curve, additional headroom above the fitted noise profile is then adaptively determined (see complete procedure below). This headroom is crucial in subsequently obtaining an accurate model fit (see Fig. 2.7c).

Next, a decay start time t_{start} is defined by discarding any short "uneven" early segments in the EDR curve, which are assumed to correspond to non-exponentially decaying early reflections. Specifically, the slopes of all segments from t_0 to t_{lim} are compared amongst each other and t_{start} is defined as the start of the first segment whose slope is within one standard deviation of the mean slope. To avoid modelling over too short of a decay segment, t_{start} is finally ensured to be smaller than or equal to the estimated t_{mix} (by forcing $t_{\text{start}} = t_{\text{mix}}$ if necessary)[9].

Once the exponentially decaying section of the EDR curve has thus been bounded by t_{start} and t_{lim}, it is verified that it covers a set minimum dynamic range $\Delta P^{\text{dB}}_{\text{fit}}$ such that:

$$P^{\text{dB}}_{\text{lim}}(f, \mathbf{\Omega}_s) \leq \text{EDR}_{\text{dB}}(f, \mathbf{\Omega}_s, t_{\text{start}}) - \Delta P^{\text{dB}}_{\text{fit}}. \tag{2.30}$$

If the condition is verified, the reverberation tail envelope parameters $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$ can then be estimated by fitting a modelled envelope:

$$\widehat{P}(f, \mathbf{\Omega}_s, t, \mathbf{\Lambda}) = P_0(f, \mathbf{\Omega}_s) e^{-2\gamma(f, \mathbf{\Omega}_s)t}, \tag{2.31}$$

where $\mathbf{\Lambda}$ denotes the parameter vector containing the P_0 and γ values.

Since the EDR is a reverse-integrated approximation of the envelope, the actual fitting must then be performed using:

$$\begin{aligned} \widehat{\text{EDR}}(f, \mathbf{\Omega}_s, t, \mathbf{\Lambda}) &= \int_t^\infty \widehat{P}(f, \mathbf{\Omega}_s, \tau, \mathbf{\Lambda}) d\tau \\ &= \frac{P_0(f, \mathbf{\Omega}_s)}{2\gamma(f, \mathbf{\Omega}_s)} e^{-2\gamma(f, \mathbf{\Omega}_s)t}. \end{aligned} \tag{2.32}$$

In the case of a single-slope decay, the envelope parameters can be estimated using a straightforward linear regression on the identified decay section of the EDR_{dB} (between t_{start} and t_{lim}). The y-intercept of the regression then corresponds to $\widetilde{P}^{\text{dB}}_0(f, \mathbf{\Omega}_s) = 10 \log_{10}\left[\frac{P_0(f, \mathbf{\Omega}_s)}{2\gamma(f, \mathbf{\Omega}_s)}\right]$ and the slope to $m(f, \mathbf{\Omega}_s) = \frac{-60}{T_{60}(f, \mathbf{\Omega}_s)}$.

In the generalized case allowing for multiple-slope decays as in Eq. 2.10, the modelled $\widehat{\text{EDR}}(f, \mathbf{\Omega}_s, t, \mathbf{\Lambda})$ must be fitted using a parameter search algorithm in an expectation-maximization (EM) or maximum-likelihood (ML) approach. A model MSE ε_{mod} over the decay region can then be used as a loss function (or inversely as a likelihood measure):

$$\varepsilon_{\text{mod}}(f, \mathbf{\Lambda}) = \frac{1}{N_{\text{fit}}} \sqrt{\sum_{n=n_s}^{n_e} \left[\text{EDR}_{\text{dB}}(f, t_n) - \widehat{\text{EDR}}_{\text{dB}}(f, t_n, \mathbf{\Lambda})\right]^2}, \tag{2.33}$$

where $N_{\text{fit}} = n_e - n_s + 1$ with n_s the time index such that $t_{n_s} = t_{\text{start}}$ and similarly n_e such that $t_{n_e} = t_{\text{lim}}$. Note that in general the parameter search space is of dimension $2C$, since for each exponential decay both $P_{0,j}(f, \mathbf{\Omega}_s)$ and $\gamma_j(f, \mathbf{\Omega}_s)$ must be estimated. In the interest of computational efficiency, it is therefore necessary to limit the length of each search dimension.

[9] One could also choose to default to $t_{\text{start}} = t_{\text{mix}}$ in general, which would indeed follow from the considerations detailed in Sec. 2.1.1; however, imposing this choice across all frequencies increases the risk of not having enough dynamic range in the decay curve for successful modelling, especially at higher frequencies. Furthermore, an adaptive definition of t_{start} can allow modeling to be performed over longer decay segments, even at low to mid frequencies, potentially increasing the accuracy of the fit.

To this end, the identified decay region (from t_{start} to t_{lim} [or equivalently n_s to n_e]) is re-segmented by the adaptive RDP algorithm and linear regressions are performed on each new segment. The resulting range of $\widetilde{P}_0^{\text{dB}}(f, \mathbf{\Omega}_s)$ and $m(f, \mathbf{\Omega}_s)$ values is then used to define the parameter search space, and the EM (or ML) is initialized using to cover the entire range using the available number of slopes. For example, in a two-slope search, the initial guess for the first slope uses the largest $\widetilde{P}_{0,j}(f, \mathbf{\Omega}_s)$ and $\gamma_j(f, \mathbf{\Omega}_s)$ values (giving the steepest possible slope) and conversely the smallest values are used for the second (giving the flattest possible slope).

Note also that the parameter search is performed in terms of $\widetilde{P}_{0,j}(f, \mathbf{\Omega}_s)$, which must therefore be "corrected" back to the actual envelope $P_{0,j}(f, \mathbf{\Omega}_s)$, as evidenced by Eq. 2.32.

Regardless of the number of slopes being used, an appropriate value for the noise floor headroom is essential to the accurate modeling of the late tail. As discussed above and shown in Fig. 2.7c, the point $\{P_{\text{lim}}, t_{\text{lim}}\}$ must be ensured to belong to the actual exponential decay region of the EDR curve. Beginning with a minimum headroom value such that $P_{\text{head}}^{\text{dB}} \leq P_{\text{lim}}^{\text{dB}}(f, \mathbf{\Omega}_s) - \widehat{N}_{\text{dB}}(f, \mathbf{\Omega}_s, t_{\text{lim}})$, the condition is iteratively increased until the headroom "runs into" the set minimum dynamic range for fitting, $\Delta P_{\text{fit}}^{\text{dB}}$ (i.e. $\widehat{N}_{\text{dB}}(f, \mathbf{\Omega}_s, t_{\text{lim}}) + P_{\text{head}}^{\text{dB}} \leq \text{EDR}_{\text{dB}}(f, \mathbf{\Omega}_s, t_{\text{start}}) - \Delta P_{\text{fit}}^{\text{dB}}$ must also be verified).

The resulting set of fits are compared using a likelihood function based on the Akaike Information Criterion (AIC) [90] (dropping the frequency and direction dependences for clarity):

$$\mathcal{L}_{\text{fit}}\left(P_{\text{head}}^{\text{dB}}\right) = -2\log\left[\varepsilon_{\text{mod}}\left(\mathbf{\Lambda}_{\text{mod}}, P_{\text{head}}^{\text{dB}}\right)\right] + \log\left[N_{\text{fit}}\left(P_{\text{head}}^{\text{dB}}\right)\right], \tag{2.34}$$

where $\mathbf{\Lambda}_{\text{mod}}$ is the estimated parameter set for the fit obtained with a given $P_{\text{head}}^{\text{dB}}$ value. The fit that maximizes $\mathcal{L}_{\text{fit}}$ is then retained as the best for the current frequency bin.

The generalization to multiple-slope decays additionally implies that the number of slopes to consider when modeling must be determined *a priori*. In order to make multiple-slope phenomena more apparent and furthermore to simplify the connection to coupled-volume theory [5] (i.e. in order to avoid having to consider the frequency-dependent extensions to the coupling theory), this is performed in a broadband manner ahead of the above bin-by-bin EDR analysis.

Essentially, the analysis procedure described above is iteratively applied to the broadband EDC of each DRIR signal, assuming an increasing number of slopes at each step. For each resulting model, a second likelihood function is calculated according to:

$$\mathcal{L}_{\text{slope}}(\mathbf{\Omega}_s, C) = \mathcal{L}_{\text{fit}}^{\max}(\mathbf{\Omega}_s, C) - 2C, \tag{2.35}$$

where $\mathcal{L}_{\text{fit}}^{\max}(\mathbf{\Omega}_s, C)$ is the fitting likelihood that was maximized as above. The $\mathcal{L}_{\text{slope}}(\mathbf{\Omega}_s, C)$ values are then averaged over the S look directions of the DRIR, and if the resulting $\overline{\mathcal{L}}_{\text{slope}}(C) > \overline{\mathcal{L}}_{\text{slope}}(C-1)$, the following iteration (for $C+1$) is tested. An additional condition on the relative differences between the modelled parameters may also be useful here in order to ensure that truly separate decays are being identified. For example, a C-slope model may only be acceptable if the relative differences between the T_{60} values of each pair of consequent decays are all above a certain threshold. Otherwise, $C-1$ is retained as the number of slopes with which to perform the above analysis.

In summary, then, the late reverberation decay envelope modeling procedure consists of:

1. *Determining the number of slopes to model*: Iterative broadband EDC fittings are performed with increasing numbers of slopes and compared using the likelihood function $\overline{\mathcal{L}}_{\text{slope}}(C)$.

2. *Segmenting the dB-scale curve of a single EDR bin*: For each directional signal, the EDR is analyzed bin-by-bin in dB scale. For each bin, the first step is to segment the decay curve using an adaptive RDP algorithm.

3. *Identifying the noise floor*: The theoretical archetype of a non-decaying background noise floor is fitted to the EDR bin by successively matching it to each segment.

4. *Determining the bounds of the exponential decay region*: The noise floor limiting point $\{P_{\text{lim}}, t_{\text{lim}}\}$ is defined using a given amount of headroom $P_{\text{head}}^{\text{dB}}$ above the fitted noise profile. The exponential decay start point t_{start} is identified by discarding uneven early segments.

5. *Modeling the exponential decay envelope*: If the true exponential decay section of the EDR bin, as bounded by $[t_{\text{start}}, t_{\text{lim}}]$, covers a dynamic range greater than $\Delta P_{\text{fit}}^{\text{dB}}$, then it is modelled either through linear regression (in the case of a single-slope decay) or by a maximum-likelihood parameter search (in the generalized case of multi-slope decays).

Steps 4 and 5 are then repeated for increasing values of $P_{\text{head}}^{\text{dB}}$ and the fit corresponding to the maximum $\mathcal{L}_{\text{fit}}\left(P_{\text{head}}^{\text{dB}}\right)$ is retained.

This analysis procedure will be evaluated in Ch. 4 on synthesized and deliberately noised SRIR simulations, and in Ch. 5 on real-world SMA measurements.

2.5.2/ Late Tail Isotropy Analysis

Though the choice of a modeling approach allowing for direction-dependent reverberation properties was initially solely guided by intuition, a growing body of recent work on the analysis of isotropy in SRIRs has come to further support this view. As mentioned in Sec. 1.2.3, a method for evaluating directional energy variations in SRIRs was recently proposed by Berzborn and Vorländer [68] through the use of directional energy decay curves (DEDC).

Through a collaborative research project with Benoit Alary (then at the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland) between fall 2018 and spring 2020, this idea was further developed into the space-time-frequency energy decay deviation (EDD) measure [91]. For a DRIR decomposed onto S look directions $\mathbf{\Omega}_s$ as described above, the EDD is defined as:

$$\text{EDD}(f, \mathbf{\Omega}_s, t) = \text{EDR}_{\text{dB}}(f, \mathbf{\Omega}_s, t) - \overline{\text{EDR}}_{\text{dB}}(f, t) \;\forall\; t \in [t_{\text{mix}}, t_{\text{lim}}], \tag{2.36}$$

where $\overline{\text{EDR}}_{\text{dB}}(f, t)$ is the spatial mean of $\text{EDR}_{\text{dB}}(f, \mathbf{\Omega}_s, t)$ over all look directions at each time step. Note that the mean is performed in dB scale in order to obtain a more perceptually relevant measure; comparisons with a linear energy average (as in Berzborn and Vorländer [68]) will be included in Ch. 4.

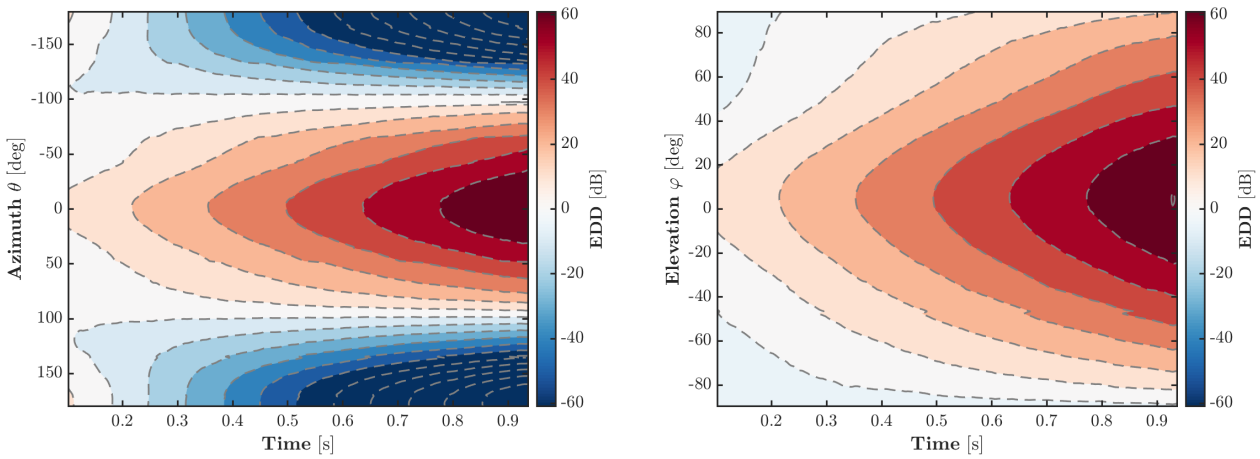
Interpretation of the EDD is thus relatively straightforward: a perfectly isotropic late reverberation tail would present an $\text{EDD}(f, \mathbf{\Omega}_s, t) = 0 \;\forall\; f, \mathbf{\Omega}_s, t$. Positive values represent space-time-frequency points where the decay curve contains more energy than the spatial mean, while conversely negative values represent points with less energy than the mean.

Though in theory the EDD could be calculated by beamforming over an arbitrary number of look directions like an SRP (since we are once again only interested in evaluating power estimates), this analysis is usually performed in the context of a DRIR generated in such a way as to preserve the underlying signal properties of the late field (as above).

In fact, as will become apparent when evaluating the late reverberation analysis-synthesis methods in Ch. 4 and Ch. 5, the exponential decay envelope properties are highly sensitive to the lobe overlap effects that stem from the limited PWD directivities. It therefore seems pertinent to evaluate the EDD in the same directional context as the other late reverberation analysis methods.

To obtain EDD values to within arbitrary precision, the same spherical cubic Hermite interpolation scheme used for the Herglotz early reflection detection method can be subsequently applied [87]. For visualization purposes, it is often practical to interpolate onto the azimuthal and elevational planes and

then average the results over a frequency band of interest in order to obtain an intuitive pair of figures such as the ones given as an example in Fig. 2.8 for the anisotropic DRIR described in Sec. 2.3.



(a) Azimuthal plane (at $\varphi = 0°$). (b) Elevational plane (at $\theta = 0°$).

Figure 2.8: An example of EDD visualization using the late reverberation tail of the simulated SRIR described in Sec. 2.3 (and used to demonstrate the mixing time estimation algorithm in Fig. 2.5). The "full" space-time-frequency $\mathrm{EDD}(f, \boldsymbol{\Omega}_s, t)$ is interpolated onto the azimuthal (a, left) and elevational (b, right) planes, and then averaged over the frequency band $[f_{\mathrm{dir},1}, f_{\mathrm{alias}}]$. The effect of the cardioid $T_{60}(\boldsymbol{\Omega}_s)$ distribution is clearly visible through the increasing decay deviation at $(0°, 0°)$.

The specific "frequency band of interest" is once again defined here in the context of SMA-measured SRIRs as the range between the first-order directivity limit $f_{\mathrm{dir},1}$ and the aliasing frequency f_{alias} (see Sec. 2.4.2 above). The use of this particular frequency range thus ensures that the directional information displayed is as accurate as possible (not affected by omnidirectional low frequencies or spatial aliasing high frequencies) while providing a broadband visualization.

Finally, one of the most important uses of the EDD in the context of this research project is in identifying the overall degree of isotropy of an SRIR's late reverberation tail. As we will see in the following chapter, this measure can condition many choices with respect to the subsequent treatment and manipulation methods that can be applied.

In this thesis, the simplest possible choice of measure is made: the range of the "broadband" (i.e. averaged over the aforementioned frequency range) DRIR EDD (i.e. not interpolated) is averaged over the length of the late tail. Mathematically:

$$\mathcal{I}_{\mathrm{EDD}} = \frac{1}{N_{\mathrm{late}}} \sum_{n=n_{\mathrm{mix}}}^{n_{\mathrm{lim}}} \left\{ \max_{\boldsymbol{\Omega}_s} \left[\overline{\mathrm{EDD}}(\boldsymbol{\Omega}_s, t_n) \right] - \min_{\boldsymbol{\Omega}_s} \left[\overline{\mathrm{EDD}}(\boldsymbol{\Omega}_s, t_n) \right] \right\}, \tag{2.37}$$

where n_{mix} and n_{lim} are the time frame indices such that $t_{n_{\mathrm{mix}}} = t_{\mathrm{mix}}$ and $t_{n_{\mathrm{lim}}} = t_{\mathrm{lim}}$, respectively, $N_{\mathrm{late}} = n_{\mathrm{lim}} - n_{\mathrm{mix}} + 1$, and $\overline{\mathrm{EDD}}(\boldsymbol{\Omega}_s, t_n)$ is the broadband EDD as averaged over the frequency range described above. The resulting value $\mathcal{I}_{\mathrm{EDD}}$ is in positive dB by definition, since $\max_{\boldsymbol{\Omega}_s} \left[\overline{\mathrm{EDD}}(\boldsymbol{\Omega}_s, t_n) \right] > 0$ and $\min_{\boldsymbol{\Omega}_s} \left[\overline{\mathrm{EDD}}(\boldsymbol{\Omega}_s, t_n) \right] < 0$ by construction.

The EDD and the isotropy measure $\mathcal{I}_{\mathrm{EDD}}$ defined above will both be evaluated on synthesized SRIR simulations in Ch. 4. Moreover, they also form an integral part of the discussion in Ch. 5 following their application to SMA measurements performed in unusual and acoustically complex spaces.

This chapter has presented the underlying SRIR signal model from which the various analysis, treatment, and manipulation methods developed throughout this thesis are all informed (Sec. 2.1). First the overall time-frequency sectioning of the SRIR was justified (Sec. 2.1.1), before developing the specific space-time-frequency characteristics of each signal regime (Secs. 2.1.2 and 2.1.3). The remaining sections of the chapter were dedicated to presenting the analysis methods directly stemming from the signal model. Section 2.3 discussed the imperative question of estimating the mixing time beyond which the stochastic late reverberation model is valid. Methods for detecting the direct sound and generating cartographies of the predominant early reflections were given in Sec. 2.4, while finally Sec. 2.5 dealt with the modeling of the late reverberation tail.

As such, the analytical side of the general space-time-frequency SRIR analysis, treatment, and manipulation framework, whose preliminary definition constitutes of the main objectives of this research project, is now complete. In the following chapter we turn to the investigation of necessary treatment procedures and possible manipulation strategies, before completing this thesis with a battery of validation tests (Ch. 4) and a general discussion with respect to the "real-world" applications of the framework (Ch. 5).

---

3 / Treatment and Manipulation Methods

The general SRIR signal model and the different analysis methods used to estimate its various parameters having been the subject of the previous chapter (Ch. 2), we now turn to the description of procedures making use of that knowledge to treat and/or manipulate a measured SRIR.

The use of both *treatment* and *manipulation* as separate terms here is deliberate and marks a fundamental difference in the strategic aim of the procedures referred to by each. Treatment methods are strictly corrective processes that seek to minimize, compensate, or otherwise mitigate potential sources of errors, corruption, or noise; in other words, they encompass methods that aim to eliminate unwanted components in order to recover a signal as "true" as possible to the ideal model.

Manipulations, on the other hand, refer to a more creative approach in which various aspects of the overall three-dimensional reverberation effect are sought to be modified in order to achieve a desired final sound. A major consequence of these differences is that while treatment procedures are constrained to strictly preserve and/or recreate all the physical properties of an SRIR, manipulations can be permitted, for example, to violate the geometrical reality of the measured space.

Since many of these processes have been inspired by phenomena encountered in measured SRIRs, a few real-world examples are provided throughout the chapter in order to illustrate the motivation behind each method. These example measurements are not necessarily detailed outright herein, as a more thorough discussion of real-world SRIRs is reserved for Ch. 5, but they nonetheless serve to highlight the observed phenomena and demonstrate either the necessity of correcting them or the possibility of manipulating them.

Finally, it should be noted that this research project focuses particularly on the manipulation of the spatial aspects of SRIRs. Indeed, this constitutes one of the main novelties of the current work with respect to the types of treatment and manipulation methods available in the literature, which have so far mostly been based on time-frequency modifications (see Sec. 1.1.3).

This chapter begins with a fundamental (pre-)treatment procedure that operates directly on the measured exponential sweep method (ESM) [20] signals of the SMA transducers (Sec. 3.1). Section 3.2 then presents three different methods for manipulating the early reflections of an SRIR, as detected using the analysis framework from Sec. 2.4. Section 3.3.1 provides a denoising procedure for eliminating the non-decaying background noise floor and replacing it with a resynthesized prolongation of the SRIR's "true" late reverberation tail; this idea is then extended to resynthesizing the entire reverberation tail from t_{mix} onwards in Sec. 3.3.2, offering the possibility of manipulating the spatial distributions of the decay parameters. The chapter ends with a description of methods allowing for the modification of the late tail's isotropy as described and parameterized by the EDD (Sec. 3.4).

3.1 / Pre-Analysis Treatment

The use of the ESM [20] for measuring RIRs is inherently exposed to four main factors that risk corrupting the resulting signal (regardless of where a single omnidirectional microphone or an entire microphone array is used). The first is the presence of a constant stationary background noise, which can be assumed to be the sum of several stationary noise sources including any electronic sensor

self-noise (e.g. from the microphone transducers). This is in fact the term $N(f, \mathbf{\Omega}, t)$ in the general SRIR signal model given in Eq. 2.1 and approximated as $\widehat{N}(f, \mathbf{\Omega}_s, t)$ in Sec. 2.5.1.

The second potential risk is the non-stationarity of the environmental conditions throughout the measurement time period: the speed of sound c is especially contingent on the temperature, atmospheric pressure, and humidity of the air in the space being measured. In the case of repeated sweep measurements, any change in conditions will result in differences between the repetitions that can corrupt the final averaged IR. In this thesis we will assume that the risk of changes in environmental conditions can be mitigated by limiting measurement sequences to time periods over which these can be considered stable (on the order of 1 to 3 minutes maximum).

As mentioned in Sec. 2.4.1, harmonic distortion from overdriving the source loudspeaker is often inevitable when high excitation levels are necessary in order to obtain an acceptable SNR over the entire audible frequency range[1]. The main danger stemming from harmonic distortion is the generation of overlapping preceding copies of the "deconvolved" IR, but as once again noted in Sec. 2.4.1 this can usually be mitigated by carefully parameterizing the exponential sweep sequence. Furthermore, successfully detecting the direct sound implies that any preceding noise can either be replaced by silence or entirely removed, such that all subsequent analysis is performed with respect to t_0.

The final risk factor to consider is the possible presence of non-stationary noise "events", i.e. salient momentary sounds compared to the stationary background noise. ESM measurements are especially sensitive to impact-type sounds such as steps, closing doors, or falling objects: when convolving the measured sweep sequence with the corresponding inverse sweep signal, such impulsive sounds generate copies of the inverse sweep in the resulting IR. Not only can these be audible, they can also strongly affect the shape of the noise floor, making the constant-power noise profile $\widehat{N}(f, \mathbf{\Omega}_s, t)$ difficult to fit.

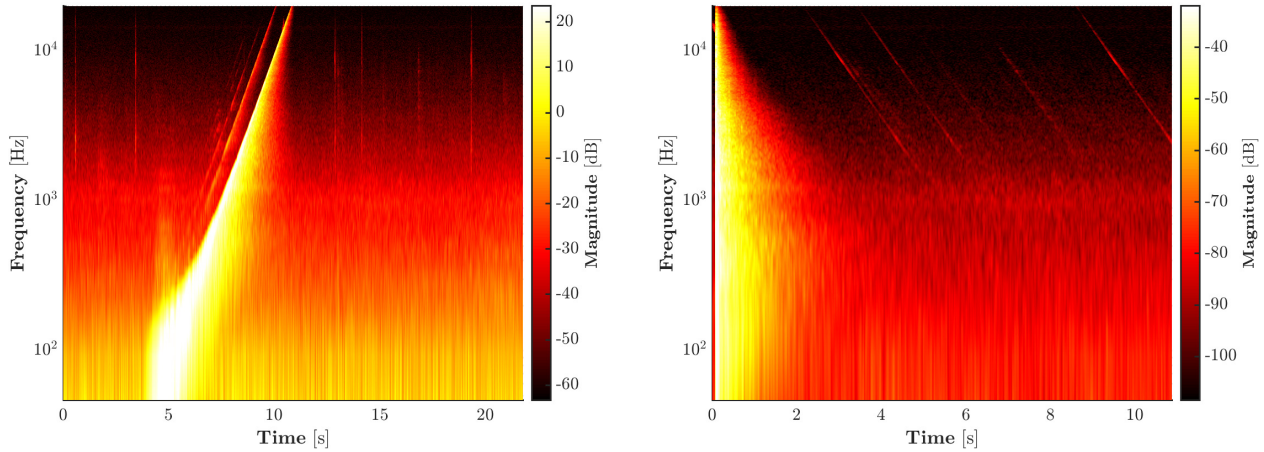
Non-stationary noise in general is an additional issue in the context of repeated sweep sequences. The main idea behind repeating sweep measurements and averaging over the repetitions is to take advantage of the fact that the IR signals will sum coherently (between themselves, not to be confused with the plane wave coherence properties within a given IR), thereby gaining ~ 6 dB each time, while the stationary background noise will sum incoherently ~ 3 dB at a time: one can therefore gain 3 dB of broadband SNR for every additional repetition in a measurement sequence.

Any non-stationary noise events, however, will not "average out" incoherently but will instead accumulate in the final averaged IR signal (though their amplitude relative to the IR could diminish with a large enough number of repetitions, such an approach would risk violating the long-term stationarity of the environment conditions). Figure 3.1 presents an example ESM measurement in which several impulsive artefacts have accumulated in the raw sweep measurement (as averaged over four repetitions, Fig. 3.1a). In particular, Fig. 3.1b demonstrates how these impulsive sounds generate copies of the inverse sweep signal in the final "deconvolved" IR (the raw sweep measurement convolved with the corresponding inverse sweep).

For these reasons it is important to treat the raw ESM measurement sequence *before* performing the various analysis techniques described in Ch. 2 in an attempt to minimize the effects of such non-stationary noise, and especially of the impulsive type shown in Fig. 3.1.

The proposed pre-treatment procedure, as initially presented in Massé et al. [84], therefore seeks to reduce the presence of non-stationary noise in the repeated ESM measurement before averaging and final convolution with the inverse sweep signal. This is done by comparing the time-frequency magnitude spectrograms of the individual measured sweep repetitions amongst each other. Positive deviations from the mean magnitude spectrogram (as averaged over the repetitions) that are above a maximum allowable threshold are thus used as a discriminating criterion in order to identify the types

[1] The idea of an "acceptable" SNR is intrinsically linked to the minimum fitting range $\Delta P_{\text{fit}}^{\text{dB}}$ described in Sec. 2.5.1. RIR measurement is thus often a cyclical process in which the excitation levels are iteratively adjusted according to the observed SNRs to ensure that the $\Delta P_{\text{fit}}^{\text{dB}}$ condition can be met across the audible frequency range.



(a) Raw averaged ESM measurement signal.

(b) "Deconvolved" IR.

Figure 3.1: Auto-PSDs of an ESM measurement presenting several impulsive non-stationary noise artefacts: (a, left) the raw transducer signal (averaged over four sweep repetitions) showing the impact-type noises and (b, right) the IR obtained after convolution with the inverse sweep signal, showing how the impulsive artefacts create copies of the inverse sweep throughout the noise floor. This particular example is taken from a single microphone channel of an mh acoustics Eigenmike® SRIR measurement performed at the Grandes Serres de Pantin in Pantin, France.

of non-stationary artefacts described above.

This threshold can be written as:

$$\xi_q(f_k, t_n) = \mu_q(f_k, t_n) + \lambda_{\text{dev}} \sigma_q(f_k, t_n), \tag{3.1}$$

where $\mu_q(f_k, t_n)$ is the mean magnitude spectrogram, λ_{dev} is an empirically set deviation factor used as a sensitivity parameter, $\sigma_q(f_k, t_n)$ is the standard deviation over the available repetitions, and the subscript q indicates that the procedure is applied to each SMA transducer signal independently.

Artefact magnitude values identified as greater than $\xi_q(f_k, t_n)$ in each individual sweep measurement are then replaced with the corresponding mean over the remaining repetitions, i.e. averaged over all repetitions excluding the "current" one in which the artefact values have been detected.

To summarize, the pre-treatment procedure operates as such:

1. Calculate magnitude spectrograms $|X_q(f_k, t_n)|$ on the raw ESM measurement signals for each SMA transducer and separate the J individual sweep repetitions, $|X_{q,j}(f_k, t_n)|$.
2. Calculate the mean magnitude spectrograms $\mu_q(f_k, t_n)$ and their standard deviations $\sigma_q(f_k, t_n)$ over the available repetition spectrograms $|X_{q,j}(f_k, t_n)|$, $j \in \{1, 2, \ldots, J\}$.
3. Calculate the thresholds $\xi_q(f_k, t_n)$ according to Eq. 3.1.
4. Identify the artefact time-frequency points $\{f_{k'}, t_{n'}\}_j$ such that $|X_{q,j}(f_{k'}, t_{n'})| > \xi_q(f_{k'}, t_{n'})$.
5. Replace the magnitude of the detected points with the corresponding values from the mean magnitude spectrum $\tilde{\mu}_{q,j}(f_{k'}, t_{n'}) = \frac{1}{J-1} \sum_{j'=1}^{J} |X_{q,j'}(f_{k'}, t_{n'})| (1 - \delta_{j,j'})$.
6. Reconstruct the sweep measurement STFTs with the "corrected" magnitudes and proceed to signal averaging and convolution with the inverse sweep as for the usual ESM.

Real-world examples of the types of impulsive artefacts discussed here and the ability of the above procedure to attenuate them will be presented in Sec. 5.1. The effect of applying the pre-treatment

process on the subsequent performance of the EDR analysis, most notably in terms of fitting the background noise profile, will also be evaluated.

3.2 / Spatiotemporal Early Reflection Filtering

This section presents some experimental manipulation methods applicable to the early segment of the SRIR, following the echo detection procedure described in Sec. 2.4 and the subsequent generation of a space-time reflection map. The first two manipulation approaches here are simply based on adjusting the relative levels of the detected reflections in order to either a) constrain them to the exponential decay envelope that governs the late reverberation tail (thereby reducing their salience with respect to the late reverberation, Sec. 3.2.1) or b) on the contrary accentuate their salience by applying a "global" space-time gain map (Sec. 3.2.2).

The third and final manipulation technique (Sec. 3.2.3) is a solely spatial procedure based on the directional echo energy density (DEED) map described in Sec. 2.4.3. The general idea is to "redistribute" the detected reflections through rotations on the sphere in order to "flatten" the echo density map (i.e. to obtain an isotropic DEED).

Due to the experimental nature of these manipulations, the evaluation of their influence on the perception of room reverberation has unfortunately fallen outside the scope of this thesis. Nonetheless, their application to real-world SMA measurements will be discussed in Ch. 5 and will form a proof-of-concept of sorts, in the hope that such preliminary results may provide motivation for future work on their perceptual aspects.

As a last note, it should be pointed out that the direct sound detection presented in Sec. 2.4.1 allows for precise manipulation of the direct-to-reverberant sound ratio (DRR), a widely used RIR descriptor with known perceptual influences [92] [93].

3.2.1 / Constraining Echoes to an Exponential Decay Envelope

The main objective in constraining the energies of detected echoes to the theoretical late reverberation decay envelope "as it would be" in the early segment (i.e. extrapolated backwards from t_{mix} to t_0) is to reduce the salience of highly prominent early reflections with respect to both the direct sound and the late reverberation. Figure 3.2 provides an example of an RIR with such salient early reflections; the broadband late decay model is superimposed (red solid line) to further illustrate the deviations from the exponential envelope in the early segment (delimited by the mixing time t_{mix} [dashed blue line], as estimated using the method proposed in Sec. 2.3).

To perform this operation, the directional late exponential decay parameters $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$, as estimated using the approach described in Sec. 2.5, are first used to synthesize an extrapolated "early" equivalent of the reverberation tail. The simplest way to achieve this is to generate a zero-mean Gaussian white noise sequence (following Jot et al. [16]) for each direction $\mathbf{\Omega}_s$ to which the space-time-frequency decay envelope model $\widehat{P}(f, \mathbf{\Omega}_s, t)$ can then be applied[2]:

$$\widehat{H}_{\text{late}}(f, \mathbf{\Omega}_s, t) = \sqrt{\widehat{P}(f, \mathbf{\Omega}_s, t)}\widehat{H}_{\mathcal{N}}(f, \mathbf{\Omega}_s, t) \;\forall\; f, \mathbf{\Omega}_s, t, \tag{3.2}$$

where $\widehat{P}(f, \mathbf{\Omega}_s, t)$ is calculated using Eq. 2.10 and $\widehat{H}_{\mathcal{N}}(f, \mathbf{\Omega}_s, t)$ is the time-frequency representation of the Gaussian white noise sequence generated for the direction $\mathbf{\Omega}_s$. See Sec. 3.3 below for more details on synthesizing the late reverberation field.

However, since the mixing time is by definition a broadband and omnidirectional measure (see Sec. 2.3), it may not in general match the refined $t_{\text{start}}(f, \mathbf{\Omega}_s)$ determined when modelling

[2] This is, in fact, also the approach taken in order to resynthesize the late reverberation tail in Sec. 3.3.

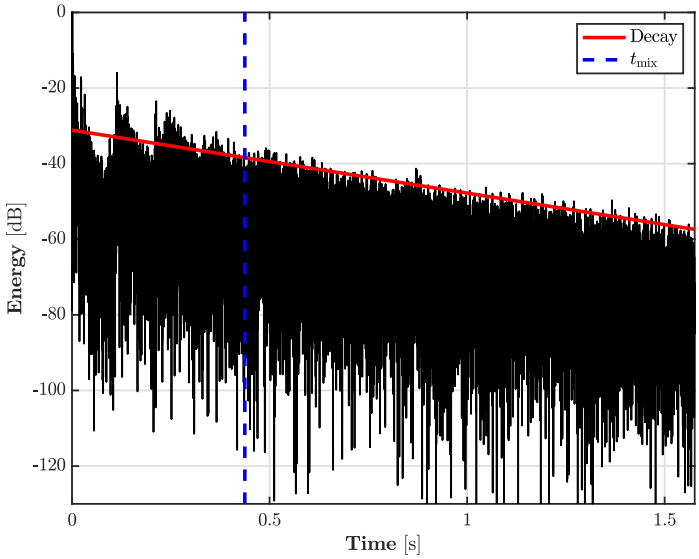


Figure 3.2: Example of a broadband omnidirectional RIR with highly salient early reflections. The solid red line represents the broadband exponential decay model envelope fitted to the late reverberation tail following the analysis method presented in Sec. 2.5.1. The blue dashed line represents the mixing time t_{mix} as estimated using the approach given in Sec. 2.3.

the late exponential decay envelope over the directional EDRs (see Sec. 2.5.1). As a result, there may exist some spectral mismatches between the measured DRIR and the resynthesized signal's envelopes at t_{mix}; to correct these, a spectral compensation can be calculated as:

$$P_{\text{comp}}(f, \boldsymbol{\Omega}_s) = \frac{\text{EDR}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})}{\widehat{\text{EDR}}_{\text{late}}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})}, \tag{3.3}$$

where $\text{EDR}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})$ is the directional EDR of the measured DRIR [as used to estimate the decay parameters $P_0(f, \boldsymbol{\Omega}_s)$ and $T_{60}(f, \boldsymbol{\Omega}_s)$], and $\widehat{\text{EDR}}_{\text{late}}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})$ is the EDR calculated on the resynthesized signal $\widehat{H}_{\text{late}}(f, \boldsymbol{\Omega}_s, t)$ (Eq. 3.2). The "corrected" exponentially decaying signal towards which the DRIR's early segment is to be adjusted is then described by:

$$\widehat{H}_{\text{ref}}(f, \boldsymbol{\Omega}_s, t) = \sqrt{P_{\text{comp}}(f, \boldsymbol{\Omega}_s)}\, \widehat{H}_{\text{late}}(f, \boldsymbol{\Omega}_s, t), \tag{3.4}$$

and the space-time-frequency gain map to be applied to the measured DRIR is defined as:

$$M^{\text{dB}}_{\text{exp}}(f, \boldsymbol{\Omega}_s, t) = 20\nu_{\text{exp}} \log_{10} \frac{\left|\widehat{H}_{\text{ref}}(f, \boldsymbol{\Omega}_s, t)\right|}{\left|\widetilde{H}(f, \boldsymbol{\Omega}_s, t)\right|}, \ \forall\ f, \boldsymbol{\Omega}_s, \ \forall\ t \in [t_0, t_{\text{mix}}]\,, \tag{3.5}$$

where $\nu_{\text{exp}} \in [0, 1]$ is a control parameter that determines the extent to which the map is applied. The final modified DRIR is then obtained by:

$$\widehat{H}_{\text{exp}}(f, \boldsymbol{\Omega}_s, t) = \begin{cases} M_{\text{exp}}(f, \boldsymbol{\Omega}_s, t)\widetilde{H}(f, \boldsymbol{\Omega}_s, t) & \text{for } t_0 \le t < t_{\text{mix}}, \\ \widetilde{H}(f, \boldsymbol{\Omega}_s, t) & \text{for } t \ge t_{\text{mix}}, \end{cases} \ \forall\ f, \boldsymbol{\Omega}_s, \tag{3.6}$$

where $M_{\text{exp}}(f, \boldsymbol{\Omega}_s, t) = 10^{\frac{M^{\text{dB}}_{\text{exp}}(f, \boldsymbol{\Omega}_s, t)}{20}}$ is the linear amplitude equivalent to $M^{\text{dB}}_{\text{exp}}(f, \boldsymbol{\Omega}_s, t)$.

Note that the modification gain map $M_{\text{exp}}(f, \boldsymbol{\Omega}_s, t)$ is defined using the relative space-time-frequency magnitudes of the resynthesized and original signals instead of the envelope model or EDRs in order to recreate the stochastic short-term energy fluctuations characteristic of the late reverberation tail.

On the other hand, the phase values of the early segment are left untouched in order to preserve the spatial coherence and DoAs of the discrete early reflections.

3.2.2/ Accentuating Reflection Salience

In a manner somewhat opposite to that described just above, the salience of the early reflections with respect to the late exponential decay can be accentuated instead of "corrected", by applying a relative gain ν_{acc} or attenuation $1/\nu_{\text{acc}}$ depending on whether their energy is above or below the modelled envelope. The reference exponentially decaying signal $\widehat{H}_{\text{ref}}(f, \boldsymbol{\Omega}_s, t)$ is first constructed in the same manner as above (Eq. 3.4), and the modification gain map is this time defined by:

$$M_{\text{acc}}(f, \boldsymbol{\Omega}_s, t) = \begin{cases} \nu_{\text{acc}}(\boldsymbol{\Omega}_s, t) & \text{if } \dfrac{\left|\widetilde{H}(f, \boldsymbol{\Omega}_s, t)\right|}{\left|\widehat{H}_{\text{ref}}(f, \boldsymbol{\Omega}_s, t)\right|} > 1, \\ \dfrac{1}{\nu_{\text{acc}}(\boldsymbol{\Omega}_s, t)} & \text{if } \dfrac{\left|\widetilde{H}(f, \boldsymbol{\Omega}_s, t)\right|}{\left|\widehat{H}_{\text{ref}}(f, \boldsymbol{\Omega}_s, t)\right|} < 1, \end{cases} \quad \forall\ f, \boldsymbol{\Omega}_s,\ \forall\ t \in [t_0, t_{\text{mix}}]. \tag{3.7}$$

The modified DRIR is then obtained in the same way as Eq. 3.6 above.

Note that the accentuation gain parameter $\nu_{\text{acc}}(\boldsymbol{\Omega}_s, t)$ is written as a potential function of both direction and time in order to allow for different mapping strategies. For example, the applied factor can be made to evolve over the time range $t \in [t_0, t_{\text{mix}}]$ such that it reaches 0 dB at t_{mix}, i.e. with earlier reflections being more affected than later ones. A natural choice for such a map would be a logarithmic (linear in dB) decrease starting from a chosen initial dB value $\nu_{\text{acc}}(\boldsymbol{\Omega}_s, t_0)$.

The modification map can also be made to evolve in direction by defining a given pattern over the sphere for the initial gain/attenuation value. A cardioid pattern is used in this way as an example in Sec. 5.3.3 to demonstrate the capabilities of this manipulation scheme (alongside the time evolution described above and the exponential decay "correction" manipulation from Sec. 3.2.1).

3.2.3/ Redistributing Echoes According to Directional Density

The final early reflection manipulation procedure to be proposed here operates in a solely directional manner rather than the purely energetic methods presented above. Conceptually, the idea is to iteratively redistribute the predominant detected echoes over the sphere in the aim of "flattening" the DEED map described at the end of Sec. 2.4.3. In other words, if certain directions are privileged (i.e. contain more echoes from t_0 to t_{mix} than others), this approach will seek to reduce their prominence by repositioning the most energetic echoes to directions with lower densities.

Working within the directional STFT representation used for the echo detection analysis described in Sec. 2.4, a rotation is defined between the DoA of the most energetic echo and $\text{argmin}_{\boldsymbol{\Omega}_j}\left[\rho_N(\boldsymbol{\Omega}_j)\right]$, the minimum of the directional echo density map. Note that the choice can be made here whether to include the direct sound in the echo redistribution scheme or not.

The rotation is applied directly to the STFT frame containing the given reflection in the SH domain using a Wigner-D matrix. Since the rotation of the entire STFT frame also affects any other echoes present within it (besides the one used to define the rotation), these simultaneous reflections "along for the ride" are removed from consideration for the following iterations (in order to avoid re-rotating an already-rotated frame, which would modify the results of previous iterations).

The DEED map $\rho_N(\boldsymbol{\Omega}_j)$ is then updated according to the applied rotation, and for each following iteration the next most energetic echo is used to define the corresponding rotation (using the minimum of the updated echo density). This procedure (rotation, DEED map update, next most energetic reflection selection) is repeated until there are either no more (valid) reflections to redistribute, no

more STFT frames to rotate, or the DEED map has converged (i.e. a given number of subsequent iterations have had little to no effect on its distribution).

This concludes the last of the three proposed methods for manipulating early reflections. Once again, a true perceptual validation of their use is unfortunately beyond the capacity of this research project, but their effect on real-world SRIRs (Ch. 5) will nonetheless be demonstrated, illustrated, and discussed as a proof-of-concept. Additional ideas that may be interesting to experiment with in the future are also discussed in the Conclusion.

3.3/ Late Reverberation Tail Resynthesis

Following the modelling of the late energy decay envelope through the EDR analysis procedure described in Sec. 2.5.1, an SRIR's reverberation tail can subsequently be resynthesized under the signal model presented in Sec. 2.1.3. As presented in the Introduction, one of the most immediately pertinent applications of such resynthesis is in the correction of the non-decaying background measurement noise floor (see also Sec. 2.5.1). Resynthesizing the late reverberation allows the noise floor to be replaced with a modelled prolongation of the measured SRIR's tail (see Sec. 3.3.1 below). The measurement's limited SNR can thus be theoretically increased to an arbitrarily large dynamic range (though it is restricted in practice by the quantization noise floor of digital audio file formats).

The ability to resynthesize the full late reverberation tail, i.e. beginning at the estimated mixing time (t_{mix}) instead of the noise floor (t_{lim}), also enables a plethora of additional manipulations. As an example, Sec. 3.3.2 provides a framework for adjusting the directional distributions of both the initial late spectrum $P_0(f, \boldsymbol{\Omega}_s)$ and the reverberation time $T_{60}(f, \boldsymbol{\Omega}_s)$, based on their estimated values.

3.3.1/ Denoising

The proposed procedure for denoising measured SRIRs by late reverberation tail resynthesis was originally published for the isotropic diffuse field case in Massé et al. [84], before being extended to anisotropic, direction-dependent cases in Massé et al. [78]. Fundamentally, the approach is the same regardless of the spatial representation used: the non-decaying noise floor, defined as the SRIR signal at all $t > t_{\text{lim}}$, is directly replaced with a strict reconstruction of the late reverberation tail, generated according to the P_0 and T_{60} values estimated from the EDR analysis.

The choice of spatial representation, however, *does* depend on the isotropy of the late energy decay envelope. As demonstrated in Sec. 1.2.2 following Jarrett et al. [61] (and also shown in Epain and Jin [62] and Massé et al. [84]), the SH domain preserves the spatial incoherence of an isotropic diffuse field. Additionally, isotropic power distributions are easily represented in the SH domain (see Appx. A.3): the decay parameters of an isotropic late reverberation tail can therefore be analyzed from EDRs calculated on SH domain signals. Both of the fundamental properties for reconstructing the late field are thus satisfied and the analysis/resynthesis/denoising procedure can be performed directly on the SH domain SRIR signals [84].

On the other hand, in the general case allowing for direction-dependent late reverberation characteristics (as assumed by the signal model from Sec. 2.1.3), neither of these properties are preserved. First, as previewed in Sec. 1.2.2 and reiterated in Sec. 2.3, the SH domain cannot properly represent the spatial incoherence of a stochastic plane wave field under an anisotropic power distribution: the incoherence between individual SH components that is a characteristic feature of diffuse fields (Eq. 1.24) does not hold under an anisotropic power distribution.

Second, modelling a power envelope $P_{l,m}(f, t)$ directly in the SH domain is only valid for isotropic distributions (see once again Appx. A.3). In order to ensure that direction-dependent decay characteristics are accurately reconstructed [i.e. in order to model $P(f, \boldsymbol{\Omega}_s, t)$ as described in Ch. 2] it is therefore

necessary to operate on the spatially decomposed DRIR (i.e. the directional PWD representation of the SRIR as described in Sec. 2.2).

Following these considerations, a choice is then possible depending on the estimated isotropy of the SRIR's late reverberation tail. This can be done either through an isotropy measure like that proposed by Nolan et al. [63] ($\mathcal{I}_{\text{PW}}$, Eq. 1.26), or through the EDD described in Sec. 2.5.2; in fact, these two approaches will be compared in Sec. 4.2.4. If an SRIR is determined to have a suitably isotropic tail, it can be processed directly in the SH domain (i.e. by acting on its HOA representation). Conversely, if the late decay envelope is deemed sufficiently anisotropic, analysis and resynthesis must happen under the directional PWD representation (i.e. on the DRIR signals).

The threshold conditioning such a choice between representation domains (i.e. the notions of "suitable isotropy" and "sufficient anisotropy" referred to above) is evaluated in Sec. 4.3.1 based on their respective capacity to reconstruct the late reverberation tail under increasingly anisotropic conditions, both in terms of spatial incoherence preservation and decay envelope modelling. Eventually this choice could also be made with respect to a perceptual just-noticeable difference (JND) in isotropy; though work performed during this research project in collaboration with Benoit Alary (Aalto University, Espoo, Finland) has offered evidence that late reverberation tail anisotropy is indeed audible [94], a thorough evaluation of JNDs has been left to future work.

In either case, once the spatial representation has been chosen, the EDR analysis method presented in Sec. 2.5.1 can be applied to the corresponding set of signal channels (representing either SH components or PWD look directions). The resulting estimated exponential decay envelope parameters [either $P_0^{l,m}(f)$ and $T_{60}^{l,m}(f)$ per SH component or $P_0(f,\boldsymbol{\Omega}_s)$ and $T_{60}(f,\boldsymbol{\Omega}_s)$ per look direction] are then used to regenerate a modelled version according to Eq. 2.10 [either $\widehat{P}_{l,m}(f,t)$ or $\widehat{P}(f,\boldsymbol{\Omega}_s,t)$].

As previously mentioned (Sec. 3.2.1), the simplest way to reconstruct the individual RIR signals (i.e. for each SH component or each look direction) is as a zero-mean, unit-power Gaussian white noise to which the corresponding time-frequency decay envelope is then applied under an STFT representation [16]. Using a Gaussian process simultaneously ensures that the instantaneous "plane wave" phases are indeed stochastic and that the time-frequency spectrogram is white (i.e. fluctuating around a constant unit-power envelope).

The regenerated decay envelope can thus be directly applied to the magnitude of the Gaussian white noise's STFT and the final signal reconstructed as such:

$$\widehat{H}_{\text{late}}(f,\boldsymbol{\Omega}_s,t) = \sqrt{\widehat{P}(f,\boldsymbol{\Omega}_s,t)}\widehat{H}_{\mathcal{N}}(f,\boldsymbol{\Omega}_s,t) \;\forall\; f,\boldsymbol{\Omega}_s,t, \tag{3.8}$$

where $\widehat{H}_{\mathcal{N}}(f,\boldsymbol{\Omega}_s,t)$ is the time-frequency representation of the synthesized zero-mean Gaussian white noise signal for the look direction $\boldsymbol{\Omega}_s$. An equivalent expression can also be written in the SH domain (per component) for the isotropic case.

Finally, for each frequency bin of each SH component or each look direction, the identified noise time frames [i.e. $\forall\, t > t_{\text{lim}}(f,\boldsymbol{\Omega}_s)$ or $t_{\text{lim}}^{l,m}(f)$] of the measured SRIR [either $\widetilde{H}(f,\boldsymbol{\Omega}_s,t)$ or $\widetilde{H}_{l,m}(f,t)$] are replaced with the corresponding time frames from $\widehat{H}_{\text{late}}(f,\boldsymbol{\Omega}_s,t)$ [or $\widehat{H}_{\text{late}}^{l,m}(f,t)$].

3.3.2/ Full Tail Resynthesis

Instead of "simply" denoising from t_{lim} onwards, a more ambitious treatment option is to resynthesize the entire late reverberation tail starting at the mixing time t_{mix}. Under the direction-dependent DRIR analysis/resynthesis framework, this allows the decay envelope's space-time-frequency characteristics to be manipulated by adjusting the analyzed $P_0(f,\boldsymbol{\Omega}_s)$ and $T_{60}(f,\boldsymbol{\Omega}_s)$ distributions.

Whereas in the denoising procedure the original and reconstructed decay envelopes are guaranteed to match at $t_{\text{lim}}(f,\boldsymbol{\Omega}_s)$, since by definition those points are chosen to be within the exponential decay

segment (see Sec. 2.5.1) and the modelled envelope is used unchanged for resynthesis, this is in general not the case when making changes to the decay parameters and "handing off" at t_{mix}.

Changes in the $P_0(f, \boldsymbol{\Omega}_s)$ and $T_{60}(f, \boldsymbol{\Omega}_s)$ distributions, combined with mismatches between the exponential decay start time $t_{\text{start}}(f, \boldsymbol{\Omega}_s)$ and t_{mix} (see once again Sec. 2.5.1, and also as mentioned in Sec. 3.2.1), can result in highly audible discontinuities or "jumps" in the power envelope at t_{mix}. Even without the spectral mismatches evoked in Sec. 3.2.1, changing the $P_0(f, \boldsymbol{\Omega}_s)$ and $T_{60}(f, \boldsymbol{\Omega}_s)$ values would in any case result in envelope discontinuities.

In order to ensure that the envelopes match at t_{mix} while preserving as much of the SRIR's original quality as possible, a logarithmic (i.e. linear in dB) space-time-frequency power compensation map can be applied to the early segment. That is, if the "original" measured envelope is represented at t_{mix} by $\text{EDR}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})$ and the modified decay by $\widehat{\text{EDR}}_{\text{mod}}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})$ (i.e. the EDR calculated on the resynthesized late reverberation tail), then the target power compensation at t_{mix} is:

$$P_{\text{comp}}(f, \boldsymbol{\Omega}_s, t_{\text{mix}}) = \frac{\widehat{\text{EDR}}_{\text{mod}}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})}{\text{EDR}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})}. \tag{3.9}$$

Then the full map can be obtained by linearly interpolating in dB from $\log_{10}\left[P_{\text{comp}}(f, \boldsymbol{\Omega}_s, t_0)\right] = 0$ to $10\log_{10}\left[P_{\text{comp}}(f, \boldsymbol{\Omega}_s, t_{\text{mix}})\right]$ as given by Eq. 3.9. The final modified DRIR can thus be described by:

$$\widehat{H}_{\text{mod}}(f, \boldsymbol{\Omega}_s, t) = \begin{cases} \sqrt{P_{\text{comp}}(f, \boldsymbol{\Omega}_s, t)}\widetilde{H}(f, \boldsymbol{\Omega}_s, t) & \text{for } t_0 \leq t < t_{\text{mix}}, \\ \sqrt{\widehat{P}_{\text{mod}}(f, \boldsymbol{\Omega}_s, t)}\widehat{H}_{\mathcal{N}}(f, \boldsymbol{\Omega}_s, t) & \text{for } t \geq t_{\text{mix}}, \end{cases} \quad \forall\ f, \boldsymbol{\Omega}_s. \tag{3.10}$$

Note that the "hand-off" must be performed at t_{mix} as opposed to t_{start} in order to further ensure that the SRIR has achieved sufficient spatial incoherence overall for the stochastic model used to resynthesize the late reverberation tail to be valid. Achieving exponential decay at a given frequency [which is essentially what $t_{\text{start}}(f, \boldsymbol{\Omega}_s)$ represents] does not necessarily imply that the underlying field has itself become incoherent.

The specific manner in which the $P_0(f, \boldsymbol{\Omega}_s)$ and $T_{60}(f, \boldsymbol{\Omega}_s)$ distributions are adjusted is theoretically arbitrary, and any number of strategies may thus be used to govern how they are to be manipulated. In this work, we will focus on a basic implementation that follows a similar logic to the early reflection manipulation procedures described in Sec. 3.2 and is also somewhat tied to the EDD and its own capacity for isotropy manipulation (see Sec. 3.4 below).

The basic idea, once again, is to either attenuate or accentuate deviations in the $P_0(f, \boldsymbol{\Omega}_s)$ and $T_{60}(f, \boldsymbol{\Omega}_s)$ distributions, relative to their spatial means. In a sense, the former can be thought of as "isotropifying" the late reverberation tail while the latter "directifies" it.

"Isotropifying" or Correcting Anisotropy

To isotropify, a simple tuning parameter $\upsilon_{\text{iso}} \in [0, 1]$ can be used to interpolate from no correction to total compensation, i.e. where $P_0(f, \boldsymbol{\Omega}_s)$ and $T_{60}(f, \boldsymbol{\Omega}_s)$ are equal to their spatial means. Interpolation can be done linearly in dB for $P_0(f, \boldsymbol{\Omega}_s)$ and in seconds for $T_{60}(f, \boldsymbol{\Omega}_s)$:

$$\begin{aligned} \log_{10}\left[\widehat{P}_0^{\text{iso}}(f, \boldsymbol{\Omega}_s)\right] &= \log_{10}\left[P_0(f, \boldsymbol{\Omega}_s)\right] + \upsilon_{\text{iso}}\log_{10}\left[\frac{\overline{P}_0(f)}{P_0(f, \boldsymbol{\Omega}_s)}\right] \text{ and} \\ \widehat{T}_{60}^{\text{iso}}(f, \boldsymbol{\Omega}_s) &= T_{60}(f, \boldsymbol{\Omega}_s) + \upsilon_{\text{iso}}\left[\overline{T}_{60}(f) - T_{60}(f, \boldsymbol{\Omega}_s)\right], \end{aligned} \tag{3.11}$$

where $\overline{P}_0(f)$ and $\overline{T}_{60}(f)$ are the respective spatial means.

Accentuating Anisotropy ("Directifying")

Directification, on the other hand, requires a slightly more sophisticated approach. As opposed to isotropification, which can be neatly defined between "doing nothing" ($\nu_{\text{iso}} = 0$) and "fully isotropifying" ($\nu_{\text{iso}} = 1$), there is no natural upper bound to use as a limiting reference when attempting to accentuate spatial variations in $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$. Furthermore, isotropification avoids aberrant values by constraining $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$ to the range between their original values and their spatial means; in a way, the method is thus regularized by its own upper bound. Directification, on the other hand, has no such built-in upper bound, reference, or regularization.

For this reason, we choose to define the directification procedure with respect to frequential averages of $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$ in addition to their spatial means. In an attempt to obtain representative values over both direction and frequency, the averaging is performed over third-octave bands in the range $f \in [f_{\text{aliasDI}}, f_{\text{alias}}]$, where f_{alias} is the SMA's aliasing frequency and f_{aliasDI} is the first frequency bin below $f_{\text{maxDI}} = \text{argmax}[\text{DI}(f)]$ such that $\text{DI}(f_{\text{aliasDI}}) = \text{DI}(f_{\text{alias}})$. In other words, $[f_{\text{aliasDI}}, f_{\text{alias}}]$ forms a band where $\text{DI}(f) \geq \text{DI}(f_{\text{alias}})\ \forall\ f \in [f_{\text{aliasDI}}, f_{\text{alias}}]$. The non-linear averaging achieved by first calculating the mean values over the bins within each third-octave band is then an attempt to compensate the more-or-less logarithmic human perception of frequencies.

Denoting $\{\overline{P}_0(\mathbf{\Omega}_s), \overline{T}_{60}(\mathbf{\Omega}_s)\}$ and $\{\underline{P}_0, \underline{T}_{60}\}$ the frequency and frequency-direction averages respectively, the modified $\widehat{P}_0^{\text{dir}}(f, \mathbf{\Omega}_s)$ and $\widehat{T}_{60}^{\text{dir}}(f, \mathbf{\Omega}_s)$ distributions can then be written as:

$$\begin{aligned} \log_{10}\left[\widehat{P}_0^{\text{dir}}(f, \mathbf{\Omega}_s)\right] &= \log_{10}\left[P_0(f, \mathbf{\Omega}_s)\right] + \upsilon_{\text{dir}} \log_{10}\left[\frac{\overline{P}_0(\mathbf{\Omega}_s)}{\underline{P}_0}\right] \text{ and} \\ \widehat{T}_{60}^{\text{dir}}(f, \mathbf{\Omega}_s) &= T_{60}(f, \mathbf{\Omega}_s)\left\{1 + \upsilon_{\text{dir}}\left[\frac{\overline{T}_{60}(\mathbf{\Omega}_s)}{\underline{T}_{60}} - 1\right]\right\}, \end{aligned} \tag{3.12}$$

where the multiplicative parameter $\upsilon_{\text{dir}} > 0$ controls the amount of directification to be applied.

These manipulation methods, both related and yet somewhat opposite, are evaluated in Sec. 5.4.4 through their application to real-world SRIR measurements. In both instances, their effect on the space-time-frequency structure of the late reverberation tail is also compared to that of their direct EDD-based counterparts, presented below.

Finally, as mentioned above, there are in theory any number of other manipulation strategies possible. For example, the decay parameters could be interpolated towards a target directivity pattern (e.g. cardioid, figure-8, etc.). The framework presented above also allows for frequency-envelope adjustments, potentially informed by perceptual spectral descriptors. These ideas will be further discussed in the following chapters but their implementation and evaluation is left to future work.

3.4/ EDD-Based Late Tail Isotropy Manipulation

The isotropy of the late reverberation tail can also be "directly" manipulated with respect to the EDD measure presented in Sec. 2.5.2. Instead of resynthesizing the full tail with modified decay parameters, a space-time-frequency gain map can be defined using the EDD and then directly applied to the STFT representations of the DRIR signals.

This approach is less "destructive" than full tail resynthesis as it only operates on the energy decay envelope and does not affect the underlying signal properties. It may therefore be a "safer" alternative in cases with significant overlap between the late part of the early reflection segment and the early part of the late reverberation tail, in which full tail resynthesis may cut out late-arriving echoes (examples illustrating this are provided in Sec. 5.4.4).

As for the isotropy manipulations above, a compensation envelope must once again be applied to

the early reflection segment in order to ensure a smooth transition at the mixing time. In the current approach, however, $P^{\text{EDD}}_{\text{comp}}(f, \mathbf{\Omega}_s, t)$ can be simply be defined to match the EDD-based modification envelope at t_{mix}, i.e. interpolating linearly in dB from $\log_{10}[P^{\text{EDD}}_{\text{comp}}(f, \mathbf{\Omega}_s, t_0)] = 0$ to:

$$\log_{10}\left[P^{\text{EDD}}_{\text{comp}}(f, \mathbf{\Omega}_s, t_{\text{mix}})\right] = \log_{10}\left[\widetilde{E}^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t_{\text{mix}})\right], \tag{3.13}$$

where

$$\widetilde{E}^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t) = \frac{E^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t)}{\overline{E}^{\text{EDD}}_{\text{mod}}(f, t)}, \text{ with } E^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t) = 10^{\widehat{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t)/10}. \tag{3.14}$$

In other words, $\widetilde{E}^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t)$ is the energy envelope corresponding to the dB-scale modification map $\widehat{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t)$ defined through the strategies presented below, but further spatially normalized such that the total energy distributed over the sphere at each time-frequency point remains constant [$\overline{E}^{\text{EDD}}_{\text{mod}}(f, t)$ is the spatial mean of $E^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t)$].

Similarly to Eq. 3.10, the modified DRIR can then be expressed as:

$$\widehat{H}^{\text{EDD}}_{\text{mod}} = \begin{cases} \sqrt{P^{\text{EDD}}_{\text{comp}}(f, \mathbf{\Omega}_s, t)}\widetilde{H}(f, \mathbf{\Omega}_s, t) & \text{for } t_0 \leq t < t_{\text{mix}}, \\ \sqrt{\widetilde{E}^{\text{EDD}}_{\text{mod}}(f, \mathbf{\Omega}_s, t)}\widetilde{H}(f, \mathbf{\Omega}_s, t) & \text{for } t \geq t_{\text{mix}}, \end{cases} \quad \forall\ f, \mathbf{\Omega}_s. \tag{3.15}$$

Although the EDD-based approach may indeed be "safer" than the tail resynthesis method, it is also more constrained in its range of valid manipulation possibilities. In order to preserve the exponentially decaying envelope model for the late reverberation, any modification must be applied constantly (in time) with respect to the measured EDD. A time-dependent modification strategy would inevitably require knowledge of the decay model parameters (as for the tail resynthesis method above) and is thus beyond the scope of an approach based solely on the EDD[3].

We therefore close this chapter by presenting two global EDD manipulation strategies similar to those used in the full reverberation tail resynthesis framework: the first seeks to "isotropify" or "correct" any anisotropy revealed by the EDD, while the second conversely aims to accentuate the measured EDD values and thereby "directify" the late tail.

3.4.1/ "Isotropifying" or Correcting Anisotropy

Since by definition the EDD already represents dB-scale energy deviations to a ("perceptual", dB-scale) spatial mean, it is relatively straightforward to "isotropify" the late reverberation tail by reversing the sign on its measured values. A scaling parameter $\upsilon^{\text{EDD}}_{\text{iso}} \in [0, 1]$ can also be used to control the extent to which the modification is applied:

$$\widehat{\text{EDD}}^{\text{iso}}_{\text{mod}}(f, \mathbf{\Omega}_s, t) = -\upsilon^{\text{EDD}}_{\text{iso}}\text{EDD}(f, \mathbf{\Omega}_s, t). \tag{3.16}$$

The modifications are then applied to the DRIR signals according to Eqs. 3.14 and 3.15.

[3] In theory, a purely frequency-dependent manipulation strategy could be used, such as one informed by perceptual spectral descriptors. As for the tail resynthesis method, however, such an approach is currently left to future work.

3.4.2/ Accentuating Anisotropy ("Directifying")

To accentuate any existing anisotropy as described by the EDD, i.e. either positive or negative dB values relative to the spatial mean (0 dB, see Fig. 2.8), a global dB-scale factor $\upsilon_{\text{dir}}^{\text{EDD}} > 0$ can be used in the following manner:

$$\widehat{\text{EDD}}_{\text{mod}}^{\text{dir}}(f, \boldsymbol{\Omega}_s, t) = \upsilon_{\text{dir}}^{\text{EDD}} \operatorname{sgn}\left[\text{EDD}(f, \boldsymbol{\Omega}_s, t)\right] - \frac{1}{S}\sum_{s=1}^{S} \upsilon_{\text{dir}}^{\text{EDD}} \operatorname{sgn}\left[\text{EDD}(f, \boldsymbol{\Omega}_s, t)\right], \tag{3.17}$$

where $\operatorname{sgn}(\zeta)$ is the signum function. In other words, at each space-time-frequency point, the existing energy deviation to the spatial mean is increased by a set dB amount $\upsilon_{\text{dir}}^{\text{EDD}}$.

However, since $\upsilon_{\text{dir}}^{\text{EDD}}$ is a constant and since in general the EDD will not present an equal number of positive and negative elements, the resulting modification map must be re-made spatially zero-mean by removing its own dB-scale spatial average (i.e. the second term of Eq. 3.17). Note that this differs from the energy preservation condition described in Eq. 3.14 as it only concerns dB-scale deviation values relative to the dB-scale spatial average and not the total energy gain (or attenuation) to be applied over the sphere at a specific time-frequency point. The modification map $\widehat{\text{EDD}}_{\text{mod}}^{\text{dir}}(f, \boldsymbol{\Omega}_s, t)$ is thus once again applied to the DRIR signals through Eqs. 3.14 and 3.15.

Both the "isotropification" and "directification" procedures are evaluated on real-world SRIR measurements in Sec. 5.4.4, and again in both instances a qualitative comparison to the full tail resynthesis approach described in Sec. 3.3.2 is included.

This chapter has presented the SRIR treatment and manipulation procedures made possible by the space-time-frequency model(s) and analysis techniques detailed in Ch. 2. In order to further differentiate the current research project from existing reverberation processing techniques, these have been particularly focused on manipulating various spatial aspects of SRIRs, such as the distribution of their early reflections (Sec. 3.2) or the isotropy of the late reverberation tail (Sec. 3.3).

Treatment procedures for mitigating the influence of non-stationary measurement artefacts (Sec. 3.1) and compensating for the presence of non-decaying background noise (Sec. 3.3.1) have also been presented in this chapter. This particular work had previously been published in Massé et al. [84] and Massé et al. [78], albeit in slightly different forms; it has been updated and restructured in this chapter to better reflect its latest iteration as well as its place within the general context of this thesis.

This chapter concludes the purely theoretical aspects of this dissertation. The following chapter presents the results from a series of validation tests designed to evaluate the different analysis, treatment, and manipulation methods detailed in Chs. 2 and 3. Finally, Ch. 5 closes this thesis with a wide-ranging discussion based on the application of these same methods to an extensive collection of SRIR measurements performed with an mh acoustics Eigenmike® SMA in a highly varied selection of acoustic spaces and locations.

---

4 / Validation Tests

The aim of this chapter is to verify the main SRIR analysis and treatment methods presented throughout Chs. 2 and 3. The studies contained herein are intended to evaluate the performance, discuss the limitations, and parameterize the implementation of these techniques that have so far been described purely theoretically. They are organized here in an order that attempts to introduce the various tools (e.g. specific types of simulations) and draw the necessary conclusions (e.g. determine which spatial representation works best in a certain case) in logical and natural succession. As such, the exact sectioning differs slightly from that of the previous two chapters.

The chapter opens (Sec. 4.1) with a thorough investigation of the directional SRIR representation (or simply *directional room impulse response*, DRIR) that underpins most of the analysis and treatment methods that are to be subsequently tested. It is therefore imperative to first understand the optimal design choices available as well as their respective advantages and inherent limitations.

These can then be adapted to the different analysis techniques: early reflection detection (Sec. 4.2.1, including a subsection on the direct sound), mixing time estimation (Sec. 4.2.2), multi-slope late reverberation decay analysis (Sec. 4.2.3) and late tail isotropy evaluation (Sec. 4.2.4). Finally, anisotropic late reverberation denoising (Sec. 4.3.1) is also studied at the end of the chapter.

Various SRIR simulations have been designed for these tests; they are each described in detail when first needed and then referred to as necessary. Most (but not all) of these synthesized examples simulate the 32-capsule mh acoustics Eigenmike® SMA measurement of a spatial RIR sound field. This is simply to ensure continuity with the real-world measurements presented in Ch. 5 (the following and final chapter), all of which have been measured with the Eigenmike.

4.1 / Directional SRIR Representation

We thus begin this chapter by evaluating the complete directional representation of an SRIR as described in Sec. 2.2: in other words, one that preserves as many of the measurement's space-time-frequency properties as possible. As detailed throughout Chs. 2 and 3, the resulting DRIR plays a fundamental role in many of the analysis, treatment, and manipulation methods developed throughout this research project. It is therefore critical that a robust definition of the DRIR be carefully described and thoroughly evaluated against its various desirable properties.

To recall and summarize the main considerations presented in Sec. 2.2, these are:

1. *Maximum directivity*: We would like each directional signal to represent, as much as possible, only its look direction and the region immediately surrounding it. The most straightforward choice here would be to use the natural PWD beamformer; however, due to its relatively strong rear lobe, other options may be more appropriate. Thus the maximum front-to-back ratio (MFBR) [45, p. 110], Dolph-Chebyshev (D.-C.) [47], and proposed maximum weighted directivity index (MWDI; Appx. A.2) beamformers will also be evaluated.

2. *Minimum overlap*: Whatever the choice of beamformer, the resulting directivities, and especially their main lobes, should overlap as little as possible. This is measured through the unique

coverage factor (UC) defined in Eq. 2.19, as well as the overall average directivity energy ratio $\overline{W}_{\mathrm{ER}}$ (see Eq. 2.18 and accompanying paragraph).

3. *Flat coverage*: The total combination of all look direction directivities over the sphere should not privilege any particular directions over any other. This is calculated through the dB-scale standard deviation of the total coverage directivity function (Eq. 2.21).

4. *Linear independence*: The decomposition should preserve the linear independence of the SH basis. As a result, the decomposition matrix must be full rank and the look direction layout grid must be of size $S = (L+1)^2$.

5. *Spatial (in)coherence preservation*: The above considerations should combine to help preserve the spatial (or directional) coherence properties of the measured signal throughout the length of the SRIR. Most notably, an incoherent late reverberation field should be properly represented as such in order to allow for mixing time estimation as in Sec. 2.3 and late reverberation tail reconstruction as in Sec. 3.3. This is evaluated using the spatial incoherence measure presented in Sec. 2.3 (inspired by the CoMEDiE diffuseness measure [62]).

The remainder of this section will thus be spent examining a range of possible DRIR definitions with respect to these properties, in search of a spatial decomposition that will satisfy as many of them as much as possible. This examination begins with an overview of different spherical grid point geometries available for defining the beamforming look direction layouts (Sec. 4.1.1). Four different beamformer designs are then compared, first in terms of their intrinsic directivity properties (Sec. 4.1.2) and then in the context of their application to the various look direction layout grids (Sec. 4.1.3). Finally, the possible combinations of look direction layout and beamformer design (i.e. what will be referred to as a particular "spatial decomposition") are evaluated with respect to their effect on the overall spatial coherence properties of the measured sound field (Sec. 4.1.4). These steps then provide the insights necessary to conclude on the specifics of DRIR generation (Sec. 4.1.5).

4.1.1/ Look Direction Layout Grid Geometries

To begin our evaluation, we first present the five spherical point grids to be used as look direction layouts. These have been chosen for answering the design objective of providing "well distributed" points over the sphere, albeit under different constraint or optimization formulations. The aforementioned Sloan-Womersley [95] and Fliege-Maier [79] grids are variations on the question of improving numerical integration on the sphere (such as in, for example, the discretized SH transform of Eq. 1.3); in particular, the Sloan-Womersley approach is based on interpolating spherical polynomials.

The three additional grids are taken from the work of Sloane and Conway [96] (among others) on spherical packing, covering, and lattice problems. Each design gives a solution to a slightly different question: for example, the spherical covering grid minimizes the maximum distance or *covering radius* between each point and its nearest neighbour, while conversely the spherical packing grid maximizes the minimum distance between nearest neighbours. In a slightly different approach, the spherical maximum volume grid maximizes the volume of the convex hull defined by the S points.

These designs have also been chosen due to the fact that they give valid solutions for $S = (L+1)^2$ points for all "reasonable" integer values of L that may be encountered both in simulation and real-world measurement. Indeed, this is not necessarily the case for all t-design strategies.

Since one of the objectives for this section is to minimize the overlap between the main lobes of the look direction directivities, a critical feature of the layout grids is the angular separation between neighbouring points. Taking the natural PWD beamformer as an example, if two look directions are less than the peak-to-null half-width Θ_{PN} apart, then their main lobes will significantly overlap as

illustrated in Fig. 2.2, and the independence between the resulting beams will be affected. On the other hand, the closer the angular separation is to the equal energy point Θ_{eqE}, the flatter the combined power response will be (see Fig. 2.2 and accompanying text once again).

Layout	**Mean Separation** [deg]	**Minimum Separation** [deg]
Sloan-Womersley	40.5°	39.5°
Fliege-Maier	40.9°	39.6°
Spherical Covering	38.7°	36.8°
Spherical Maximum Volume	40.1°	38.6°
Spherical Packing	41.6°	41.6°

Table 4.1: Mean and minimum angular separations between nearest neighbours for five different spherical grid designs of $S = 25$ points each (i.e. corresponding to SH order $L = 4$).

When extended to a full set of S look directions over the sphere, and considering the various possible beamformer designs, this trade-off quickly becomes non-trivial. Before proceeding to the thorough evaluation of the 20 different combinations afforded by the five chosen spherical point grids and four beamformer designs, it will first be useful to briefly analyze the angular separations between nearest neighbours in each grid, if only as an initial reference point. To this end, the mean and minimum nearest neighbour separations are displayed in Tab. 4.1 for each grid design's $S = 25$-point solution, which corresponds to the number of look directions for a maximum SH order of $L = 4$.

Note that these values are all contained between $\Theta_{\text{PN}} = 45°$ and $\Theta_{\text{eqE}} = 35°$ for an $L = 4$ natural PWD beamformer. Considering the natural PWD beamformer presents the thinnest main lobe width of the four chosen beamformers, we can therefore expect main lobe overlap and beam independence to play a significant role in the determination of an appropriate spatial decomposition (though one could indeed parameterize the Dolph-Chebyshev design to obtain a thinner main lobe, this will not be the case here due to the inherent trade-off with the side lobe level, as discussed below).

4.1.2/ Beamformer Directivities

Since each possible spatial decomposition is the combination of a look direction layout grid and a beamformer design, it is important to complement the discussion above with a brief overview of the main properties of each (broadband) directivity function. Table 4.2 thus lists the main lobe peak-to-null half-width Θ_{PN}, equal energy point Θ_{eqE}, "plain" and weighted directivity indices (DI and WDI, respectively), and front-to-back ratio (FBR) for each of the four beamformer designs in consideration.

As mentioned above, the natural PWD beamformer presents the thinnest main lobe width, though this is partly due to the Dolph-Chebyshev design being parameterized to match the main lobe width of the maximum WDI solution. This choice is motivated by the fact that, in principle, the MWDI offers an optimized trade-off between DI and side lobe suppression (with rear lobes further suppressed than frontal ones); the comparison with a Dolph-Chebyshev design of equal main lobe width, which finds the lowest possible constant side lobe level in response, therefore seems particularly pertinent.

Indeed, this is confirmed by the DI, WDI, and FBR measures for both beamformers, which are nearly identical. While the Dolph-Chebyshev's "plain" DI is slightly higher, both the WDI and FBR are favourable to the maximum WDI design. This is rather to be expected for the WDI, whose maximization is the entire goal of the MWDI beamformer. In fact, it is almost surprising to see that the Dolph-Chebyshev design can almost achieve maximum WDI (although we did need to solve the MWDI problem in order to set the Dolph-Chebyshev's target main lobe width).

Beamformer	$\boldsymbol{\Theta}_{\mathrm{PN}}$ [deg]	$\boldsymbol{\Theta}_{\mathrm{eqE}}$ [deg]	**DI** [dB]	**WDI** [dB]	**FBR** [dB]
Natural PWD	43.9°	35.0°	9.42	18.9	27.6
Dolph-Chebyshev	53.7° (set)	38.5°	9.11	21.8	42.6
Maximum FBR	94.4°	49.6°	7.84	18.9	107
Maximum WDI	53.7°	39.1°	9.05	21.9	47.4

Table 4.2: Principal broadband directivity function characteristics for each of the four chosen beamformer designs, with SH order $L = 4$. The equal energy point Θ_{eqE} is described in Sec. 2.2 and the weighted DI (WDI) is defined in Appx. A.2. Note that the Dolph-Chebyshev design used here is obtained by setting the main lobe width equal to that of the maximum WDI beamformer.

In a sense, the maximum WDI design can be thought of as a compromise between the natural PWD and maximum FBR beamformers. Indeed, the weighting function used to define the WDI has the effect of promoting response energy towards the look direction, and can be seen as a spatially "spread out" or "smoothed" version of the FBR's denominator. This view is reinforced by the fact that the MWDI solution's properties all lie between the MFBR and the natural PWD designs.

Finally, these results also confirm that the natural PWD beamformer represents the maximum DI for a given SH order, while the maximum FBR implementation also delivers on its design objective.

4.1.3/ Directivity Overlap and Spherical Coverage

To evaluate the coverage characteristics of the 20 possible spatial decompositions (i.e. combinations of beamformer design and look direction layout grid), we rely on the three measures defined in Sec. 2.2: the dB-scale standard deviation in the total combined power response $W(\boldsymbol{\Omega})$ (σ_W, Eq. 2.21), the unique coverage factor (UC, Eq. 2.19), and the average directivity energy ratio ($\overline{W}_{\mathrm{ER}}$, Eq. 2.18).

In particular, we are looking for a spatial decomposition that minimizes the combined response's standard deviation σ_W while simultaneously maximizing the UC and $\overline{W}_{\mathrm{ER}}$. A complete tabulation of the results obtained for each measure on each possible beamformer-grid combination is given in Tab. 4.3, with the values representing the best performance for each measure emphasized in bold.

	$\boldsymbol{\sigma_W}$ [dB]				**UC** [%]				$\boldsymbol{\overline{W}_{\mathrm{ER}}}$ [dB]			
Layout	Nat.	D.-C.	MFBR	MWDI	Nat.	D.-C.	MFBR	MWDI	Nat.	D.-C.	MFBR	MWDI
Sloan-Womersley	0.477	0.314	0.113	0.299	68.6	71.6	2.68	71.1	**−9.08**	−10.0	−12.5	−10.2
Fliege-Maier	0.400	0.197	**0.057**	0.177	68.9	71.7	2.29	71.3	−9.17	−10.0	−12.5	−10.1
Sph. Covering	0.456	0.251	0.098	0.232	69.4	71.9	2.81	71.4	−9.35	−10.0	−12.5	−10.2
Sph. Max. Vol.	0.437	0.276	0.137	0.260	68.8	71.6	3.76	70.9	−9.17	−9.98	−12.5	−10.1
Sph. Packing	0.552	0.355	0.177	0.338	69.2	**72.0**	3.66	71.5	−9.43	−10.1	−12.5	−10.2

Table 4.3: Coverage evaluation for the five chosen look direction layout grids and the four selected beamformer designs: natural PWD (Nat.), Dolph-Chebyshev (D.-C.), maximum front-to-back ratio (MFBR), and maximum weighted DI (MWDI), with SH order $L = 4$ and thus grid size $S = 25$. The best performing values for each of the three measures (combined response standard deviation [σ_W], unique coverage factor [UC], and average directivity energy ratio [$\overline{W}_{\mathrm{ER}}$]) are marked in bold.

These results highlight certain patterns that confirm previously discussed characteristics of both the spherical point grids and the beamformer designs. For instance, we have seen in Tab. 4.2 that the natural PWD beamformer presents both the thinnest main lobe Θ_{PN} and the smallest equal energy

point Θ_{eqE} (the former inevitably influencing the latter). It is therefore not surprising to see the natural PWD beamformer consistently yield the highest standard deviations in coverage (σ_W), regardless of the specific look direction layout grid.

Conversely, the maximum FBR beamformer has the widest main lobe (and as a consequence the largest Θ_{eqE}), which results in an extremely flat combined response (low σ_W) on all layout grids. However, its weak DI (and WDI) gives it consistently the worst UC and $\overline{W}_{\text{ER}}$.

Table 4.2 also shows how the maximum WDI beamformer can be seen as a compromise between the maximum FBR and natural PWD designs. This is reflected in Tab. 4.3 through its σ_W and $\overline{W}_{\text{ER}}$ results, which all lie between those for the MFBR and natural PWD beamformers. However, the MWDI solution outperforms the natural PWD beamformer in terms of UC, which provides a first hint that other, specific characteristics of the directivity function can influence overlap and coverage.

Taking a closer look at the UC results, we notice that the Dolph-Chebyshev design is the highest performing of all, but also that it is directly comparable to the MWDI beamformer across the board. It furthermore appears to perform slightly better with respect to the UC factor as well as $\overline{W}_{\text{ER}}$, but the MWDI design provides flatter combined responses overall. Interestingly enough, both the Dolph-Chebyshev and MWDI beamformers outperform the natural PWD in terms of UC, which suggests that the side lobe level control offered by these designs is a particular advantage when considering spherical coverage; it also demonstrates the utility of the UC factor as an evaluation measure.

Finally, in order to obtain an overall picture of the 20 different spatial decompositions' coverage performances, Tab. 4.4 proposes a "combined coverage property score" for each grid-beamformer combination. These scores are obtained by scaling the values reported in Tab. 4.3 from 0 to 10 for each one of the σ_W, UC, and $\overline{W}_{\text{ER}}$ measures, where 0 corresponds to the worst performance (i.e. the highest value for σ_W and the lowest values for UC and $\overline{W}_{\text{ER}}$) and conversely 10 corresponds to the best performance (lowest value for σ_W, highest values for UC and $\overline{W}_{\text{ER}}$). The mean of the three scores thus obtained for each possible spatial decomposition then give the results displayed in Tab. 4.4.

Layout	**Natural**	**Dolph-Chebyshev**	**Max. FBR**	**Max. WDI**
Sloan-Womersley	7.01	7.35	2.97	7.24
Fliege-Maier	7.45	8.15	3.33	**8.16**
Spherical Covering	6.93	7.79	3.08	7.70
Spherical Maximum Volume	7.20	7.63	2.86	7.59
Spherical Packing	6.19	7.00	2.59	6.99

Table 4.4: Combined coverage property scores for each of the 20 possible combinations of beamformer design and look direction layout grid, with SH order $L = 4$ and thus grid size $S = 25$. Scores are obtained by scaling the results of each measure from 0 (worst performance, i.e. the maximum for σ_W and minimum for UC and $\overline{W}_{\text{ER}}$) to 10 (best performance: minimum for σ_W, maximum for UC and $\overline{W}_{\text{ER}}$), and then taking the mean over the three properties for each combination.

From these, it is confirmed that the Dolph-Chebyshev and MWDI designs offer the best performances over all five look direction layout grids. It is also clear that the Fliege-Maier configuration gives the best results for all four beamformers. Overall, the highest-scoring spatial decomposition is the MWDI beamformer applied to a Fliege-Maier grid, marginally beating out the Dolph-Chebyshev design on the same look direction layout. In terms of spherical coverage and directivity overlap, at least according to the metrics defined in Sec. 2.2, these are therefore the two implementations that should be retained for further evaluation.

4.1.4/ Spatial Incoherence Preservation

To complete our assessment of the different spatial decomposition possibilities, we now focus our attention on their capacity to preserve the spatial incoherence of simulated archetypal late reverberation plane wave fields. Late field incoherence preservation is used here as an evaluation tool under the assumption that it is representative of the SRIR's more general spatial coherence properties; that is, if the incoherence of the reverberation tail is properly preserved, so too will the high coherence of the early reflections and any variations between the two. In other words, the incoherence profiles used for mixing time estimation and echo detection (see Secs. 2.3 and 2.4) will be as accurate as possible.

In order to generate such profiles, the incoherence measure must therefore correctly increase with the number of uncorrelated incident plane waves whilst remaining unaffected by any anisotropy in their directional power distribution. Simulated SMA measurements of an increasing number of uncorrelated plane waves subjected to an increasingly anisotropic cardioid power distribution will thus be used to examine these characteristics for each one of the spatial decompositions detailed in the previous section. The CoMEDiE measure [62], calculated directly in the SH domain, is used as a reference.

The SMA simulation itself is based on the 32-capsule mh acoustics Eigenmike®, which enables up to 4$^{\text{th}}$-order SH/HOA encoding. The SMA measurement is simulated by extending Eq. A.3 to an arbitrary number N_{PW} of incident plane waves placed on a synthesis DoA layout grid of size D_{synth}:

$$X_{\text{PW}}(f, \mathbf{\Omega}_q, t) = \sum_{l=0}^{L_{\text{synth}}} b_l(f) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}_q) \sum_{j=1}^{N_{\text{PW}}} \sum_{d=1}^{D_{\text{synth}}} \alpha_d \delta_{d,p_j} \sqrt{P(\mathbf{\Omega}_{p_j})} Y^*_{l,m}(\mathbf{\Omega}_d) e^{i\hat{\phi}(f,\mathbf{\Omega}_{p_j},t)}, \tag{4.1}$$

where $\mathbf{\Omega}_q$ are the SMA transducer positions and $\hat{\phi}(f, \mathbf{\Omega}_s, t) \sim \mathcal{U}[0, 2\pi)$ provides a random phase. The Kronecker delta δ_{d,p_j} "selects" or "turns on" the incident PW signal from a given DoA for a set of N_{PW} randomly chosen unique indices $p \sim \mathcal{G}^{N_{\text{PW}}}_{D_{\text{synth}}}$, where $\mathcal{G}^{N_{\text{PW}}}_{D_{\text{synth}}}$ denotes the subset of all size N_{PW} sequences without repetition within the complete set $\{1, 2, \ldots, D_{\text{synth}}\}$ (sometimes referred to as "partial permutations"), and p_j denotes the j^{th} element of p. We note that this implies $N_{\text{PW}} \leq D_{\text{synth}}$.

Equation 4.1 can perhaps be more conveniently expressed in matrix form:

$$\mathbf{x}_{\text{PW}}(f, t) = \mathbf{Q}(f)\mathbf{D}\mathbf{M}\mathbf{p}_{\text{PW}}(f, t), \tag{4.2}$$

where $\mathbf{Q}(f)$ is a $Q \times (L_{\text{synth}} + 1)^2$ matrix with elements $b_l(f)Y_{l,m}(\mathbf{\Omega}_q)$, with Q being the number of SMA transducers, $\mathbf{D}$ is a $(L_{\text{synth}} + 1)^2 \times D_{\text{synth}}$ matrix with elements $\alpha_d Y^*_{l,m}(\mathbf{\Omega}_d)$, $\mathbf{M}$ is a sparse $D_{\text{synth}} \times N_{\text{PW}}$ mapping matrix with elements δ_{d,p_j}, and finally $\mathbf{p}_{\text{PW}}(f, t)$ is an $N_{\text{PW}} \times 1$ column vector containing the directional signals $\sqrt{P(\mathbf{\Omega}_{p_j})} e^{i\hat{\phi}(f,\mathbf{\Omega}_{p_j},t)}$.

In practice, just as for the late reverberation tail resynthesis in Sec. 3.3.2, this signal model is implemented using a zero-mean Gaussian noise sequence, from which a time-frequency representation is obtained through a discrete short-term Fourier transform (STFT). For reference, the STFT applied here makes use of a 1024-sample length Nuttall [97] sliding window with a hop size of 128 samples (i.e. 87.5% overlap) and no additional zero-padding in the fast Fourier transform (FFT). At a sampling frequency of $f_s = 48$ kHz, this corresponds to a window length of 21.3 ms with a 2.67 ms step and a lowest detectable frequency of 46.9 Hz.

Note that, due to the spatial discretization and the independence of the stochastic phase with respect to frequency, N_{PW} technically represents a plane wave density rather than an absolute number. However, this definition guarantees that ideal diffuse field conditions are recovered as expected when $N_{\text{PW}} \to D_{\text{synth}}$ and $P(\mathbf{\Omega}_{p_j}) \to 1 \; \forall \; \mathbf{\Omega}_{p_j}$ (assuming a large enough D_{synth}).

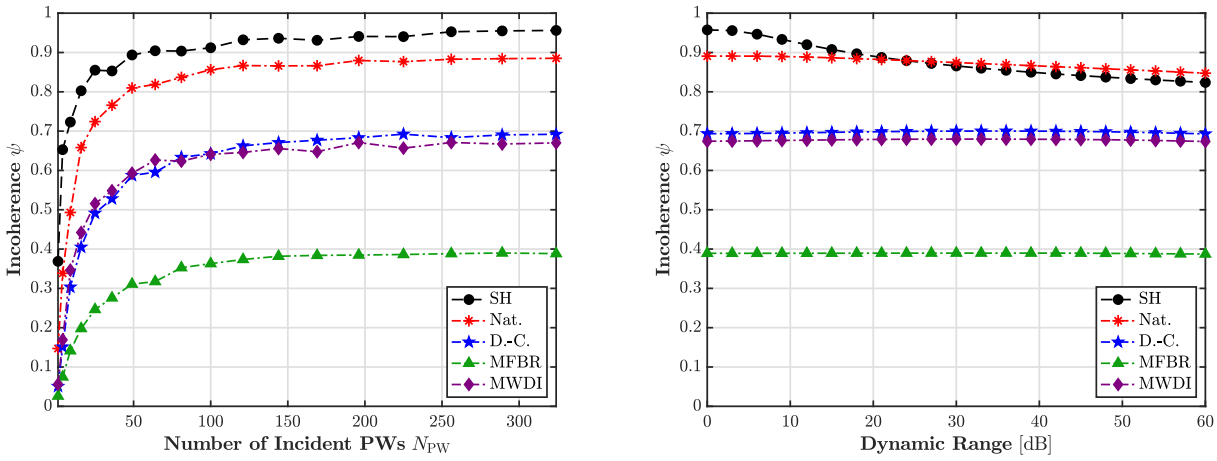
To implement this simulation, we quantize the incident DoAs to a large layout grid with well-determined quadrature weights α_d. Of the five point grids presented above, only the Fliege-Maier

and Sloan-Womersley solutions are thus designed for numerical integration over the sphere. The Sloan-Womersley grid is chosen here since it offers a smaller average angular separation between points (see Sec. 4.1.1 above, and which is furthermore true at any given SH order). This implies that the Sloan-Womersley synthesis DoA layout will create denser plane wave fields as $N_{\mathrm{PW}} \to D_{\mathrm{synth}}$.

The size of the synthesis grid is then chosen to be large enough to ensure that the SMA simulation does not generate any "artificial" spatial aliasing at any frequency up to the Shannon-Nyquist frequency: in other words, only the "true" spatial discretization due to the SMA should contribute to any aliasing. This can be calculated by taking Eq. 1.15 and forcing $f_{\mathrm{alias}} = f_s/2$:

$$L_{\mathrm{synth}} \geq \left\lceil \frac{\pi f_s r_s}{c} \right\rceil, \tag{4.3}$$

and thus the synthesis grid must be of size $D_{\mathrm{synth}} \geq (L_{\mathrm{synth}} + 1)^2$. For the 32-capsule Eigenmike ($Q = 32$, $r_s = 4.2$ cm) and using a sampling frequency of $f_s = 48$ kHz, this gives $L_{\mathrm{synth}} \geq 19$ and therefore $D_{\mathrm{synth}} \geq 400$. To keep computing costs as low as possible, the minimum $L_{\mathrm{synth}} = 19$ is finally chosen (implying $D_{\mathrm{synth}} = 400$).



(a) Spatial incoherence as a function of the number of incident plane waves.

(b) Spatial incoherence as a function of the dynamic range of a cardioid power distribution.

Figure 4.1: Covariance matrix eigen-decomposition spatial incoherence measure ψ_{dir} as a function of increasing number of incident plane waves (a, left) and increasing cardioid power distribution dynamic range (b, right), calculated in the SH domain (black points) and for four beamformed directional representations: natural PWD (red asterisks), Dolph-Chebyshev (blue stars), MFBR (green triangles), and MWDI (purple diamonds). The cardioid power distributions are applied to a maximally dense field of $N_{\mathrm{PW}} = D_{\mathrm{synth}}$ simulated plane waves. Synthesis is performed according to Eq. 4.2.

Figure 4.1 subsequently shows the evaluation of the incoherence measure ψ_{dir}, as discussed in Sec. 2.3, on short (120 ms) signals synthesized as above, first with an increasing number of incident plane waves N_{PW} under an isotropic power distribution $P(\mathbf{\Omega}_{p_j}) = 1 \; \forall \; \mathbf{\Omega}_{p_j}$ (Fig. 4.1a, left), and then with an increasingly anisotropic power distribution applied to a maximally dense $N_{\mathrm{PW}} = D_{\mathrm{synth}}$ plane wave field (Fig. 4.1b, right). As an arbitrary choice, the anisotropic power distribution is given the form of a "front-back" cardioid directivity pattern [i.e. centred on $\mathbf{\Omega}_0 = (0°, 0°)$]. The Fliege-Maier look direction layout grid used to generate the directional representations is then rotated such that one of its points faces this particular DoA.

A reference incoherence is first calculated in the SH domain on the HOA encoding of the simulated SMA signals from Eq. 4.2, which is thus equivalent to the CoMEDiE diffuseness measure (black points, Fig. 4.1). Then the spatial incoherence measure is evaluated on the directional representations obtained through each one of the four beamformers described above, applied to a Fliege-Maier look direction

layout of size $S = 25$. The covariance matrices are directly calculated in the time domain.

These results offer several important insights. First, it is clear that the CoMEDiE diffuseness calculated in the SH domain performs the best out of all five measures when analyzing a perfectly diffuse field, i.e. a maximally dense field of uncorrelated plane waves under an isotropic power distribution, as in the last point of Fig. 4.1a and the first point of Fig. 4.1b. This follows directly from the fact that the SH-domain analysis is unaffected by the directivity overlap present in the directional representations, and is really only limited by the SMA's orthogonality error (see Sec. 1.2.1 and Rafaely [35]).

However, as hinted at in the discussion on covariance matrices in Sec. 2.3, SH-domain diffuseness is far more sensitive to changes in incoherent field isotropy, as evidenced by Fig. 4.1b. Furthermore, it also seems to slightly over-estimate incoherence for low densities of incident plane waves (Fig. 4.1a), as it does not drop below a value of $\psi_{\text{cov}} \simeq 0.35$ for $N_{\text{PW}} = 1$. The combination of these two characteristics can lead to issues when generating incoherence profiles, for example, since the presence of coherent early reflections and the appearance of late tail anisotropy could become indistinguishable.

Of the directional decompositions, the natural PWD beamformer presents the highest incoherence preservation overall: this follows from the fact that it offers the best possible DI and therefore the greatest independence between directional beams (as evidenced e.g. by its $\overline{W}_{\text{ER}}$ scores in Tab. 4.3). However, it is also the most affected (besides the reference SH-domain diffuseness) by increasing anisotropy, which is most likely a consequence of its spherical coverage and directivity overlap limitations. As seen in Sec. 4.3.1 below, a major part of these limitations stems from the significant rear lobe of its directivity function[1] (this is also discussed in Sec. 2.2, and is indeed the main motivation behind testing the three other beamformer designs).

As the dynamic range of the cardioid power distribution increases, the rear lobe of a look direction facing away from the maximum power (i.e. toward the minimum) begins to dominate the entire beam. But the signal content carried by that rear lobe is nearly identical to the signal represented by the main lobe of the beam facing toward the maximum power. Thus the total signal content in these two directional beams (facing toward and facing away from the maximum power) becomes increasingly similar and therefore increasingly coherent.

On the other hand, the Dolph-Chebyshev and MWDI beamformers offer constant incoherence values regardless of the field's isotropy. This can be traced to their higher FBRs (see Tab. 4.2) and better UC factors (see Tab. 4.3) compared to the natural PWD beamformer. Unfortunately, due to their wider main lobes and lower DIs, their absolute incoherence preservation is significantly weaker. Overall, the Dolph-Chebyshev beamformer appears to perform slightly better than the MWDI design.

In terms of tracking an increasing number of incident uncorrelated plane waves (Fig. 4.1a), we can note that, relative to their respective maximum incoherence values, each directional representation presents a similar profile. In general, these appear slightly more progressive than the SH-domain reference, as they all drop to $\psi_{\text{dir}} < 0.15$ for $N_{\text{PW}} = 1$ while reaching their maxima around $N_{\text{PW}} \approx 200$.

Finally, and as could be expected following the results of Sec. 4.1.3, we should note that the MFBR beamformer is clearly the worst performing design across the board (at least in the current context of directional SRIR decompositions as defined in this thesis).

4.1.5/ Complete DRIR Generation

Considering the results above, we are now faced with the dilemma that no single spatial decomposition comprehensively outperforms the rest. Indeed, there appears to be an unavoidable trade-off to be made between optimal spherical coverage, absolute incoherence preservation, and robustness to anisotropy. The first is crucial in obtaining accurate directional energies or power envelope estimates, which in turn

[1] This is also evidenced by the fact that the natural PWD beamformer presents the worst FBR of the four chosen beamformer designs, as given in Tab. 4.2.

is massively important when analyzing the late reverberation tail's energy decay envelope. But the second is also important in resynthesizing the late field: since the synthesized signals are perfectly incoherent (as they are generated independently for each beam, see Sec. 3.3), the measurement signals to which they are matched should thus be as incoherent as possible to avoid any audible artefacts in the reconstructed SRIR. Finally, the third is necessary for generating accurate incoherence profiles for mixing time estimation and early reflection detection, as noted in the previous section.

The compromise, therefore, is to choose the spatial decomposition that is best adapted to a given analysis, treatment, or manipulation application. For example, energy estimation in the early reflection detection procedure should use DRIRs generated with the MWDI beamformer, since we have seen it offers the best overall spherical coverage. Incoherence profiles could use the Dolph-Chebyshev design, as it is the most robust against changes in incoherent field isotropy while maintaining a relatively high maximum incoherence (though the difference with the MWDI beamformer's performance is minimal and the natural PWD implementation's sensitivity to anisotropy is most likely negligible in the context of an incoherence profile, the Dolph-Chebyshev design is still technically the best choice here).

Some applications, however, require all three characteristics to be considered simultaneously. For late reverberation tail resynthesis, for instance, both accurate directional power estimates and maximum incoherence are important, as noted above. Since the resynthesized prolongation of the tail is matched to the given DRIR, whose analysis itself provides the decay parameters for resynthesis (see Sec. 3.3 again), these two properties cannot be uncoupled. In other words, the DRIR whose late tail is to be replaced or prolonged with a resynthesized version must also be the one that was analyzed.

Furthermore, the DRIR's late field incoherence should be maximized regardless of the late reverberation tail's isotropy (or lack thereof). Of course, following Fig. 4.1b, we know this cannot perfectly be the case. For these reasons, further evaluation of the natural PWD, Dolph-Chebyshev, and MWDI beamformers will be necessary throughout Sec. 4.3.1 in order to make a final choice of DRIR representation for the particular case of late reverberation tail resynthesis.

4.2/ Analysis Methods

In this section, the analysis methods described throughout Ch. 2 for the estimation of various fundamental model parameters are each examined in turn. The specific analysis procedures are thoroughly evaluated with respect to their performance as applied to simulated SMA measurements of synthesized archetypal sound fields.

4.2.1/ Early Reflection Detection

Fundamental DoA Estimation

The performance of the two chosen DoA estimation approaches, namely the steered response power (SRP) and Herglotz inversion methods described in Sec. 2.4, is first evaluated with respect to their accuracy in localizing single incident impulses over the sphere. This represents the most fundamental error measurement available in the context of echo detection, since the ability to detect individual impulses inherently conditions the performance of all subsequent applications.

The test signals used herein are therefore ideal impulses (i.e. discrete time-domain Dirac distributions). In order to avoid any time-frequency aliasing when performing the SMA simulation, the synthesized impulses are first low-passed with a cutoff frequency of 16 kHz. Furthermore, since the geometrical acoustics view is a "high frequency" description by definition (see the brief discussion on the subject in Sec. 2.1.2), the signals are high-passed with a 125 Hz cutoff. The signals are synthesized with a 1024-sample temporal support, corresponding to 21.3 ms at a sampling rate of $f_s = 48$ kHz.

To obtain an overall characterization of the estimation error for the localization of single impulses, a large synthesis spherical point grid, of size D_{synth} (defined with respect to L_{synth} as in Sec. 4.1.4 above, Eq. 4.3), is "scanned" by successively placing a single test impulse at each DoA in turn. Similarly to Eqs. 4.1 and 4.2, the SMA measurement for an impulse placed at $\mathbf{\Omega}_j$ is simulated as:

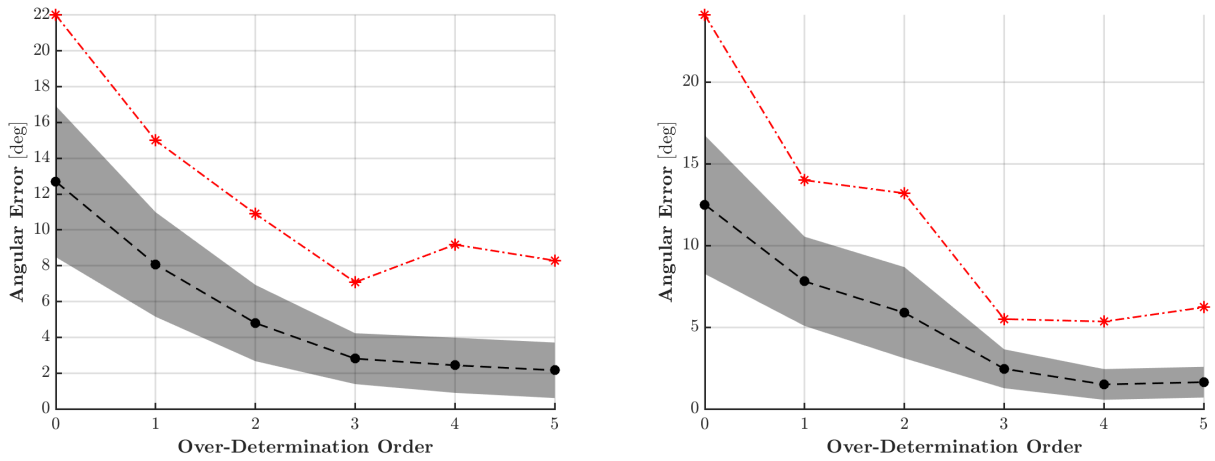
$$\mathbf{x}_{\text{imp},j}(f,t) = \mathbf{Q}(f)\mathbf{D}\mathbf{m}_j p_{\text{imp}}(f,t), \tag{4.4}$$

where $\mathbf{Q}(f)$ and $\mathbf{D}$ are the same as in Eq. 4.2, $\mathbf{m}$ is a $D_{\text{synth}} \times 1$ mapping vector (i.e. a reduction of the mapping matrix $\mathbf{M}$) with elements $\delta_{d,j}$, and $p_{\text{imp}}(f,t)$ is the synthesized archetypal impulse signal as described above. This simulation is then performed in turn for every $\mathbf{\Omega}_j$ with $j = \{1, 2, \ldots, D_{\text{synth}}\}$.

For each simulated DoA, the two echo localization methods (SRP and regularized Herglotz inversion) are applied. The full 1024-sample signal length is treated as a single frame on which an FFT is performed, and the localization maps are then generated from the three most directive frequency bins (i.e. the three bins closest to the Eigenmike's 4^{th}-order natural PWD maximum directivity frequency $f_{\text{maxDI}} \approx 3.45$ kHz, as shown in Fig. 1.5b). The mean localization error, as well as its standard deviation and maximum, can be subsequently evaluated over the sphere to form a comprehensive assessment of each method's performance.

With respect to the regularization of the Herglotz inversion, two different strategies can be considered: either a straightforward adaptive approach in which the regularization parameter χ is estimated by generalized cross-validation (GCV) for each analyzed signal individually, or a "global" optimization in which a single χ value is imposed for all localizations and chosen so as to simultaneously minimize the mean error and its standard deviation over the sphere.

Additionally, and as mentioned in Sec. 2.4, an appropriate layout grid for the D_{Herg} Herglotz inversion sampling points on the sphere must also be determined. Figure 4.2 thus shows the mean error as well as its standard deviation and maximum over the sphere for the optimized Herglotz method as a function of increasing spatial sampling grid order (resulting in an increasingly under-determined matrix inversion). Both Sloan-Womersley (Fig. 4.2a, left) and Fliege-Maier layouts (Fig. 4.2b, right) are evaluated, with a constant final 99^{th}-order (10^4-point) Sloan-Womersley interpolation grid used throughout. Once again, the simulated SMA is the 32-capsule, 4.2 cm radius Eigenmike.



(a) Sloan-Womersley grid. (b) Fliege-Maier grid.

Figure 4.2: Mean angular errors (black points and dashed line) over the sphere for single impulse localization using the regularized under-determined Herglotz inversion method with Sloan-Womersley (left) and Fliege-Maier (right) spatial sampling grids of increasing orders. The shaded grey region represents one standard deviation around the mean error, while the red stars (with dash-dotted line) show the maximum error.

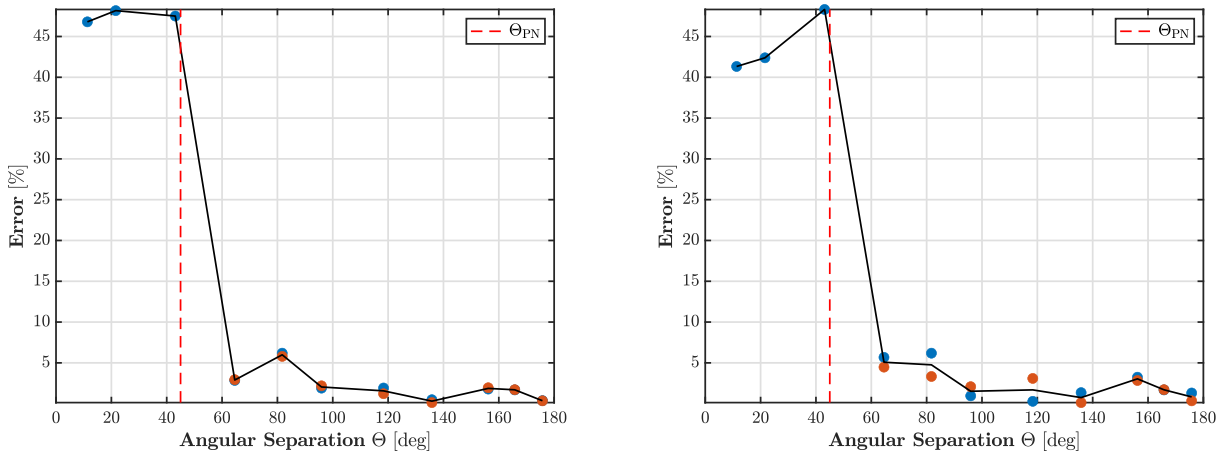
Following these results, the order $L+4$ over-determined Fliege-Maier grid is consequently chosen for the regularized Herglotz inversion, with an imposed optimized global regularization parameter $\chi \simeq 0.142$. As mentioned above, the value for χ is obtained by minimizing both the mean error and its standard deviation over the sphere; in this case, the geometric mean of the two quantities is defined as a cost function and a basic slice-sampling algorithm is used for the parameter search.

Finally, the aforementioned 10^4-point, 99^{th}-order Sloan-Womersley layout is used throughout these validation tests for both the SRP steering directions and the Herglotz inversion interpolation in order to achieve maximum precision. In practice, however, the size of this grid is such that computing costs can quickly become impractical. As such, a simple optimization strategy will be described below in order to obtain a trade-off between localization precision and computation length.

To conclude the examination of single impulse localization, Tab. 4.5 presents the mean estimation errors, as well as their standard deviation and maximum, over the D_{synth} simulated impulse DoAs for both the SRP and interpolated regularized Herglotz inversion approaches. Though the results are comparable, the SRP seems to produce slightly more accurate DoA estimations in this particular case.

	Single Echo DoA Estimation Errors		
	Mean	**Standard Dev.**	**Maximum**
SRP	0.83°	0.33°	1.58°
Herglotz (Optimized Reg.)	1.57°	0.83°	5.35°

Table 4.5: DoA estimation errors for individually localized single incident impulses distributed over the sphere.



(a) SRP. (b) Herglotz inversion.

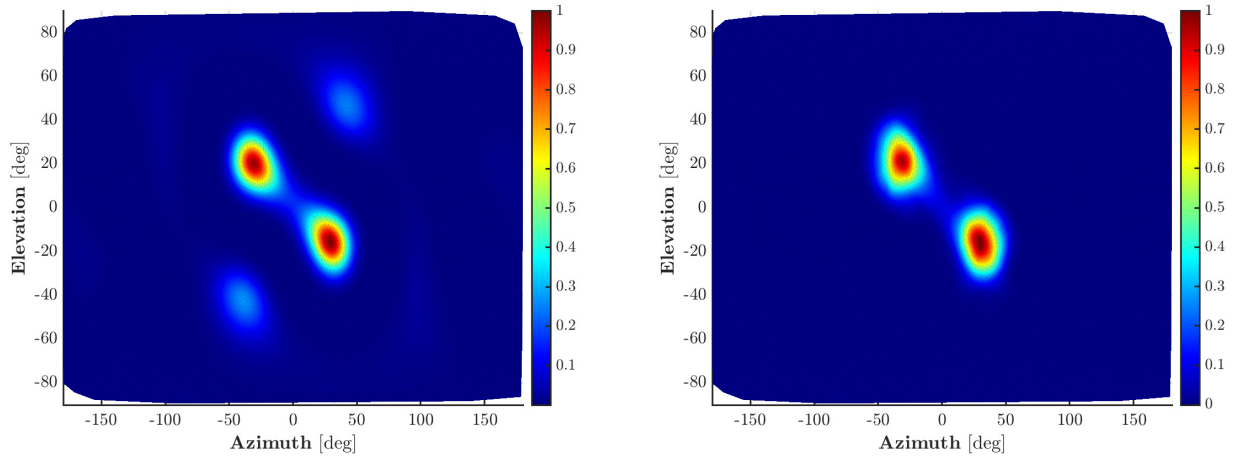
Figure 4.3: DoA estimation errors for two simultaneous incident impulses with varying angular separations. Simulated mh acoustics Eigenmike, 4^{th}-order HOA encoding.

Since the choice of using these two methods was partially motivated by their ability to detect simultaneous correlated reflections, two other fundamental tests are pertinent: the first tracks the methods' estimation errors with respect to the angular separation between two simultaneous incident impulses (Fig. 4.3), while the second evaluates their performance against an increasing number of simultaneous impulses (Fig. 4.5). These are simulated by simply extending the mapping vector in Eq. 4.4 to include a set of indices corresponding to the desired configuration.

In the first case, the angular separation Θ between the two incident impulses is controlled by varying their respective DoAs $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, from $\boldsymbol{\Omega}_{1,1} = [-90°, 45°]$ to $\boldsymbol{\Omega}_{1,N} = [-2°, 1°]$ and $\boldsymbol{\Omega}_{2,1} = [90°, -45°]$ to

$\mathbf{\Omega}_{2,N} = [2°, -1°]$. For each DoA pair, the two localization errors (coloured points) and their mean (solid black line) are shown in Fig. 4.3 as percentages of the corresponding synthesized angular separation; the 4$^{\text{th}}$-order PWD main-lobe peak-to-null half-width Θ_{PN} is also included (dashed red line). Once again the results are comparable, with most errors below 5% for both methods and a slight overall advantage to the SRP.

These results also verify the discussion from Sec. 1.2.1, in which Θ_{PN} represents a resolution limit with regard to the PWD formalism: below an angular separation of Θ_{PN}, simultaneous impulses cannot be resolved by either method and the estimation error increases to around half the synthesized angular separation. Since the Herglotz inversion approach is fundamentally based on the same mathematical description (see Sec. 2.1.2), it follows that the limitation should also appear despite the regularized under-determination.



(a) SRP. (b) Herglotz inversion.

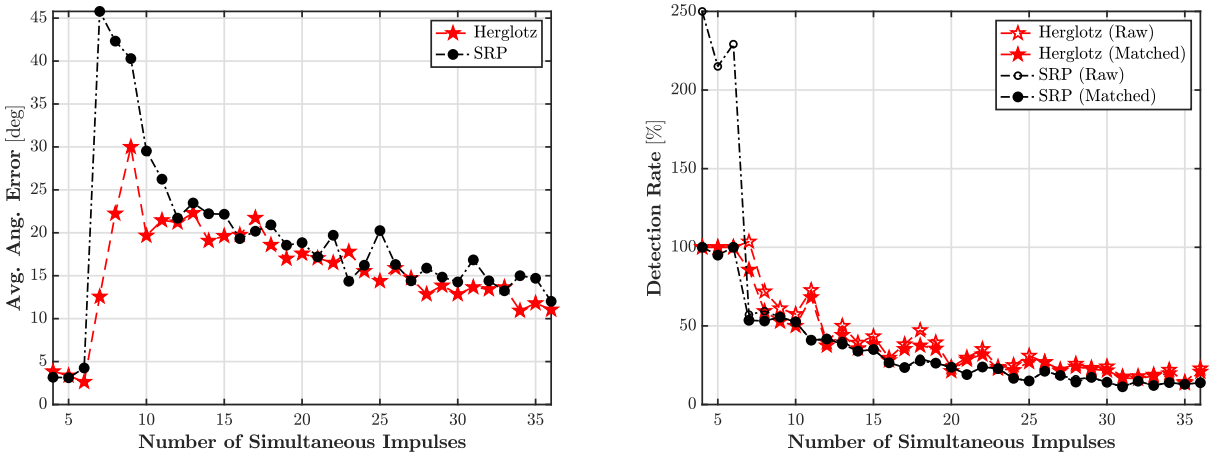
Figure 4.4: Localization maps for the SRP (left) and interpolated regularized Herglotz inversion (right) DoA estimation methods, with two simultaneous incident impulses synthesized $\Theta = 63°$ apart. Simulated mh acoustics Eigenmike measurement with 4$^{\text{th}}$-order HOA encoding.

However, another limitation of the straightforward beamformed SRP approach is the appearance of "ghost" peaks in the localization map due to the inevitable overlap between side lobes, even at the most highly directive frequencies. This is illustrated in Fig. 4.4 using the SRP and Herglotz localization maps for a simulated 63° angular separation. As such, the SRP method appears potentially more prone to the detection of false positives, thereby highlighting the need to evaluate the two methods' performance against an increasing number of simultaneous impulses.

Because of these potential mismatches between the number of true incident impulses and those detected, it is necessary to include a "matching" step when calculating localization errors on multiple simultaneous echoes. This ensures that the errors are calculated between corresponding synthesized-detected pairs, and that false positives (or missing/undetected impulses) are identified as such. Consequently, Fig. 4.5 shows the average angular error over the matched echoes as well as the resulting detection rates, both before and after matching, for three different synthesis DoA layout grids.

In general, the regularized Herglotz inversion seems to "match" better in the sense that it tends to show less disparity between the matched and unmatched/raw detection rates while maintaining lower localization errors even as the number of incident impulses greatly increases. This also confirms the impression given above with respect to the generation of false positives under the SRP method.

The fundamental localization capabilities of both the SRP and regularized Herglotz inversion methods are therefore generally comparable, each with their own advantages and weaknesses. Whereas the strict DoA estimation accuracy of the SRP method is slightly better than that of the Herglotz



(a) Average angular error.

(b) Detection rate.

Figure 4.5: Mean DoA estimation error (left) and detection rate (right) as a function of the number of synthesized simultaneous incident impulses. For each number of simulated reflections, detection and estimation is performed using four different layout grids for the synthesized DoAs: spherical covering, maximum volume, packing, and minimum potential grids based on Sloane and Conway [96]. The angular error and detection rate results for the four grids are then averaged together.

inversion, for example, the latter appears more robust when faced with multiple simultaneous impulses. Continuing with the specific parameterizations obtained and validated in this section, we now look to extend our assessment of the two methods to a more realistic SRIR context by applying them to image-source (IS) simulations in different modelled spaces.

Combined DoA-ToA-Energy Estimations on Simulated IS SRIRs

The IS simulations to be used in the following tests are each generated by the EVERTims auralization engine [98] [99]. The modelled spaces include three fictional "test rooms" and two real spaces: the historic Fogg Art Museum lecture hall at Harvard University [100] and the recent (2006) 300-seat auditorium at the Morgan Library & Museum in New York.

Once again, since this research project is focused on a "blind" measurement analysis and manipulation approach, the geometrical and architectural details of these spaces are not of any particular interest in this context, and indeed their choice is completely arbitrary. In fact, only the resulting space-time-frequency distributions of the IS-modelled reflections are of any importance for the current evaluation; detailed descriptions of the simulated spaces will therefore be foregone.

Concretely, the EVERTims simulations output the DoAs, ToAs, and octave-band linear amplitudes for every modelled reflection up to a given maximum IS order. Just as in the fundamental localization tests above, the IS model echoes are then spatially quantized to a L_{synth}-order Sloan-Womersley grid of incident DoAs, and an archetypal impulse signal is once again constructed by band-pass filtering a sample-rate Dirac delta between 125 Hz and 16 kHz.

However, since the IS simulations give frequency-dependent octave-band linear amplitudes (translating the frequency-dependent absorption properties of the space), this archetypal signal must be further filtered using an octave-band shelving filter bank. Then each constructed reflection can be added to the complete sample-rate space-time-frequency signal $\mathbf{p}_{\text{IS}}(f,t)$, with the appropriate time delays corresponding to each echo's ToA. To avoid numerical instabilities, $\mathbf{p}_{\text{IS}}(f,t)$ is initialized using a virtual zero-mean Gaussian white noise floor set at $P^{\text{dB}}_{\text{noise}} = -90$ dB.

Once all the modelled echoes have thus been synthesized and placed, the resulting space-time signal is used to simulate an SMA measurement in a similar manner to Eq. 4.4 but extended to a full set of

N_{ech} echoes. This is perhaps best described as:

$$\mathbf{h}_{\text{IS}}(f,t) = \mathbf{Q}(f)\mathbf{D}\mathbf{p}_{\text{IS}}(f,t), \tag{4.5}$$

where $\mathbf{Q}(f)$ and $\mathbf{D}$ are once again the same as in Eqs. 4.2 and 4.4, and $\mathbf{p}_{\text{IS}}(f,t)$ is thus a $D_{\text{synth}} \times 1$ vector containing the total synthesized space-time-frequency signal model of the IS sound field. The simulated measurement is then encoded to HOA as $\mathbf{h}_{\text{IS}}^{\text{HOA}}(f,t) = \mathbf{Y}_{\text{HOA}}(f)\mathbf{h}_{\text{IS}}$, where $\mathbf{Y}_{\text{HOA}}$ is the SMA's HOA-encoding matrix with elements $\beta_q a_l(f) Y_{l,m}^*(\mathbf{\Omega}_q)$ [β_q are the quadrature weights corresponding to the SMA's transducer positions $\mathbf{\Omega}_q$ and $a_l(f)$ are the HOA encoding filters]. Then the complete echo detection and DoA-ToA-energy estimation procedure described in Sec. 2.4.2 can be applied to $\mathbf{h}_{\text{IS}}^{\text{HOA}}(f,t)$ using both the SRP and Herglotz inversion methods.

As noted above, the maximally precise 99$^{\text{th}}$-order Sloan-Womersley grid can quickly become prohibitively expensive in terms of computing cost when repeatedly used such as in the frame-by-frame analysis of an SRIR. The two computational bottlenecks in the echo detection procedure are the cubic Hermite spherical interpolation [87] and spherical surface peak detection (Appx. A.4) processes.

Without opening a technical discussion on algorithmic efficiency, since there has admittedly been very little work done to optimize the underlying functions used in this work, there is nonetheless a simple way to obtain a trade-off between the final localization precision and the total computing cost of these two processes. For each order of the Sloan-Womersley grid, the average angular distance between neighbouring points can be used as a measure of precision, while the interpolation and peak detection algorithms can be applied to a single synthesized impulse (as in the first test above) to obtain a characteristic computation length (note that the SRP method does not require interpolation, though the energy and phase estimation procedures do).

These two measures can then be averaged together – and potentially weighted to emphasize either better precision or lower computing cost – to obtain an order-dependent cost function. The order that minimizes this cost function can finally be considered to provide the best trade-off. For single-impulse localization on a simulated Eigenmike measurement, a 38$^{\text{th}}$-order Sloan-Womersley grid was thus determined to be the optimal compromise, and is therefore the layout used for the SRP steering directions and the final Herglotz inversion interpolation leading to the results given in Tab. 4.6.

Following the procedure described in Sec. 2.4.2, echoes are detected frame-by-frame on the simulated IS SRIRs, with each detection giving a DoA, ToA, and energy estimate. Once again, an STFT was used to obtain a full space-time-frequency representation of the SRIR. As previously noted, the parameterization of this STFT is especially important in the context of echo detection.

First, in order to avoid repeated detections of any single reflection, a straight rectangular window (a.k.a. the boxcar or Dirichlet window) should be used with zero overlap (in other words, the hop size should be equal to the window length). Though the rectangular window offers limited band rejection in the frequency domain, this is of less concern in the present application since high frequency resolution is not an imperative. Indeed, the only spectral information exploited outright is the interpolated phase used for ToA estimation, on which a linear regression is performed in any case[2].

Another important frequential constraint to consider is that we would like to combine the localization maps over several frequency bins around the maximum directivity frequency, as described in Sec. 2.4.2. Assuming the combination of at least three bins (the one containing the maximum directivity frequency and its two neighbours), it should therefore be ensured that these are thin enough so that the highest one does not "bleed" too far over the spatial aliasing limit f_{alias}.

In general, however, allowing for a lower frequency resolution has the major benefit of enabling shorter time windows to be used. Since we have seen that the DoA estimation methods are both

[2]In fact, the frequency-domain smoothing resulting from the rectangular window's important side lobes is somewhat of a desirable quality here, since it will regularize the data points fed to the linear regression.

similarly limited in their ability to detect multiple simultaneous reflections, using shorter windows avoids "overcrowding" any given frame with more echoes that can be detected.

For example, as noted above, the Eigenmike's 4^{th}-order natural PWD maximum directivity frequency is $f_{\text{maxDI}} \approx 3.45$ kHz, and its spatial aliasing frequency is $f_{\text{alias}} \simeq 5.19$ kHz. At a sampling rate $f_s = 48$ kHz, a 64-sample window would therefore allow us to use three bins centred on 3 kHz, 3.75 kHz, and 4.5 kHz. Assuming our choice of window sample length is restricted to powers of 2 (to optimize the FFT while avoiding zero-padding), this is the shortest possible option that avoids both the spatial aliasing limit and the weak directivity at low frequencies.

The only significant downside to choosing the shortest window possible is that the spatial incoherence estimation suffers as a result, since the mathematical expectation (i.e. the covariance matrix) is approximated by time averaging. In other words, longer windows result in better incoherence estimations. The solution, hinted at in Sec. 2.4.2, is therefore to group several short STFT frames into a larger incoherence estimation window (e.g. according to a given integer factor), and assume that the constituent frames of any point of low incoherence (i.e. high coherence) all contribute equally. This strategy does run the added risk of including frames that do not contain any coherent reflections in the analysis, but such an effect can be mitigated by keeping the "combination factor" (i.e. the number of frames to combine) as low as possible.

With these considerations in mind, a final window size of 128 samples is chosen (2.67 ms at $f_s = 48$ kHz), with a combination factor of 3 for incoherence estimation (i.e. a combined window size of 384 samples, or 8 ms at $f_s = 48$ kHz). This choice is the next power of 2 above the shortest value derived previously (64 samples), and allows for three bins centred at 3 kHz, 3.38 kHz, and 3.75 kHz to be used in generating the localization maps with less spectral bleeding above $f_{\text{alias}} \simeq 5.19$ kHz.

Two different DRIR representations are chosen to respectively calculate the incoherence profile $\psi_{\text{dir}}^{\text{early}}(t)$ and provide the directional spectra for the energy estimation and phase regression methods (see Sec. 2.4.2). Both make use of the Fliege-Maier look direction layout grid, rotated so that one point faces the DoA of the direct sound, but the former applies the natural PWD beamformer while the latter exploits the maximum WDI design.

Since the simulated IS SRIRs used for this test do not include an incoherent late reverberation tail, their incoherence profiles will simply alternate between frames where synthesized reflections are present and ones containing only background noise. Instead of attempting to estimate a mixing time on an SRIR that, by construction, does not contain one, the reference values $\overline{\psi}_{\text{inc}}^{\text{late}}$ and $\sigma_{\psi}^{\text{late}}$ (see Sec. 2.4) are substituted by values calculated over the noise floor of the simulated SRIR.

A complete presentation of the synthesized, detected, and matched error echo cartography figures, as well as the incoherence profiles, is relegated in Appx. B.1 to avoid overcrowding this chapter. The average DoA, ToA, and total energy errors over all matched echoes are simply presented here in Tab. 4.6 for each of the five IS SRIR simulations, and these global results are discussed below.

Echo matching is performed by calculating a likelihood between every synthesized (j) and detected (k) reflection, defined as:

$$\mathcal{L}_{j,k}^{\text{pos}} = \begin{cases} \Theta_{j,k}^{-1} & \text{if } -\frac{3T_w}{2} \leq (\tilde{t}_k^n - t_j) < \frac{3T_w}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{4.6}$$

where $\Theta_{j,k}$ is the central angle (or equivalently the unit great circle distance) between the synthesized DoA $\boldsymbol{\Omega}_j$ and the estimated $\widetilde{\boldsymbol{\Omega}}_k$, $\tilde{t}_k^n$ is the centre time of the detected echo's analysis window, t_j is the synthesized reflection's ToA, and T_w is the length of the analysis window ($T_w = 2.7$ ms in the current example).

In other words, echoes are matched by DoA within an extended time window of length $3T_w$ around the centre time of the analysis frame they were localized in (which corresponds to the length of the

	Average Echo Detection Errors (Raw Mean)											
	DoA		**ToA** [ms]		**Energy** [dB]		$\mathbf{N}_{\text{det}}$ [%]		$\mathbf{N}_{\text{match}}$ [%]		**MEL** [%]	
	SRP	Herg.	SRP	Herg.	SRP	Herg.	SRP	Herg.	SRP	Herg.	SRP	Herg.
Test 1	12.9°	8.83°	1.87	1.67	9.89	9.73	−79.7	−81.2	−84.9	−87.0	5.30	6.65
Test 2	3.20°	3.05°	1.27	1.29	3.85	4.61	−10.5	−2.33	−36.0	−36.0	11.0	4.70
Test 3	6.14°	5.45°	1.49	1.50	7.57	7.45	−66.7	−67.9	−72.6	−74.2	9.25	6.89
Fogg	6.87°	4.67°	1.44	1.53	7.71	5.94	−43.8	−53.5	−65.9	−66.8	12.2	10.1
Morgan	6.00°	5.37°	1.17	1.19	3.76	4.10	−8.89	−7.78	−27.8	−27.8	0.440	4.63
Average	7.02°	5.47°	1.45	1.44	6.56	6.37	−41.9	−42.5	−57.4	−58.4	7.64	6.59

Table 4.6: Mean echo detection errors for five different SRIRs simulated using image sources (IS), and analyzed with the steered response power (SRP) and regularized Herglotz inversion methods. N_{det} and N_{match} represent the total number of detected and matched echoes, respectively; their errors are made relative to the true synthesized N_{synth}. The "matching energy loss" (MEL) is the percentage ratio of the total discarded (unmatched) echo energy to the total detected echo energy.

incoherence estimation window). This extended window is simply used to allow for the "clustering" space-time localization effects that will be further discussed below. Matches are furthermore forced to be unique: any detected echoes matched to a synthesized reflection that has already been paired with a larger $\mathcal{L}^{\text{pos}}_{j,k}$ are discarded as false positives.

Finally, the three model errors (DoA, ToA, and energy) can be calculated for each remaining validated matching pair. The mean values over all valid matching pairs for each of the five IS SRIR simulations are thus reported in Tab. 4.6, for both the SRP and Herglotz inversion methods. The angular DoA errors are calculated as $\Theta_{j,k}$ in Eq. 4.6, i.e. central angles or unit great circle distances. The ToA errors are averaged as absolute differences $|t_j - \tilde{t}_k|$, where $\tilde{t}_k$ is the ToA estimated by phase regression, and similarly the energy errors are averaged as absolute dB values.

In addition to the DoA-ToA-energy estimation errors, Tab. 4.6 also includes some information pertaining to the echo matching itself. The relative errors in the total number of echoes detected (N_{det}) and the final number of matched echoes (N_{match}) are thus reported alongside the "matching energy loss" (MEL). The MEL quantifies the amount of energy discarded during the matching process, and expresses it as a percentage ratio between the total energy of the discarded reflections and the total detected energy. The lower the percentage, the less energy was discarded; in other words, the less energy was carried by false positive detections (within the limits of the matching process, which cannot completely rule out all false positives).

The first conclusion that can be drawn from Tab. 4.6 is that the overall average DoA estimation errors for both methods are below 9°; in other words, since the maximum angular error is 180°, the relative error is less than 5%. In fact, only the SRP method on Test Room 1 presents an error above this mark. Note that this performance is achieved despite having resorted to the less precise 38^{th}-order (1521-point) Sloan-Womersley grid for the localization maps, as described above. This grid has an average angular separation between nearest neighbours of 5.63°, which implies that for any given DoA estimation, up to 2.81° of its error could be inherently due to quantization.

It is also interesting to note that overall it is the Herglotz inversion method that performs marginally better in this application, despite the fundamental localization tests above suggesting that the SRP provides more accurate estimates. This is perhaps due to the Herglotz inversion's greater robustness in detecting multiple simultaneous reflections, as evidenced by the results given in Fig. 4.5.

Despite the use of an extended time window for matching the detected and synthesized echoes (see Eq. 4.6), the average ToA estimation errors are all smaller than the length of the analysis window $T_w = 2.67$ ms. It should be noted that, by construction (see Sec. 2.4.2), the ToA estimates are not allowed to extend beyond the time bounds of the current frame. If the echo has been localized in the "correct" time frame, therefore, it follows that its maximum possible error is T_w.

However, as noted above, it is possible that a given echo may not, in fact, be localized in the correct time frame. This is mostly due to what could be referred to as a "clustering" effect, whereby multiple reflections grouped closely together both in space and time are erroneously identified as a single echo. This effect is clearly visible in many of the reflection cartography figures in Appx. B.1, and is greatly responsible for the N_{det} and N_{match} errors reported in Tab. 4.6. On average, under half the actual synthesized reflections were detected (just over 40%), with variations depending on the echo density of each particular SRIR (a point to be further discussed below).

This clustering effect is essentially due to the limited spatial resolution offered by the SMA. As shown in Fig. 4.3, the main lobe peak-to-null half-width Θ_{PN} of the natural PWD directivity function appears to be a fundamental resolution limit in terms of localization separability. When simultaneous echoes are within Θ_{PN} of each other, the localization maps for both methods present a single peak around their directional barycentre (hence the near 50% error below Θ_{PN} in Fig. 4.3).

Furthermore, such clustering appears to have an effect in the time domain as well. The lack of spatial resolution "combines" directionally separate signals, which results in an averaging of both their spatial and temporal information. This relationship between the directional and temporal detection limits requires further study and would merit its own series of fundamental simulation tests.

On top of averaging the space-time properties of any reflections under the limit of resolvability, the clustering effect also results in the combination of their energies. As such, the detected energies for clustered reflections are generally over-estimated. This can be readily seen in the energy error cartography figures given in Appx. B.1, and also helps explain why, of the three model properties, the energy estimation errors reported in Tab. 4.6 are by far the most important and widely varying.

In consequence, a general trend can be expected with respect to the synthesized echo density of each SRIR: the denser the SRIR, the more important the clustering effects will be. In other words, both N_{det} and N_{match} should be further below the true N_{synth} for denser SRIRs, resulting in greater estimation errors. To illustrate this, the synthesized and detected echo densities (ρ, in echoes per ms) are presented in Fig. 4.6 for each of the five simulated SRIRs.

When compared to the results from Tab. 4.6, the expected trend is indeed confirmed: the discrepancy between the synthesized and detected/matched densities increases with ρ_{synth}. Furthermore, the average energy estimation errors also increase with the true echo density, which again suggests that these are heavily affected by the clustering effects described above. On the other hand, the MEL does not follow the same trend, which at least demonstrates that the echo detection procedure does not necessarily generate more important false positives when the SRIR's echo density is higher.

The detection (and matching) densities also seem to "max out" at a certain value regardless of the true ρ_{synth}; this ceiling is most likely fundamentally tied to the spatial resolution of the given SMA in the same way as the clustering effects. More work is needed to determine how these limitations could be overcome with higher resolution SMAs, a point that will be further addressed in the Conclusion.

In general, the echo detection procedure presented in Sec. 2.4.2 therefore allows for a decently accurate characterization of an SRIR's prominent early reflections. To close out this section, we complement the above analysis of early reflections with a special focus on characterizing the direct sound which, as noted in Sec. 2.4.1, merits its own particular treatment due to its singular importance in both the analytical and perceptual description of reverberation effects.

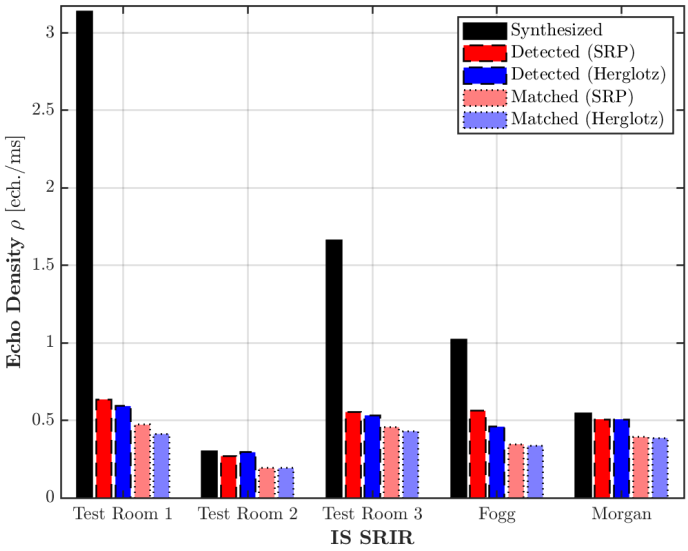


Figure 4.6: Synthesized and estimated echo densities for five different SRIRs simulated using image sources (IS), and analyzed with the steered response power (SRP) and regularized Herglotz inversion methods. All densities are calculated using the maximum synthesized ToA $t_j^{\max}$ in milliseconds: $\rho_{\{\text{synth,det,match}\}} = N_{\{\text{synth,det,match}\}}/(10^3 t_j^{\max})$.

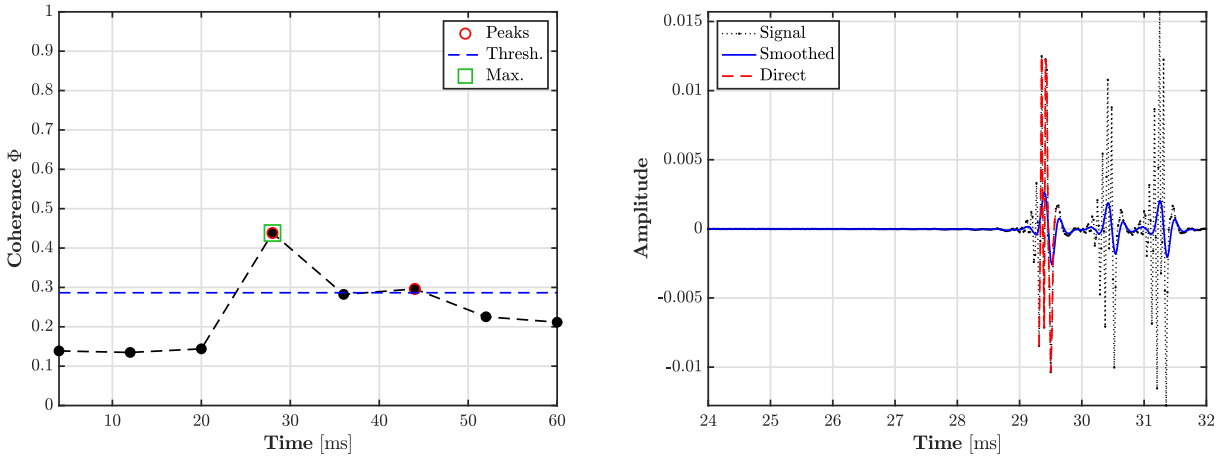
Direct Sound Detection

The direct sound detection method presented in Sec. 2.4.1 was mainly motivated by, and is therefore more adapted to, an application to real-world measurements and the specific challenges encountered therein (mostly characterized by the presence of signal artefacts preceding the true direct sound, from excessive harmonic distortion or non-stationary background noise). The basic functioning of the procedure can nonetheless be demonstrated on the IS SRIR simulations described above.

Figure 4.7 summarizes the signal detection procedure, first using the spatial coherence measure (a, left) to identify the frame within which to then estimate its time-domain position (b, right) in the omnidirectional HOA channel $h_{0,0}(t)$. This position can in fact be technically characterized in the time domain by three sample values: the start point n_{start}, the peak value point n_{peak} (i.e. the sample corresponding to the maximum amplitude), and the end point n_{end}. Note that these have been omitted from Fig. 4.7b for clarity.

Direction of arrival estimation is then performed within a 128-sample rectangular window centred on n_{peak}, using both the SRP and regularized Herglotz inversion methods as parameterized above. The localization maps are once again generated using a 1521-point (38^{th}-order) Sloan-Womersley spherical point grid. The ToA is simply taken here as n_{peak}/f_s (i.e. without the phase regression technique used in the complete echo detection approach), and the energy is estimated in the same manner as for the echo detection (i.e. using the MWDI beamformed DRIR representation, see above). Errors in the direct sound DoA, ToA, and energy estimations for the five previously described SRIR simulations are reported in Tab. 4.7.

For these five simulated SRIR examples, the SRP and regularized Herglotz inversion methods give identical localizations, with all DoA errors below the quantization threshold of 2.81° for the 38^{th}-order Sloan-Womersley grid (based on its average angular separation of 5.61° between nearest neighbours). The ToAs are all slightly over-estimated, but the errors are negligible: the average of 0.187 ms corresponds to just under 9 samples at the simulated $f_s = 48$ kHz, or only 6.4 cm at a sound speed of $c = 343$ m/s (i.e. at 20° C), which is much smaller than the actual size of the loudspeakers used in real-world measurements (see the following Ch. 5).



(a) Spatial coherence.

(b) Temporal peak detection.

Figure 4.7: Direct sound detection on the "Test Room 1" IS-simulated SRIR. First a measure of spatial coherence $\Phi(t) = 1 - \psi_{\text{dir}}(t)$ (a, left) is used to identify a time window within which to search for the actual direct sound signal. The exact position of the direct sound's impulsive signal within the frame is then determined using the procedure described in Sec. 2.4.1 on the time-domain omnidirectional HOA-SRIR channel $h_{0,0}(t)$ (b, right).

Direct Sound Detection Errors

	DoA		**ToA** [ms]	**Energy** [dB]	
	SRP	Herg.		SRP	Herg.
Test Room 1	2.29°	2.29°	0.112	3.70	3.70
Test Room 2	2.29°	2.29°	0.198	5.05	5.05
Test Room 3	1.67°	1.67°	0.252	2.56	2.56
Fogg	2.29°	2.29°	0.117	−0.180	−0.180
Morgan	2.80°	2.80°	0.258	1.45	1.45
Average	2.27°	2.27°	0.187	2.52	2.52

Table 4.7: Direct sound detection errors for five different SRIRs simulated using image sources (IS), and analyzed with the steered response power (SRP) and regularized Herglotz inversion methods. Note that, as opposed to the full echo detection strategy, the direct sound ToA is simply estimated using the omnidirectional peak signal sample n_{peak} and therefore does not depend on the choice of localization method.

The energy estimations suffer from the same phenomena as in the complete echo detection procedure, and, with the notable exception of the Fogg simulation, are all over-estimated by more than a decibel. However, three of the five estimation errors are below 3 dB (i.e. a 100 % relative error in linear scale energy), and all of them are significantly lower than the average errors from the full echo detections given in Tab. 4.6.

Overall, the direct sound detection procedure appears to give an accurate characterization of the SRIR measurement's basic geometrical property, i.e. the relative configuration of the source and receiver positions (without additional information, their absolute positions cannot be placed within the space based on this analysis alone). As will be seen in Ch. 5, the presence of additional uncertainties as to the absolute measured sound pressure levels in real-world measurements also serves to temper the

significance of the energy estimation inaccuracies: what is truly important is capturing the space-time evolution of the relative energy levels throughout the early segment of the SRIR, as in the complete echo cartographies shown in Appx. B.1.

This concludes the validation tests for the analysis of the early reflections; as noted above, the use of these methods will next be evaluated in the context of real-world measurements in Sec. 5.3. For now, we turn to assessing the mixing time estimation algorithm described in Sec. 2.3.

4.2.2/ Mixing Time Estimation

In order to evaluate the mixing time estimation method from Sec. 2.3, the simulated SRIRs must include both discrete early reflections and an incoherent late reverberation tail (since t_{mix} describes the transition point between the two). This section therefore begins with a description of how such "complete" SRIR simulations can be obtained from the IS-only versions from Sec. 4.2.1.

Following the results of Sec. 4.1.4 on spatial incoherence preservation in both the SH domain and various DRIR representations, it is to be expected that the mixing time estimation algorithm will be sensitive to the different parameterization choices made when generating the time-dependent incoherence profiles. A brief summary of the specific choices made in this work is therefore also given.

Finally, the mixing time estimation results obtained using both the SH-domain CoMEDiE diffuseness measure (ψ_{cov}) and the proposed DRIR spatial incoherence measure (ψ_{dir}) are presented.

Adding Synthesized Reverberation Tails to IS SRIR Simulations

Working from one of the IS SRIR simulations described in Sec. 4.2.1 above, a complete "hybrid" SRIR can be obtained by adding a late reverberation tail synthesized as a field of zero-mean Gaussian noise signals subjected to an exponentially decaying energy envelope whose space-time-frequency properties are made to match those of the IS-only SRIR. This synthesized reverberation tail is then faded in to a given t_{mix} value, at which point the original IS SRIR is cut off.

More specifically, the target t_{mix} is set as half of the broadband t_{lim} obtained by analyzing the omnidirectional component of the HOA-domain IS-only SRIR's broadband reverse-integrated energy decay curve (EDC) using the method described in Sec. 2.5.1 (i.e. treating the EDC as a single EDR bin). The T_{60} reverberation time given by this analysis is also used to parameterize the exponentially decaying energy envelope that is applied to the zero-mean Gaussian noise signals.

A fully isotropic late tail can then be generated by using this omnidirectional T_{60} value to define an exponentially decaying energy envelope applied to zero-mean Gaussian noise signals synthesized per SH component. To ensure a "smooth" transition [and a "correct" initial power spectrum $P_0(f)$], the EDRs of the IS SRIR and the synthesized tail are subsequently matched per SH order at the time frame corresponding to the given mixing time. In other words, the ratio of the the EDRs at that time frame is used to define an energy envelope that is then applied to the STFT of the synthesized tail.

In keeping with the objectives of this thesis, a highly anisotropic tail can also be created by making the reference omnidirectional T_{60} evolve over the sphere, for example according to a cardioid directivity pattern. In this work, a cardioid pattern making the T_{60} evolve from half to 1.5 times the reference omnidirectional value is used. Furthermore, the cardioid is steered so that its maximum value faces the (known) DoA of the IS SRIR's direct sound.

In the anisotropic case, the energy decay envelope must therefore be applied to the zero-mean Gaussian noise signals in a directional manner; as a result, the EDR matching at the t_{mix} frame must also be performed in the same way. Since absolute spatial incoherence must be preserved as much as possible when transforming back to the SH domain, the DRIR representation is defined using the natural PWD beamformer (this point is further developed in Sec. 4.3.1 below). The Fliege-Maier grid is again used for the look direction layout, and is rotated to align a point with the direct sound's DoA.

For both the isotropic and anisotropic tails, the fade-in envelope is chosen to be exponentially increasing at twice the decay rate. In other words, the reverberation tail signals are made to fade in at the same rate at which they will then be decaying starting at t_{mix} (since the fade-in envelope is applied directly to the decaying tail signals).

Note that neither of these reverberation tails are by any means an accurate prediction of the sound field that is truly generated in the simulated room. As noted earlier when presenting the IS-only simulations, only the generation of "pseudo-realistic" examples with known parameter values is of any real interest within the context of this work.

It should also be noted that these examples are no longer "true" SMA measurement simulations: although this was the case for the original IS-only SRIRs, the late reverberation tails are then directly synthesized and matched. For the anisotropic simulation, this can result in greater anisotropy than that possible from an equivalent SMA measurement, since the spatial discretization and rigid sphere scattering effects are no longer considered. But the presence of extreme anisotropy only ends up giving the mixing time estimation method a more rigorous test, and bears no consequence on its validation.

Parameterizing Incoherence Profiles

As opposed to the incoherence measures shown in Sec. 4.1.4, in which the covariance matrices are calculated directly in the time domain, the incoherence profiles used to estimate mixing times in this section are based on a time-frequency STFT representation (see Sec. 2.3; the covariance matrices are averaged over both frequency bins and time frames). Working in the STFT domain offers a more refined trade-off between expectation value estimations and total window length (made up of overlapping frames), though it comes at a higher computational cost than the time-domain implementation. It also has the added practicality of fitting naturally into the general STFT-based space-time-frequency analysis-treatment-manipulation framework considered throughout this thesis.

The incoherence profiles generated for this work are calculated from an STFT using 1024-sample Nuttall windows with 87.5% overlap (i.e. a hop size of 128 samples) at a sampling rate $f_s = 48$ kHz. Expectation values are then estimated using a "frame averaging factor" $N_{\text{exp}} = 8$ that defines the number of STFT frames over which to average. This results in a total incoherence profile window length of 1920 samples, or 40 ms at $f_s = 48$ kHz.

To generate the DRIR representation, the MWDI beamformer (Appx. A.2) is chosen for its robustness with respect to anisotropy, as demonstrated in Sec. 4.1.4. Though it cannot achieve as high a maximum incoherence value as, for example, the natural PWD beamformer, this robustness is necessary in order to verify the mixing time estimation algorithm's assumption that the late reverberation tail is represented by a stable maximum value on the incoherence profile. Once again, the look direction layout is a Fliege-Maier grid rotated so that one point faces the DoA of the direct sound.

Finally, the mixing time estimation algorithm's re-segmentation control parameter λ_{reSeg} is set to $\lambda_{\text{reSeg}} = 0.14$, and the maximum of the re-segmentation scores is taken in order to obtain a "safe" t_{mix} (see Sec. 2.3 once again for more details on the estimation algorithm).

Results

The complete mixing time estimation incoherence profile plots for each of the ten synthesized SRIRs (both isotropic and anisotropic tails fitted to each of the five original IS simulations) are shown in Appx. B.2. As noted above, both the SH-domain CoMEDiE diffuseness and the proposed DRIR spatial incoherence measures are used to generate incoherence profiles for estimation. Tab. 4.8 thus reports the absolute and relative errors obtained for both measures on each of the ten examples.

As could be expected from Sec. 4.1.4, the DRIR spatial incoherence measure is far more robust with respect to anisotropy, providing valid t_{mix} estimations for all ten SRIRs regardless of tail isotropy.

	t_{mix} Estimation Errors			
	SH-CoMEDiE ψ_{cov}		**DRIR Inc.** ψ_{dir}	
	Abs. [ms]	Rel. [%]	Abs. [ms]	Rel. [%]
Test Room 1 (Iso.)	17.4	9.82	−19.9	−11.2
Test Room 1 (Aniso.)	×	×	14.7	8.31
Test Room 2 (Iso.)	0.0729	0.0506	−2.59	−1.80
Test Room 2 (Aniso.)	−5.26	−3.65	8.07	5.61
Test Room 3 (Iso.)	41.4	23.0	20.1	11.2
Test Room 3 (Aniso.)	257	143	14.7	8.19
Fogg (Iso.)	4.07	3.77	1.41	1.30
Fogg (Aniso.)	×	×	4.07	3.77
Morgan (Iso.)	14.7	19.4	−11.9	−15.7
Morgan (Aniso.)	×	×	14.7	19.4

Table 4.8: Mixing time (t_{mix}) estimation results for ten SRIRs synthesized by matching both isotropic and anisotropic late reverberation tails to the IS-simulated SRIRs from Sec. 4.2.1. Estimations are obtained by applying the algorithm described in Sec. 2.3 to incoherence profiles generated using both the SH-domain CoMEDiE diffuseness measure (left) and the proposed DRIR spatial incoherence measure (right). DRIRs are defined here using the MWDI beamformer (Appx. A.2). Examples where the estimation algorithm failed to return a valid t_{mix} are marked with a × symbol.

In comparison, the SH-domain CoMEDiE measure only returns a valid estimate in two anisotropic cases, and only one of those (Test Room 2) is of any acceptable accuracy. In fact, the DRIR spatial incoherence measure is even more accurate than the SH-domain CoMEDiE in three out of the five isotropic examples (i.e. 60%), despite the latter appearing more sensitive to changes in plane wave density in the first validation tests in Sec. 4.1.4.

Additionally, seven out of the ten DRIR spatial incoherence results are over-estimated, which corresponds to the choice made to search for a "safe" t_{mix} (i.e. overestimating the mixing time presents less of a risk in terms of accidentally including discrete early reflections in the late reverberation tail, especially in applications such as full tail resynthesis [Sec. 3.3.2] and hybrid reverberation [15]). This also corresponds to the design of the synthesized SRIRs, in which the incoherent reverberation tail is the only field present from t_{mix} onward.

Six out of the ten relative errors are below 10%, and none are above 20%; in absolute terms, all the errors are below the 24 ms characteristic time of the human auditory system [5] (i.e. the value initially used by Polack [7] to define a "mixing" sound field). In general, then, it appears that the algorithm described in Sec. 2.3 provides a robust and accurate estimate of an SRIR's mixing time, at least under the definition and identification objective of a mixing time that refers to the moment the stochastic late reverberation model (Sec. 2.1.3) completely describes the SRIR (above f_{Sch}).

4.2.3/ Multiple Slope Analysis

This short section briefly demonstrates the capacity of the late reverberation tail analysis methods presented in Sec. 2.5.1 to detect and model multi-slope decays. In order to evaluate these capabilities, double-slope ($C = 2$) decays can be readily simulated based on the classic theory of coupled volume

acoustics as presented e.g. by Cremer and Müller [5].

This approach is based on frequency-independent properties (decay rates, absorption coefficients, coupling factor, etc.) and the traditional diffuse view of the late reverberation field (i.e. isotropic and incoherent). It is also limited to the coupling of only two volumes, in which a double-slope can appear for certain source/receiver configurations.

Extending the theory to include frequency-dependent factors and allow for more than two coupled volumes is far from trivial and would necessitate a dedicated research effort; for this reason, the current validation test will focus solely on detecting the presence of double-slope decays using the framework described in Sec. 2.5.1 without attempting to investigate a complete frequency- and direction-dependent modelling approach.

Indeed, the multi-slope detection procedure from Sec. 2.5.1 is based on a broadband EDC (instead of a frequency-dependent EDR) analysis for a similar reason: multi-slope decays are known to appear clearly in a broadband view, but their frequency dependence is a less than straightforward affair. This point is to be further explored in Sec. 5.4.1, where the traditional view of coupled volumes as isotropic double-slopes will be confronted to the observation of directional analyses in which they can be separated into two highly anisotropic single-slope decays.

To generate the simulations for the current evaluation, coupled volume theory is used in the configuration known to produce the most identifiable double-slope decays: when both source and receiver are placed in the volume with the shortest reverberation time. In this case, the broadband isotropic (direction-independent) power envelope can be written, following Eq. 2.10, as:

$$P_1(t) = P_{0,\mathrm{I}1} e^{-2\gamma_\mathrm{I} t} + P_{0,\mathrm{II}1} e^{-2\gamma_\mathrm{II} t}, \tag{4.7}$$

where $P_{0,\mathrm{I}1}$, γ_I, $P_{0,\mathrm{II}1}$, and γ_II are the initial powers and decay rates *corresponding* to the two volumes but *as affected by the coupling*. They can be written in terms of the "natural" decay rates of the two volumes, γ_1 and γ_2 (i.e. as they would be if they were uncoupled):

$$\gamma_{\mathrm{I,II}} = \frac{1}{2}\left(\gamma_1 + \gamma_2\right) \pm \sqrt{\frac{1}{4}\left(\gamma_1 - \gamma_2\right)^2 + \Gamma^2 \gamma_1 \gamma_2} \tag{4.8}$$

and

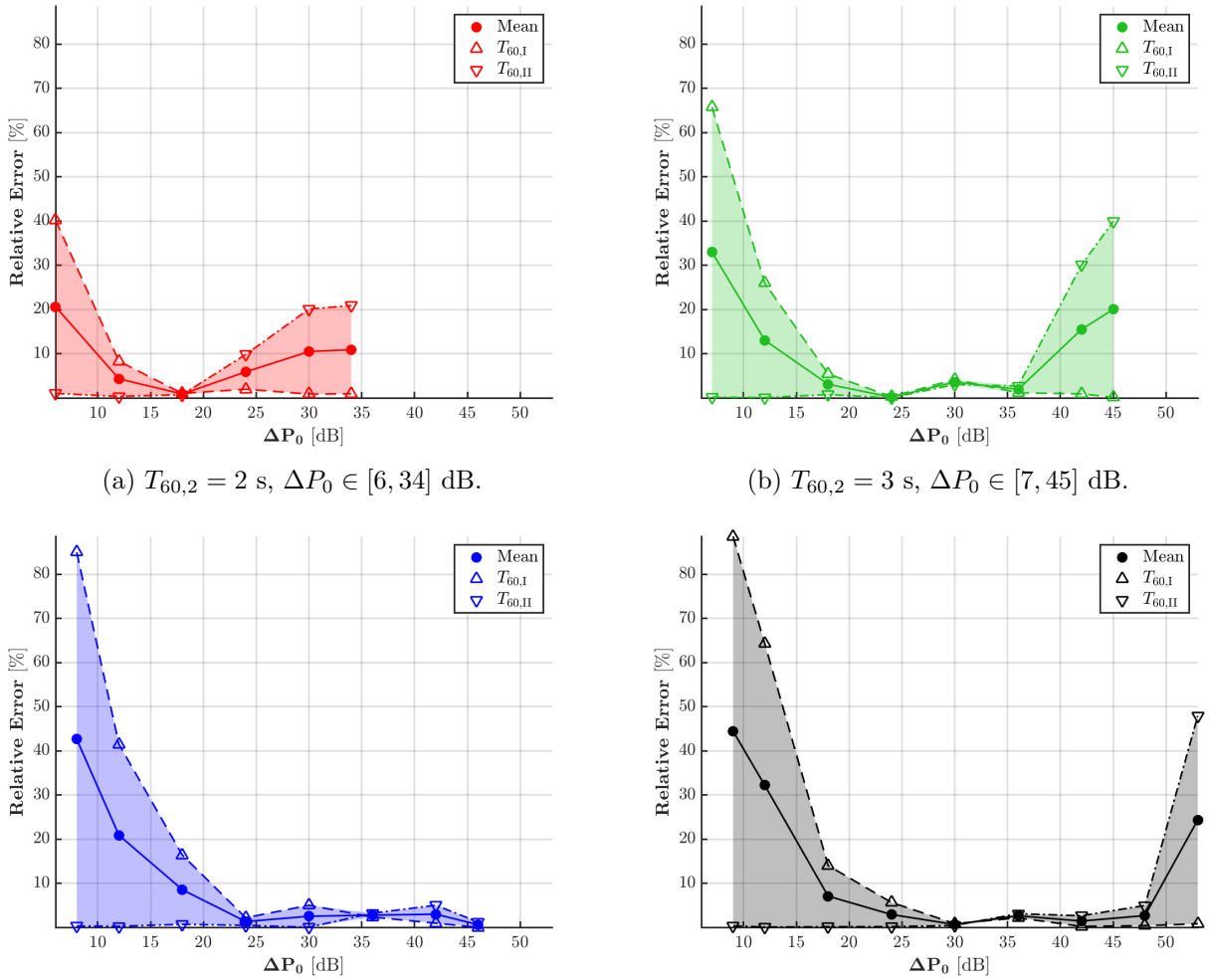
$$\begin{cases} P_{0,\mathrm{I}1} = \dfrac{P_{0,1}\left[1 - \Gamma^2 / \left(1 - \gamma_\mathrm{II}/\gamma_1\right)\right]}{1 - \Gamma^2 / \left(1 - \gamma_\mathrm{I}/\gamma_2\right)\left(1 - \gamma_\mathrm{II}/\gamma_1\right)} \\ P_{0,\mathrm{II}1} = \dfrac{P_{0,1}\Gamma^2\left[1 - 1 / \left(1 - \gamma_\mathrm{I}/\gamma_2\right)\right]}{1 - \gamma_\mathrm{II}/\gamma_1 - \Gamma^2 / \left(1 - \gamma_\mathrm{I}/\gamma_2\right)} \end{cases}, \tag{4.9}$$

where Γ is the "coupling factor" that controls the extent to which the two volumes are coupled and $P_{0,1}$ is the initial power of the source in the first volume.

As stated above, it is assumed in this configuration that $T_{60,1} < T_{60,2}$ and it therefore follows that $\gamma_\mathrm{I} > \gamma_1 > \gamma_2 > \gamma_\mathrm{II}$, since $\gamma = 3\ln(10)/T_{60}$. Since the initial source power $P_{0,1}$ is arbitrary and can, for simulation's sake, be considered unity, the more interesting quantity to consider is $\Delta P_0 = P_{0,\mathrm{I}1}/P_{0,\mathrm{II}1}$.

Having set γ_1 and γ_2 through $T_{60,1}$ and $T_{60,2}$ respectively, the resulting ΔP_0 can thus be set by tuning the coupling strength Γ. In this validation test, a constant $T_{60,1} = 1$ s is maintained while $T_{60,2}$ is made to vary from 2 s to 5 s by increments of 1 s. For each $\{T_{60,1}, T_{60,2}\}$ pair, the range of ΔP_0 values for which a double-slope decay is detected by the analysis procedure from Sec. 2.5.1 is therefore evaluated. In addition, the relative errors in the $T_{60,\mathrm{I}}$ and $T_{60,\mathrm{II}}$ estimates are also reported (i.e. corresponding to γ_I and γ_II, respectively).

As in Sec. 4.2.2 (and throughout this work), the late reverberation tails are synthesized as zero-mean Gaussian noise signals subjected to the appropriate decay envelope, given here by Eq. 4.7. In this



(a) $T_{60,2} = 2$ s, $\Delta P_0 \in [6, 34]$ dB.

(b) $T_{60,2} = 3$ s, $\Delta P_0 \in [7, 45]$ dB.

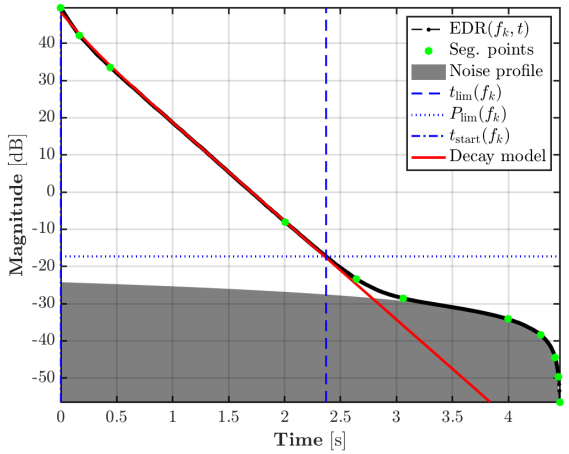
(c) $T_{60,2} = 4$ s, $\Delta P_0 \in [8, 46]$ dB.

(d) $T_{60,2} = 5$ s, $\Delta P_0 \in [9, 53]$ dB.

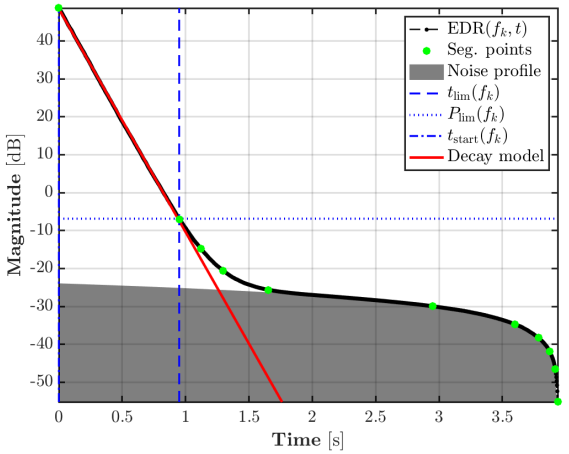
Figure 4.8: Double-slope detection results on simulated coupled volume late reverberation tails. The reverberation time of the first volume (containing both source and receiver) is kept constant at $T_{60,1} = 1$ s while $T_{60,2} = \{2, 3, 4, 5\}$ is increased (a-d). The simulated decay envelopes are generated according to Eqs. 4.7 to 4.9, the diffuse (isotropic and incoherent) late sound fields are synthesized as zero-mean Gaussian noise signals, and an Eigenmike SMA simulation is performed. The solid lines represent the mean relative errors in the estimated broadband coupled reverberation times $T_{60,I}$ and $T_{60,II}$, whose specific errors are given by the dashed lines with upward triangles and dash-dot lines with downward triangles, respectively. Results are displayed over the range of ΔP_0 values for which double-slopes were successfully detected at each given $T_{60,2}$.

case, an SMA measurement is simulated (based once more on the Eigenmike) using a Sloan-Womersley spatial synthesis grid of size $D_{\text{synth}} = 400$ in the same manner as in Sec. 4.1.4. A non-decaying isotropic zero-mean Gaussian noise floor set at -90 dBFS is also included. Since the synthesized reverberation tails are isotropic, the broadband analysis is then simply performed on the omnidirectional component of the HOA-encoded SMA simulation.

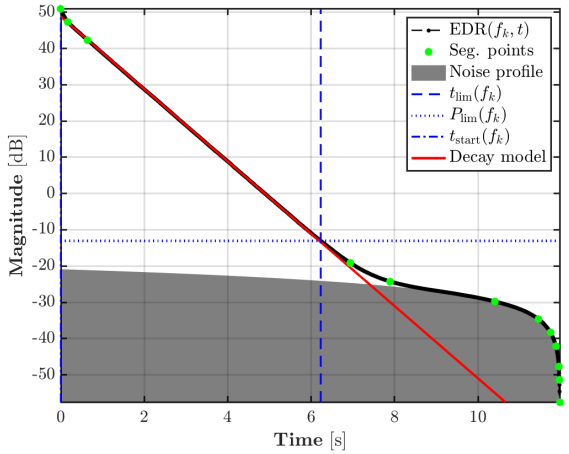
Figure 4.8 shows the results for the relative errors in the $T_{60,I}$ and $T_{60,II}$ estimations over the full ΔP_0 ranges of the double-slope decays detected on the simulations described above (a set constant $T_{60,1} = 1$ s and a varying $T_{60,2} = \{2, 3, 4, 5\}$ s). These error curves clearly demonstrate that the detectable ΔP_0 range generally increases with the $T_{60,2}/T_{60,1}$ ratio (and, therefore, with the coupled $T_{60,II}/T_{60,I}$ ratio as well). In all cases, there appears to be a "sweet spot" near the centre of the detectable ΔP_0 range where both decays are modelled with minimal error. These correspond to examples where the "turning point" is clearly positioned in the middle of the decay (see Fig. 4.9d), allowing the analysis method to



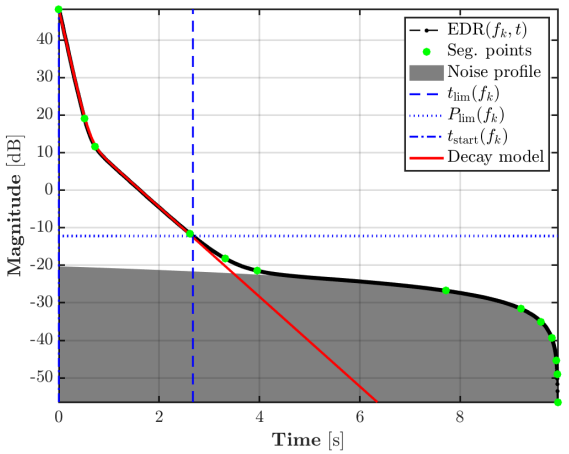
(a) $T_{60,2} = 2$ s, $\Delta P_0 = 6$ dB. Two slopes detected, $T_{60,\text{I}}$ error = 40.2%, $T_{60,\text{II}}$ error = 1.02%.



(b) $T_{60,2} = 2$ s, $\Delta P_0 = 36$ dB. One slope detected.



(c) $T_{60,2} = 5$ s, $\Delta P_0 = 6$ dB. One slope detected.



(d) $T_{60,2} = 5$ s, $\Delta P_0 = 36$ dB. Two slopes detected, $T_{60,\text{I}}$ error = 2.15%, $T_{60,\text{II}}$ error = 3.09%.

Figure 4.9: Broadband EDC analyses on simulated coupled-volume late reverberation tails. The reverberation time $T_{60,1} = 1$ s of the first volume (containing both source and receiver) is kept constant while $T_{60,2}$ is varied [2 s in (a-b), 5 s in (c-d)]. The simulated decay envelopes are generated according to Eqs. 4.7 to 4.9, the diffuse (isotropic and incoherent) late sound fields are synthesized as zero-mean Gaussian noise signals, and an Eigenmike SMA simulation is performed.

fit both slopes with similar precision.

At the edges of the detectable ΔP_0 range, meanwhile, the turning point occurs either too early (Fig. 4.9c), or too late (Fig. 4.9b). In the former case, the first decay is too subtle and only a single slope is able to be modelled. (It would be interesting, as a topic of future research, to examine how these limits of detectability corresponds to that of human perceptibility.) In the latter, the turning point becomes indistinguishable from the EDC's integrated transition to the noise floor, and the analysis either goes back to detecting a single slope ($T_{60,2} = \{2, 3\}$, e.g. Fig. 4.9c), or begins to over-detect an additional decay ($T_{60,2} = \{4, 5\}$).

The specific bounds to the detectable ΔP_0 ranges are in fact determined by the noise floor. Somewhat counter-intuitively, a higher noise floor can actually favour the detection of a double-slope with a small ΔP_0, since the relative decay length of the first slope is no longer negligible with respect to the second. However, higher noise floors also exacerbate the phenomenon of the turning point "disappearing" into the integrated noise profile.

As a final note on the topic, the T_{60} estimation errors also appear to be related to the "prominence"

of their corresponding slope. This ties in to the discussion above on the sweet spots in the detectable ΔP_0 ranges: the longer a given slope is "visible" in the decay, the better the analysis method will be able to fit it. As a result (see Fig. 4.8), the errors in the $T_{60,\mathrm{I}}$ estimates are higher for smaller ΔP_0 values, while conversely the $T_{60,\mathrm{II}}$ estimation errors increase at higher ΔP_0 (the latter can also be seen as the noise profile negatively influencing the model fit).

These results confirm that the multi-slope detection procedure described in Sec. 2.5.1 as part of the larger late reverberation tail analysis framework is indeed capable of discriminating between broadband single and double-slope decays in the sense of classic Cremer-Müller coupled volume theory. Though in theory it is also capable of detecting higher numbers of decay slopes, further work is needed to methodically extend the Cremer-Müller approach to more complex configurations. Additionally, detected double-slope decays are subsequently analyzed over the full frequency spectrum in the EDR analysis (using $C = 2$), but it is as yet unclear how valid or necessary this assumption truly is.

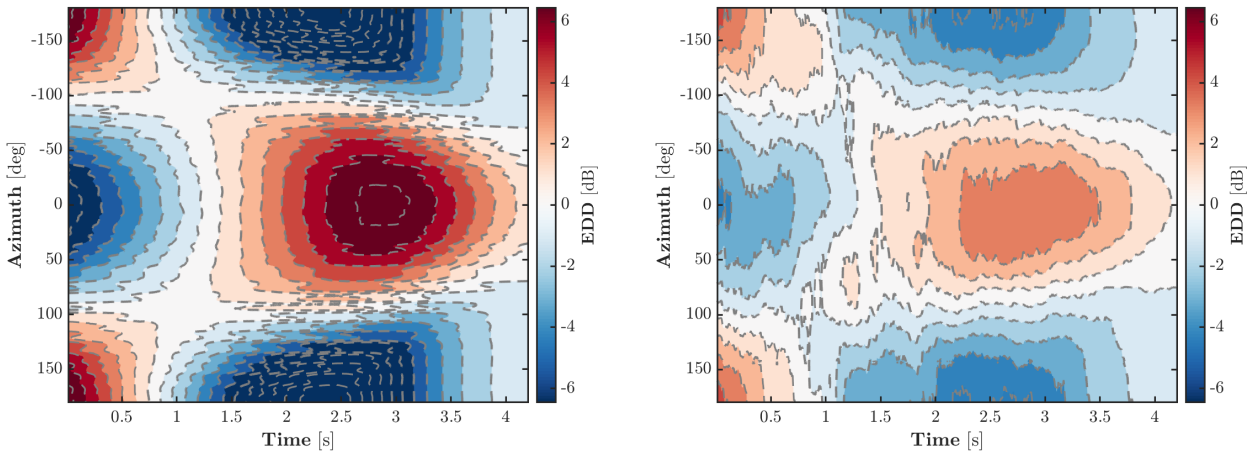
4.2.4/ Evaluating Isotropy Using the EDD

To close out this section on the evaluation of the various analysis methods presented in Ch. 2, the use of the energy decay deviation (EDD) described in Sec. 2.5.2 for the evaluation of late reverberation tail isotropy (or lack thereof) is investigated. To do so, SMA measurements of highly anisotropic reverberation tails are once again simulated using cardioid distributions for both $P_0(\mathbf{\Omega}_d)$ and $T_{60}(\mathbf{\Omega}_d)$, where $\mathbf{\Omega}_d$ is the appropriate spatial synthesis grid for the given SMA, as described in Sec. 4.1.4.

For the current validation test, a $P_0(\mathbf{\Omega}_d)$ distribution with a dynamic range of 18 dB is chosen, while $T_{60}(\mathbf{\Omega}_d)$ is made to vary from 2 s to 4 s (note that the simulation is broadband/frequency-independent, at least in terms of synthesis). In a slight variation on the previous cardioid-based late reverberation tail simulations, the $T_{60}(\mathbf{\Omega}_d)$ distribution is directionally inverted with respect to $P_0(\mathbf{\Omega}_d)$; in other words, the minimum $T_{60} = 2$ s is placed at the maximum P_0 and vice-versa. This design is inspired by the observed behaviour of certain real-world coupled volume configurations: as we will see in the following chapter (Sec. 5.4.1), a directional analysis of these cases can lead to the re-interpretation of a "double-slope" reverberation tail as two separate and distinct single decays.

As can be seen in Fig. 4.10, this simulation design choice also gives a more complex evolution to the EDD. In Fig. 4.10a (left), the EDD is calculated directly on the synthesized spatial signals and interpolated to the azimuthal plane by cubic Hermite spherical interpolation [87]. The Eigenmike SMA measurement simulation is then performed (see Sec. 2.5.2 once again) and the resulting signals are encoded to HOA. The EDD shown in Fig. 4.10b (right) is finally calculated from a DRIR representation generated using the MWDI beamformer, following the conclusions of Sec. 4.1 (and as usual on a size $S = 25$ Fliege-Maier look direction layout grid [79]). To further optimize the representation of directional information, it is averaged over five frequency bins around the MWDI beamformer's maximum directivity frequency $f_{\mathrm{maxDI}} = 2309$ Hz (the underlying EDR having been calculated with 1024-sample windows at a sampling frequency of $f_s = 48$ kHz, this corresponds to the frequency range $f_{\mathrm{EDD}} \in [2203, 2391]$ Hz). The EDD is finally displayed on the azimuthal plane through cubic Hermite spherical interpolation in the same way as Fig. 4.10a. Note that in both cases the EDD is cut off at the noise floor (as detected by the EDR analysis method from Sec. 2.5.1).

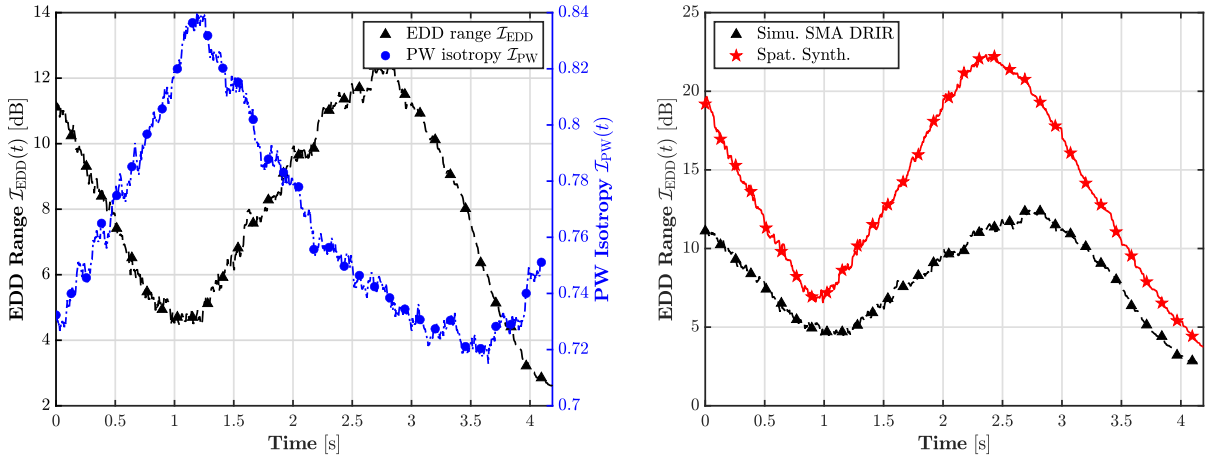
The most important conclusion to draw from Fig. 4.10 is that a DRIR-based EDD can correctly reproduce the temporal evolution of an anisotropic late reverberation tail, though the specific limitations of DRIR generation (see Sec. 4.1) mean that the resulting dynamic range is greatly reduced. To further examine this point, Fig. 4.11 looks at using the EDD's dynamic range as an isotropy measure $\mathcal{I}_{\mathrm{EDD}}$, as proposed in Sec. 2.5.2 (Eq. 2.37, though without the temporal average). Specifically, Fig. 4.11b (right) compares the EDD range measure $\mathcal{I}_{\mathrm{EDD}}(t)$ for both the synthesized spatial signals (red stars) and the simulated SMA measurement DRIR (black triangles): again, the temporal evolution is correctly



(a) Synthesized spatial signal.

(b) Simulated SMA measurement.

Figure 4.10: Azimuthal EDDs for a simulated anisotropic late reverberation tail synthesized using cardioid distributions for both $P_0(\mathbf{\Omega}_s)$ and $T_{60}(\mathbf{\Omega}_s)$. The $P_0(\mathbf{\Omega}_s)$ distribution has an 18 dB dynamic range while $T_{60}(\mathbf{\Omega}_s)$ varies from 2 s at the maximum P_0 to 4 s at its minimum; in other words, the $T_{60}(\mathbf{\Omega}_s)$ distribution is directionally inverted with respect to $P_0(\mathbf{\Omega}_s)$. (a, left) The EDD as calculated directly on the synthesized spatial signals and interpolated to the azimuthal plane by cubic Hermite spherical interpolation. (b, right) The EDD calculated from a DRIR representation of an Eigenmike SMA measurement simulated from the same spatial signals. This DRIR is generated using the MWDI beamformer on a 25-point Fliege-Maier look direction layout grid. The EDD is averaged over five frequency bins around the MWDI beamformer's maximum directivity frequency $f_{\text{maxDI}} \simeq 2309$ Hz ($f_{\text{EDD}} \in [2203, 2391]$ Hz), and is once again shown on the azimuthal plane by cubic Hermite spherical interpolation. Both EDDs are cut off at the noise floor.



(a) EDD range $\mathcal{I}_{\text{EDD}}$ (left axis) vs. PW isotropy $\mathcal{I}_{\text{PW}}$ (right axis).

(b) Spatial synthesis vs. simulated SMA measurement DRIR EDD ranges.

Figure 4.11: Isotropy measures evaluated over the length of the anisotropic reverberation tail described and shown in Fig. 4.10. (a, left) The EDD range measure $\mathcal{I}_{\text{EDD}}(t)$ (Eq. 2.37, Sec. 2.5.2; black triangles) is compared to the PW isotropy measure $\mathcal{I}_{\text{PW}}(t)$ (Eq. 1.26, Sec. 1.2.2; blue dots). Both are calculated on the simulated SMA measurement DRIR. (b, right) Comparison of the EDD range measures $\mathcal{I}_{\text{EDD}}(t)$ as calculated directly on the synthesized spatial signals (red stars) and the simulated SMA measurement DRIR (black triangles again), respectively.

reproduced but with a much reduced dynamic range (around half its maximum).

Figure 4.11a (left), on the other hand, proposes a comparison between the proposed $\mathcal{I}_{\text{EDD}}(t)$ (black triangles again) and the plane-wave isotropy measure $\mathcal{I}_{\text{PW}}(t)$ (Eq. 1.26, Sec. 1.2.2; blue dots) [63], as averaged over all frequencies and calculated over 4192-sample short-term windows, with both

measures calculated on the simulated SMA measurement DRIR. Both measures clearly carry the same information, though the EDD range is inverted (indeed, one could argue that it should rather be referred to as a measure of *an*isotropy).

Interestingly, whereas the PW isotropy measure could be expected to avoid the limitations of the beamformed DRIR, its time-frequency representation is in fact far less clear than the EDD's (see Fig. B.26 in Appx. B.3). For example, much of the initial increase in PW isotropy seems to be carried by frequencies above the SMA's spatial aliasing limit f_{alias}, which somewhat undermines confidence in the measure's pertinence. On the other hand, the effect of spatial aliasing can be clearly interpreted on the time-frequency EDD range (see Fig. B.26b in particular), and it does not affect its dimensional reduction to the time-dependent curve shown in Fig. 4.11. It therefore appears that the EDD is able to provide a much richer characterization of the late reverberation tail's directional properties.

This concludes the series of validation tests used to verify the SRIR analysis methods presented in Ch. 2, covering all aspects from direct sound and early reflection detection (Sec. 4.2.1) to mixing time estimation (Sec. 4.2.2), multiple slope detection (Sec. 4.2.3), and late tail isotropy evaluation (this section, Sec. 4.2.4). Having thus evaluated the performance, discussed the limitations, and optimized the implementation of these analysis methods, we will finish this chapter by investigating one of the main SRIR processing applications to have emerged from the work done in this research project.

4.3/ Treatment Procedures

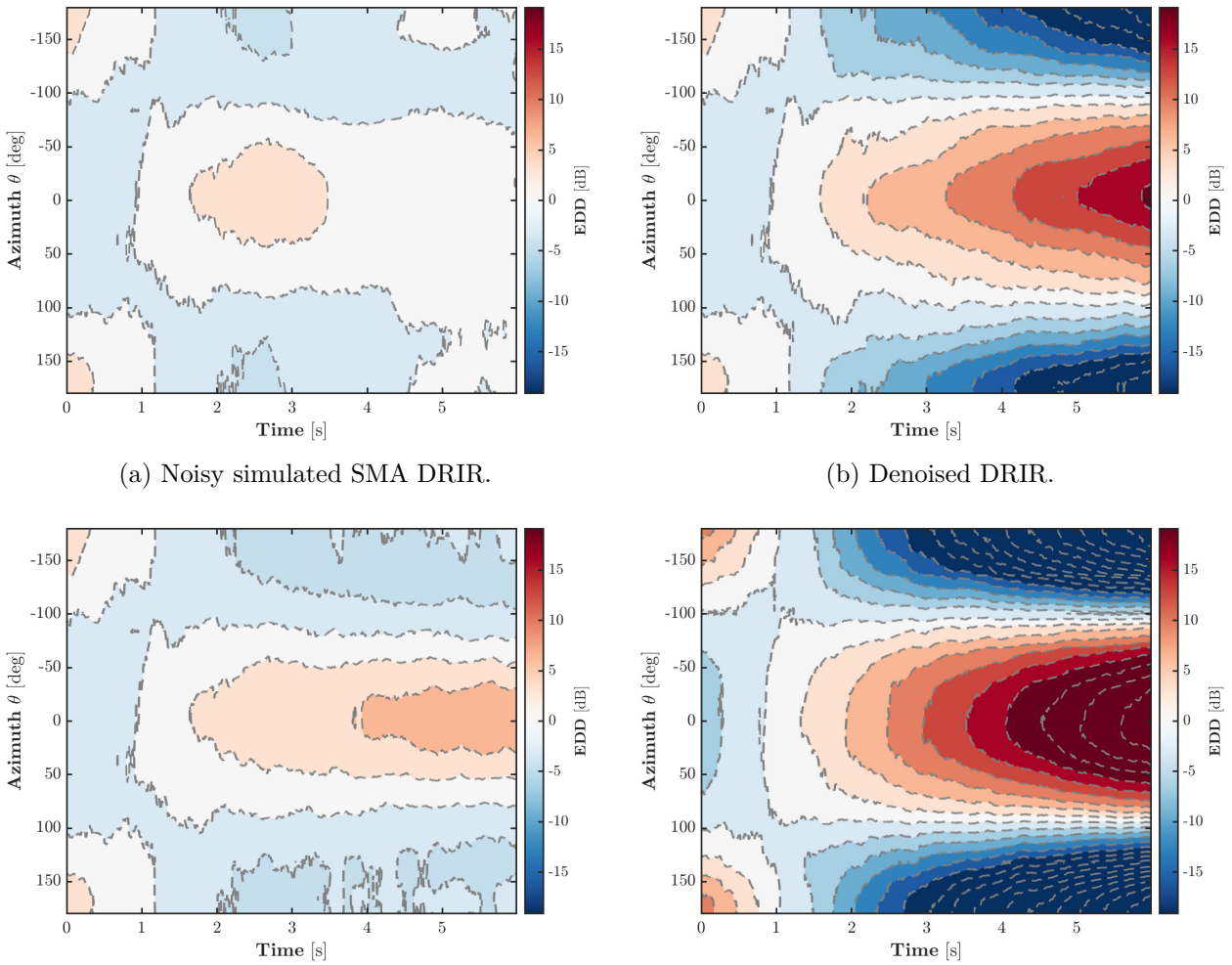
The majority of the processes described in Ch. 3 are either intended to treat certain particularities only encountered in real-world measurements (e.g. the pre-analysis treatment from Sec. 3.1), or are otherwise inscribed in a more creative approach to SRIR manipulation for which the only references (i.e. "ground truths") are the analysis results obtained through the methods validated above. In both cases, "verification" is an ill-defined concept, and a qualitative evaluation of their application on real examples is a far more suitable investigation. For these reasons, demonstrations of the pre-analysis treatment and the various manipulation methods have been relegated to the final chapter.

4.3.1/ Denoising Anisotropic SRIRs

The generalized direction-dependent denoising method presented in Sec. 3.3.1, however, can be tested on the very same type of simulated late reverberation tail as described in the previous section (Sec. 4.2.4). In fact, the SMA SRIR measurement simulation used in this evaluation is the exact same as the previous example: a cardioid $P_0(\boldsymbol{\Omega}_d)$ distribution with a dynamic range of 18 dB and an inverted cardioid $T_{60}(\boldsymbol{\Omega}_d)$ distribution varying from 2 s at the maximum P_0 to 4 s at its minimum.

The denoising procedure is applied directly following the description in Sec. 3.3.1 for general, direction-dependent treatment (i.e. the option chosen when the reverberation tail is known to be anisotropic). The DRIR used here, in contrast to Sec. 4.2.4 above, is obtained with the natural PWD beamformer in order to preserve the incoherence of the "original" reverberation tail (i.e. the part above the noise floor that will not be resynthesized) as much as possible (see further the conclusions from Sec. 4.1.4). The Fliege-Maier look direction layout grid is maintained, on the other hand.

Though the simulated SRIR is synthesized in a broadband manner, the frequency-dependent limitations of the DRIR representation (see Sec. 4.1 once more) imply that the directional properties of the late reverberation tail are only able to be reconstructed accurately in a limited frequency range. This investigation will therefore remain focused on the average results over the five frequency bins surrounding the maximum directivity frequency $f_{\text{maxDI}} \approx 3.45$ kHz for the natural PWD beamformer. The EDR analysis and tail resynthesis processes can be expected to be just as robust over the full frequency spectrum, but there is little reason to verify results pertaining to directional properties at



(a) Noisy simulated SMA DRIR.

(b) Denoised DRIR.

(c) Noiseless SMA DRIR.

(d) Synthesized spatial field.

Figure 4.12: Azimuthal EDDs for (a, top left) the simulated noisy Eigenmike SMA SRIR measurement, (b, top right) the denoised SMA SRIR simulation, (c, bottom left) a noiseless version of the simulated SMA SRIR, and (d, bottom right) the ideal late reverberation tail as originally synthesized in the spatial domain. The late reverberation tail is synthesized using cardioid distributions for both $P_0(\mathbf{\Omega}_s)$ and $T_{60}(\mathbf{\Omega}_s)$. The $P_0(\mathbf{\Omega}_s)$ distribution has an 18 dB dynamic range while $T_{60}(\mathbf{\Omega}_s)$ varies from 2 s at the maximum P_0 to 4 s at its minimum; in other words, the $T_{60}(\mathbf{\Omega}_s)$ distribution is directionally inverted with respect to $P_0(\mathbf{\Omega}_s)$. The DRIRs used to calculate the first three EDDs (a-c) are generated using the natural PWD beamformer on a 25-point Fliege-Maier look direction layout grid. All four EDDs are averaged over five frequency bins around the natural PWD beamformer's maximum directivity frequency $f_{\text{maxDI}} \approx 3.45$ kHz ($f_{\text{EDD}} \in [3328, 3516]$ Hz), and is once again shown on the azimuthal plane by cubic Hermite spherical interpolation.

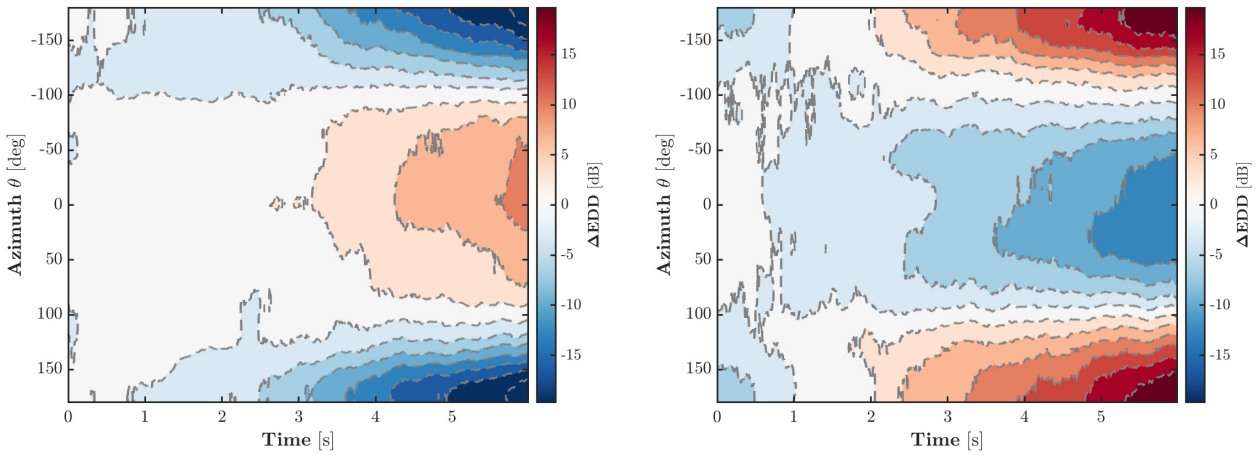
frequencies where these are demonstrably known to be inaccurate.

Figure 4.12 summarizes the four principal DRIR representations obtained through this validation test using their azimuthal EDDs (i.e. their EDDs taken around $f_{\text{maxDI}} \approx 3.45$ kHz and projected to the azimuthal plane by cubic Hermite spherical interpolation). Figure 4.12a shows the simulated noisy mh acoustics Eigenmike SMA SRIR measurement to which the denoising procedure is applied, and Fig. 4.12b subsequently shows the result of this application. In other words, Fig. 4.12a is exactly equivalent to Fig. 4.10b above, though with a different displayed dynamic range (set here to the dynamic range of the denoised EDD), a longer displayed length (most of the noise floor is shown here), and the different choice of beamformer used to generate the DRIR representation (natural PWD here vs. MWDI in Fig. 4.10b).

Figures 4.12c and 4.12d then show the two references that can be used to evaluate the result of the

denoising process: a noiseless version of the SMA SRIR measurement simulation (i.e. equivalent to Fig. 4.12a but without the addition of the non-decaying noise floor) and the ideal late reverberation field as synthesized in the spatial domain $\mathbf{\Omega}_d$ (see Eq. 4.1), respectively.

Several observations can already be made considering Fig. 4.12. The most immediate is that the denoising process produces a DRIR whose space-time properties lie somewhere between the simulated SMA measurement of the noiseless SRIR (Fig. 4.12c) and the ideal synthesized late reverberation tail (Fig. 4.12d). Indeed, as the true dynamic range of the late reverberation tail increases with time, the combined directivity limitations of the SMA measurement, HOA encoding, and DRIR beamforming become more and more significant. This is well illustrated by Fig. 4.12c and the stark contrast it draws to the true synthesized spatial field in Fig. 4.12d. Since the late reverberation tail resynthesis is performed in the DRIR representation, it is not itself affected by these limitations, although the $P_0(\mathbf{\Omega}_s)$ and $T_{60}(\mathbf{\Omega}_s)$ estimates used to parameterize it are (as shown below). As a result of this trade-off, the denoised reverberation tail ends up somewhat closer to the true sound field than the noiseless SRIR, but without ever being able to match it entirely.



(a) EDD error to noiseless SMA DRIR. Avg. abs. error: 3.88 dB (signed max.: −25.0 dB).

(b) EDD error to synthesized spatial field. Avg. abs. error: 4.26 dB (signed max.: 19.5 dB).

Figure 4.13: Azimuthal plane EDD errors (ΔEDD) between the denoised simulated Eigenmike SMA SRIR measurement (Fig. 4.12b) and: (a, left) the noiseless version of the simulated SMA SRIR (Fig. 4.12c), and (b, right) the ideal late reverberation tail as originally synthesized in the spatial domain (Fig. 4.12d). Positive ΔEDD values signify that the denoised EDD was greater than the reference EDD (i.e. the absolute energy deviation, whether itself positive or negative, was greater than that of the reference), and vice-versa.

Figure 4.13 offers a quantitative assessment of these "errors" by showing the differences between the denoised EDD and that of the noiseless SMA SRIR simulation (Fig. 4.13a) and the synthesized spatial sound field (Fig. 4.13b), respectively. These confirm that the denoised SRIR lies somewhere between the two references, as discussed above. Compared to the simulated noiseless SMA measurement (Fig. 4.13a), the denoised EDD has a greater dynamic range as the reverberation tail is prolonged throughout what was originally the noise floor (i.e. negative energy deviations are more negative and vice-versa); hence the errors are of the same sign as the EDD.

On the other hand, the reverse is true when the denoised EDD is confronted to the true spatial field (Fig. 4.13b): the errors are thus of opposite sign to the EDD. In the latter case, there are also differences in the segment before the noise floor, which correspond to the error between the noiseless SMA SRIR simulation and the true sound field (i.e. Fig. 4.12c and Fig. 4.12d). Note that since, prior to the noise floor, the noiseless (Fig. 4.12c), noisy (Fig. 4.12a), and denoised (Fig. 4.12b) SRIRs are all

(more or less) identical, it follows that the error shown in Fig. 4.13a should indeed be negligible in the first part of the reverberation tail[3].

It is also interesting to note that the average and maximum errors are of approximately the same magnitude in both Figs. 4.13a and 4.13b (3.88 dB vs. 4.26 dB and −25.0 dB vs. 19.5 dB, respectively), suggesting that the denoised SMA SRIR's directional dynamic range is almost exactly midway in between the two reference signals. This is further confirmed by looking at the EDD range isotropy measure $\mathcal{I}_{\text{EDD}}$, shown here in Fig. 4.14. Evaluated throughout the different SRIRs (denoised, noiseless, and synthesized), the EDD range $\mathcal{I}_{\text{EDD}}(t)$ takes on the same initial form as in the previous section (Fig. 4.11b, Sec. 4.2.4), but without the later influence of the isotropic noise floor. The differences between the three SRIRs in terms of the evolution of their EDD's dynamic range over the full length of the reverberation tail thus become apparent.

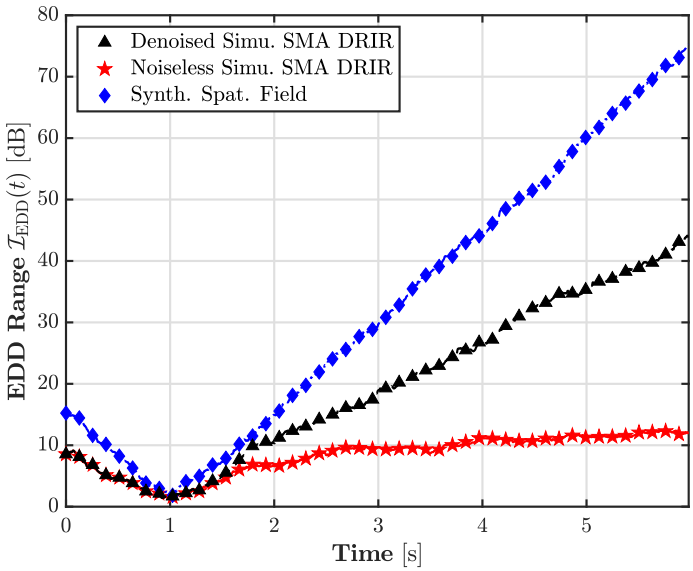


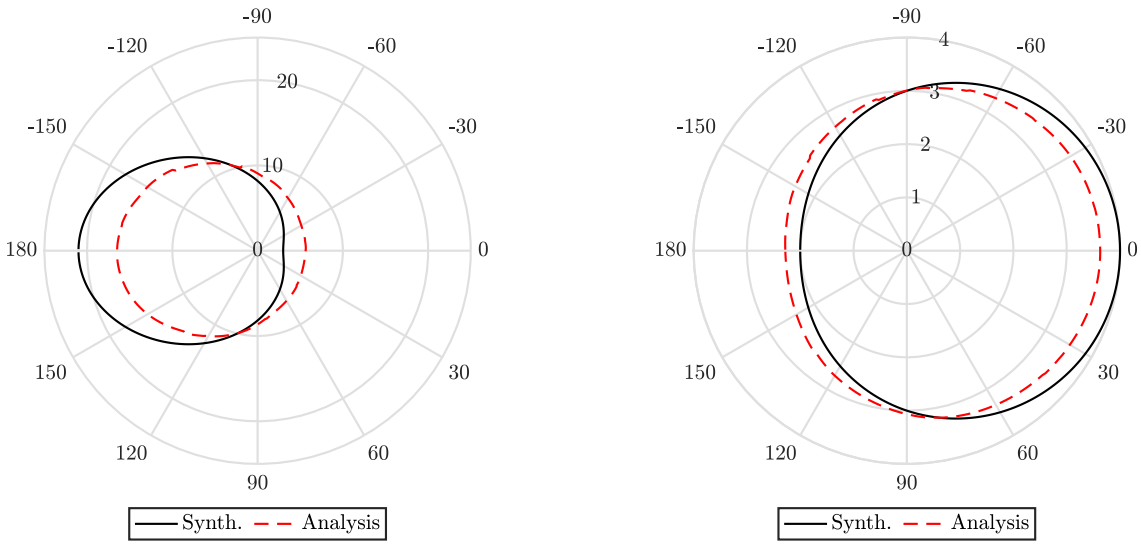
Figure 4.14: EDD range isotropy measures $\mathcal{I}_{\text{EDD}}(t)$ shown throughout the denoised SMA SRIR measurement simulation (black triangles), the noiseless reference simulation (red stars), and the synthesized spatial sound field (blue diamonds) of a highly anisotropic late reverberation tail, as calculated from the EDDs shown in Fig. 4.12 (i.e. not over the full frequency spectrum as in Fig. 4.11).

The fact that the denoised EDD cannot fully match that of the true synthesized sound field is itself mostly due to the limitations imposed by the SMA measurement, HOA encoding, and DRIR generation on the directional properties of the noisy SRIR, upon which the late reverberation tail's decay properties are subsequently estimated. In other words, the reduced directional dynamic range of the SMA SRIR measurement (and, crucially, of its time evolution, as evidenced by Fig. 4.14) is inherently reflected in the directional EDRs on which the modelling analysis is performed, and thus inevitably compounded onto the $P_0(\mathbf{\Omega}_s)$ and $T_{60}(\mathbf{\Omega}_s)$ estimates.

The specific $P_0(\mathbf{\Omega}_s)$ and $T_{60}(\mathbf{\Omega}_s)$ estimation results obtained for the current example are given in Fig. 4.15, in comparison to the values used to synthesize the ideal spatial late reverberation field. Once again, these are averaged over the five frequency bins surround the maximum directivity frequency $f_{\text{maxDI}} \approx 3.45$ kHz and covering the frequency band $f_{\text{EDD}} \in [3328, 3516]$ Hz, and projected onto the azimuthal plane for display by cubic Hermite spherical interpolation [87].

These results further support the interpretation developed throughout this section: the direction-dependent denoising procedure can correctly reconstruct the general directional properties of an anisotropic reverberation tail, but total accuracy is limited by the directivity of the DRIR representation.

[3] Due to of the frequency-dependent nature of the EDR analysis, and notably the noise floor fitting, some bins may have been resynthesized starting earlier or later than others, thereby resulting in the slight differences observed in Fig. 4.13a.



(a) Initial power P_0.
Avg. error: 1.67 dB (max.: 4.88 dB).

(b) Reverberation time T_{60}.
Avg. error: 6.85 % (max.: 15.8 %).

Figure 4.15: Azimuthal plane projections (by cubic Hermite spherical interpolation) of the EDR analysis $P_0(\mathbf{\Omega}_s)$ and $T_{60}(\mathbf{\Omega}_s)$ estimation results, averaged over $f_{\text{EDD}} \in [3328, 3516]$ Hz. Polar plot radii represent (a, left) dB and (b, right) seconds.

More specifically, it fails to match the true directional dynamic range of the measured sound field. The exact error values as reported in Fig. 4.15 depend on the degree of anisotropy in the underlying distributions: the more anisotropic the tail, the more the analysis method will struggle to model it accurately. As such, these errors are simply included here for reference.

Finally, it should be noted that the results discussed up to this point have only dealt with energetic aspects of the denoising procedure. However, as justified at the start of this section, the natural PWD beamformer has specifically been chosen for DRIR generation in this case because it offers the greatest overall incoherence preservation (see once again the results of Sec. 4.1.4). We therefore end this evaluation with Fig. 4.16, an overview of the spatial incoherence profiles $\psi_{\text{dir}}(t)$ for each of the four simulated SRIR representations described in Fig. 4.12.

The limitations of the DRIR representation are reflected once more in these incoherence profiles. As we know from Sec. 4.1.4, the incoherence measure ψ_{dir} calculated on a DRIR generated with the natural PWD beamformer is sensitive to changes in isotropy. This can be seen clearly in the noiseless DRIR's incoherence profile (blue diamonds), which steadily decreases throughout the late reverberation tail just as its directional dynamic range increases (see Fig. 4.14, and note also that the incoherence profiles are calculated over all frequencies and not just the most directive ones as is the case for the EDDs).

The same phenomenon is observed in the first part of the noisy DRIR (green circle), as should be expected, but the presence of the isotropic diffuse noise floor raises incoherence in the latter part of the signal. The ideal synthesized sound field, meanwhile, remains maximally incoherent throughout the SRIR (black pentagrams), regardless of changes in isotropy: this is simply due to the directional signals being synthesized as independent zero-mean Gaussian noise sequences, thereby perfectly fulfilling the condition of Eq. 2.16 (Sec. 2.1.3) with respect to the diagonality of its covariance matrix.

The denoised DRIR's incoherence profile (red hexagrams) then navigates between these two behaviours as its signal content shifts from the "original" SMA SRIR simulation, with its aforementioned limitations, to a resynthesized reverberation tail presenting the same stochastic properties as the ideal field (since each DRIR channel is resynthesized independently, its covariance matrix will also be diagonal). Most importantly, Fig. 4.16 demonstrates that the denoising procedure preserves the

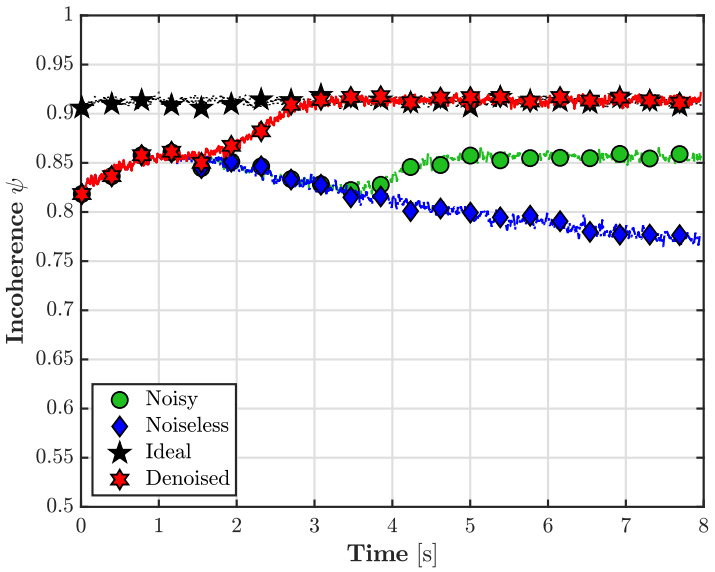


Figure 4.16: Spatial incoherence profiles $\psi_{\text{dir}}(t)$ for each of the four different representations of the SRIR simulation as described in Fig. 4.12: noisy SMA SRIR measurement simulation (green circles), noiseless SMA SRIR measurement simulation (blue diamonds), denoised SMA SRIR measurement simulation (red hexagrams), and ideal/synthesized spatial sound field (black pentagrams).

fundamental spatial incoherence conditions of the late reverberation tail.

In closing, we should reiterate that the results presented throughout this section have been restricted to the maximally directive frequency band $f_{\text{EDD}} \in [3328, 3516]$ around the $f_{\text{maxDI}} \approx 3.45$ kHz (except for the incoherence profiles above, which are broadband). Further work is therefore necessary to improve the directional analysis, not only in terms of its absolute performance but also with respect to its behaviour and regularity over the entire frequency spectrum (though there are some fundamental limitations that cannot be fully avoided). This approach could also be complemented by an investigation into the degree of reproduction accuracy that would be required for the reconstructed late reverberation tail to be perceptually indistinguishable from the true spatial sound field.

These results, and especially the fact that the DRIR resynthesis allows us to overcome the limitations of the SMA SRIR measurement, further suggest that it could be interesting to consider modelling the late reverberation tail on as short a portion of the decay as possible, in order to then resynthesize a much greater portion of the late sound field (compared to "simply" denoising as proposed here). As a somewhat counter-intuitive illustration of this, it can be seen from Figs. 4.12 and 4.14 that modelling the late tail's decay on the noiseless DRIR would actually have been less accurate than the results obtained from the noisy simulation. The questions of which segment to model and how short is too short or just short enough are non-trivial, however, and merit their own research effort.

This chapter has provided a comprehensive series of validations tests that have been used to verify, evaluate, and parameterize the major analysis and treatment methods presented throughout Chs. 2 and 3. First, in Sec. 4.1, came an investigation of the DRIR representation that underpins many (if not all) the subsequent analysis and treatment methods (as well as the manipulation techniques to be seen in the final chapter). The use of a Fliege-Maier spherical point grid as a look direction layout was settled upon, and the various advantages and disadvantages of different beamformer design choices was discussed.

Next, a large study on the detection of early reflections was presented in Sec. 4.2.1 (including a smaller subsection on the direct sound), mainly comparing the application of the two DoA estimation methods described in Sec. 2.4 (SRP and regularized under-determined Herglotz inversion). These tests made use of simulated SMA measurements of SRIRs consisting of discrete IS-modelled reflections. As a natural follow-up, the mixing time estimation algorithm was subsequently evaluated in Sec. 4.2.2 using similar simulations to which incoherent late reverberation tails had been matched and added.

Finally, three tests were dedicated to the analysis and treatment of the late reverberation tail: Sec. 4.2.3 was dedicated to the detection and modelling of multi-slope decays such as those encountered in certain coupled-volume configurations, Sec. 4.2.4 investigated the use of the EDD in the evaluation of late tail isotropy (or lack thereof), and Sec. 4.3.1 above validated the denoising procedure presented in Sec. 3.3.1. These tests were all performed on simulated SMA measurements of late reverberation tails synthesized in the spatial domain as zero-mean Gaussian noise sequences.

Having thus been verified, these analysis and treatment methods can now be applied to real-world SMA SRIR measurements. Some examples of these applications are given in the next (and final!) chapter, along with preliminary "proof of concept" illustrations of the manipulation techniques they subsequently help make possible.

---

5 / Applications to Real-World Measurements

This final chapter aims to comprehensively demonstrate the analysis, treatment, and manipulation methods described throughout Chs. 2 and 3 (and referred to in the very title to this thesis) through their application to various examples of "real-world" SRIRs measured by the mh acoustics Eigenmike® SMA. The measured spaces cover a wide range of acoustic phenomena and reverberation characteristics, from coupled-volume configurations to semi-open spaces and obstructed source-receiver paths.

Although the validation tests performed on simulated examples in the previous chapter (Ch. 4) provide a crucial verification of the underlying signal models and analysis methods, the only true "ground truth" available in the context of the complete analysis-treatment-manipulation framework we are attempting to define is in the evaluation of its operation on such real-world cases. As such, this chapter covers every aspect presented so far, beginning with the pre-analysis treatment procedure (Sec. 5.1) and the mixing time estimation algorithm (Sec. 5.2).

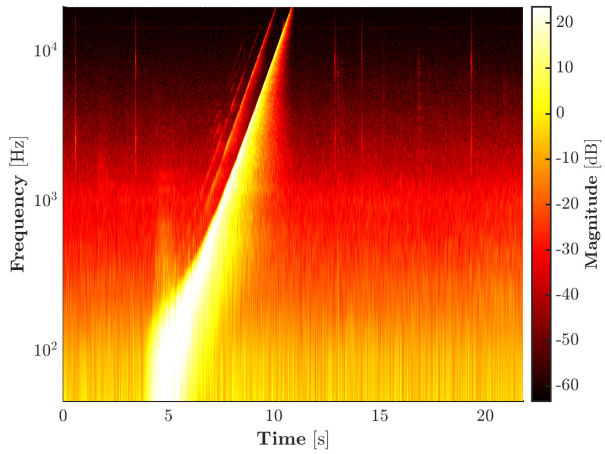
Then, following the usual structure of this dissertation, the two main sections in this chapter deal respectively with the detection of early reflections (Sec. 5.3) and the properties of directional late reverberation tails (Sec. 5.4). The former includes the detection and characterization of the direct sound (Sec. 5.3.1), the generation of space-time echo distribution cartographies (Sec. 5.3.2), and the subsequent modifications thus enabled (namely reflection salience manipulation and echo redistribution, Sec. 5.3.3). The late reverberation segment begins with a discussion on coupled-volume measurements and their manifestation as either highly isotropic double-slope decays or directionally separated anisotropic single decays (Sec. 5.4.1), before an example of the directional denoising process is given, as applied to a severely anisotropic SRIR (Sec. 5.4.2). The last two section then deal directly with late reverberation isotropy, first with respect to its measurement and evaluation (Sec. 5.4.3), and second in terms of modifying it (Sec. 5.4.4).

Finally, it should be noted that although this chapter already contains a relatively large collection of figures, many more are included in Appcs. B.4 to B.7 in order to complement the ones presented.

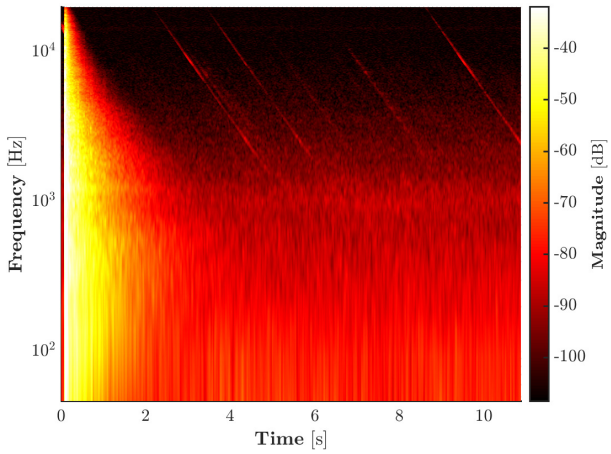
5.1 / Pre-Analysis Treatment Performance

As stated in the introduction to Sec. 4.3 in the previous chapter, it is only truly appropriate to verify the pre-analysis treatment procedure described in Sec. 3.1 on examples of real-world SMA SRIR measurements presenting the very type of non-stationary impulsive noises it was designed to address. The aim of this section is therefore to study its application to such an example and highlight the regularizing effect it can have on the subsequent EDR analysis.

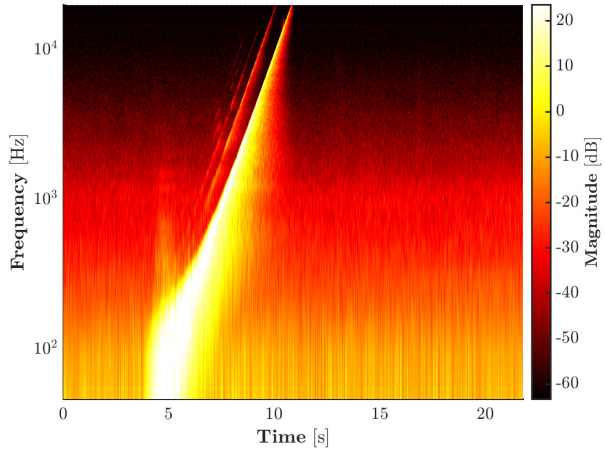
Figure 5.1 thus illustrates the application of the artefact reduction process to a single microphone channel of an ESM SRIR measurement performed at the Grandes Serres de Pantin in Pantin, France, using the 32-capsule mh acoustics Eigenmike SMA (the Grandes Serres is a mid 20$^{\text{th}}$-century industrial site formerly operated by the metalworking company Pouchard and composed of several large adjacent hangars). Figures 5.1a to 5.1d represent auto-PSDs obtained by averaging over the square magnitudes of 16 STFT frames; the underlying STFT uses 1024-sample Nuttall windows with 87.5% overlap at a $f_s = 48$ kHz sampling rate, and so the total PSD window duration is of 61.3 ms. Figures 5.1e and 5.1f



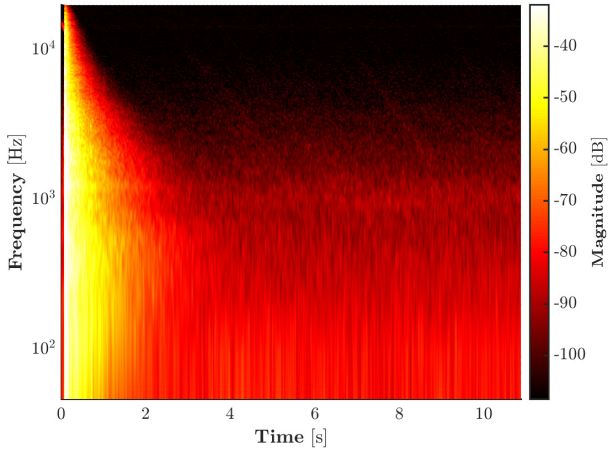
(a) Original "raw" ESM measurement.



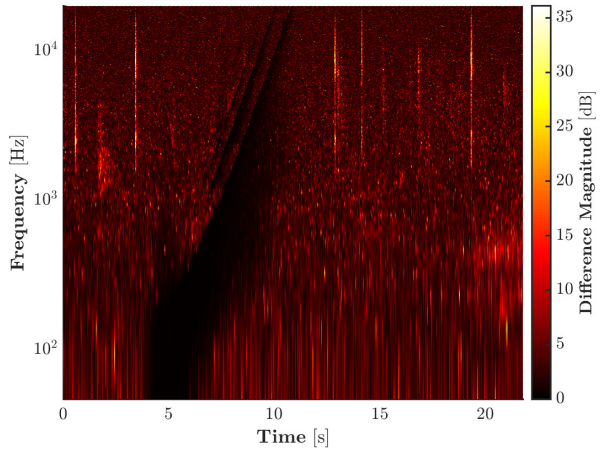
(b) Original "deconvolved" IR.



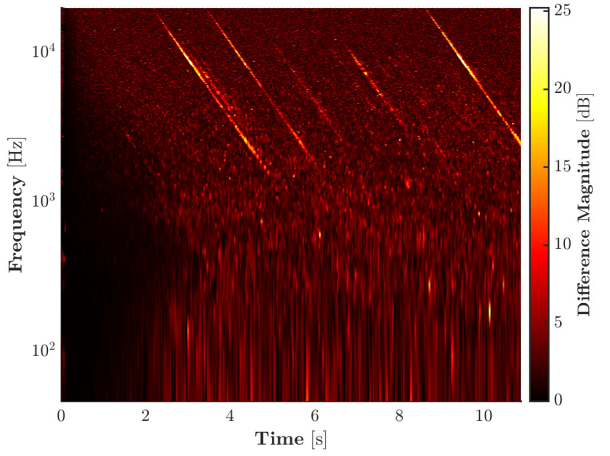
(c) ESM measurement after artefact reduction.



(d) Deconvolved IR after artefact reduction.



(e) ESM measurement difference.



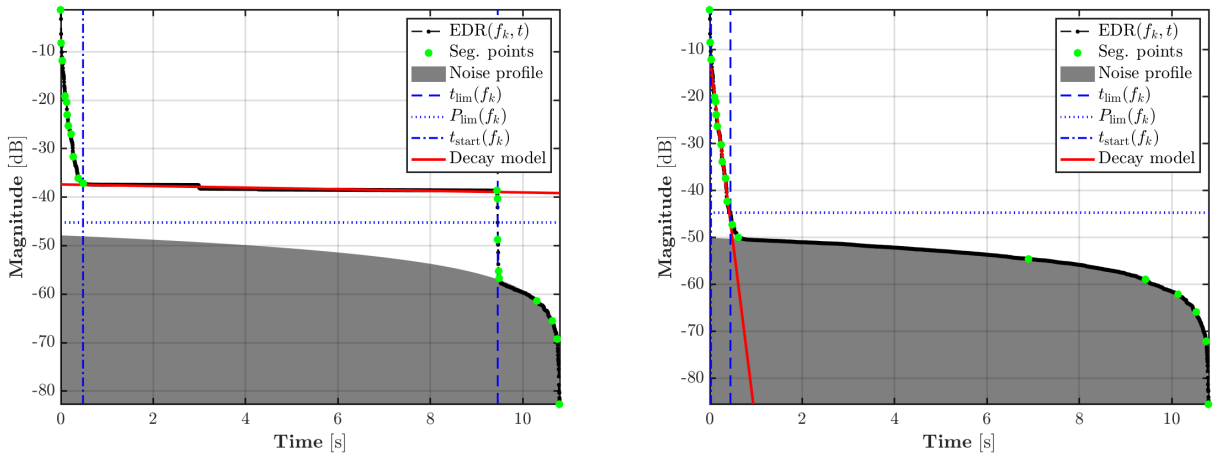
(f) Deconvolved IR difference.

Figure 5.1: Artefact reduction applied to a single microphone channel of an ESM measurement performed at the Grandes Serres de Pantin in Pantin, France, using the mh acoustics Eigenmike SMA. (a, top left) Auto-PSD of the original "raw" ESM measurement signal (averaged over five repetitions), presenting several impulsive parasite sounds. (b, top right) Auto-PSD of the original RIR obtained by inverse-sweep convolution with the raw ESM measurement signal in (a) (i.e. without artefact reduction applied). (c, centre left) Auto-PSD of the ESM measurement signal after artefact reduction. (d, centre right) Auto-PSD of the RIR obtained by inverse-sweep convolution with the artefact-reduced ESM measurement signal in (c). (e, bottom left) PSD difference between (a) and (c). (f, bottom right) PSD difference between (b) and (d).

are then dB-scale differences between two PSDs (i.e. linear-scale ratios).

In Fig. 5.1a, several impulsive artefacts can be seen occurring over the course of the ESM measurement signal, as averaged over five sweep repetitions; Fig. 5.1b then shows how these parasitic impulses are transformed into repeated inverse-sweep artefacts when the ESM measurement is convolved with the time-reversed and amplitude-corrected excitation signal to generate an RIR, as per Farina [20]. Figures 5.1c and 5.1d demonstrate the effect of the artefact reduction procedure on the raw ESM measurement and resulting RIR, respectively, and finally Figs. 5.1e and 5.1f serve to further highlight the time-frequency points corresponding to the non-stationary noise artefacts by displaying only the magnitude differences between the untreated and treated signals.

The removal of impulsive measurement noises and the subsequent reduction in inverse-sweep-type artefacts as revealed in Fig. 5.1 is crucial in ensuring that the reverse-integration of the RIR's noise floor approaches the theoretical profile fitted to each EDR bin in order to identify the noise floor limiting point $\{P_{\text{lim}}, t_{\text{lim}}\}(f, \mathbf{\Omega}_s)$ (see Fig. 2.7, Sec. 2.5.1). To illustrate this, Fig. 5.2 compares the results obtained by applying the late reverberation decay modelling process from Sec. 2.5.1 to a single bin of the Grandes Serres de Pantin DRIR's EDR, first without (Fig. 5.2a, left) and then with (Fig. 5.2b, right) the artefact reduction pre-analysis treatment.



(a) EDR bin modelling *without* artefact reduction.

(b) EDR bin modelling *with* artefact reduction.

Figure 5.2: Late reverberation decay modelling applied to a single bin of the Grandes Serres de Pantin DRIR's EDR, at the look direction $\mathbf{\Omega}_s = (-18.5°, 48.6°)$ and frequency $f_{180} = 8391$ Hz: (a, left) *without* applying artefact reduction, and (b, right) *with* artefact reduction applied (see also Fig. 5.1).

Not only is the SNR greatly increased, but the "accident" in the curve (i.e. the deviation from the theoretical reverse-integration of a constant power) caused by the non-stationarity of the inverse-sweep artefact is entirely removed. This can have a critical regularizing effect on the late reverberation analysis process (see Sec. 2.5.1 again), as evidenced by Figs. 5.2a and 5.2b: without artefact reduction, the noise floor is fitted far below where it should be, which has dire consequences on the decay modelling (the non-stationary noise risks getting identified as the exponentially decaying late reverberation tail).

In an attempt to quantify the amount of artefact reduction that has been performed, we can define an "artefact-to-total-energy ratio" (ATE) as the total artefact energy removed by replacing detected outlying points with the mean magnitude value (according to the description given in Sec. 3.1) versus the total signal energy in a given frame:

$$\text{ATE}_q(t) = \frac{\sum_{k=0}^{K} \left|\widetilde{H}_q(f_k,t)\right|^2}{\sum_{k=0}^{K} \left|H_q(f_k,t)\right|^2}, \text{ with } \widetilde{H}_q(f_k,t) = \begin{cases} H_q(f_k,t) - \mu_q(f_k,t), & |H_q(f_k,t)| > \xi_q(f_k,t) \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

Thus $\mathrm{ATE}_q(t)$ represents the total relative amount of energy removed in each time frame during artefact reduction. In the current example (the Grandes Serres SRIR Eigenmike channel shown in Fig. 5.1), this measure averaged to $\overline{\mathrm{ATE}}_q = 12.6\%$ or -8.98 dB over the five sweep repetitions.

The artefact reduction pre-analysis treatment procedure is therefore a highly effective tool that allows the use of repeated ESM sequences in SRIR measurement[1] to be fully exploited by minimizing its major drawbacks (increased danger of exposure to non-stationary noise events as measurement time lengthens) while maintaining its primary advantage (increased SNR by coherent summing of measured signal and incoherent summing of background noise).

5.2/ Mixing Time Estimation

In this section, we present a real-world evaluation of the mixing time estimation algorithm (Sec. 2.3) through a few examples of its application to SRIRs measured in a variety of different acoustic spaces, each with their own distinctive characteristics. By comparing the results obtained from these examples to the simulated SRIR study in Sec. 4.2.2, the real-world applicability of the mixing time estimation procedure can therefore be evaluated despite the lack of verifiable "ground truths".

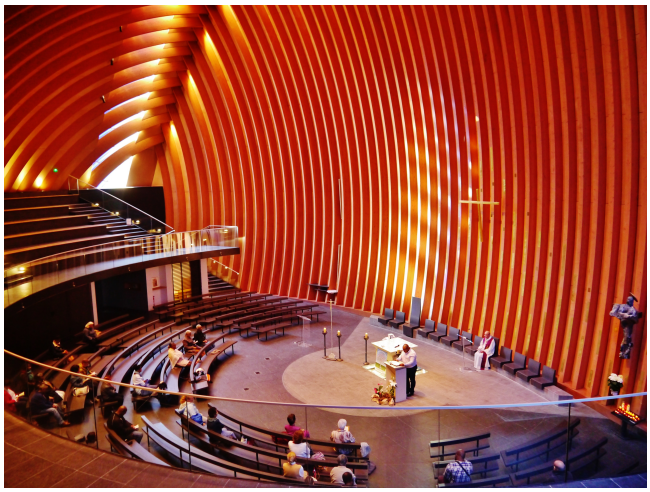
The use of both the SH-domain CoMEDiE diffuseness measure ψ_{cov} and the DRIR-based spatial incoherence measure ψ_{dir} will additionally be compared. As opposed to the simulation-based validation test of Sec. 4.2.2 (which used the MWDI design), the DRIR representation used here is generated with the natural PWD beamformer in order to obtain a more sensitive incoherence profile[2]. Furthermore, real-world SRIRs are very rarely as extremely anisotropic as the examples synthesized in the previous chapter; in the rare cases where such anisotropy is observed, the MWDI beamformer can always be used for a secondary ψ_{dir} estimation.

The first such example is an SRIR measured in the modern Notre-Dame Cathedral in Créteil, France, and the results of its mixing time estimation are summarized in Fig. 5.3. A view of the measured space is provided in Fig. 5.3a (top left), along with the broadband omnidirectional $[H_{0,0}(f,t)]$ decay analysis used to estimate a preliminary noise floor limiting time (Fig. 5.3b, top right). The hemispherical volume favours the appearance of focalization and flutter echo effects, which can be seen in the opening ~ 200 ms of both the CoMEDiE (Fig. 5.3c, bottom left) and spatial incoherence (Fig. 5.3c, bottom right) profiles. However, the SH-domain covariance measure $\psi_{\mathrm{cov}}(t)$ reveals the presence of two further "notches", likely corresponding to late-arriving echoes overlapping with the late reverberation tail. These notches lead to a re-segmentation of the diffuseness profile and a final estimated $\tilde{t}^{\mathrm{cov}}_{\mathrm{mix}} = 403$ ms. On the other hand, the effect of these echoes on the spatial incoherence measure $\psi_{\mathrm{dir}}(t)$ is negligible, which results in an under-estimated $\tilde{t}^{\mathrm{dir}}_{\mathrm{mix}} = 187$ ms.

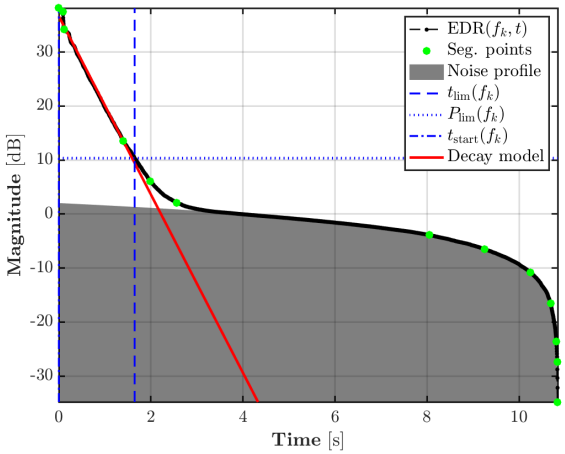
The next example is an SRIR measured in the cloister at the Dominicains de Haute-Alsace cultural centre and former convent in Guebwiller, France, for which the results are given in Fig. 5.4. This space is further explored in Secs. 5.4.2 to 5.4.4 due to its unique characteristics as a semi-open volume: the grassy inner garden (20×23 m) is open to the sky while the surrounding, 3 m-wide and partially enclosed stone walkway is highly reverberant (see Figs. 5.4a and 5.4b). This SRIR demonstrates that it is in fact possible to achieve incoherent late reverberation conditions in such semi-open spaces (and in this case with both source and receiver in garden). It also gives another example of the DRIR ψ_{dir} measure (Fig. 5.4d, bottom right) providing an under-estimation of t_{mix} compared to the SH-domain ψ_{cov} (Fig. 5.4c, bottom left), though the underlying reason appears to be slightly different: whereas in Fig. 5.3d it was clearly the measure itself that was not sensitive enough to the incoming reflections, it seems plausible that an additional adaptive RDP segmentation step could "catch" the notch occurring

[1] Or simply RIR measurement in general, for that matter, since the operation is performed microphone-by-microphone.

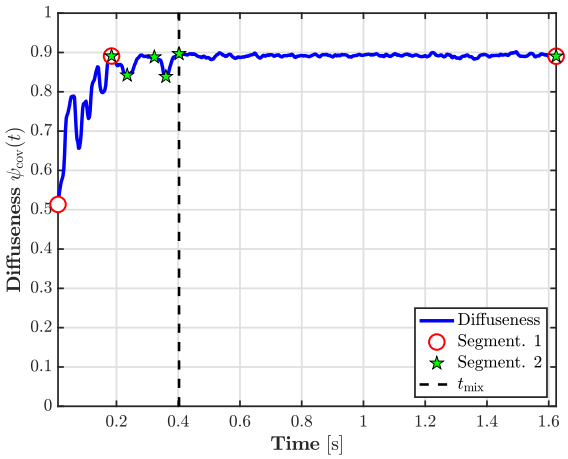
[2] Due to the presence of background noise and semi-stochastic cluster reverberation effects, perceptually noticeable discrete reflections may not appear as explicitly in incoherence profiles as the IS-simulated echoes from Ch. 4.



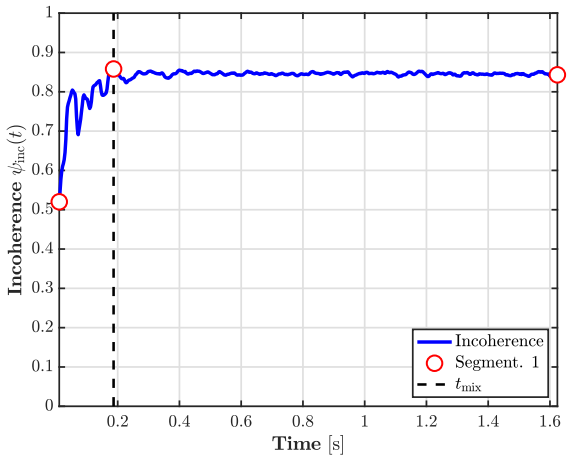
(a) View of the measured space in the Notre-Dame Cathedral, Créteil, France.



(b) Broadband $H_{0,0}(f,t)$ decay analysis, $\widetilde{T}_{60} = 3.64$ s.



(c) SH-domain CoMEDiE diffuseness ψ_{cov}, $\tilde{t}^{\text{cov}}_{\text{mix}} = 403$ ms.



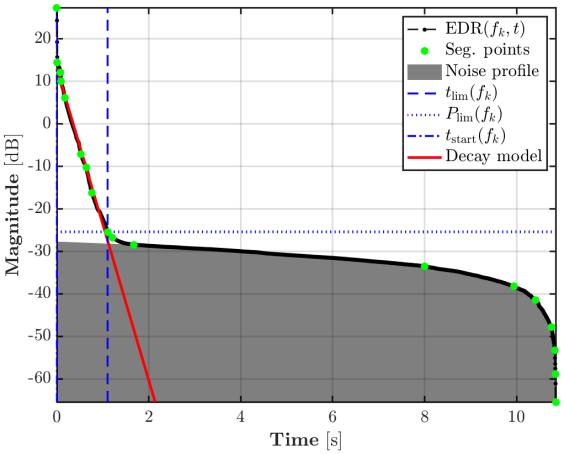
(d) DRIR spatial incoherence ψ_{dir}, $\tilde{t}^{\text{dir}}_{\text{mix}} = 187$ ms.

Figure 5.3: Mixing time estimation for an SRIR measured in the Notre-Dame Cathedral in Créteil, France, with a 32-capsule mh acoustics Eigenmike SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4$^{\text{th}}$-order HOA. The DRIR representation used in (d) is obtained with a natural PWD beamformer on a 25-point Fliege-Maier look direction grid.

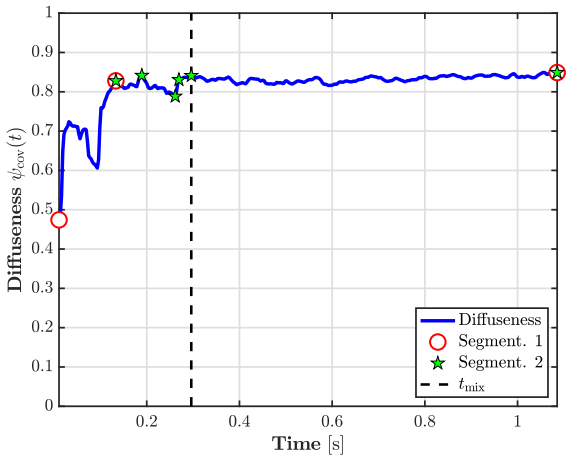
just before 300 ms in both Figs. 5.4c and 5.4d. In other words, perhaps the re-segmentation routine within the mixing time estimation algorithm (see Sec. 2.3) should be iterative.

In general, when compared to the simulated SRIRs from Sec. 4.2.2 (diffuseness and incoherence profile figures in Appx. B.2), the real-world examples shown here all possess significantly longer late reverberation tails. This is mostly due to the type of room originally simulated by the EVERTims auralization engine [98], as well as the approach chosen to fit, match, and add synthesized reverberation tails to the IS-SRIRs (see the full simulation description in Sec. 4.2.2). Additionally, and as mentioned above, the simulated anisotropic reverberation tails were purposefully created to have extreme directional energy variations even at frequencies where such anisotropy is not possible in an SMA measurement. Regardless of these differences, the diffuseness and incoherence profiles as well as the different steps of the estimation process remain very similar in appearance. The simulated examples can therefore be considered sufficiently accurate facsimiles of real-world SMA SRIR measurements, which in turn further verifies the results of the validation tests in Sec. 4.2.2 and by extension the performance of the mixing time estimation method itself.

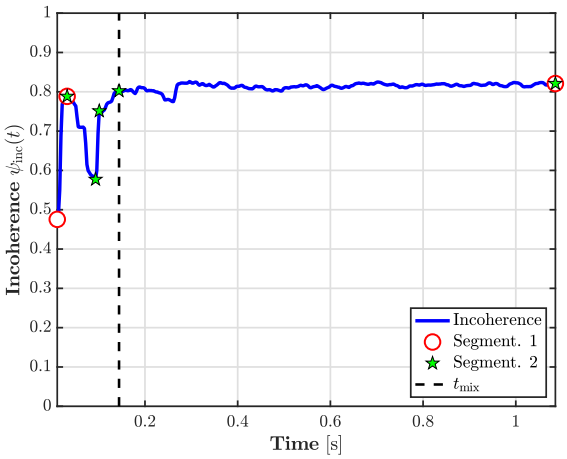
(a) View of the measured convent cloister garden at the Dominicains de Haute-Alsace, Guebwiller, France.



(b) Broadband $H_{0,0}(f,t)$ decay analysis, $\widetilde{T}_{60} = 1.63$ s.



(c) SH-domain CoMEDiE diffuseness ψ_{cov}, $\tilde{t}^{\text{cov}}_{\text{mix}} = 296$ ms.



(d) DRIR spatial incoherence ψ_{dir}, $\tilde{t}^{\text{dir}}_{\text{mix}} = 144$ ms.

Figure 5.4: Mixing time estimation for an SRIR measured in the cloister at the Dominicains de Haute-Alsace cultural centre and former convent in Guebwiller, France, with a 32-capsule mh acoustics Eigenmike SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4$^{\text{th}}$-order HOA. The DRIR representation used in (d) is obtained with a natural PWD beamformer on a 25-point Fliege-Maier look direction layout grid.

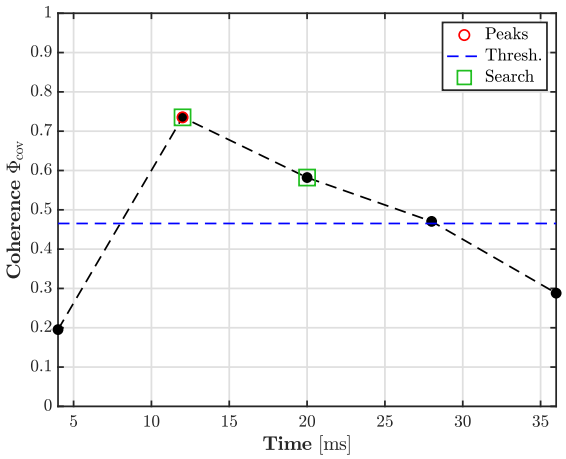
5.3/ Early Reflection Detection

This section provides two brief examples highlighting the capabilities of the early reflection analysis tools described in Sec. 2.4 (and further studied in Sec. 4.2.1) when applied to real-world SMA SRIRs.

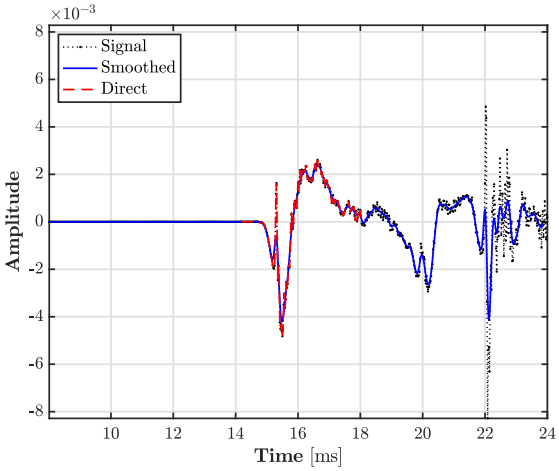
5.3.1/ Direct Sound

To demonstrate the direct sound detection algorithm described in Sec. 2.4.1, an SRIR measured in the Church of St. Eustache, in the 1$^{\text{st}}$ *arrondissement* of Paris, is selected as an example as it presents the potentially pathological case of having "late" early reflections with greater energy than the direct sound itself. This phenomenon can arise in certain source/receiver configurations where the direct sound's path is obstructed; in fact, there may be some ambiguity as to whether this truly represents a "direct sound" since, if the path is truly blocked, the SMA would first receive a (low-order) reflection. Without additional information, it is impossible to determine whether the first observed impulse is an echo or a filtered/weakened version of the direct sound. In any case, the framework defined in this

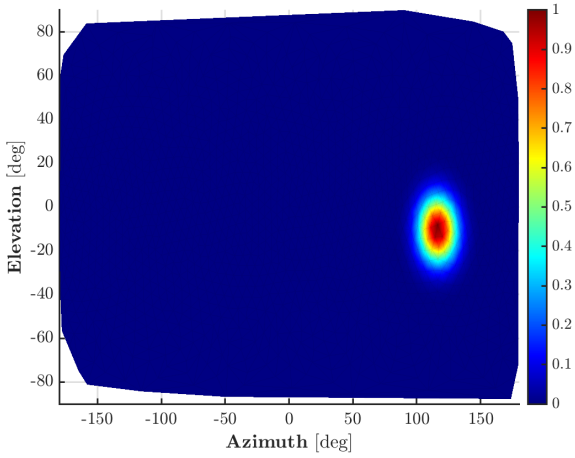
(a) View of the measured space in the Church of St. Eustache, Paris, France.



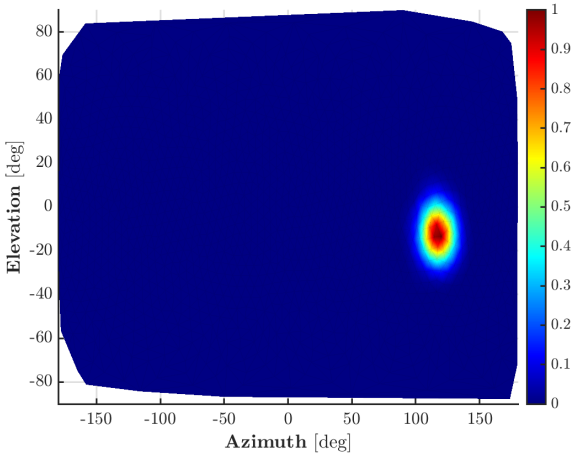
(b) SH-domain (CoMEDiE) coherence, $\Phi_{\text{cov}} = 1 - \psi_{\text{cov}}$.



(c) Direct sound signal detection on $H_{0,0}(t)$, $\tilde{t}_0 \simeq 15.5$ ms.



(d) SRP DoA estimation, $\widetilde{\mathbf{\Omega}}_{\text{DS}}^{\text{SRP}} = (116°, -8.14°)$.



(e) Herglotz inversion DoA estimation, $\widetilde{\mathbf{\Omega}}_{\text{DS}}^{\text{Herg}} = (120°, -13.4°)$.

Figure 5.5: Direct sound detection for an SRIR measured in the Church of St. Eustache in Paris, France, with a 32-capsule mh acoustics Eigenmike SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4$^{\text{th}}$-order HOA. Note that the time scales in (b) and (c) are relative to the start of the actual SRIR measurement (as opposed to almost every other time-domain figure in this dissertation, which are relative to t_0, the ToA of the direct sound).

thesis only requires the initial signal to be detected and characterized, i.e. to have its ToA and DoA estimated (to give the reference time t_0 and parameterize the DRIR representation, respectively).

The result of the direct sound detection procedure is summarized in Fig. 5.5. Figure 5.5b shows the SH-domain coherence measure Φ_{cov} calculated from the CoMEDiE diffuseness as $\Phi_{\text{cov}} = 1 - \psi_{\text{cov}}$ over the time period $t \in [0, t_{\text{max}}^{\text{DSdet}}]$ within which to search for the direct sound (see Sec. 2.4.1 again). The first (and only, in this case) peak coherence value (red circle) and any adjacent frames with coherence values above the threshold (green squares) form the constrained time range within which to look for the direct impulse. Figure 5.5c then shows the actual time-domain impulse signal detection on the omnidirectional SH component of the HOA-encoded SRIR, $H_{0,0}(t)$. The louder second reflection at around $t \approx 22$ ms is thus clearly visible. Indexing the direct sound's ToA at its instantaneous energy peak finally gives an estimate of $\tilde{t}_0 \simeq 15.5$ ms.

The DoA estimation results are presented in Figs. 5.5d and 5.5e for the SRP and regularized under-determined Herglotz inversion methods, respectively. Their estimated DoAs are almost identical, differing by only 6.22°, which is just barely above the average angular quantization step for the 38$^{\text{th}}$-order Sloan-Womersley spherical point grid, $\varepsilon_{\mathbf{\Omega}}^{\text{SW}} = 5.63°$. The angular average between the two estimates gives a DoA of $\overline{\mathbf{\Omega}}_{\text{DS}} \simeq (118°, -10.8°)$; this averaged value is the one that will subsequently be used to parameterize the DRIR representation.

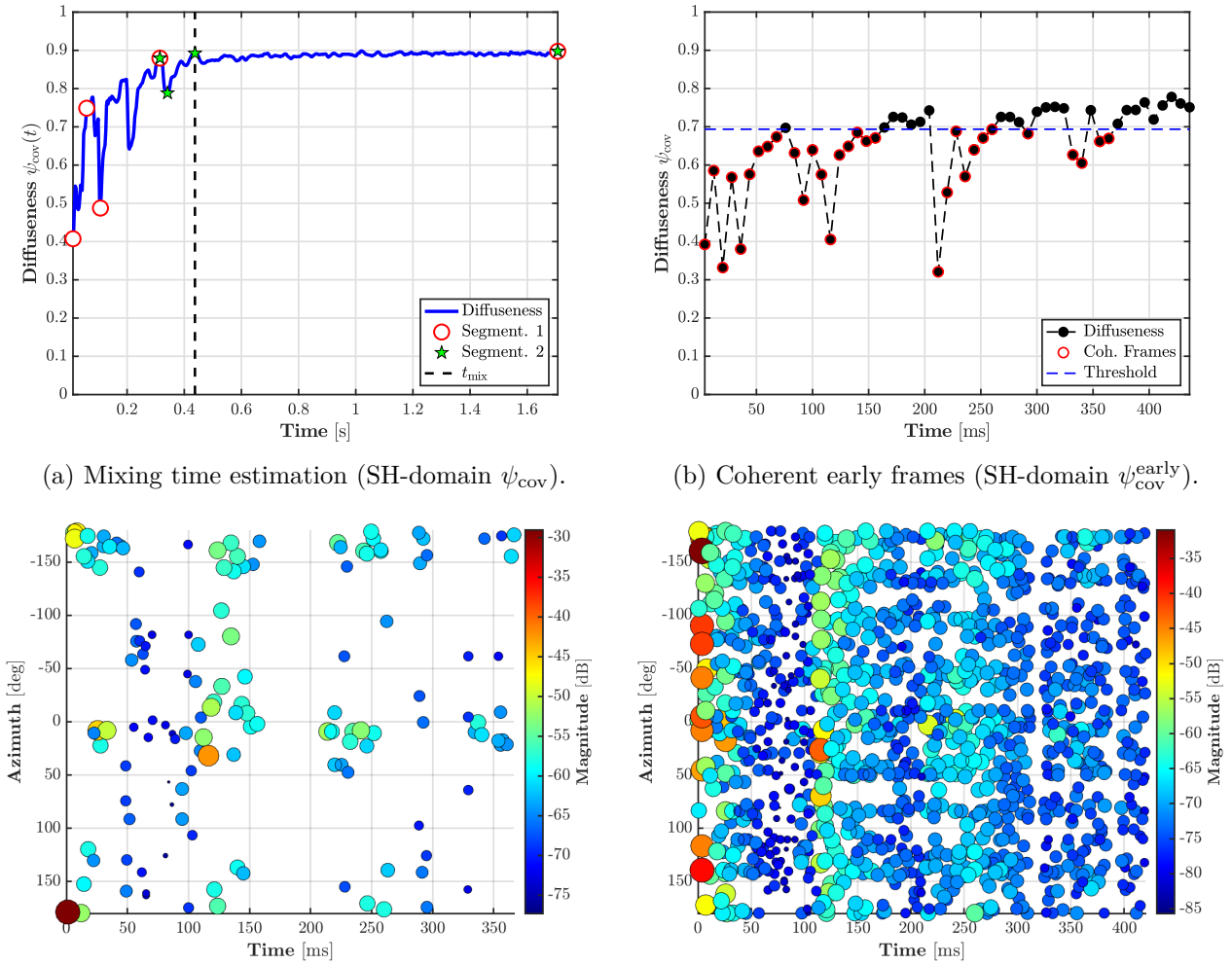
Note that the accuracy of the direct sound's ToA greatly depends on all the various latencies in the measurement signal processing chain having been properly compensated and/or accounted for. Although some of these are relatively straightforward to keep track of (e.g. buffers, in/out signal vector sizes, etc.), others may not be (interfaces, cables, clock synchronizations, etc.). In most musical production applications, however, the final t_0 value is simply an arbitrarily set parameter (often called the "pre-delay"). Furthermore, there is little interest here in obtaining accurate characterizations of the measurement's geometry, i.e. identifying the source-receiver distance as well as their relative positions (using the direct sound's DoA). For these reasons, we will be satisfied in this work with a "correct" detection of the first impulse as it presents itself; in any case, the final analyzed/treated/manipulated SRIR will be stored without any pre-delay, since the t_0 can always be added back in and even arbitrarily modified in the convolution processor used to reproduce the reverberation effect.

5.3.2/ Echo Cartographies

We now return to an SRIR measured in the Notre-Dame Cathedral in Créteil, France, in order to illustrate the complete early reflection cartography procedure first presented in Sec. 2.4.2. Although measured in the same space, this is a different SRIR to the one used in Fig. 5.3 (both source and receiver positions are on the upper "mezzanine" level in the current case, whereas they were both on the ground floor in the former example – see Fig. 5.3a). In fact, this is the SRIR whose omnidirectional channel $H_{0,0}(t)$ was shown in Fig. 3.2 (Sec. 3.2.1) to motivate the reflection salience modification ideas.

The results of the three main steps in the echo detection process are given in Fig. 5.6: mixing time estimation using the SH-domain CoMEDiE diffuseness measure ψ_{cov} (Fig. 5.6a), identification of low diffuseness frames susceptible of containing relevant echoes (Fig. 5.6b), and finally combined DoA/ToA/energy estimation using both the SRP and Herglotz inversion localization methods (Figs. 5.6c and 5.6d). The SH-domain CoMEDiE diffuseness measure is chosen as we know it to be more sensitive to changes in spatial coherence (see Sec. 4.1.4), and it appears to provide a more robust t_{mix} estimate (see Sec. 5.2 above). The 38$^{\text{th}}$-order Sloan-Womersley spherical point grid is once again used to generate the localization maps for DoA estimation.

This example helps bring to light the relationship between the arrival of coherent early reflections and the corresponding "dips" in the diffuseness profiles. More specifically, those dips in diffuseness in



(a) Mixing time estimation (SH-domain ψ_{cov}).

(b) Coherent early frames (SH-domain $\psi_{\text{cov}}^{\text{early}}$).

(c) SRP echo detection.

(d) Herglotz inversion echo detection.

Figure 5.6: Early reflection cartography for an SRIR measured in the Notre-Dame Cathedral in Créteil, France, with a 32-capsule mh acoustics Eigenmike SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4$^{\text{th}}$-order HOA. Additional views of the space-time maps in (c) and (d) are available in Appx. B.5, along with several examples of different SRIR measurements taken in various spaces.

both Figs. 5.6a and 5.6b can be seen to align with the predominant "groups"[3] of detected echoes in Figs. 5.6c and 5.6d. As expected following the results of Sec. 4.2.1, the Herglotz inversion method (Fig. 5.6d) detects a great number of reflections overall, and more simultaneously, compared to the SRP approach (Fig. 5.6c). However, this also makes it much more susceptible to identifying noise or semi-stochastic cluster signals as echoes – though as we will see below (Sec. 5.3.3), this is not necessarily a problem with respect to the echo redistribution manipulation technique.

Once again (picking up from the brief discussion at the end of Sec. 5.3.1 above), it is not the aim of this research project to infer a characterization of the space's geometry from the SRIR measurement; other recent work by Tukuljac et al. [101] and Lovedee-Turner and Murphy [102], for example, has addressed this particular objective in various interesting ways. As such, although the detected echo groups appear to be generally organized in a repeating front-back pattern, this discussion will avoid attempting to draw connections between the analyzed directional properties of the SRIR and any

[3]The word "cluster" would probably be more appropriate here, or would at least transcribe the idea more accurately, but it has unfortunately become a victim of its own success and has already been used to refer to two different concepts in this thesis: the semi-stochastic signal content that builds up over the course of the early segment to eventually form the late reverberation tail, and the combination of closely arriving individual reflections detected as a single echo.

knowledge of the space's geometry/architecture.

Instead, as is the stated goal of this thesis, this modelling of the early segment is intended to allow for the space-time-energy distribution(s) of the detected reflections to be modified and manipulated according to any arbitrary design strategy. One of these, of course, is the aforementioned echo redistribution method that is used as a proof of concept in this dissertation, but a possible future topic of research would be find ways to enable their control with respect to some of the perceptual room acoustics and reverberation descriptors presented back in Sec. 1.1.3.

Finally, note that additional views of the space-time maps shown in Figs. 5.6c and 5.6d (namely elevation-time and azimuth-elevation) are available in Appx. B.5, along with a collection of results from several other SRIR measurements performed in a variety of different spaces (including the Dominicains de Haute-Alsace cloister shown in Fig. 5.4 and the Church of St. Eustache from Fig. 5.5).

5.3.3/ Modification Examples

This section now closes with some "proof of concept" applications of the early reflection manipulation strategies mentioned above and detailed in Sec. 3.2. The SRIR measured in the Notre-Dame Cathedral in Créteil (France, see Figs. 5.3 and 5.6) will continue to be used as an example throughout, with additional examples (once again the Dominicains de Haute-Alsace cloister, Fig. 5.4 and the Church of St. Eustache, Fig. 5.5) included in Appx. B.6.

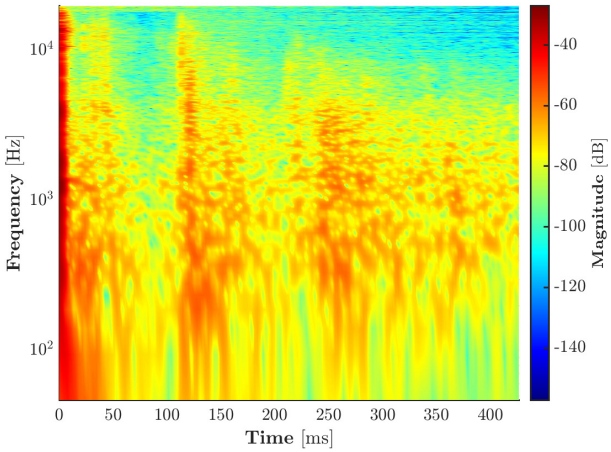
Reflection Salience Manipulation

We begin by presenting the two related early segment modification processes from Secs. 3.2.1 and 3.2.2, namely the energy envelope-based strategies of either constraining echoes to the analyzed late reverberation decay or accentuating their salience with respect to it. As a reminder, these methods both make use of the late reverberation tail analysis results and in particular the modelled directional exponential energy decay envelope. The first, "exponential constraint", aims to re-adjust the energies of the early reflections such that their envelope matches an extrapolation of the late reverberation tail's decay over the early segment. Conversely, "salience accentuation" attempts to increase the energy differences, both positive and negative (in dB, i.e. corresponding to energy ratios > 1 and < 1, respectively), between the echoes and the "backward" extrapolation of the late envelope.

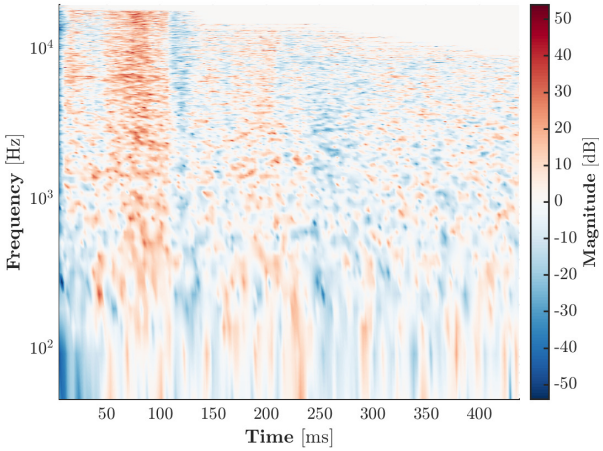
Figure 5.7 illustrates the effect of these manipulations on the example SRIR from the Notre-Dame Cathedral in Créteil, France (see once again Figs. 5.3 and 5.6 for reference). The chosen view is that of the energy spectrogram (square magnitude of the STFT) for the DRIR look direction facing the DoA of the direct sound, $\mathbf{\Omega}_{\text{DS}}$. Once again the DRIR is obtained with a size $S = (L+1)^2$ (i.e. $S = 25$ for the 4^{th}-order HOA-encoded Eigenmike SRIR) Fliege-Maier look direction layout grid and the natural PWD beamformer, in order to preserve the SRIR's spatial coherence properties as much as possible.

Figure 5.7a (top) thus shows the original early segment $|H_{\text{early}}(f, \mathbf{\Omega}_{\text{DS}}, t)|^2$, while Figs. 5.7c and 5.7e (middle and bottom right) show the exponentially constrained $|\widehat{H}^{\text{early}}_{\text{exp}}(f, \mathbf{\Omega}_{\text{DS}}, t)|^2$ and accentuated salience $|\widehat{H}^{\text{early}}_{\text{acc}}(f, \mathbf{\Omega}_{\text{DS}}, t)|^2$ modified signals, respectively. Figures 5.7b and 5.7d (middle and bottom left) then give the modification maps $M_{\text{exp}}(f, \mathbf{\Omega}_{\text{DS}}, t)$ and $M_{\text{acc}}(f, \mathbf{\Omega}_{\text{DS}}, t)$ defined by Eq. 3.5 (Sec. 3.2.1) Eq. 3.7 (Sec. 3.2.2), respectively.

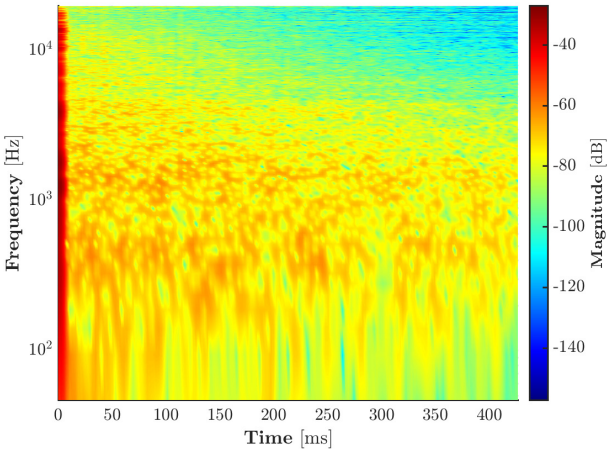
In terms of the control parameters ν_{exp} and $\nu_{\text{acc}}(\mathbf{\Omega}_s, t)$, somewhat extreme settings have been used for this illustrative example: $M_{\text{exp}}(f, \mathbf{\Omega}_{\text{DS}}, t)$ simply uses $\nu_{\text{exp}} = 1$ (i.e. completely constraining the echoes to the exponential energy decay) while $\nu_{\text{acc}}(\mathbf{\Omega}_s, t)$ is defined using a cardioid pattern steered toward $\mathbf{\Omega}_{\text{DS}}$ and varying from 18 dB at its maximum to 6 dB at the opposite point. As described in Sec. 3.2.2, the salience accentuation map is also made to evolve over time, starting from the aforementioned cardioid pattern and linearly reaching 0 dB everywhere at t_{mix} (i.e. logarithmically



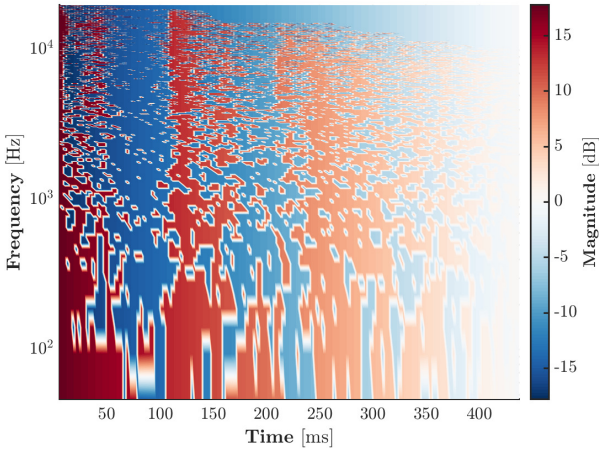
(a) Original early $|H_{\text{early}}(f, \boldsymbol{\Omega}_{\text{DS}}, t)|^2$.



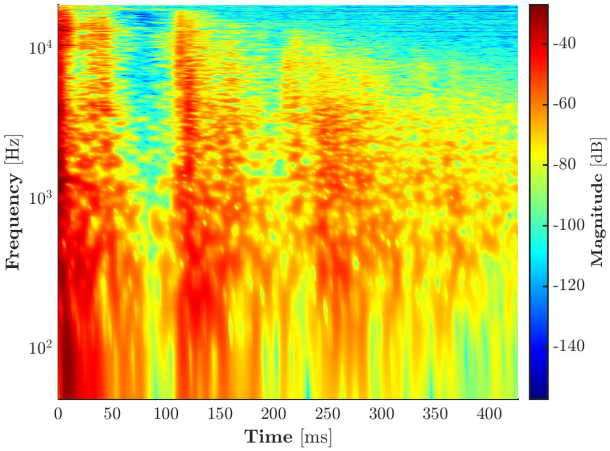
(b) Modification map $M_{\text{exp}}(f, \boldsymbol{\Omega}_{\text{DS}}, t)$ to constrain echoes to the late energy decay envelope.



(c) Modified $|\widehat{H}^{\text{early}}_{\text{exp}}(f, \boldsymbol{\Omega}_{\text{DS}}, t)|^2$ with echoes constrained to the late energy decay envelope.



(d) Modification map $M_{\text{acc}}(f, \boldsymbol{\Omega}_{\text{DS}}, t)$ for accentuating salience with respect to the late energy decay.



(e) Modified $|\widehat{H}^{\text{early}}_{\text{acc}}(f, \boldsymbol{\Omega}_{\text{DS}}, t)|^2$ with increased echo salience.

Figure 5.7: Early reflection salience manipulation for the SRIR measured in the Notre-Dame Cathedral in Créteil, France, with a 32-capsule mh acoustics Eigenmike SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and encoded to 4$^{\text{th}}$-order HOA. The DRIR representation is once again obtained with a Fliege-Maier look direction layout grid rotated to face the DoA of the direct sound, $\boldsymbol{\Omega}_{\text{DS}}$; the natural PWD beamformer is used to preserve the SRIR's spatial coherence properties as much as possible. $M_{\text{exp}}(f, \boldsymbol{\Omega}_s, t)$ (b, middle left) is defined using $\nu_{\text{exp}} = 1$, and similarly $M_{\text{acc}}(f, \boldsymbol{\Omega}_s, t)$ (d, bottom left) is defined using $\nu_{\text{acc}}(\boldsymbol{\Omega}_s, t)$ starting in a cardioid pattern varying from 18 dB at $\boldsymbol{\Omega}_{\text{DS}}$ to 6 dB opposite (see Eq. 3.5, Sec. 3.2.1, and Eq. 3.7, Sec. 3.2.2).

in energy). In other words, earlier reflections are more affected than later ones, ensuring a smooth transition at t_{mix}. Note that in both manipulation strategies, the direct sound itself is left unchanged.

Echo Redistribution

It is, in a sense, somewhat counter-intuitive to consider the salience manipulation methods presented above as early reflection modification procedures since they act on the complete "continuous" space-time-frequency signal using information from the late reverberation analysis. In other words, the echo detection process from Sec. 2.4 remains completely unused. This is, at least in part, the principal interest behind the echo redistribution strategy described in Sec. 3.2.3 as a proof of concept for early segment SRIR modifications based on reflection cartographies such as the ones shown in Fig. 5.6.

Using the Créteil Notre-Dame Cathedral example once more, and in particular the regularized under-determined Herglotz inversion echo detection results, we can notice in Figs. 5.6d and 5.8c how the predominant reflections seems (mostly) aligned in space-time groups along the front-back axis [$\theta = (0°, \pm 180°)$, i.e. the Cartesian x-axis]. They also mostly appear on or below the azimuthal plane [$\varphi = 0°$, or the (x, y)-plane], which is further confirmed by the DEED given in Fig. 5.8a.

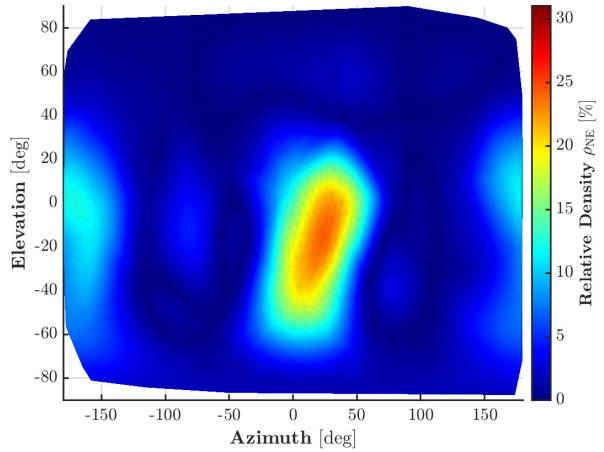
The DEED in Fig. 5.8b subsequently shows the theoretical effect of the echo redistribution procedure; in other words, it represents the DEED obtained after the final iteration/frame rotation step. As expected, it is "flatter" than the original (Fig. 5.8a), and the two density peaks that do remain are in regions where the original was at a minimum. This corresponds directly to the description of the manipulation method from Sec. 3.2.3. Furthermore, the individual reflections most likely "responsible" for the new DEED distribution can be identified by comparing the azimuth-elevation view of the detected echoes before (Fig. 5.8c) and after (Fig. 5.8d) redistribution.

Again, Figs. 5.8d and 5.8f represent the theoretical or predicted "new" positions of the early reflections after having been iteratively rotated through the echo redistribution method. As such, they contain exactly the same echoes as Figs. 5.8c and 5.8e (the latter of which is by definition identical to Fig. 5.6d above). On the other hand, Fig. 5.9 compares this theoretical distribution to the echo detection results obtained by re-analyzing the redistributed SRIR (i.e. the actual HOA-encoded signal upon which the rotations were performed frame-by-frame in the SH domain).

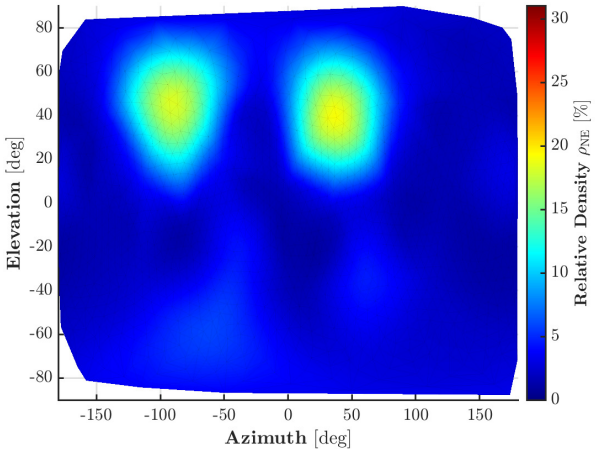
Although neither the theoretical nor the re-analyzed echo distributions provide a "ground truth" in and of themselves, the relative agreement between them suggests that the manipulation procedure does indeed achieve what it has been designed to do. The differences between the two sets of figures also reveals an interesting side-effect of the echo redistribution's tendency to "spread" reflections over the sphere (by targeting minima in the DEED), since, mainly for reasons of stability and sensitivity (see Secs. 4.1.4, 4.2.1 and 5.3.2), the SH-domain CoMEDiE measure ψ_{cov} is used to determine the frames within which to detect echoes.

When applied to frames containing multiple echoes with comparable energies, the redistribution process appears to result in increased diffuseness over those frames: this phenomenon is most visible when comparing Figs. 5.9e and 5.9f between 125 and 150 ms. In the latter case, the higher diffuseness resulted in those frames not being analyzed for echo (re-)detection. This is somewhat surprising, since the detected reflections are supposed to be coherent and the CoMEDiE measure, being based on covariance, should be sensitive to this property.

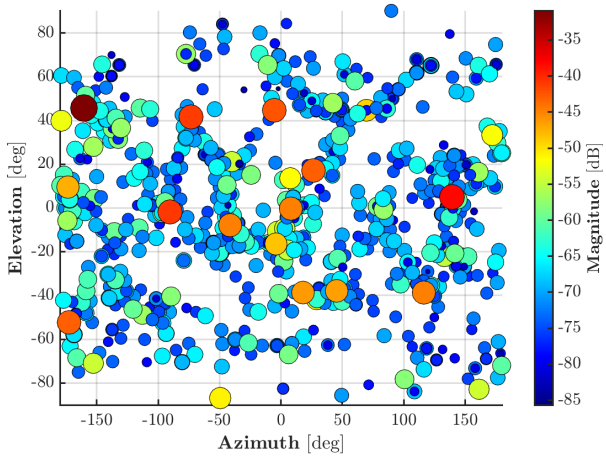
However, the diffuseness profiles used to determine which frames to analyze (e.g. Fig. 5.6b) only indicate if there is *some* coherent content in the frames making up the diffuseness window; the actual DoA localization algorithms can still end up detecting directional peaks with incoherent underlying signal content. Additionally, when looking at Fig. 5.6b once more, one could argue that the windows covering the time range in question should have been considered highly incoherent and not have been analyzed in the first place. As such, the echo redistribution procedure may simply have been spreading



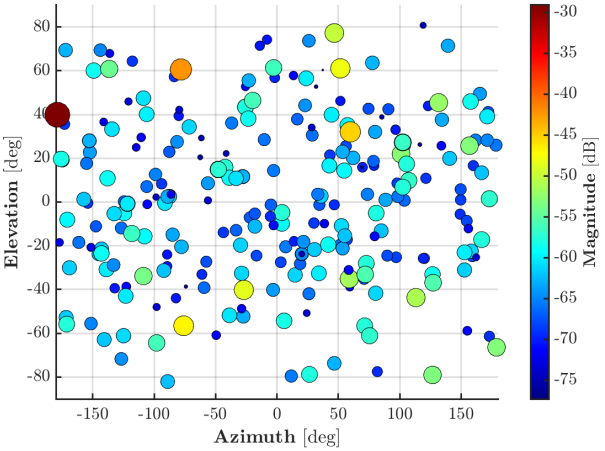
(a) DEED from Herglotz inversion echo detection.



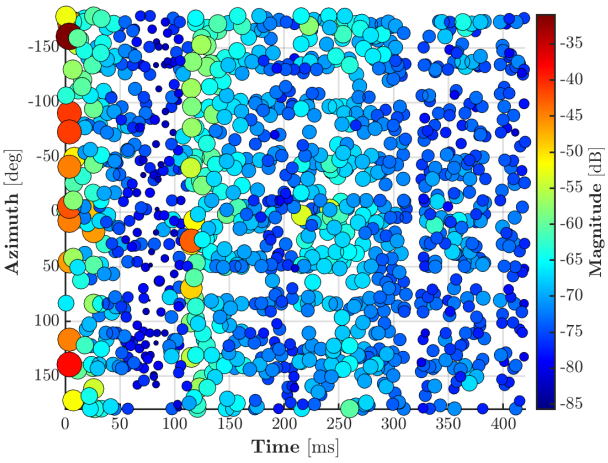
(b) Theoretical DEED after echo redistribution.



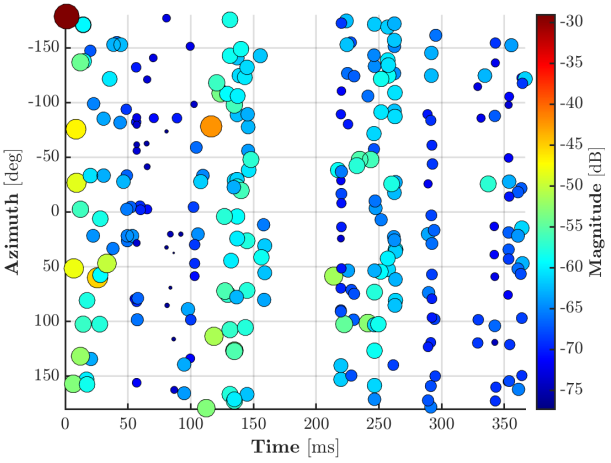
(c) Herglotz inversion echo detection.



(d) Theoretical echo redistribution.

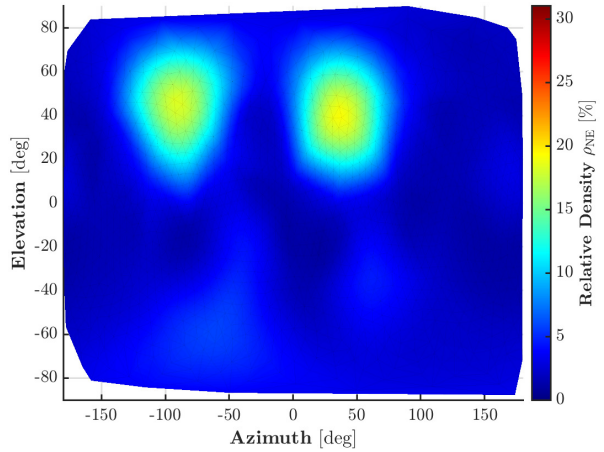


(e) Herglotz inversion echo detection.

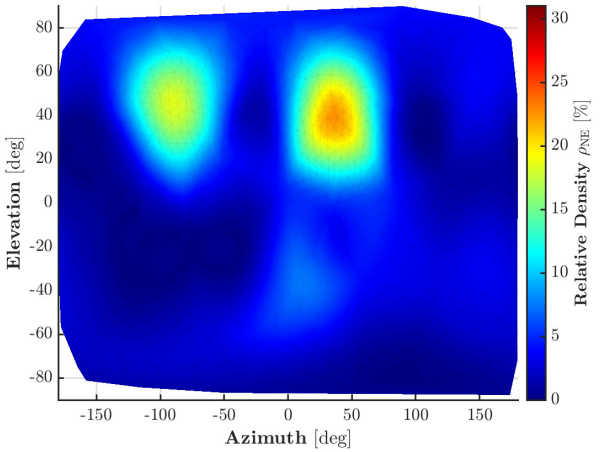


(f) Theoretical echo redistribution.

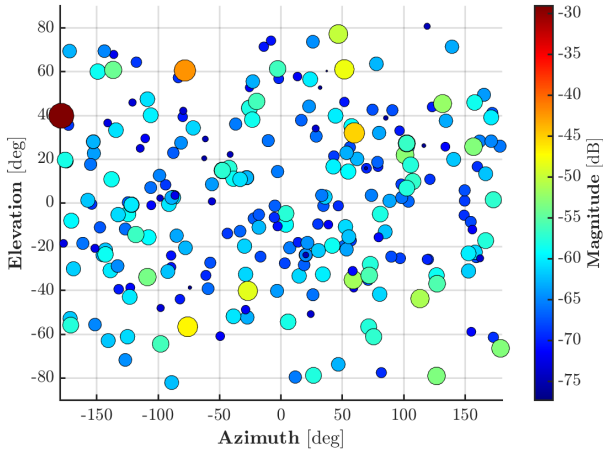
Figure 5.8: Theoretical performance of the echo redistribution procedure (Sec. 3.2.3) on the early reflections detected by regularized under-determined Herglotz inversion on the Notre-Dame Cathedral SRIR (Créteil, France, see Figs. 5.3 and 5.6). The left-hand figures (a,c,e) represent the results obtained on the original SRIR while the right-hand side (b,d,f) shows how each corresponding figure is modified by the echo redistribution method.



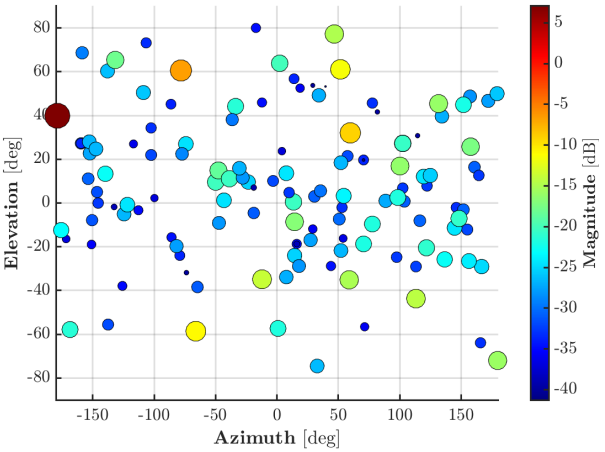
(a) Theoretical DEED after echo redistribution.



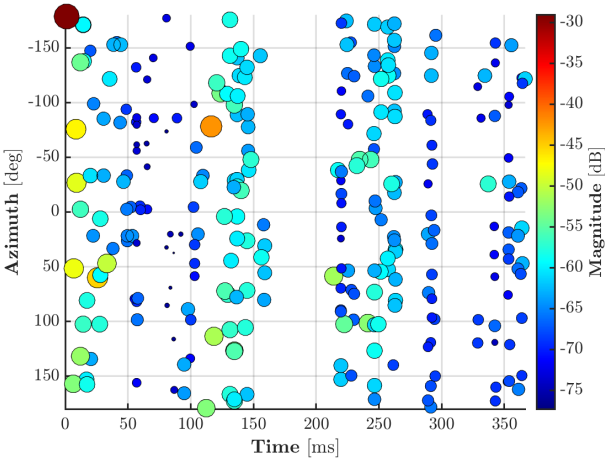
(b) Herglotz inversion re-analyzed DEED.



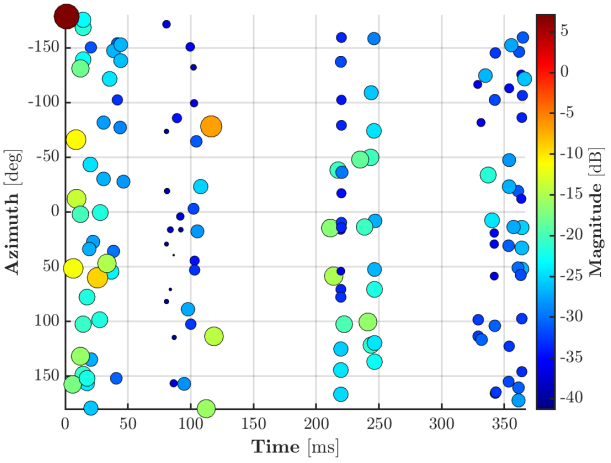
(c) Theoretical echo redistribution.



(d) Herglotz inversion echo re-detection.



(e) Theoretical echo redistribution.



(f) Herglotz inversion echo re-detection.

Figure 5.9: Comparison of the theoretical echo redistribution (a,c,e, left, identical to Figs. 5.8b, 5.8d and 5.8f above) and the regularized under-determined Herglotz inversion echo re-detection (b,d,f, right, i.e. as applied to the result of the redistribution procedure) on the Notre-Dame Cathedral SRIR (Créteil, France). The results shown on the right-hand side are therefore obtained with no *a priori* knowledge of the echo redistribution (i.e. the left-hand side).

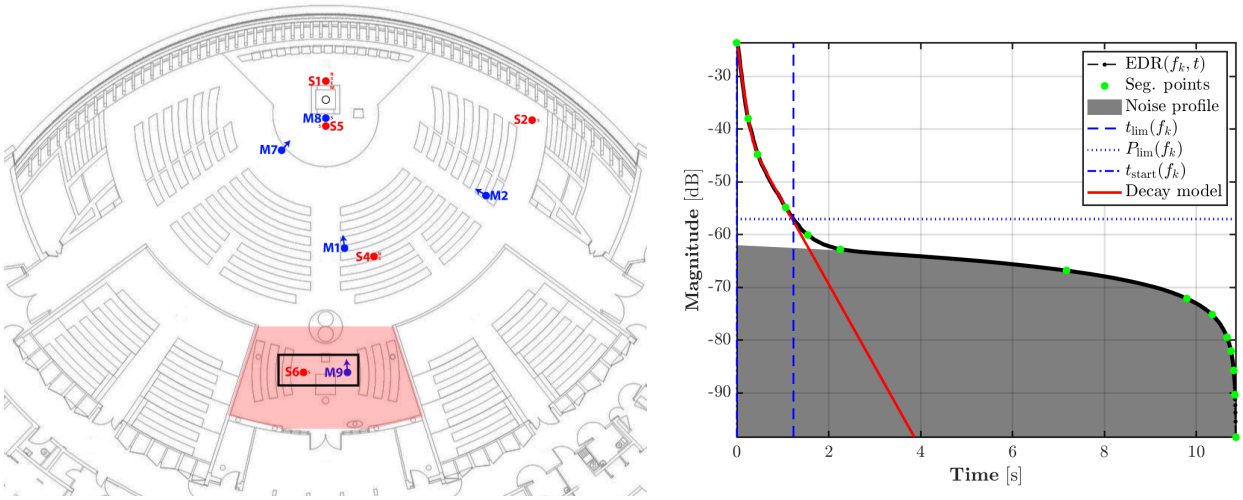
incoherent signals over the sphere, resulting in increased diffuseness. Such observations highlight the need for the careful parameterization of the control parameter λ_{coh} (see Sec. 2.4), which has been set to $\lambda_{\text{coh}} = 3$ throughout this work.

To conclude this presentation of the echo redistribution strategy (and the early reflection manipulation methods more generally), we can note that in its current form the process lacks the ability to operate along the time dimension. In other words, once could imagine a full three-dimensional echo redistribution that can also fill in temporal gaps in echo density. Of course, this would also require an updated definition of echo density over both space and time (as opposed to the purely directional DEED). Nevertheless, even the preliminary proof of concepts presented in this section demonstrate that the analysis and manipulation tools developed throughout this research project provide the ability to deeply modify the space-time characteristics of the early reflections.

5.4/ Directional Late Reverberation Properties

This chapter (and thus the principal content of this dissertation) ends with the following overview of the late reverberation analysis (Sec. 2.5), treatment (i.e. denoising, Sec. 3.3.1), and manipulation methods (Sec. 3.3.2 and Sec. 3.4) presented throughout this thesis. In particular, their application to real-world SMA SRIR measurements is used to motivate some general discussions relating to their various capabilities and limitations as well as their potential uses and developments.

5.4.1/ Coupled Volumes as Double-Slope Decays vs. Anisotropic Single Decays



(a) Measurement plan.

(b) Double-slope detection ($C = 2$).

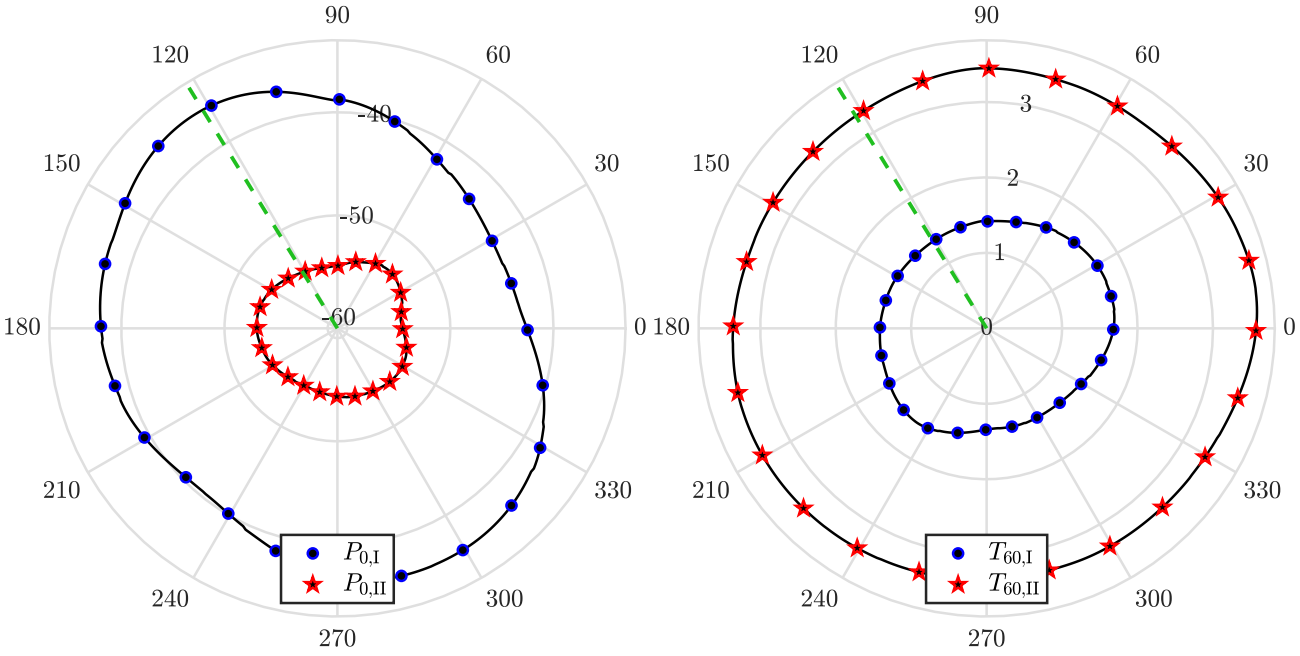
Figure 5.10: Measurement configuration (a, left) and double-slope detection (b, right) for the coupled-volume SRIR measured at the Notre-Dame Cathedral in Créteil, France, using a 32-capsule mh acoustics Eigenmike SMA. The measurement plan highlights the small chapel inset off of the main nave (i.e. the predominant acoustic space) within which the SRIR was measured (source position S6, receiver position M9). Note that the chapel is also covered overhead by the mezzanine level seen in Fig. 5.3a. The multi-slope detection is performed using the algorithm described in Sec. 2.5.1.

One of the most interesting phenomena to be revealed by the use of a late reverberation tail envelope model allowing for both direction-dependent and multiple-slope decays is the observation of a certain ambiguity in the analysis of coupled-volume configurations. Indeed, the classic theory from Cremer and Müller [5] (see Sec. 4.2.3) makes fundamental use of the traditional diffuse field hypothesis. As such, directional variations are completely ignored, and the characteristic double-slope decay must therefore be considered an omnidirectional phenomenon (hence the specific choice of evaluating the

multi-slope detection algorithm on omnidirectional simulations in Sec. 4.2.3). Furthermore, the diffuse field hypothesis places limitations on where the source and receiver can be positioned in order for a double-slope decay to be observed as predicted by the theory.

The directional analysis methods developed in this work allow for a less restrictive view to be considered instead. We thus present two examples in particular here that each demonstrate remarkably different late reverberation characteristics despite both being measured in similar coupled-volume configurations. Coincidentally (or not), both examples are SRIRs measured in churches with small chapels set just off of, and thereby acoustically coupled to, the main nave.

The first is once again a measurement performed at the Notre-Dame Cathedral in Créteil, France, although a different example to the one previously presented in this chapter: whereas the first was measured with both source and receiver at the mezzanine level of the main nave (see Fig. 5.3a), the current SRIR was measured from inside the small chapel tucked in underneath the mezzanine (see Fig. 5.10a). In other words, this measurement theoretically fulfills the requirements for the generation of a classic Cremer-Müller double-slope decay.



(a) $\overline{P}_0(\theta)$, radial axis in [dB]. (b) $\overline{T}_{60}(\theta)$, radial axis in [s].

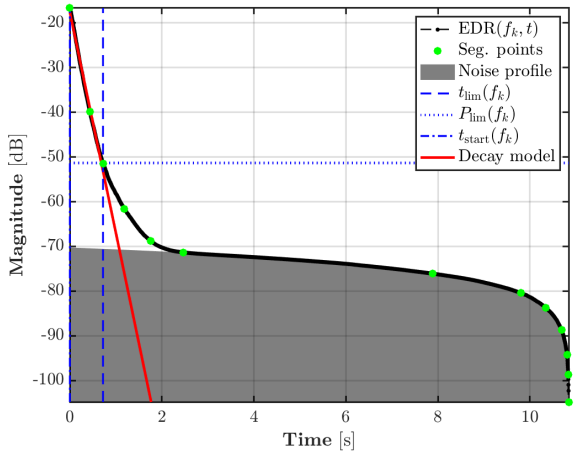
Figure 5.11: Azimuthal projections of the estimated double-slope $\overline{P}_0$ (a, left) and $\overline{T}_{60}$ (b, right) for the coupled-volume SRIR measured at the Notre-Dame Cathedral in Créteil, France, using a 32-capsule mh acoustics Eigenmike SMA. Both quantities are averaged over their third-octave band values (themselves averaged from the individual frequency bin estimates) over the frequency range $f \in [809, 5203]$ Hz, i.e. the PWD DI band with DI ≥ 9.16 dB, its value at f_{alias}. The green dashed line represents θ_{DS}, the azimuthal DoA of the direct sound.

The double-slope is successfully detected using the algorithm described at the end of Sec. 2.5.1; the result of the broadband EDC fit is shown in Fig. 5.10b for the look direction with the highest $\mathcal{L}_{\mathrm{slope}}(\boldsymbol{\Omega}_s, C)$ value (see Eq. 2.35). Figure 5.11 then shows the estimated double-slope $\overline{P}_0(\theta)$ (Fig. 5.11a, left) and $\overline{T}_{60}(\theta)$ (Fig. 5.11b, right), averaged over third-octave bands in the frequency range $f \in [809, 5203]$ Hz and projected onto the azimuthal plane ($\varphi = 0°$) by cubic Hermite spherical interpolation. This specific frequency range corresponds to the PWD DI band with DI ≥ 9.16 dB, which is its value at f_{alias} (see Fig. 1.5b for reference). Note that the third-octave band values are themselves obtained by averaging over the individual frequency bin estimates within each band.

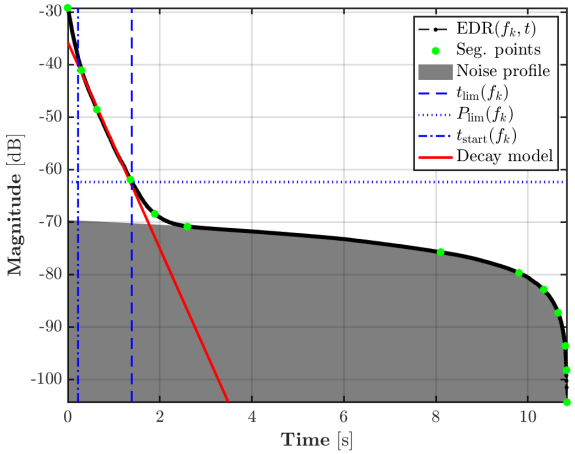
These results appear to confirm that this particular measurement configuration does indeed produce a double-slope decay as predicted by Cremer and Müller [5]. This view is further supported by

comparing the estimated mixing time $\tilde{t}_{\text{mix}} \simeq 424$ ms, as averaged between SH diffuseness (ψ_{cov}) and DRIR incoherence (ψ_{dir}) estimates, to the "turning point" between the two individual slopes, $\bar{t}_{\text{turn}} \simeq 640$ ms (calculated from the frequency-averaged analysis results). Since the turning point represents the moment the second slope starts entirely describing the reverberation tail, the fact that the t_{mix} estimate is over 200 ms shorter indicates that the SRIR reaches maximum incoherence while the first decay is still predominant. As such, it is absolutely necessary to treat the late reverberation tail's energy envelope as a double-slope decay in this example.

(a) View of the main acoustic space (the nave) in the Christuskirche, Karlsruhe, Germany.



(b) Slope number estimation at $\mathbf{\Omega}_{\text{DS}} = (-16.8°, 5.77°)$: $C = 1$, $t_{\text{start}} = 0$ ms, $t_{\text{lim}} = 720$ ms.



(c) Slope number estimation at $\mathbf{\Omega}_s = (160°, 15.2°)$: $C = 1$, $t_{\text{start}} = 221$ ms, $t_{\text{lim}} = 1.39$ s.

Figure 5.12: View of the main acoustic space (a, top) and estimation of the number of slopes in the late reverberation tail (b and c, bottom) for the coupled-volume SRIR measured at the Christuskirche in Karlsruhe, Germany, using a 32-capsule mh acoustics Eigenmike SMA. The SRIR is measured at the entrance to a small chapel located through an open doorway just off the nave.

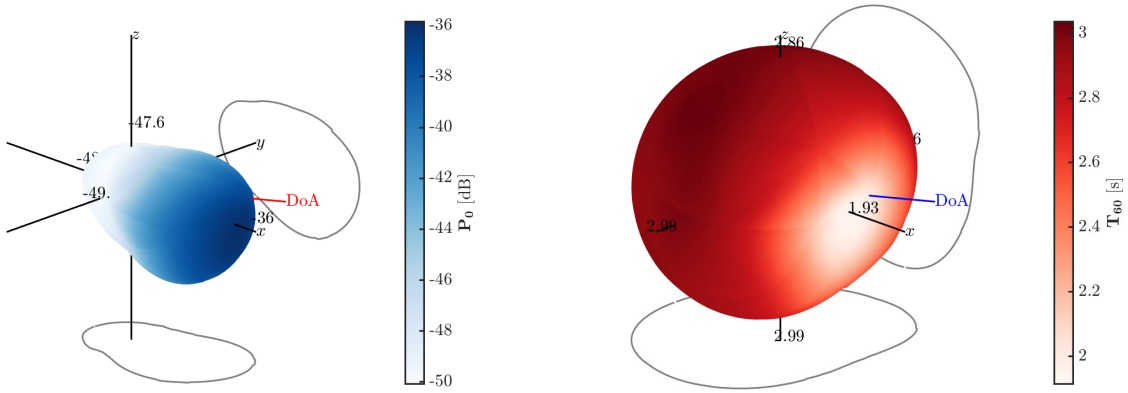
However, the measurement of SRIRs in coupled-volume configurations does not always satisfy the above requirements for the generation of double-slope decays. Indeed, Cremer and Müller [5] demonstrate a range of source/receiver placement options that, although they do result in a coupling of the two acoustic spaces, do not produce the characteristic "kinked" energy envelope. As noted above, one of the particularly interesting uses of SMA SRIR measurements is in exploiting the directional analysis techniques presented throughout this thesis in order to investigate some pathological cases not

covered by the Cremer-Müller approach, e.g. due to their inherent anisotropy.

An example of one such case is offered by an SRIR measured at the Christuskirche in Karlsruhe, Germany, a late 19$^{\text{th}}$-century church built in Gothic Revival style with a large dome-like nave and several small annexed chapels (see Fig. 5.12a, top). The SRIR in question is measured with the SMA placed at the entrance to one of these chapels – that is, within the acoustic aperture through which the two volumes are coupled. The source is then placed inside the chapel, exciting the less reverberant space first and thus fulfilling one of the prerequisites for the generation of a double slope.

In this case, we obtain a combined mixing time estimate of $\tilde{t}_{\text{mix}} \simeq 573$ ms (again, the average of $\tilde{t}^{\text{cov}}_{\text{mix}} \simeq 659$ ms and $\tilde{t}^{\text{dir}}_{\text{mix}} \simeq 488$ ms). The estimation of the number of slopes, shown in Figs. 5.12b and 5.12c (bottom left and right, respectively), returns $C = 1$; it would therefore appear that this configuration does not give rise to a double-slope decay. A closer look at the decay curves in Figs. 5.12b and 5.12c, however, reveals that in each case the detection algorithm has chosen a single exponential best fit by ignoring different parts of the energy envelope.

What is most interesting is that the chosen decay section to model (i.e. between t_{start} and t_{lim}) varies depending on the look direction in question. When facing the direct sound from the source inside the small chapel, $\mathbf{\Omega}_{\text{DS}} \simeq (-16.8°, 5.77°)$, the early part of the curve is considered ($t \in [0, 720]$ ms), and the later part is assimilated to the integrated transition to the noise floor profile. On the opposite side of the sphere, at $\mathbf{\Omega}_s = (160°, 15.2°)$, a later segment is chosen ($t \in [221, 1390]$ ms), and the early part is considered a non-exponential early reflection regime. In both cases, the mixing time ($\tilde{t}_{\text{mix}} \simeq 573$ ms) occurs part way through the detected single exponential decay.



(a) $\overline{P}_0(\mathbf{\Omega})$. (b) $\overline{T}_{60}(\mathbf{\Omega})$.

Figure 5.13: Directional $\overline{P}_0(\mathbf{\Omega})$ (a, left) and $\overline{T}_{60}(\mathbf{\Omega})$ (b, right) estimates, interpolated from DRIR look directions $\mathbf{\Omega}_s$ to the full sphere by cubic Hermite spherical interpolation [87], for the coupled-volume SRIR measured at the Christuskirche in Karlsruhe, Germany, using a 32-capsule mh acoustics Eigenmike SMA. Both quantities are averaged over their third-octave band values (themselves averaged from the individual frequency bin estimates) over the frequency range $f \in [809, 5203]$ Hz, which corresponds to the PWD DI band with DI ≥ 9.16 dB, its value at f_{alias}.

These observations suggest that, contrary to the previous example, this SRIR measured on the threshold between two different acoustic spaces is better described by two highly anisotropic (i.e. directionally separated) single-slope decays rather than a (more or less) omnidirectional double slope. Furthermore, since the multi-slope detection algorithm works on broadband EDC curves, it is possible that either the omnidirectional low frequencies or the spatial aliasing high frequencies are contributing to the appearance of the "secondary" slopes seen in Figs. 5.12b and 5.12c. Note that the DRIR representation in question is obtained using the natural PWD beamformer.

Directional views of the $\overline{P}_0(\mathbf{\Omega})$ and $\overline{T}_{60}(\mathbf{\Omega})$ estimates, shown in Fig. 5.13 interpolated to the full

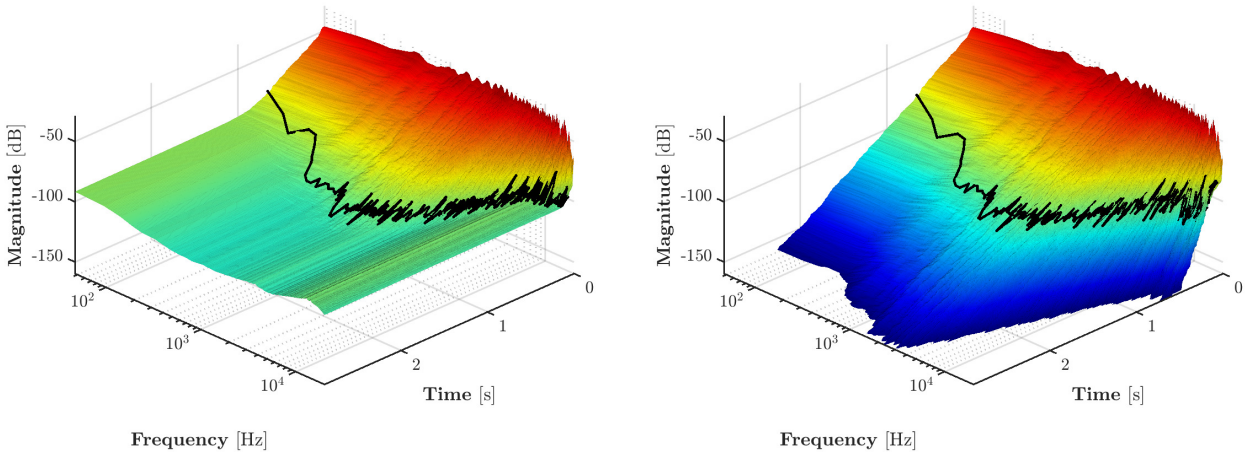
sphere from the analyzed DRIR look directions $\mathbf{\Omega}_s$ by cubic Hermite spherical interpolation [87], appear to support this. Once again, the frequency-dependent estimates have been averaged over third-octave bands in the $f \in [809, 5203]$ Hz range (as described above). In both cases, the results are consistent with the discussion above: Fig. 5.13a shows a far greater initial power in the hemisphere facing the less reverberant space excited by the source, and this difference in reverberation is confirmed by the distribution of the reverberation time $[\overline{T}_{60}(\mathbf{\Omega})]$ in Fig. 5.13b.

As evidenced by these two examples, the mixing time estimation, DRIR generation, and late reverberation analysis tools developed in this research project (see Ch. 2) allow for a wide variety of complex room acoustics phenomena to be investigated through SMA SRIR measurements. Indeed, the above results could still be discussed in far more detail, for example with respect to the effect of the source's directivity on the SRIR's directional properties, but in an effort to remain somewhat concise such considerations will be left to future investigations.

5.4.2/ Denoising

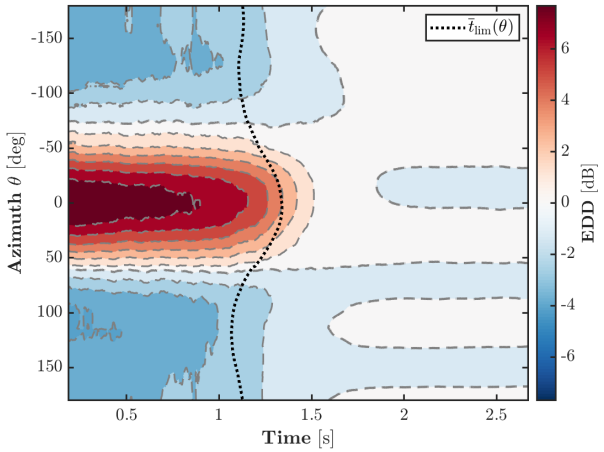
The aim of this section is to provide a real-world example of the direction-dependent denoising procedure first described in Sec. 3.3.1 and then later evaluated in Sec. 4.3.1. The chosen SRIR is a measurement performed in a small neo-Gothic chapel within the Dominicains de Haute-Alsace convent complex (where the cloister SRIR from Fig. 5.4 was also measured). This SRIR is measured with the Eigenmike SMA placed just inside the entrance to the shoebox-shaped chapel, with the doors wide open, and the source loudspeaker is positioned near the altar on the opposite end of the room: a high degree of anisotropy can therefore be expected from this configuration.

Once again, the SRIR is pre-treated using the method presented in Secs. 3.1 and 5.1, and encoded to $L = 4^{\text{th}}$-order HOA. The DRIR representation chosen for denoising makes use of the natural PWD beamformer on the size $S = (L+1)^2 = 25$ Fliege-Maier layout of look directions (rotated such that one of the look directions faces the DoA of the direct sound, $\mathbf{\Omega}_{\text{DS}}$). This choice is made in order to preserve the spatial incoherence of the late reverberation field as much as possible, and follows directly from the discussions in Secs. 4.1.4 and 4.3.1.

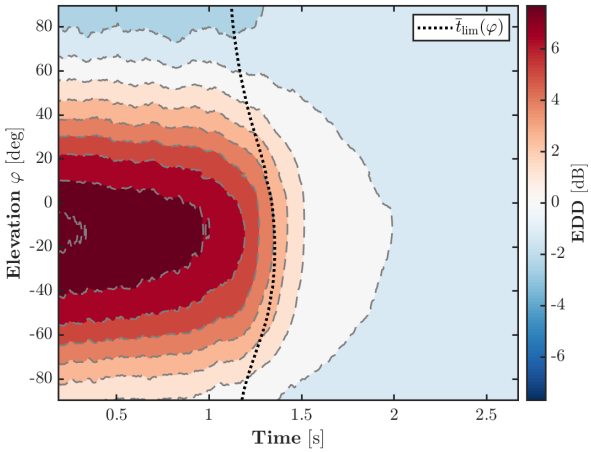


(a) Noisy EDR$(f, \mathbf{\Omega}_{\text{DS}}, t)$. (b) Denoised EDR$(f, \mathbf{\Omega}_{\text{DS}}, t)$.

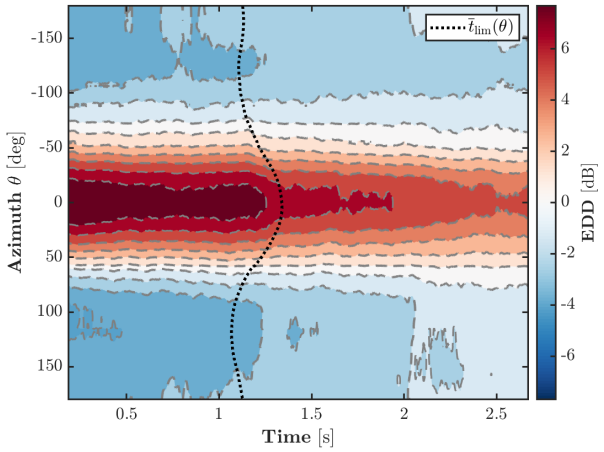
Figure 5.14: Directional EDRs taken at the DRIR look direction facing the DoA of the direct sound, $\mathbf{\Omega}_{\text{DS}}$, for (a, left) the noisy SRIR measured in the neo-Gothic chapel at the Dominicains de Haute-Alsace convent, and (b, right), the DRIR obtained after application of the tail resynthesis denoising procedure. The original SRIR is once again measured with a 32-capsule mh acoustics Eigenmike SMA and encoded to 4^{th}-order HOA; the DRIR representation is generated using the natural PWD beamformer on a 25-point Fliege-Maier look direction layout grid. The superimposed black solid line represents the noise floor limiting time $t_{\text{lim}}(f, \mathbf{\Omega}_{\text{DS}})$.



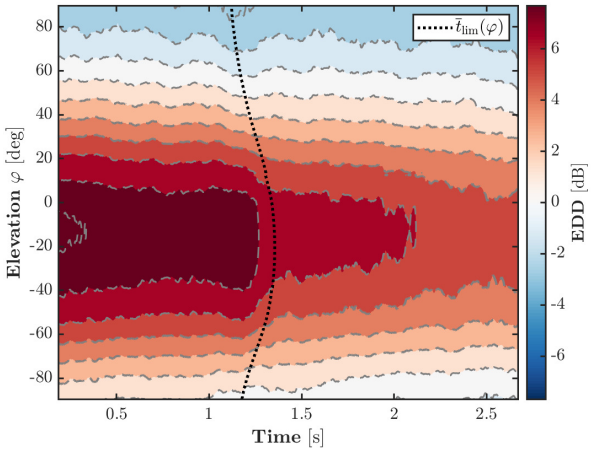
(a) Noisy azimuthal EDD ($\varphi = 0°$).



(b) Noisy elevational EDD ($\theta = 0°$).



(c) Denoised azimuthal EDD ($\varphi = 0°$).



(d) Denoised elevational EDD ($\theta = 0°$).

Figure 5.15: Azimuthal and elevational plane EDDs for (a,b, top) the noisy SRIR measured in the neo-Gothic chapel at the Dominicains de Haute-Alsace convent, and (c,d, bottom) the DRIR obtained after application of the tail resynthesis denoising procedure. The original SRIR is once again measured with a 32-capsule mh acoustics Eigenmike SMA and encoded to 4$^{\text{th}}$-order HOA; the DRIR representation is generated using the natural PWD beamformer on a 25-point Fliege-Maier look direction layout grid. The DRIR's EDD($\mathbf{\Omega}_s, t$) is calculated by first summing the directional EDR($f, \mathbf{\Omega}_s, t$) over the third-octave bands in the frequency range $f \in [809, 5203]$ Hz, which corresponds to the PWD DI band with DI ≥ 9.16, its value at f_{alias}. The EDD($\mathbf{\Omega}_s, t$) is then projected onto the azimuthal and elevational planes by cubic Hermite spherical interpolation [87]. Finally, the black dashed lines represent the noise floor limiting time $\bar{t}_{\text{lim}}(\mathbf{\Omega})$, similarly averaged over third-octave bands and interpolated to the azimuthal and elevational planes.

Mixing time analysis gives an average estimate of $\tilde{t}_{\text{mix}} \simeq 181$ ms for this SRIR ($\tilde{t}^{\text{cov}}_{\text{mix}} \simeq 272$ ms and $\tilde{t}^{\text{dir}}_{\text{mix}} \simeq 90.7$ ms). The multiple slope detection algorithm confirms that the late reverberation tail is described by a single exponential decay, as could be expected (since only a single acoustic space is excited, even though it is made highly anisotropic by the open doors). The effect of the tail resynthesis method on a single look direction of the DRIR representation is then illustrated in Fig. 5.14 through the comparison of the EDR calculated before (Fig. 5.14a, left) and after (Fig. 5.14b) denoising. The look direction in question is the one oriented towards the DoA of the direct sound, $\mathbf{\Omega}_{\text{DS}}$.

Additionally, the noise floor limiting times $t_{\text{lim}}(f, \mathbf{\Omega}_{\text{DS}})$ detected at each frequency bin during the EDR analysis process (Sec. 2.5.1) are indicated by the superimposed solid black line. The EDR plots are furthermore restricted to a total dynamic range of 130 dB.

In order to investigate the effect of the denoising process on the spatial structure of the DRIR,

Fig. 5.15 subsequently presents both the azimuthal (a and c, left) and elevational plane (b and d, right) EDDs, before (a and b, top) and after (c and d, bottom) tail resynthesis. These EDDs are calculated by first summing the directional $\mathrm{EDR}(f, \mathbf{\Omega}_s, t)$ over the aforementioned third-octave bands in the frequency range $f \in [809, 5203]$ (the PWD DI band with $\mathrm{DI} \geq 9.16$, its value at f_{alias}), and then applying Eq. 2.36 (but without the frequency dependence). As also described in Sec. 2.5.2, the resulting DRIR $\mathrm{EDD}(\mathbf{\Omega}_s, t)$ is then interpolated for visualization onto the azimuthal and elevational planes by cubic Hermite spherical interpolation [87]. For reference, the azimuthal plane is defined at $\varphi = 0°$, i.e. on the (x, y)-plane, and the elevational plane at $\theta = 0°$, i.e. on the (x, z)-plane.

Once again, since we are interested here in illustrating the effect of the denoising procedure, $\bar{t}_{\mathrm{lim}}(\mathbf{\Omega})$ values averaged and interpolated in the same way as the EDDs are shown by an overlain dashed black line (still in Fig. 5.15). Whereas Fig. 5.14 allows the time-frequency reconstruction of the envelope to be visually verified, as in e.g. Noisternig et al. [103] and Massé et al. [84], Fig. 5.15 seeks to do the same for the late reverberation's space-time properties, as in Massé et al. [78].

We can note here how $\bar{t}_{\mathrm{lim}}(\mathbf{\Omega})$ is shorter in directions with an energy deficit to the spatial mean (i.e. a negative EDD), and conversely longer where the EDD is generally positive. This follows directly from the fact that, for example, consistently lower energies in a given direction implies either a lower $\overline{P}_0(\mathbf{\Omega})$ or a shorter $\overline{T}_{60}(\mathbf{\Omega})$, or both, compared to the spatial means.

Besides these slight variations in $\bar{t}_{\mathrm{lim}}(\mathbf{\Omega})$, however, the noise floor in this example is highly isotropic. This allows the method's capabilities in reconstructing the late tail's directional properties to be clearly demonstrated: indeed, by comparing the denoised EDDs before and after $\bar{t}_{\mathrm{lim}}(\mathbf{\Omega})$, it appears that the spatial distribution of the energy envelope has been correctly prolonged (within the limitations discussed with respect to the simulated example in Sec. 4.3.1). Combined with the validation of the time-frequency reconstruction from Fig. 5.14 (assuming similar performance over all S look directions), this example serves to further support the DRIR-based approach to SRIR denoising.

5.4.3/ Evaluating Late Tail Isotropy

As mentioned several times throughout this dissertation, the fact that the signal model and thus the analysis-treatment framework are both made (potentially) direction-dependent does not imperatively mean that these capabilities need to be exploited at all times. Indeed, we have seen that the DRIR approach is subject to some inherent limitations, even under perfectly isotropic conditions (most notably in terms of spatial incoherence preservation – see Sec. 4.1.4). In cases where a direction-dependent view can be deemed unnecessary, i.e. where the anisotropy of the late reverberation tail is negligible, it may therefore be advantageous to perform the denoising treatment entirely in the SH domain (as originally proposed in Massé et al. [84]).

The critical question then becomes: under what criteria can we make such a choice of analysis-treatment domain? Or in other words: how do we determine if an SRIR is anisotropic enough (or will become, once the reverberation tail is prolonged) to warrant directional processing? A preliminary approach to this problem is presented in Alary, Massé, Schlecht, Noisternig, and Välimäki [94], a paper that relates the results of a perceptual study conducted in collaboration between the Department of Signal Processing and Acoustics at Aalto University (Finland) and the STMS-lab at Ircam in order to assess the audibility of anisotropic energy decays in late reverberation tails.

Though this work has provided some initial insights into the perception of late reverberation anisotropy (most importantly demonstrating that directional decay characteristics can indeed be heard), additional investigations would be needed in order to determine, for instance, a JND based on a given objective measure. This JND could then be used as a discriminating criterion in the aforementioned choice of analysis-treatment domain (i.e. between the SH-domain HOA-encoded SRIR and the beamformed DRIR representation).

A potential choice of objective measure for such a perceptual evaluation could be the EDD range isotropy measure $\mathcal{I}_{\text{EDD}}$ defined by Eq. 2.37 (Sec. 2.5.2). For example, Fig. 5.16 shows the $\mathcal{I}_{\text{EDD}}(t)$ for the noisy (black triangles) and denoised (red stars) versions of the SRIR measured in the neo-Gothic chapel at the Dominicains de Haute-Alsace convent and used to demonstrate the denoising procedure in the previous section. The EDD ranges here are calculated from the DRIRs' frequency-averaged $\text{EDD}(\boldsymbol{\Omega}_s, t)$ [see Fig. 5.15], and the thin blue dashed line shows the spatial mean $\bar{t}_{\text{lim}}$ [from the directional, frequency-averaged $\bar{t}_{\text{lim}}(\boldsymbol{\Omega}_s)$, see also Fig. 5.15].

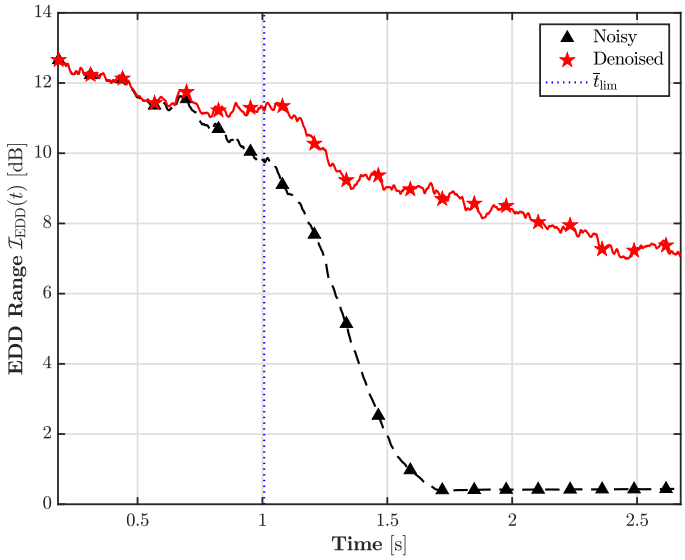


Figure 5.16: EDD range isotropy measure $\mathcal{I}_{\text{EDD}}(t)$ for the noisy (black triangles) and denoised (red stars) SRIR measured in the neo-Gothic chapel at the Dominicains de Haute-Alsace convent, i.e. the same example as presented in Sec. 5.4.2 above. The EDD ranges are calculated on the DRIRs' $\text{EDD}(\boldsymbol{\Omega}_s, t)$ defined as in Fig. 5.15. The thin dotted blue line shows the spatial mean $\bar{t}_{\text{lim}}$ calculated from the frequency-averaged $\bar{t}_{\text{lim}}(\boldsymbol{\Omega}_s)$ also defined under Fig. 5.15.

In this example, the mean $\overline{\mathcal{I}}^{\text{orig}}_{\text{EDD}}$ value over $\tilde{t}_{\text{mix}} \leq t < \bar{t}_{\text{lim}}$ (i.e. before the noise floor) is 11.4 dB. Such a value may be a useful unidimensional description of the level of anisotropy in a measured SRIR, and could eventually be used to choose an analysis-treatment domain. Determining a JND would then involve testing the perceptibility of anisotropic reverbs against this measure.

More generally, Fig. 5.16 also helps confirm the (an)isotropy trends observed in Fig. 5.15: since the former is calculated on the DRIR-based, non-interpolated $\text{EDD}(\boldsymbol{\Omega}_s, t)$, it can be used to verify that the space-time characteristics observed on the azimuthal and elevational plane projections are consistent with those contained in the full three-dimensional representation. In other words, it ensures (visually) that the cubic Hermite spherical interpolation does not introduce any artificial phenomena.

For the current SRIR, the influence of the highly isotropic noise floor observed in the EDDs of the original noisy measurement (Fig. 5.15) is plainly evidenced in Fig. 5.16 by the sudden drop in the $\mathcal{I}^{\text{orig}}_{\text{EDD}}(t)$ (black triangles) beyond the average $\bar{t}_{\text{lim}}$. On the other hand, the $\mathcal{I}^{\text{denoised}}_{\text{EDD}}(t)$ (red stars) appears to follow and prolong the more or less steady decrease in anisotropy already present throughout the original measurement's late decay (i.e. before the noise floor). This steady decrease in $\mathcal{I}^{\text{denoised}}_{\text{EDD}}(t)$ even seems almost linear, except for a slight deviation around the average $\bar{t}_{\text{lim}}$ that may be due to the spatial variations in $\bar{t}_{\text{lim}}(\boldsymbol{\Omega}_s)$ as illustrated in Fig. 5.15.

5.4.4/ Manipulating Late Tail Isotropy

We finally close this chapter, and in doing so reach the end of the thesis work to be presented in this dissertation, with an example of the late reverberation tail isotropy modification methods presented in Sec. 3.3.2 and Sec. 3.4. These involved two different manipulation strategies ("isotropification" and "directification") implemented in two different contexts: first by altering the analyzed space-frequency $P_0(f, \mathbf{\Omega}_s)$ and $T_{60}(f, \mathbf{\Omega}_s)$ distributions and completely resynthesizing the reverberation tail, and second by defining a modification map based on the space-time-frequency EDD.

The SRIR on which these manipulations are demonstrated is the measurement performed at the Grandes Serres de Pantin (Pantin, France) that was previously used in Sec. 5.1 to illustrate the pre-analysis treatment procedure (Fig. 5.1). This SRIR is particularly interesting as it was measured with the receiver (the 32-capsule Eigenmike SMA) on the threshold of a large doorway connecting two abandoned industrial hangars (i.e. a similar configuration to the coupled-volume Christuskirche measurement presented in Sec. 5.4.1). The Eigenmike was facing the thin edge of the wall containing the doorway and the source loudspeaker was placed in the less reverberant, semi-open hangar to its left, resulting in an estimated direct sound DoA of $\mathbf{\Omega}_{\text{DS}} \simeq (89.1°, -14.9°)$ [again averaged between the SRP and Herglotz inversion estimations as described in Sec. 5.3.1 above]. Figures 5.17a and 5.18a illustrate this measurement configuration from two opposite angles (though it should be noted that Fig. 5.18a shows the source in a different position to that of the SRIR in question, which was measured with the loudspeaker setup in Fig. 5.17a).

For each of the two manipulation strategies (isotropification and directification), the two different approaches (tail resynthesis or EDD compensation/accentuation) are subsequently compared below.

Isotropification

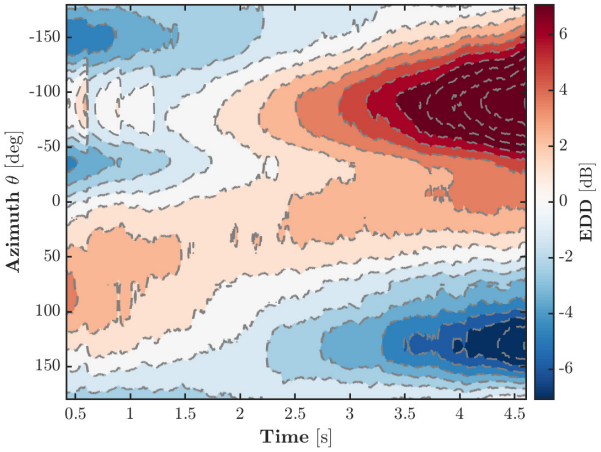
Figure 5.17 (b–d) shows the effect of applying the isotropification procedure with a control parameter value of $\upsilon_{\text{iso}} = 1$ for both the tail resynthesis (c, bottom left) and EDD compensation methods (d, bottom right, with $\upsilon_{\text{iso}}^{\text{EDD}} = 1$ as well) through the azimuthal plane EDDs of the modified SRIRs. The EDD of the "original" denoised SRIR is also given in Fig. 5.17b (top right) for reference. Corresponding elevational plane EDD and EDD range isotropy [$\mathcal{I}_{\text{EDD}}(t)$] figures, both omitted here to avoid overcrowding the current section, are available in Appx. B.7.

The EDDs are calculated in the same way as in Sec. 5.4.2, i.e. on a 25-point DRIR representation using a Fliege-Maier look direction layout and the natural PWD beamformer, and averaged over third-octave bands in the $f \in [f_{\text{aliasDI}}, f_{\text{alias}}] \simeq [809, 5203]$ Hz frequency range. The mixing time estimate $\tilde{t}_{\text{mix}} \simeq 415$ ms is once again obtained by averaging the results of the SH-domain diffuseness ψ_{cov} and directional incoherence ψ_{dir} estimations.

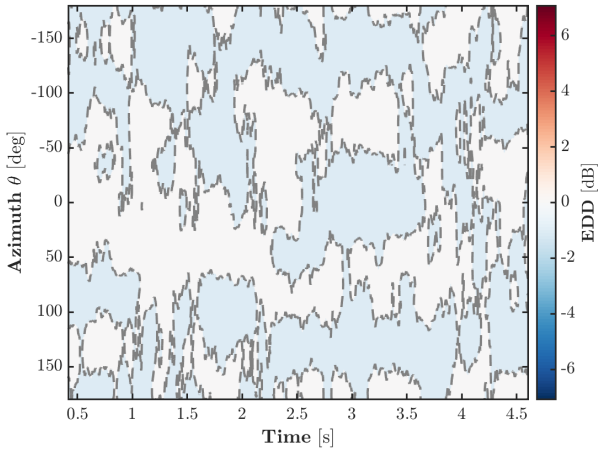
Both approaches clearly succeed in isotropifying the SRIR, though it is interesting to note that the EDD compensation method appears to retain some directional trends (and even seems to overcompensate around $\theta \approx 125°$ starting at $t \geq 3$ s). Tail resynthesis, on the other hand, completely undoes the original directional characteristics of the late reverberation tail. However, the reference EDD (Fig. 5.17b) shows evidence of some late arriving echoes that the t_{mix} estimation failed to take into account (at around $\theta \approx -90°$, between $0.415 \leq t \leq 1.25$ s). Naturally, the tail resynthesis method erases these, severely altering the SRIR's character beyond the stated goals of this manipulation; EDD compensation, by modifying only the energy envelope, preserves them.

More generally, the fact that the t_{mix} was underestimated with respect to such late arriving echoes in this particular spatial configuration raises some questions as to the pertinence of a global, omnidirectional mixing time in such multi-volume or otherwise complex spaces. In the present example, there is clear evidence that there exists a subspace, mostly in the left hemisphere ($0° \leq \theta \leq 180° \; \forall \; \varphi$), that could be correctly described as a fully stochastic late reverberation tail starting from perhaps even

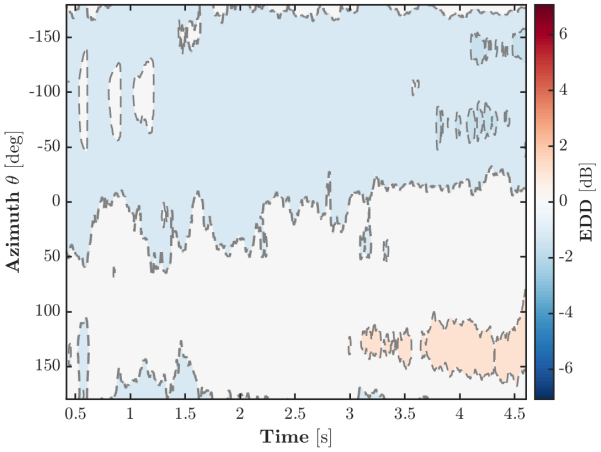
(a) View of the excited space with measurement source and receiver positions. Reverse angle in Fig. 5.18a.
Photo: Nadine Schütz.



(b) Reference denoised EDD, $\mathbf{\Omega}_{\mathrm{DS}} \simeq (89.1°, -14.9°)$.



(c) Isotropification by tail resynthesis, $\upsilon_{\mathrm{iso}} = 1$.



(d) Isotropification by EDD compensation, $\upsilon_{\mathrm{iso}}^{\mathrm{EDD}} = 1$.

Figure 5.17: Azimuthal plane EDDs demonstrating the effect of isotropification as applied by (c, bottom left) tail resynthesis and (d, bottom right) EDD compensation to the SRIR measured at the Grandes Serres de Pantin in Pantin, France. The azimuthal EDD of the "original" denoised SRIR is included in (b, top right) as a reference. Elevational plane EDDs are also available in Appx. B.7.

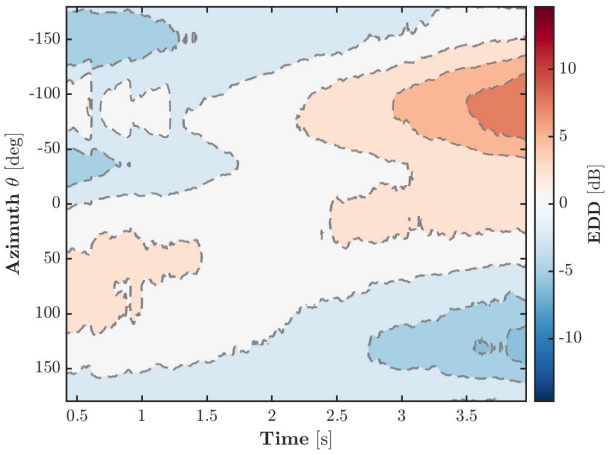
earlier than the t_{mix} estimate. Future research could therefore explore the possibility of identifying when such subspaces are present and then characterizing them with "localized" mixing times.

Directification

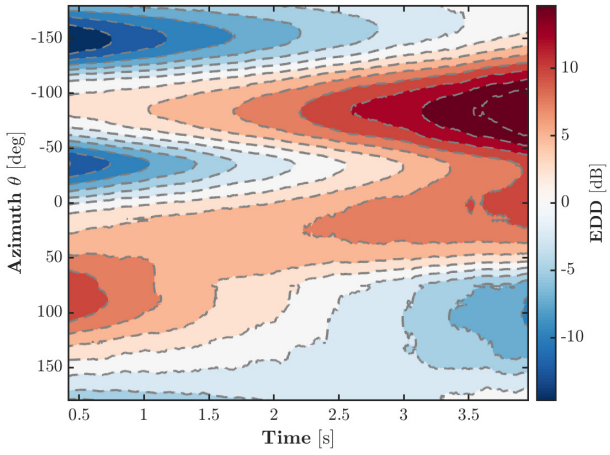
In contrast, Fig. 5.18 (b–d) shows the effect of the directification procedure. The tail resynthesis approach (Fig. 5.18c) is applied with a multiplicative control parameter value of $\upsilon_{\mathrm{dir}} = 2$ (see Sec. 3.3.2), while the EDD is accentuated by $\upsilon_{\mathrm{dir}}^{\mathrm{EDD}} = 18$ dB (Fig. 5.18d, see Sec. 3.4.2). Note that the reference EDD (Fig. 5.18b) is the same as in Fig. 5.17b but displayed over different time and EDD ranges.

Here, the difference between the two methods (tail resynthesis and EDD accentuation) is particularly interesting. Although, once again, the tail resynthesis implementation removes the late arriving echoes present after the estimated t_{mix}, it appears to better preserve the overall space-time evolution of the late reverberation tail's directional energy decay properties, at least in terms of their relative directional trends. On the other hand, the EDD accentuation seems to affect the early moments of the decay (i.e

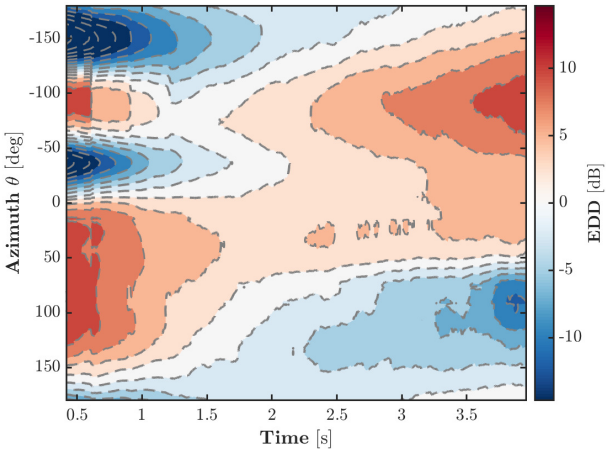
(a) View of the coupled volume (source position does not correspond to the current SRIR).
Photo: Nadine Schütz.



(b) Reference denoised EDD, $\mathbf{\Omega}_{\text{DS}} \simeq (89.1°, -14.9°)$.



(c) Directification by tail resynthesis, $\upsilon_{\text{dir}} = 2$.



(d) Directification by EDD accentuation, $\upsilon_{\text{dir}}^{\text{EDD}} = 18$ dB.

Figure 5.18: Azimuthal plane EDDs demonstrating the effect of directification as applied by (c, bottom left) tail resynthesis and (d, bottom right) EDD compensation to the SRIR measured at the Grandes Serres de Pantin in Pantin, France. The azimuthal EDD of the "original" denoised SRIR is included in (b, top right) as a reference. Elevational plane EDDs are also available in Appx. B.7.

around those infamous late echoes) more than it does the later part, thereby modifying the EDD's overall "shape". Whether or not this *should* be a property of the directification effect is, however, entirely a matter of design choice.

Such is, finally, one of the main goals of this thesis: that the space-time-frequency analysis-treatment-manipulation framework developed through this research project allow for unrestricted creative control of the spatialized reverberation effect. The two ideas presented above, as well as the early reflection manipulations demonstrated in Sec. 5.3.3, provide an initial proof of concept to this end. As mentioned in the Introduction (and further outlined in the Conclusion to follow), a natural extension would then be the addition of a perceptive control layer capable of "translating" the technical signal treatments into intuitive parameters.

Thus concludes the principal content of this dissertation. This final chapter has offered concrete examples of the analysis, treatment, and manipulation methods theoretically described throughout Chs. 2 and 3 by applying them to real-world SMA SRIR measurements. These begin with the pre-analysis treatment procedure, whose ability to mitigate the impact of impulsive non-stationary noises is demonstrated in Sec. 5.1. Several examples of mixing time estimations are then shown in Sec. 5.2, and the results obtained using both the SH-domain diffuseness ψ_{cov} and DRIR incoherence ψ_{dir} measures are compared.

Section 5.3 is dedicated to the detection of early reflections, with Sec. 5.3.1 focusing on the direct sound, Sec. 5.3.2 on generating space-time echo cartographies, and Sec. 5.3.3 providing proofs of concepts for the proposed echo modification methods. As per the usual structure of this thesis, these are followed by a significant segment on late reverberation (Sec. 5.4), which begins with a discussion on coupled-volume measurements and their manifestation as either highly isotropic double-slope decays or directionally separated anisotropic single decays (Sec. 5.4.1). A straightforward example of the directional denoising process as applied to a severely anisotropic SRIR is then given in Sec. 5.4.2. Finally, the last two sections deal directly with late reverberation tail isotropy: first with respect to measuring and evaluating it (Sec. 5.4.3), and second in terms of modifying it (Sec. 5.4.4).

The various demonstrations of the proposed manipulation methods (early reflection salience modification and echo redistribution, Sec. 5.3.3, and late reverberation tail isotropification and directification, Sec. 5.4.4), in particular, represent one of the main achievements of this research project. Although the fundamental evaluations on simulated examples from Ch. 4 serve to validate the underlying signal models and analysis methods, the implementation of these manipulation methods paves the way for the definition of a comprehensive top-down SRIR control framework, ideally informed by a perceptual model. A roadmap towards this objective is outlined in the Conclusion to follow.

Conclusion

This dissertation has attempted to encapsulate the various work carried out over the course of the author's doctoral research project. Its principal objective, as stated in the Introduction, was the construction of a complete space-time-frequency framework enabling the analysis, treatment, and manipulation of SMA SRIR measurements (schematically represented in Fig. II, a view reprised here below in Fig. III). To close out this thesis, we now offer in the following conclusion: a short summary of the dissertation itself, a list of the different publications that have resulted from this research (both peer-reviewed journal papers and articles accompanying conference presentations), a brief discussion of the work accomplished with regard to the initial goals, and finally an overview of the various topics that could potentially extend this work in the future.

Summary

This dissertation begins with a thorough overview of the theoretical underpinnings required in order to define the space-time-frequency framework's underlying signal models (Ch. 1). These range from the historical geometrical approach to room acoustics, and the first statistical methods used to probe the phenomenon of reverberation, to the state of the art in SMA and SRIR processing. A brief presentation of the perceptual evaluation of room reverberation effects is also given.

Chapter 2 sets the essential foundations for the analysis framework, starting with a global view of the SRIR and then developing specific signal models for each of the individual acoustical regimes identified within. Most notable is the temporal sectioning of the SRIR into an "early" segment composed of highly coherent, spatiotemporally resolvable reflections of the direct sound, and a stochastic, incoherent, and exponentially decaying "late" reverberation tail. Consequently, the first analysis method to be presented, following a description of spatial decomposition methods used to obtain a directional view of the SRIR (i.e. a DRIR), is the estimation of the moment at which the transition between these two segments operates, a property characterized as the "mixing time" (t_{mix}). The second half of the chapter is subsequently dedicated to the detection of the early reflections (including the direct sound) and the modelling of the late reverberation tail's exponential energy decay envelope.

Chapter 3 introduces the different treatment and manipulation methods made possible by the analysis and modelling of Ch. 2, and in the process clarifies the conceptual difference between treating and manipulating SRIRs. In a somewhat contradictory manner, the chapter in fact opens with a treatment procedure designed to operate directly on the SMA measurement signals and before any of the analysis processes; "out of turn", in a sense, with respect to the schematic view of Figs. II and III[4]. This is followed by the description of some potential early reflection manipulation strategies, first looking to modify their energies relative to the exponential decay envelope of the late reverberation tail, then aiming to spatially redistribute the echoes according to a measure of "directional echo energy density" (DEED). Continuing what then becomes the usual chapter structure for this dissertation, treatment and manipulation methods for the late reverberation tail are presented next. Treatment

[4]The secondary blue line surrounding the "Measurement" layer in Figs. II and III is meant to represent just this.

comes in the form of a denoising procedure that replaces the inevitable background measurement noise floor with a resynthesized prolongation of the reverberation tail. The manipulation methods involve either resynthesizing the complete tail using modified decay parameters, or directly affecting the space-time-frequency representation based on a measure of the energy decay deviation (EDD).

The aim of Ch. 4 is to present the plethora of (more or less) fundamental evaluations carried out in order to verify the different analysis methods (including the DRIR representation) theorized in Ch. 2, and the aforementioned denoising treatment procedure. In fact, a thorough investigation into the constituent aspects of the DRIR definition makes up the entire first part of the chapter. This is mainly due to the critical importance of the DRIR throughout the analysis-treatment-manipulation framework. Breaking the usual chapter structure somewhat, the complete validation tests for all the analysis methods are presented next, both for the early reflections and the late reverberation. These tests make use of simulated SRIR SMA measurements specifically designed to help assess the performance of each method. The chapter finally ends with an evaluation of the denoising treatment procedure as also applied to simulated (noisy) SRIRs.

Chapter 5 concludes the main content of the dissertation by proposing a comprehensive demonstration of the space-time-frequency analysis-treatment-manipulation framework's full capabilities through its application to "real-world" SMA SRIR measurements. As noted in the Introduction, these SRIRs have all been measured with the commercially available 32-capsule mh acoustics Eigenmike® SMA (which is, quite intentionally, the SMA simulated in Ch. 4). Here, every facet of the framework is highlighted, from the pre-analysis treatment procedure to the different manipulation techniques. Examples of spaces with particularly complex or unusual acoustic properties (at least with respect to the traditional geometrical acoustics description of room reverberation) are deliberately chosen in order to demonstrate the expanded capacities of the approach proposed in this thesis. This final chapter also serves as a general proof of concept for the SRIR modification strategies, whose successful implementation is, in a sense, one of the ultimate goals of this work.

List of Publications

In order to emphasize the scientific output of the research conducted during this doctorate, we present the complete list of published communications directly tied to the work completed within.

Journal Papers

[84] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "A Robust Denoising Process for Spatial Room Impulse Responses with Diffuse Reverberation Tails," *The Journal of the Acoustical Society of America*, 2020

[78] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Denoising Directional Room Impulse Responses with Spatially Anisotropic Late Reverberation Tails," *Applied Sciences*, 2020

[94] B. Alary, P. Massé, S. Schlecht, M. Noisternig, and V. Välimäki, "Perceptual Analysis of Directional Late Reverberation," *The Journal of the Acoustical Society of America*, 2021

Conference Proceedings

[104] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Refinement and Implementation of a Robust Directional Room Impulse Response Denoising Process, Including Applications to Highly Varied Measurement Databases," *26th International Congress on Sound and Vibration*, Montréal, Canada, 2019

[91] B. Alary, P. Massé, V. Välimäki, and M. Noisternig, "Assessing the Anisotropic Features of Spatial Impulse Responses," *EAA Spatial Audio Signal Processing Symposium*, Paris, France, 2019

[105] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Measurement, Analysis, and Denoising of Directional Room Impulse Responses in Complex Spaces," *Forum Acusticum*, Lyon, France, 2020

Discussion

At the risk of undue repetition, the original overarching objective of this research project was the definition and eventual construction of a self-contained, top-down framework for the space-time-frequency analysis, treatment, and manipulation of SRIRs measured with SMAs. The general structure of this framework was summarized schematically in the Introduction by Fig. II, and has been reprised here as Fig. III with the addition of a top-level control layer. This extension is deliberately included in order to open the discussion with respect to the specific goals achieved in this work and their role within the wider context of SRIR convolution-based artificial reverberation reproduction.

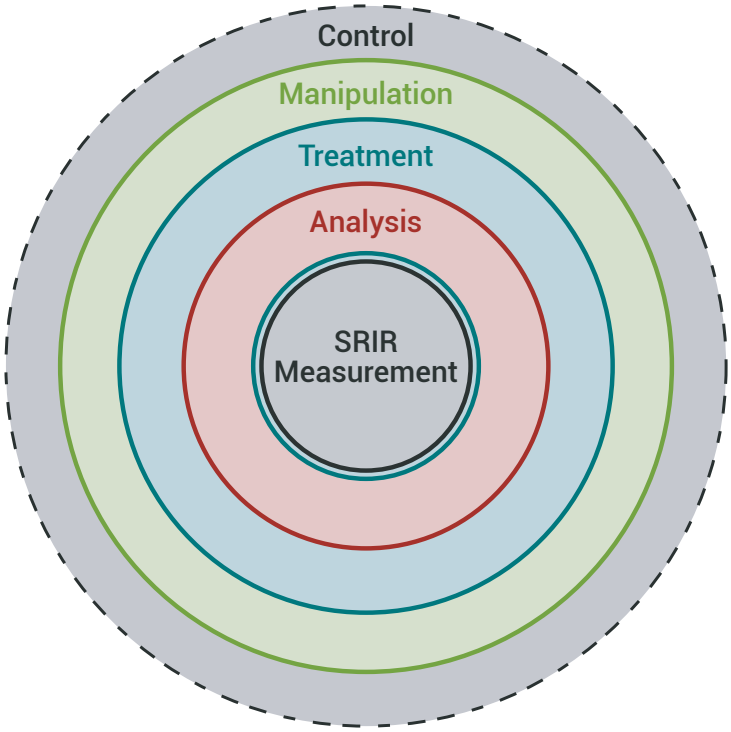


Figure III: Schematic presentation of the analysis-treatment-manipulation framework as presented in the Introduction (Fig. II), with the addition of a top-level control layer.

In particular, the idea of a high-level perceptual control system, in the vein of the "one knob" paradigm mentioned in the Introduction, remains mostly unrealized. Although some basic perceptual factors have been considered throughout this work (e.g. when evaluating averages over frequency), none of the well-known descriptors of room reverberation, such as the ones presented in Sec. 1.1.3, have been successfully tied to the objective parameters or measures obtained through the analysis and modelling of the SRIR. Nonetheless, the demonstration of manipulation techniques capable of deeply altering several space-time-frequency characteristics of the SRIR shows that the analysis and treatment methods developed in this thesis provide a plausible foundation for the subsequent definition, evaluation, and implementation of such higher-level control strategies.

The principal extension to the work accomplished in this thesis would therefore be an abstraction of the manipulation layer in Fig. III, and therefore of the different objective quantities parameterizing it (e.g. the DEED, the T_{60}, the EDD), to an outer control layer defined almost exclusively in terms of perceptual qualities (e.g. "reverberance", "envelopment", "source width", etc. – see Sec. 1.1.3 again). To this end, it would be interesting to continue in the vein of contemporary work such as e.g. Dick and Vigeant [28] but with a direct link to the analysis and manipulation methods from this thesis.

Potential Topics for Future Work

Though the discussion above highlights the most fundamental manner in which the research from this doctorate could be carried forward, there also exists a plethora of additional details, hinted at throughout this dissertation, that may also merit further study.

The first of these was mentioned several times during the evaluation of the spatial decomposition methods used to define the DRIR representation (Sec. 4.1). One of the main limitations to the definition retained in this work is the inevitable trade-off between obtaining optimized directional power estimates over the sphere (Sec. 4.1.3) and preserving the spatial coherence (or *in*coherence) properties of the SRIR (Sec. 4.1.4). Unfortunately, the results from each section lead to different conclusions: the MWDI beamformer design (Appx. A.2) seems better adapted to representing directional power envelopes, but the natural PWD beamformer is the best at preserving spatial incoherence. For the late reverberation tail resynthesis applications (e.g. Secs. 4.3.1 and 5.4.2), accuracy in the analysis of the directional decay parameters is thus sacrificed in order to ensure that the coherence properties remain as unaffected as possible.

The search for an improved directional representation of the SRIR therefore constitutes one of the main subjects to which future research could be dedicated. In this view we can cite such ideas as the spherically localized SH transform (SLSHT) [50] (which was briefly explored during this project but could not be fully investigated due to time constraints), spherical sector-based approaches [54] [67], or a regularized reconstruction of the sound field [86] as possible starting points. Perhaps a reformulation of the MWDI beamformer design, extended to include the frequency-dependent characteristics of SMA-HOA encoding, could also be envisioned. In any case, it is clear that the current state of the work produced in this thesis would greatly benefit from additional research into more advanced spatial filter-banks; furthermore, it is worth noting that any such improvements would be directly applicable to the framework presented in this dissertation.

With respect to the detection of early reflections, the development of novel localization techniques specifically geared towards the identification of multiple simultaneous and potentially correlated sources would no doubt help improve the generation of space-time echo cartographies (Secs. 4.2.1 and 5.3.2). For example, perhaps the under-determined Herglotz inversion method could be regularized using more optimized methods than the simple identity Tikhonov matrix used in this work, and perhaps the chosen regularization method could then be parameterized by considering multiple incident echoes (instead of the single echo localization conditions used in this work).

Another briefly discussed limitation is that of the multi-slope decay detection algorithm (Sec. 4.2.3) and its reliance on classic broadband acoustical coupling theory as given by Cremer and Müller [5]. Not only is this "traditional" description independent of frequency, it is also restricted to two-volume configurations. Extending the coupling formalism to include a frequency dependence as well as the possibility for higher numbers of spaces (i.e. $C > 2$) would allow for a much more robust identification of multi-slope decays. Indeed, if the EDR analysis process could be informed by an underlying coupling model when multi-slope conditions are suspected, confidence in the estimations of the coupled decay parameters, i.e. $\{P_{0,\mathrm{I}}, P_{0,\mathrm{II}}, \ldots, P_{0,C}\}(f)$ and $\{T_{60,\mathrm{I}}, T_{60,\mathrm{II}}, \ldots, T_{60,C}\}(f)$, would be far greater, and the corresponding "uncoupled" values $\{P_{0,1}, P_{0,2}, \ldots, P_{0,C}\}(f)$ and $\{T_{60,1}, T_{60,2}, \ldots, T_{60,C}\}(f)$ could even

be estimated. However, the question of the pertinence of a directional view of multi-slope decays, as posed in Sec. 5.4.1, would still be open, suggesting that the potential extension(s) described in this paragraph might also need to add a spatial dimension to the classic Cremer-Müller theory.

In fact, as we saw in Sec. 5.4.1, a slightly related issue to the topic of coupled volumes is that of directional subspaces (from the point of view of an SMA) with different reverberation characteristics and, crucially, different mixing times. Whereas the former is covered by the late reverberation tail model used in this thesis (which allows for direction-dependent decay properties), the mixing time is still considered an "omnidirectional" measure. As discussed in Chs. 1 and 2, it has already evolved from being an entirely "global" quantity, truly unique to each reverberating or "mixing" space in the sense of Polack [6], to being an RIR-dependent measure in more contemporary analysis approaches. This allows for different t_{mix} values to be observed within a single acoustic space, depending on the specific behaviour of the early reflections (and cluster) in different source/receiver positions. However, the observations related in Sec. 5.4.1 suggest that both approaches may have similar merit, and that a directional approach could also be taken with respect to the mixing time, perhaps even enabling the identification of reverberant directional subspaces.

To finish, a quick note on the evaluation of the perceptibility of anisotropic late reverberation, as mentioned in Sec. 5.4.3: although the basic capacity to perceive such anisotropy was demonstrated in Alary et al. [94], further studies would be necessary in order to determine a JND in terms of the EDD range isotropy measure proposed in Sec. 2.5.2 (Eq. 2.37). Such a JND would be useful for example in determining whether the denoising procedure (and thus the required late reverberation tail analysis as well) can be performed omnidirectionally in the SH domain (i.e. on the HOA-encoded SRIR), or if it must be done anisotropically on the DRIR representation, which entails additional processing and is subjected to the constrained spatial decomposition conditions reiterated above.

Once again, despite the limitations discussed relative to the work accomplished during this doctorate, we have managed to demonstrate that the proposed framework for the space-time-frequency analysis, treatment, and manipulation of SMA SRIR measurements has the capacity to model and subsequently modify the principal objective characteristics of spatialized reverberation effects, as reproduced by multi-channel SRIR convolution. Considering the current momentum and unrelenting demand for innovative tools in the fields of spatial audio and immersive sound, there is a decent chance that this research will be able to provide a practical basis for future developments in those disciplines (some of which have been outlined above). For now, as Porky would say:

Th-th-that's all, folks!

Bibliography

[1] M. R. Schroeder, "Die statistischen Parameter der Frequenzkurven von großen Räumen," *Acustica*, vol. 4, no. 2, pp. 594–600, 1954. (Cited on pages 5, 6, 7, and 24.)

[2] ——, "Natural-Sounding Artificial Reverberation," *Journal of the Audio Engineering Society*, vol. 10, no. 3, pp. 219–223, 1962. (Cited on pages 5, 7, 8, and 24.)

[3] ——, "New Method of Measuring Reverberation Time," *The Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965. (Cited on pages 5 and 46.)

[4] K. H. Kuttruff, *Room Acoustics*, 4th ed. London, U.K.: Spon Press, 2000. (Cited on pages 5, 24, and 28.)

[5] L. Cremer and H. A. Müller, *Principles and Applications of Room Acoustics, vol. 1.* Barking, England: Applied Science Publishers, 1982. (Cited on pages 5, 6, 49, 86, 87, 113, 114, 115, and 128.)

[6] J.-D. Polack, "Playing Billiards in the Concert Hall: The Mathematical Foundations of Geometrical Room Acoustics," *Applied Acoustics*, vol. 38, no. 2-4, pp. 235–244, 1993. (Cited on pages 5, 6, 26, and 129.)

[7] ——, "La transmission de l'énergie sonore dans les salles," Ph.D. dissertation, Université du Maine, 1988. (Cited on pages 6, 7, 46, and 86.)

[8] M. R. Schroeder and K. H. Kuttruff, "On Frequency Response Curves in Rooms: Comparison of Experimental, Theoretical, and Monte Carlo Results for the Average Frequency Spacing between Maxima," *The Journal of the Acoustical Society of America*, vol. 34, no. 76, 1962. (Cited on pages 6, 7, and 24.)

[9] K. H. Kuttruff and R. Thiele, "Über die Frequenzabhängigkeit des Schalldrucks in Räumen," *Acustica*, vol. 4, no. 2, pp. 614–617, 1954. (Cited on page 6.)

[10] M. A. Gerzon, "Synthetic Stereo Reverberation, part i," *Studio Sound*, vol. 13, pp. 632–635, 1971. (Cited on page 7.)

[11] ——, "Synthetic Stereo Reverberation, part ii," *Studio Sound*, vol. 14, pp. 24–28, 1972. (Cited on page 7.)

[12] J. Stautner and M. Puckette, "Designing Multi-Channel Reverberators," *Computer Music Journal*, vol. 6, no. 1, p. 52, 1982. (Cited on pages 7 and 8.)

[13] J.-M. Jot and A. Chaigne, "Digital Delay Networks for Designing Artificial Reverberators," in *Proceedings of the 90th Audio Engineering Society Convention*, Paris, France, 1991. (Cited on pages 7 and 8.)

[14] J. A. Moorer, "About This Reverberation Business," *Computer Music Journal*, vol. 3, no. 2, pp. 13–28, 1979. (Cited on pages 7 and 26.)

[15] T. Carpentier, M. Noisternig, and O. Warusfel, "Hybrid Reverberation Processor With Perceptual Control," in *Proceedings of the 17th International Conference on Digital Audio Effects*, Erlangen, Germany, 2014. (Cited on pages 7, 8, 9, 10, 39, and 86.)

[16] J.-M. Jot, L. Cerveau, and O. Warusfel, "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model," in *Proceedings of the 103rd Audio Engineering Society Convention*, New York, U.S.A., 1997. (Cited on pages 7, 8, 35, 46, 56, 60, and A4.)

[17] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973. (Cited on pages 8 and 10.)

[18] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université de Paris 6, 2001. (Cited on pages 8 and 13.)

[19] J. R. Driscoll and D. M. J. Healy, "Computing Fourier Transforms and Convolutions on the 2-Sphere," *Advances in Applied Mathematics*, vol. 15, pp. 202–250, 1994. (Cited on pages 8 and 10.)

[20] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *Proceedings of the 108th Audio Engineering Society Convention*, Paris, France, 2000. (Cited on pages 8, 42, 53, and 101.)

[21] P. Massé, "Réverbérateur hybride basé sur des réponses impulsionnelles directionnelles," Master's dissertation, Sorbonne Université, 2018. (Cited on pages 8 and 35.)

[22] M. Barron and A. H. Marshall, "Spatial Impression Due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," *Journal of Sound and Vibration*, vol. 77, no. 2, pp. 211–232, 1981. (Cited on page 9.)

[23] T. Okano, L. L. Beranek, and T. Hidaka, "Relations Among Interaural Cross-Correlation Coefficient ($IACC_E$), Lateral Fraction (LF_E), and Apparent Source Width (ASW) in Concert Halls," *The Journal of the Acoustical Society of America*, vol. 104, no. 1, pp. 255–265, 1998. (Cited on page 9.)

[24] J. S. Bradley and G. A. Soulodre, "Objective Measures of Listener Envelopment," *The Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2590–2597, 1995. (Cited on page 10.)

[25] L. L. Beranek, "Listener Envelopment LEV, Strength G and Reverberation time RT in Concert Halls," in *Proceedings of the 20th International Congress on Acoustics*, Sydney, Australia, 2010. (Cited on page 10.)

[26] M. Morimoto and K. Iida, "A New Physical Measure for Psychological Evaluation of a Sound Field: Front/Back Energy Ratio as a Measure for Envelopment," *The Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 2282–2282, 1993. (Cited on page 10.)

[27] T. Hanyu and S. Kimura, "New Objective Measure for Evaluation of Listener Envelopment Focusing on the Spatial Balance of Reflections," *Applied Acoustics*, vol. 62, no. 2, pp. 155–184, 2001. (Cited on page 10.)

[28] D. A. Dick and M. C. Vigeant, "An Investigation of Listener Envelopment Utilizing a Spherical Microphone Array and Third-Order Ambisonics Reproduction," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2795–2809, 2019. (Cited on pages 10, 21, and 128.)

[29] E. Kahle, "Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras," Ph.D. dissertation, Université du Maine, 1995. (Cited on pages 10 and 25.)

[30] E. Kahle and J.-P. Jullien, "Subjective Listening Tests in Concert Halls: Methodology and Results," in *Proceedings of the 15th International Congress on Acoustics*, Trondheim, Norway, 1995. (Cited on page 10.)

[31] J. Meyer and G. Elko, "A Highly Scalable Spherical Microphone Array Based On an Orthonormal Decomposition of the Soundfield," in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, U.S.A., 2002, pp. 1781–1784. (Cited on page 10.)

[32] T. D. Abhayapala and D. B. Ward, "Theory and Design of High Order Sound Field Microphones Using Spherical Microphone Array," in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, Orlando, U.S.A., 2002, pp. 1949–1952. (Cited on pages 10, 11, and 31.)

[33] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone Array Measurement System for Analysis of Directional and Spatial Variations of Sound Fields," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 1980–1991, 2002. (Cited on page 10.)

[34] B. Rafaely, "Analysis and Design of Spherical Microphone Arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005. (Cited on page 10.)

[35] ——, "Spatial Sampling and Beamforming for Spherical Microphone Arrays," in *Proceedings of the 2008 Hands-Free Speech Communication and Microphone Arrays*, Trento, Italy, 2008. (Cited on pages 10, 13, 14, and 72.)

[36] F. Zotter, "Analysis and Synthesis of Sound-Radiation with Spherical Arrays," Ph.D. dissertation, University of Music and Performing Arts, Graz, Austria, 2009. (Cited on page 10.)

[37] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Cambridge, U.S.A.: Academic Press, 1999. (Cited on pages 10, 11, 13, and A1.)

[38] G. Chardon, W. Kreuzer, and M. Noisternig, "Design of Spatial Microphone Arrays for Sound Field Interpolation," *IEEE Journal on Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 780–790, 2015. (Cited on page 11.)

[39] M. Noisternig, F. Zotter, and B. F. G. Katz, "Reconstructing Sound Source Directivity in Virtual Acoustic Environments," in *Principles and Applications of Spatial Hearing*, Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato, Eds. World Scientific Publishing Co. Pte. Ltd., 2011, pp. 357–373. (Cited on pages 11 and 31.)

[40] S. Moreau and J. Daniel, "Study of Higher Order Ambisonic Microphone," in *Proceedings of the 7th French Acoustics Congress (CFA)*, Strasbourg, France, 2004. (Cited on page 13.)

[41] J. Daniel and S. Moreau, "Further Study of Sound Field Coding with Higher Order Ambisonics," in *Proceedings of the 116th Audio Engineering Society Convention*, Berlin, Germany, 2004. (Cited on page 13.)

[42] B. Rafaely, "Plane-Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, 2004. (Cited on pages 14 and 15.)

[43] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 6th ed. Elsevier, 2005. (Cited on page 14.)

[44] 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, 2007. (Cited on page 15.)

[45] B. Rafaely, *Fundamentals of Spherical Array Processing*. Berlin: Springer, 2015. (Cited on pages 15, 34, 65, and A3.)

[46] D. L. Alon and B. Rafaely, "Beamforming with Optimal Aliasing Cancellation in Spherical Microphone Arrays," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 1, pp. 196–210, 2016. (Cited on page 15.)

[47] A. Koretz and B. Rafaely, "Dolph–Chebyshev Beampattern Design for Spherical Arrays," *IEEE Transactions on Signal Processing*, 2009. (Cited on pages 15, 34, and 65.)

[48] B. T. T. Yeo, W. Ou, and P. Golland, "On the Construction of Invertible Filter Banks on the 2-Sphere," *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 283–300, 2008. (Cited on page 15.)

[49] F. J. Narcowich, P. Petrushev, and J. D. Ward, "Localized Tight Frames on Spheres," *SIAM Journal on Mathematical Analysis*, vol. 38, no. 2, pp. 574–594, 2006. (Cited on page 15.)

[50] Z. Khalid, R. A. Kennedy, S. Durrani, P. Sadeghi, Y. Wiaux, and J. D. McEwen, "Fast Directional Spatially Localized Spherical Harmonic Transform," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2192–2203, 2013. (Cited on pages 15 and 128.)

[51] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*. Geneva, Switzerland: Springer, 2017, vol. 9. (Cited on pages 16 and 17.)

[52] L. McCormack, A. Politis, and V. Pulkki, "Sharpening of Angular Spectra Based on a Directional Re-assignment Approach for Ambisonic Sound-field Visualisation," in *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 576–580. (Cited on page 16.)

[53] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D Source Localization in the Spherical Harmonic Domain Using a Pseudointensity Vector," in *Proceedings of the 18th European Signal Processing Conference*, Aalborg, Denmark, 2010. (Cited on page 16.)

[54] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain," *IEEE Journal on Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, 2015. (Cited on pages 17, 34, and 128.)

[55] B. Jo, F. Zotter, and J.-W. Choi, "Extended Vector-Based EB-ESPRIT Method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020. (Cited on pages 17 and 40.)

[56] A. Herzog and E. A. P. Habets, "Eigenbeam-ESPRIT for DOA-Vector Estimation," *IEEE Signal Processing Letters*, vol. 26, no. 4, pp. 572–576, apr 2019. (Cited on page 17.)

[57] B. Jo and J.-W. Choi, "Spherical Harmonic Smoothing for Localizing Coherent Sound Sources," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 10, pp. 1969–1984, 2017. (Cited on page 17.)

[58] T. Padois, O. Doutres, and F. Sgard, "On the Use of Modified Phase Transform Weighting Functions for Acoustic Imaging with the Generalized Cross Correlation," *The Journal of the Acoustical Society of America*, vol. 145, no. 3, pp. 1546–1555, 2019. (Cited on pages 17 and 44.)

[59] S. Tervo, J. Pätynen, and T. Lokki, "Acoustic Reflection Localization from Room Impulse Responses," *Acta Acustica United with Acustica*, vol. 98, no. 3, pp. 418–440, 2012. (Cited on pages 17, 41, and 44.)

[60] J. Ahonen and V. Pulkki, "Diffuseness Estimation Using Temporal Variation of Intensity Vectors," in *Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, U.S.A.: IEEE, 2009, pp. 285–288. (Cited on pages 18 and 40.)

[61] D. P. Jarrett, O. Thiergart, E. A. P. Habets, and P. A. Naylor, "Coherence-Based Diffuseness Estimation in the Spherical Harmonic Domain," in *Proceedings of the 27th IEEE Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, 2012. (Cited on pages 19 and 59.)

[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, 2016. (Cited on pages 19, 30, 36, 37, 59, 66, and 70.)

[63] M. Nolan, E. Fernandez-Grande, J. Brunskog, and C.-H. Jeong, "A Wavenumber Approach to Quantifying the Isotropy of the Sound Field in Reverberant Spaces," *The Journal of the Acoustical Society of America*, vol. 143, no. 4, pp. 2514–2526, 2018. (Cited on pages 19, 29, 60, and 91.)

[64] 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. (Cited on page 19.)

[65] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007. (Cited on pages 19 and 20.)

[66] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1-2, pp. 17–28, 2013. (Cited on pages 19, 20, and 40.)

[67] 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. (Cited on pages 20 and 128.)

[68] M. Berzborn and M. Vorländer, "Investigations on the Directional Energy Decay Curves in Reverberation Rooms," in *Proceedings of the 11th European Congress and Exposition on Noise Control Engineering*, Heraklion, Crete, 2018. (Cited on pages 20, 29, and 50.)

[69] B. Alary, A. Politis, S. J. Schlecht, and V. Välimäki, "Directional Feedback Delay Network," *Journal of the Audio Engineering Society*, vol. 67, no. 10, pp. 752–762, oct 2019. (Cited on page 20.)

[70] R. Badeau, "Common Mathematical Framework for Stochastic Reverberation Models," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2733–2745, 2019. (Cited on page 20.)

[71] A. Meacham, R. Badeau, and J.-D. Polack, "Lower Bound on Frequency Validity of Energy-Stress Tensor Based Diffuse Sound Field Model," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, 2019, pp. 6012–6019. (Cited on page 21.)

[72] H. Furuya, K. Fujimoto, Y. Takeshima, and H. Nakamura, "Effect of Early Reflections from Upside on Auditory Envelopment," *Journal of the Acoustical Society of Japan*, vol. 16, no. 2, pp. 97–104, 1995. (Cited on page 21.)

[73] M. Morimoto, K. Iida, and K. Sakagami, "Role of Reflections from Behind the Listener in Spatial Impression," *Applied Acoustics*, vol. 62, no. 2, pp. 109–124, 2001. (Cited on page 21.)

[74] J.-M. Jot, "Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality," in *Proceedings of the 1997 International Computer Music Conference*, Thessaloniki, Greece, 1997, pp. 236–243. (Cited on page 25.)

[75] M. de Avelar Gomes, P. Bonifacio, E. Brandao, L. Sant'Ana, E. Bertoti, R. Catai, and H. Azikri de Deus, "Crossover frequency estimation from statistical features of a room transfer function," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, 2019, pp. 4529–4536. (Cited on page 25.)

[76] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 3rd ed. New York, U.S.A.: Springer, 2013. (Cited on pages 26, A1, and A2.)

[77] G. I. Koutsouris, J. Brunskog, C.-H. Jeong, and F. M. Jacobsen, "Combination of Acoustical Radiosity and the Image Source Method," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 3963–3974, 2013. (Cited on page 28.)

[78] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Denoising Directional Room Impulse Responses with Spatially Anisotropic Late Reverberation Tails," *Applied Sciences*, vol. 10, no. 3, p. 1033, 2020. (Cited on pages 31, 34, 35, 46, 59, 64, 119, and 126.)

[79] J. Fliege and U. Maier, "The Distribution of Points on the Sphere and Corresponding Cubature Formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999. (Cited on pages 33, 37, 66, 90, and B27.)

[80] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-Preserving Ambisonic Decoding," *Acta Acustica United with Acustica*, vol. 98, no. 1, pp. 37–47, 2012. (Cited on page 34.)

[81] J. S. Abel and P. Huang, "A Simple, Robust Measure of Reverberation Echo Density," in *Proceedings of the 121st Audio Engineering Society Convention*, San Francisco, U.S.A., 2006. (Cited on page 35.)

[82] G. Defrance and J.-D. Polack, "Measuring the Mixing Time in Auditoria," in *Proceedings of Acoustics '08*, Paris, France, 2008. (Cited on page 35.)

[83] P. Götz, K. Kowalczyk, A. Silzle, and E. A. P. Habets, "Mixing Time Prediction Using Spherical Microphone Arrays," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. EL206–EL212, 2015. (Cited on page 35.)

[84] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "A Robust Denoising Process for Spatial Room Impulse Responses with Diffuse Reverberation Tails," *The Journal of the Acoustical Society of America*, vol. 147, no. 4, pp. 2250–2260, 2020. (Cited on pages 35, 46, 54, 59, 64, 119, and 126.)

[85] D. K. Prasad, M. K. Leung, C. Quek, and S. Y. Cho, "A Novel Framework for Making Dominant Point Detection Methods Non-Parametric," *Image and Vision Computing*, vol. 30, no. 12, pp. 843–859, 2012. (Cited on page 39.)

[86] E. Fernandez-Grande, "Sound Field Reconstruction using a Spherical Microphone Array," *The Journal of the Acoustical Society of America*, vol. 139, no. 3, pp. 1168–1178, 2016. (Cited on pages 40, 128, and A4.)

[87] C. L. Lawson, "$\mathcal{C}^1$ Surface Interpolation for Scattered Data on a Sphere," *Rocky Mountain Journal of Mathematics*, vol. 14, no. 1, pp. 177–202, 1984. (Cited on pages 41, 45, 50, 78, 90, 95, 116, 117, 118, and 119.)

[88] A. H. Marshall and M. Barron, "Spatial Responsiveness in Concert Halls and the Origins of Spatial Impression," *Applied Acoustics*, vol. 62, no. 2, pp. 91–108, 2001. (Cited on page 42.)

[89] P. N. Samarasinghe, T. D. Abhayapala, and H. Chen, "Estimating the Direct-To-Reverberant Energy Ratio Using a Spherical Harmonics-Based Spatial Correlation Model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 310–319, 2017. (Cited on page 42.)

[90] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. AC-19, no. 6, pp. 716–723, 1974. (Cited on page 49.)

[91] B. Alary, P. Massé, V. Välimäki, and M. Noisternig, "Assessing the Anisotropic Features of Spatial Impulse Responses," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*, Paris, France, 2019, pp. 43–48. (Cited on pages 50 and 127.)

[92] S. Vesa, "Binaural Sound Source Distance Learning in Rooms," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 8, pp. 1498–1507, 2009. (Cited on page 56.)

[93] Y. C. Lu and M. Cooke, "Binaural Estimation of Sound Source Distance via the Direct-to-Reverberant Energy Ratio for Static and Moving Sources," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 7, pp. 1793–1805, 2010. (Cited on page 56.)

[94] B. Alary, P. Massé, S. J. Schlecht, M. Noisternig, and V. Välimäki, "Perceptual Analysis of Directional Late Reverberation," *The Journal of the Acoustical Society of America*, vol. 149, no. 5, pp. 3189–3199, 2021. (Cited on pages 60, 119, 126, and 129.)

[95] I. H. Sloan and R. S. Womersley, "Extremal Systems of Points and Numerical Integration on the Sphere," *Advances in Computational Mathematics*, vol. 21, no. 1-2, pp. 107–125, 2004. (Cited on page 66.)

[96] N. J. A. Sloane and J. H. Conway, *Spherical Packings, Lattices, and Groups*, 3rd ed. New York, U.S.A.: Springer, 1999. (Cited on pages 66 and 77.)

[97] A. H. Nuttall, "Some Windows with Very Good Sidelobe Behavior," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 1, pp. 84–91, 1981. (Cited on page 70.)

[98] M. Noisternig, B. F. G. Katz, S. Siltanen, and L. Savioja, "Framework for Real-Time Auralization in Architectural Acoustics," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 1000–1015, 2008. (Cited on pages 77 and 103.)

[99] S. Laine, S. Siltanen, T. Lokki, and L. Savioja, "Accelerated Beam Tracing Algorithm," *Applied Acoustics*, vol. 70, no. 1, pp. 172–181, 2009. (Cited on page 77.)

[100] B. F. G. Katz and E. A. Wetherill, "Fogg Art Museum Lecture Room: A Calibrated Recreation of the Birthplace of Room Acoustics," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 3009–3009, nov 2006. (Cited on page 77.)

[101] H. P. Tukuljac, T. P. Vu, H. Lissek, and P. Vandergheynst, "Joint Estimation of the Room Geometry and Modes with Compressed Sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 6882–6886. (Cited on page 107.)

[102] M. Lovedee-Turner and D. Murphy, "Three-Dimensional Reflector Localisation and Room Geometry Estimation Using a Spherical Microphone Array," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3339–3352, 2019. (Cited on page 107.)

[103] M. Noisternig, T. Carpentier, T. Szpruch, and O. Warusfel, "Denoising of Directional Room Impulse Responses Measured with Spherical Microphone Arrays," in *Proceedings of the 40th Annual German Congress on Acoustics (DAGA)*, Oldenburg, Germany, 2014, pp. 600–601. (Cited on page 119.)

[104] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Refinement and Implementation of a Robust Directional Room Impulse Response Denoising Process, Including Applications to Highly Varied Measurement Databases," in *Proceedings of the 26th International Congress on Sound and Vibration*, Montréal, Canada, 2019. (Cited on page 126.)

[105] ——, "Measurement, Analysis, and Denoising of Directional Room Impulse Responses in Complex Spaces," in *Proceedings of Forum Acusticum 2020*, Lyon, France, 2020. (Cited on page 127.)

A / Additional Math and Algorithms

A.1 / Herglotz Wavefunction Formalism

As described in Sec. 1.2.1, an incident unit-amplitude plane wave from the direction $\mathbf{\Omega}_d$ can be entirely described in the SH domain by [37, p. 227]:

$$\begin{aligned} X_{\text{in}}(k, \mathbf{r}, \mathbf{d}) &= e^{ik\mathbf{r}\cdot\mathbf{d}} \\ X_{\text{in}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) &= 4\pi \sum_{l=0}^{\infty} i^l j_l(kr) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d), \end{aligned} \tag{A.1}$$

which gives rise to the following scattered field in the case of a rigid sphere [37, p. 228]:

$$X_{\text{sc}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) = -4\pi \sum_{l=0}^{\infty} i^l \frac{h_l'(kr_s)}{j_l'(kr_s)} h_l(kr) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d), \tag{A.2}$$

resulting in the familiar total field:

$$\begin{aligned} X_{\text{PW}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) &= X_{\text{in}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) + X_{\text{sc}}(kr, \mathbf{\Omega}, \mathbf{\Omega}_d) \\ &= \sum_{l=0}^{\infty} b_l(kr) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) Y_{l,m}^*(\mathbf{\Omega}_d), \end{aligned} \tag{A.3}$$

where $b_l(kr) = 4\pi i^l \left[j_l(kr) - \frac{h_l'(kr_s)}{j_l'(kr_s)} h_l(kr) \right]$ is the mode strength/holographic function for a rigid sphere.

As further seen in Sec. 1.2.1, the implicit plane wave decomposition (PWD) hypothesis made here allows Eq. A.3 to be simplified under a sufficiently well-conditioned truncated discrete SH transform:

$$\begin{aligned} X_{l,m}(kr_s, \mathbf{\Omega}_d) &= \sum_q \alpha_q X_{\text{PW}}(kr_s, \mathbf{\Omega}_q, \mathbf{\Omega}_d) Y_{l,m}^*(\mathbf{\Omega}_q) \\ &= \sum_{l'=0}^{L} b_{l'}(kr_s) \sum_{m'=-l'}^{l'} Y_{l',m'}^*(\mathbf{\Omega}_d) \\ &\quad \times \sum_q \alpha_q Y_{l',m'}(\mathbf{\Omega}_q) Y_{l,m}^*(\mathbf{\Omega}_q) \\ &\approx \sum_{l'=0}^{L} b_{l'}(kr_s) \sum_{m'=-l'}^{l'} Y_{l',m'}^*(\mathbf{\Omega}_d) \delta_{l,l'} \delta_{m,m'} \\ &= b_l(kr_s) Y_{l,m}^*(\mathbf{\Omega}_d), \end{aligned} \tag{A.4}$$

thanks to the orthogonality of the SH basis.

Now, as presented in Sec. 2.1.2, the Herglotz wavefunction [76, p. 68]

$$v(k, \mathbf{r}) = \int_{S_r^2} g(\mathbf{\Omega}_d) e^{ik\mathbf{r}\cdot\mathbf{d}} d\mathbf{\Omega}_d, \tag{A.5}$$

describes the presence of a plane wave incident on a given (spherical, in this case) surface $\mathbf{r} = [r, \mathbf{\Omega}]$. The most important part of the Herglotz wavefunction is the *Herglotz kernel* $g(\mathbf{\Omega}_d)$ which acts somewhat like a probability function over the space of all possible incident directions $\mathbf{d} = [1, \mathbf{\Omega}_d]$. Without getting into the mathematical details, it is worth noting that the Herglotz kernel is a nice smooth, continuous (at least $\mathcal{C}^1$), and square integrable (i.e. L^2) function.

Colton and Kress [76, p. 68] show that if we describe a plane wave incident on a rigid sphere in the form of a Herglotz wavefunction:

$$\begin{aligned} v_{\text{in}}(k, \mathbf{r}_s) &= \int_{S^2_{r_s}} g(\mathbf{\Omega}_d) e^{ik\mathbf{r}_s \cdot \mathbf{d}} d\mathbf{\Omega}_d \\ v_{\text{in}}(kr_s, \mathbf{\Omega}) &= \int_{S^2_{r_s}} g(\mathbf{\Omega}_d) X_{\text{in}}(kr_s, \mathbf{\Omega}, \mathbf{\Omega}_d) d\mathbf{\Omega}_d, \end{aligned} \tag{A.6}$$

then the total field (including the rigid sphere's self-scattering) can be described in the same way:

$$\begin{aligned} v(kr_s, \mathbf{\Omega}) &= \int_{S^2_{r_s}} g(\mathbf{\Omega}_d) X_{\text{PW}}(kr_s, \mathbf{\Omega}, \mathbf{\Omega}_d) d\mathbf{\Omega}_d \\ &= \sum_{l=0}^{\infty} b_l(kr_s) \sum_{m=-l}^{l} Y_{l,m}(\mathbf{\Omega}) \int_{S^2_{r_s}} g(\mathbf{\Omega}_d) Y^*_{l,m}(\mathbf{\Omega}_d) d\mathbf{\Omega}_d. \end{aligned} \tag{A.7}$$

The above expression can thus be simplified by taking its SH transform in the same way as Eq. A.4:

$$\begin{aligned} v_{l,m}(f) &= \sum_{l'=0}^{L} b_{l'}(f) \sum_{m'=-l'}^{l'} \sum_{q} \alpha_q Y_{l',m'}(\mathbf{\Omega}_q) Y^*_{l,m}(\mathbf{\Omega}_q) \\ &\quad \times \int_{S^2_r} g(\mathbf{\Omega}_d) Y_{l',m'}(\mathbf{\Omega}_d) d\mathbf{\Omega}_d \\ &\approx \sum_{l'=0}^{L} b_{l'}(f) \sum_{m'=-l'}^{l'} \delta_{l,l'} \delta_{m,m'} \int_{S^2_r} g(\mathbf{\Omega}_d) Y^*_{l',m'}(\mathbf{\Omega}_d) d\mathbf{\Omega}_d \\ &= b_l(f) \int_{S^2_{r_s}} g(\mathbf{\Omega}_d) Y^*_{l,m}(\mathbf{\Omega}_d) d\mathbf{\Omega}_d. \end{aligned} \tag{A.8}$$

Finally, since we are looking to numerically estimate the kernel function $g(\mathbf{\Omega}_d)$, it is necessary to discretize the Herglotz wavefunction's integral over the surface of interest (the sphere, in our case). This results in a system of linear equations which can be written in matrix form:

$$\mathbf{v}_{\text{SH}}(f) = \mathbf{D}_{\text{SH}}(f)\mathbf{g}, \tag{A.9}$$

where $D^{\text{HOA}}_{n,d} = \alpha_d b_l(f) Y^*_{l,m}(\mathbf{\Omega}_d)$ is the $(L+1)^2 \times D$ matrix described in Sec. 2.1.2, α_d are the quadrature weights for the spherical point grid $\mathbf{\Omega}_d$, and $\mathbf{g} = [g(\mathbf{\Omega}_0), g(\mathbf{\Omega}_1), \ldots, g(\mathbf{\Omega}_d), \ldots, g(\mathbf{\Omega}_{D-1})]^\top$.

This formalism essentially forms the basis for the early reflection modelling and analysis techniques described throughout Ch. 2, with some additional considerations (see Sec. 2.1.2 once again) included in order to tie this approach in to the complete SRIR model used throughout this thesis.

A.2/ Maximum Weighted DI Beamformer

The weighted directivity index (WDI) is defined as

$$\text{WDI} = \frac{|w_L(\mathbf{\Omega}_s, \mathbf{\Omega}_s)|^2}{\int_{\mathcal{S}^2} \zeta(\mathbf{\Omega}_s, \mathbf{\Omega}) \, |w_L(\mathbf{\Omega}_s, \mathbf{\Omega})|^2 \, d\mathbf{\Omega}}, \tag{A.10}$$

where the weighting function $\zeta(\mathbf{\Omega}_s, \mathbf{\Omega})$ is proportional to the central angle between $\mathbf{\Omega}_s$ and $\mathbf{\Omega}$:

$$\zeta(\mathbf{\Omega}_s, \mathbf{\Omega}) = \frac{1 - \cos\left[\Theta(\mathbf{\Omega}_s, \mathbf{\Omega})\right]}{\pi}. \tag{A.11}$$

Since the problem is axisymmetric (the weighting function is identical no matter which great circle is chosen), it can be rewritten simply as a function of the central angle Θ (thereby dropping the dependence on $\mathbf{\Omega}_s$ and $\mathbf{\Omega}$). Thus the axisymmetric beampattern can be exploited as in Sec. 1.2.1:

$$\begin{aligned} w_L(\mathbf{\Omega}_s, \mathbf{\Omega}) &= \sum_{l=0}^{L} \sum_{m=-l}^{l} d_l Y_{l,m}^*(\mathbf{\Omega}_s) Y_{l,m}(\mathbf{\Omega}) \\ &= \sum_{l=0}^{L} d_l \sum_{m=-l}^{l} Y_{l,m}^*(\mathbf{\Omega}_s) Y_{l,m}(\mathbf{\Omega}) \\ &= \sum_{l=0}^{L} d_l \frac{2l+1}{4\pi} \mathcal{P}_l(\cos\Theta) \\ &= w_L(\Theta), \end{aligned} \tag{A.12}$$

where the SH addition theorem has once again been applied between the third and fourth lines.

The WDI can then be expressed as:

$$\begin{aligned} \text{WDI} &= \frac{|w_L(0)|^2}{\int_0^\pi \zeta(\Theta) \, |w_L(\Theta)|^2 \, d\Theta} \\ &= \frac{\pi \sum_{l=0}^{L} d_l (2l+1) \sum_{l'=0}^{L} d_{l'} (2l'+1)}{\int_0^\pi (1-\cos\Theta) \sum_{l=0}^{L} d_l (2l+1) \mathcal{P}_l(\cos\Theta) \sum_{l'=0}^{L} d_{l'} (2l'+1) \mathcal{P}_{l'}(\cos\Theta) d\Theta}, \end{aligned} \tag{A.13}$$

which is perhaps more nicely written in Rayleigh quotient matrix form:

$$\text{WDI} = \frac{\mathbf{d}^{\text{H}} \mathbf{A} \mathbf{d}}{\mathbf{d}^{\text{H}} \mathbf{B} \mathbf{d}}, \tag{A.14}$$

with $A_{l,l'} = \pi(2l+1)(2l'+1)$ and

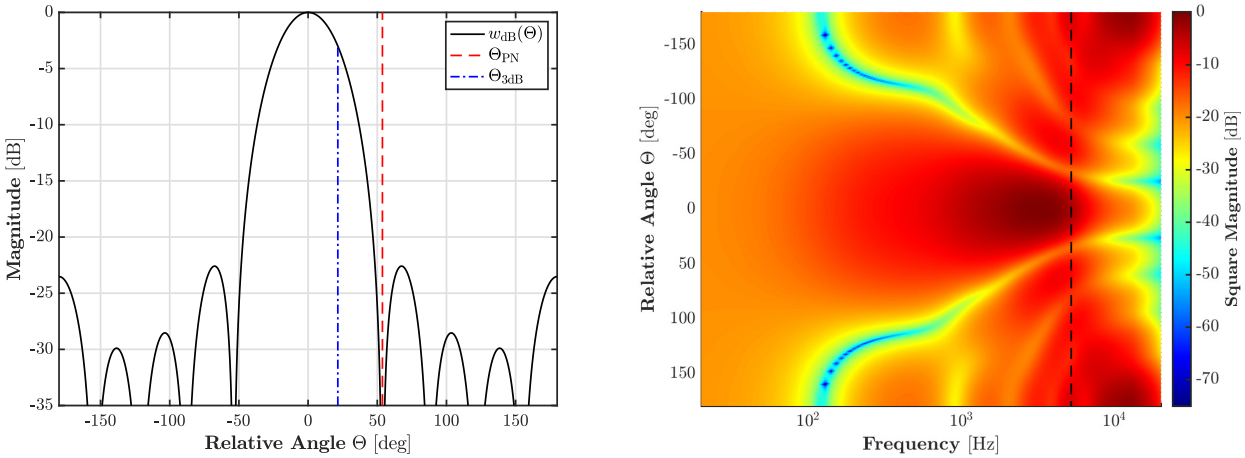
$$B_{l,l'} = (2l+1)(2l'+1) \sum_{j=0}^{l} \sum_{k=0}^{l'} p_j^l p_k^{l'} \left[\frac{1 + (-1)^{j+k}}{j+k+1} - \frac{1 + (-1)^{j+k+1}}{j+k+2} \right], \tag{A.15}$$

where the integral in the denominator of Eq. A.13 has been evaluated by writing the Legendre polynomials explicitly as $\mathcal{P}_l(z) = \sum_{j=0}^{l} p_j^l z^j$, with the Legendre polynomial coefficients p_j^l, and finally substituting $z = \cos\Theta$ (see e.g. Rafaely [45, p. 111]).

Equation A.14 then leads to the following generalized eigenvalue problem:

$$\Lambda_{\text{WDI}} \mathbf{B} \mathbf{d} = \mathbf{A} \mathbf{d}, \tag{A.16}$$

whose eigenvector corresponding the largest eigenvalue gives the vector $\tilde{\mathbf{d}}$ containing to the beamforming coefficients $\tilde{d}_l$ that maximize the WDI.



(a) Broadband directivity function. $\Theta_{\mathrm{PN}} = 53.7°$

(b) Frequency dependence of the MWDI directivity function for an HOA-encoded SMA measurement.

Figure A.1: Directivity functions for the MWDI beamformer. (a, left) 4th-order broadband. (b, right) As applied to a 4th-order HOA-encoded SMA measurement, including the spatial aliasing frequency $f_{\mathrm{alias}} \simeq 5.2$ kHz for the simulated rigid SMA with a radius of 4.2 cm (vertical dashed lines).

A.3/ Reconstructing Directional Power Distributions

Following Jot et al. [16], an EDR calculated in the SH domain can be expressed as

$$\begin{aligned} \mathrm{EDR}_{l,m}(f,t) &= \int_t^\infty \left|\widetilde{H}^{\mathrm{late}}_{l,m}(f,\tau)\right|^2 d\tau \\ &\approx \int_t^\infty \mathrm{ENV}_{l,m}(f,\tau) d\tau \\ &= \int_t^\infty \mathrm{E}\left\{\left|\widetilde{H}^{\mathrm{late}}_{l,m}(f,\tau)\right|^2\right\} d\tau, \end{aligned} \tag{A.17}$$

where $\widetilde{H}^{\mathrm{late}}_{l,m}(f,t)$ is the late reverberation tail model as in Eq. 2.12. This formulation gives

$$\begin{aligned} \mathrm{ENV}(f,t) &= \mathrm{E}\left\{\left|\widetilde{H}^{\mathrm{late}}_{l,m}(f,t)\right|^2\right\} \\ &= |b_l(f)|^2 \sum_{d=1}^{D}\sum_{j=1}^{D} \alpha_d \alpha_j \sqrt{P(f,\mathbf{\Omega}_d,t)P(f,\mathbf{\Omega}_j,t)}\, \mathrm{E}\left\{e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)} e^{-i\hat{\phi}'(f,\mathbf{\Omega}_j,t)}\right\} Y^*_{l,m}(\mathbf{\Omega}_d) Y_{l,m}(\mathbf{\Omega}_j) \\ &= |b_l(f)|^2 \sum_{d=1}^{D} \alpha_d P(f,\mathbf{\Omega}_d,t) \left|Y_{l,m}(\mathbf{\Omega}_d)\right|^2, \end{aligned} \tag{A.18}$$

since

$$\mathrm{E}\left\{e^{i\hat{\phi}(f,\mathbf{\Omega}_d,t)} e^{-i\hat{\phi}'(f,\mathbf{\Omega}_j,t)}\right\} = \delta_{\mathbf{\Omega}_d,\mathbf{\Omega}_j} \tag{A.19}$$

due to the independence of $\hat{\phi}(f,\mathbf{\Omega}_d,t)$ over space, time, and frequency.

In other words, analysis in the SH domain does not give direct access to the direction-dependent properties of the envelope: a directional representation must therefore be constructed. Rewriting Eq. A.18 as a linear system and inverting it leads to the same frequency-dependent limitations and conditions problems as noted in Sec. 1.2.1 and Secs. 2.2 and 2.4. Additionally, it can be shown that this inversion, similar to the approach taken by Fernandez-Grande [86], would result in a more or less

equivalent representation as the DRIR (see Secs. 2.2 and 4.1).

In the special case of an isotropic decay envelope, however, Eq. A.18 becomes:

$$\begin{aligned} \mathrm{E}\left\{\left|\widetilde{H}_{l,m}^{\text{late}}(f,t)\right|^2\right\} &= |b_l(f)|^2\, P(f,t) \sum_{d=1}^{D} \alpha_d \left|Y_{l,m}(\boldsymbol{\Omega}_d)\right|^2 \\ &= |b_l(f)|^2\, P(f,t) \mathcal{N}_{l,m}, \end{aligned} \tag{A.20}$$

where $\mathcal{N}_{l,m}$ is a normalization constant that depends on the chosen HOA scheme (e.g. N3D, SN3D, etc.). Since we can more or less correct for $b_l(f)$ either before or after estimation (with the appropriate encoding filters, see Sec. 1.2.1), SH-domain analysis/modelling is applicable to, and indeed *only* applicable to, isotropic/diffuse late reverberation fields.

A.4/ Spherical Surface Peak Detection

The recursive function `SurfPeaks(surfData,surfPts,Nfail,usedPts,noiseThresh)` aims to detect two-dimensional peaks in any input data (`surfData`) that covers the discretized surface of a sphere, as described by a point grid (`surfPts`).

The utility function `all` returns true if and only if the given condition is true for all elements of input array, and similarly `any` returns true if any one element of the array satisfies the condition.

The subprocess `ImmNeighb(surfPts)` finds the "immediate nearest neighbours" or "triangularly adjacent points" (in the sense of a Delaunay triangulation) for each point in a spherical grid `surfPts`.

The subprocess `OutNeighb(inPts,edgePts,surfPts)` finds the "outside neighbours" for any number of outer edge points (`edgePts`) surrounding a given set of inner points (`inPts`) on a spherical grid (`surfPts`). In other words, it finds the immediate nearest neighbours to each point in `edgePts` such that these neighbours do not belong to `inPts`.

Function `SurfPeaks(surfData,surfPts,Nfail,usedPts,noiseThresh)`:

Data: A list of N spherical surface data values `surfData` $= [x_1(\boldsymbol{\Omega}_1), \ldots, x_i(\boldsymbol{\Omega}_i), \ldots, x_N(\boldsymbol{\Omega}_N)]$, a corresponding data structure describing the N-point spherical grid `surfPts` $= [\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_i, \ldots, \boldsymbol{\Omega}_N]$, a counter of failed peak searches, `Nfail` (initialized as `Nfail` $= 0$), a Boolean list of which points on the sphere have already been identified as belonging to a peak, `usedPts` (initialized as false $\forall\ \boldsymbol{\Omega}_i$), and finally a noise threshold value `noiseThresh`.

Result: A data structure containing the M peak points and their corresponding values, `surfPeaks` $= [\{\boldsymbol{\Omega}_{p_1}, x_{p_1}(\boldsymbol{\Omega}_{p_1})\}, \ldots, \{\boldsymbol{\Omega}_{p_j}, x_{p_j}(\boldsymbol{\Omega}_{p_j})\}, \ldots, \{\boldsymbol{\Omega}_{p_M}, x_{p_M}(\boldsymbol{\Omega}_{p_M})\}]$.

```
unUsedMask ← surfPts ≠ surfPts[usedPts];
maxValPt ← argmax(surfData[unUsedMask]);
maxVal ← surfData[maxValPt];
usedPts[maxValPt] ← true;                    /* Continued on following page... */
```

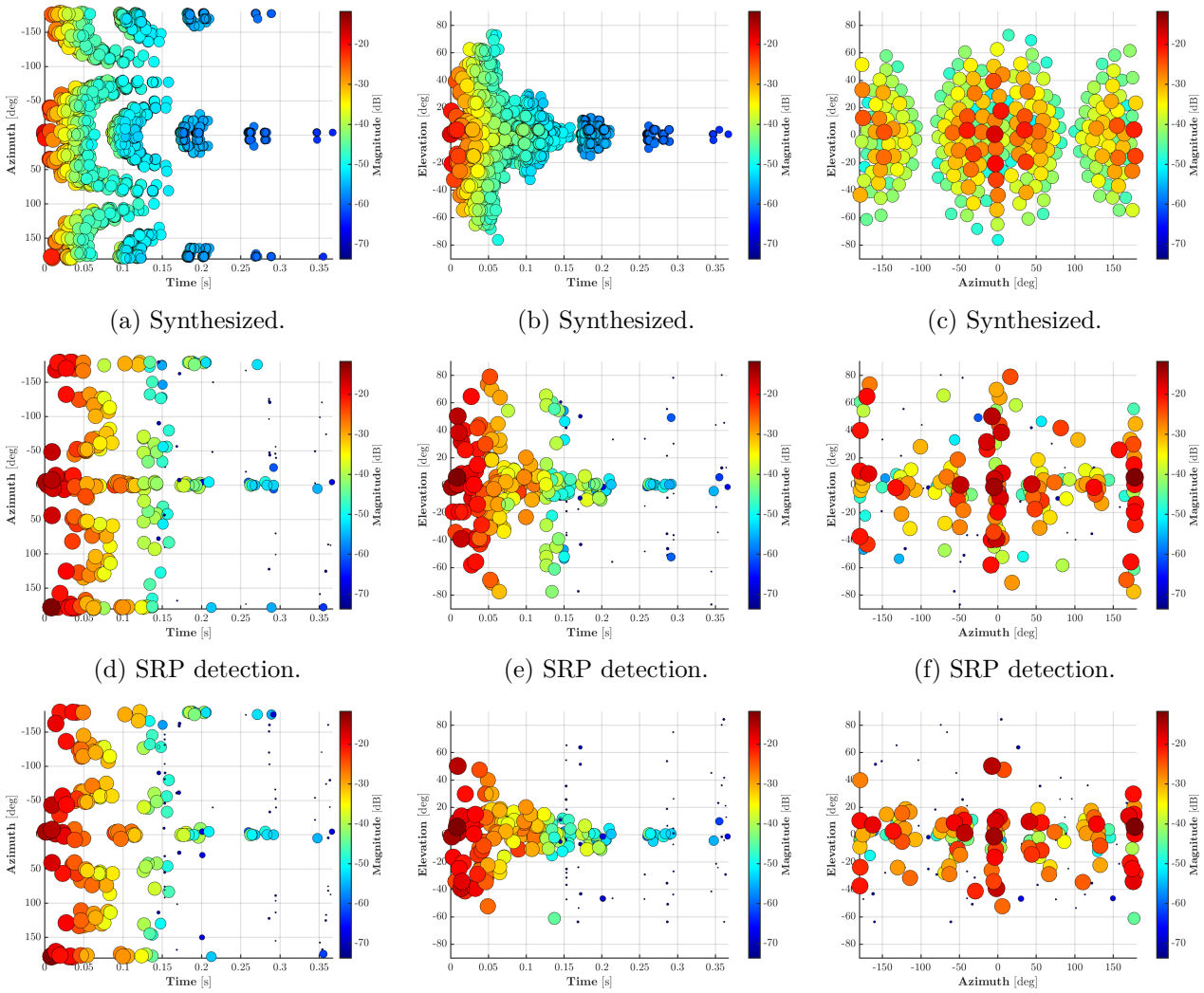
```
    immNeighb ← ImmNeighb(surfPts);      /* Nearest neighbours of each point. */
    maxValNeighb ← immNeighb[maxValPt];   /* Nearest neighbours of max point. */
    neighbVals ← surfData[maxValNeighb];
    usedPts[maxValNeighb] ← true;
    if all(neighbVals<maxVal)∧(maxVal>noiseThresh) returns true then
        {outNeighb,inOutConnect} ← OutNeighb(maxValPt,maxValNeighb,surfPts);
         /* Get immediate outside neighbours to maximum point's first immediate
         neighbours, and corresponding connection map. */
        usedPts[outNeighb] ← true;
        outVals ← surfData[outNeighb];
        inPts ← [maxValPt,maxValNeighb];      /* All points inside ring of outside
         neighbours. */
        Nout ← len(outNeighb);
        nextOut ← false ∀ i ∈ [0, 1, ..., (Nout − 1)];
        for i ← 0 to (Nout − 1) do
            inVals ← surfData[inOutConnect[i]]; /* Inner points connecting to the i^th
             outside neighbour. */
            nextOut[i] ← all(outVals[i]<inVals);   /* True iff all connecting inner
             point values greater than outside neighbour. */
        end
        while any(nextOut) do
            outNeighb ← outNeighb[nextOut];
            {outNeighb,inOutConnect,inPts} ← OutNeighb(inPts,outNeighb,surfPts);
            usedPts[inPts,outNeighb] ← true;
            outVals ← surfData[outNeighb];
            Nout ← len(outNeighb);
            nextOut ← false ∀ i ∈ [0, 1, ..., (Nout − 1)];
            for i ← 0 to (Nout − 1) do
                inVals ← surfData[inOutConnect[i]];
                nextOut[i] ← all(outVals[i]<inVals);
            end
        end
        Nfail ← 0;
    else
        maxVal ← null;
        maxInd ← null;
        Nfail ← Nfail+1;
    end
    if any(¬usedPts)∧(Nfail<3) returns true then
        otherPks ← SurfPeaks(surfData,surfPts,Nfail,usedPts,noiseThresh);
    else
        otherPks ← null;
    end
    surfPeaks ← [{surfPts[maxValPt],surfData[maxValPt]},otherPks];
end
```

Figure A.2: Spherical surface peak detection algorithm.

B / Additional Figures

B.1 / Early Reflection Detection

Test Room 1

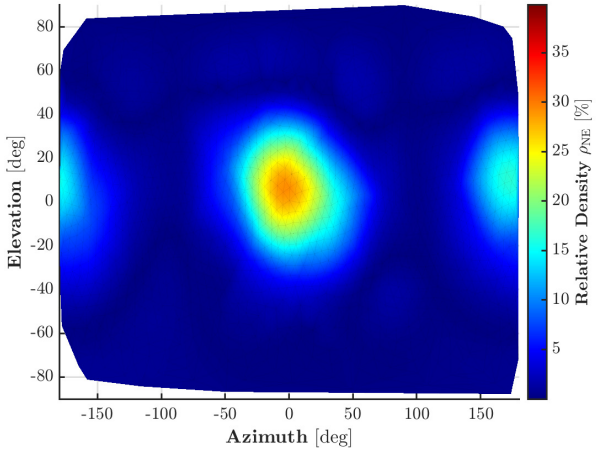


(a) Synthesized. (b) Synthesized. (c) Synthesized.

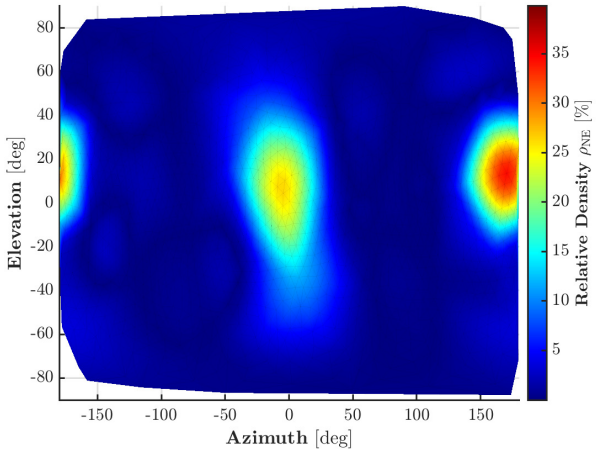
(d) SRP detection. (e) SRP detection. (f) SRP detection.

(g) Herglotz detection. (h) Herglotz detection. (i) Herglotz detection.

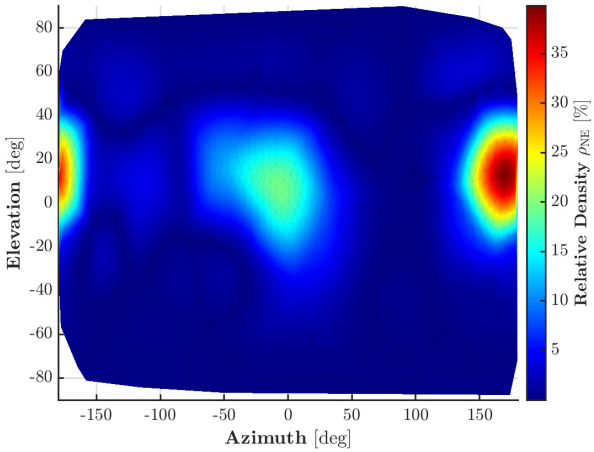
Figure B.1: Synthesized and detected echos for the EVERTims "Test Room 1" IS SRIR simulation. Point colours and sizes are both proportional to the estimated reflection energies. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4. For more details on the synthesis of these simulated SRIRs, see Sec. 4.2.1.



(a) Synthesized.

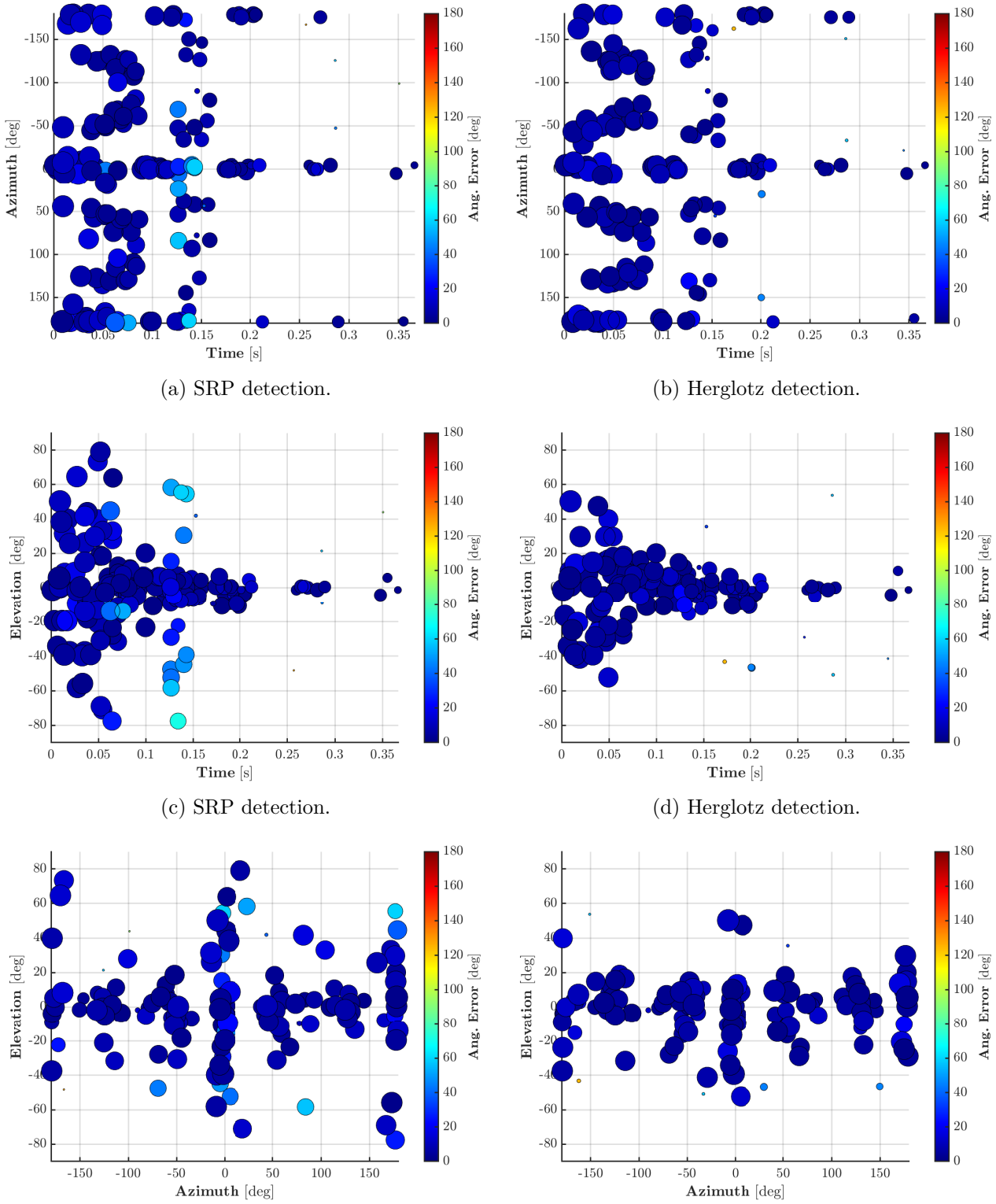


(b) SRP detection.



(c) Herglotz detection.

Figure B.2: Directional echo energy density (DEED) maps calculated on synthesized (a), SRP detected (b), and Herglotz detected (c) echoes from the EVERTims "Test Room 1" IS SRIR simulation. For more details on the DEED, see Sec. 2.4.3.



(a) SRP detection.

(b) Herglotz detection.

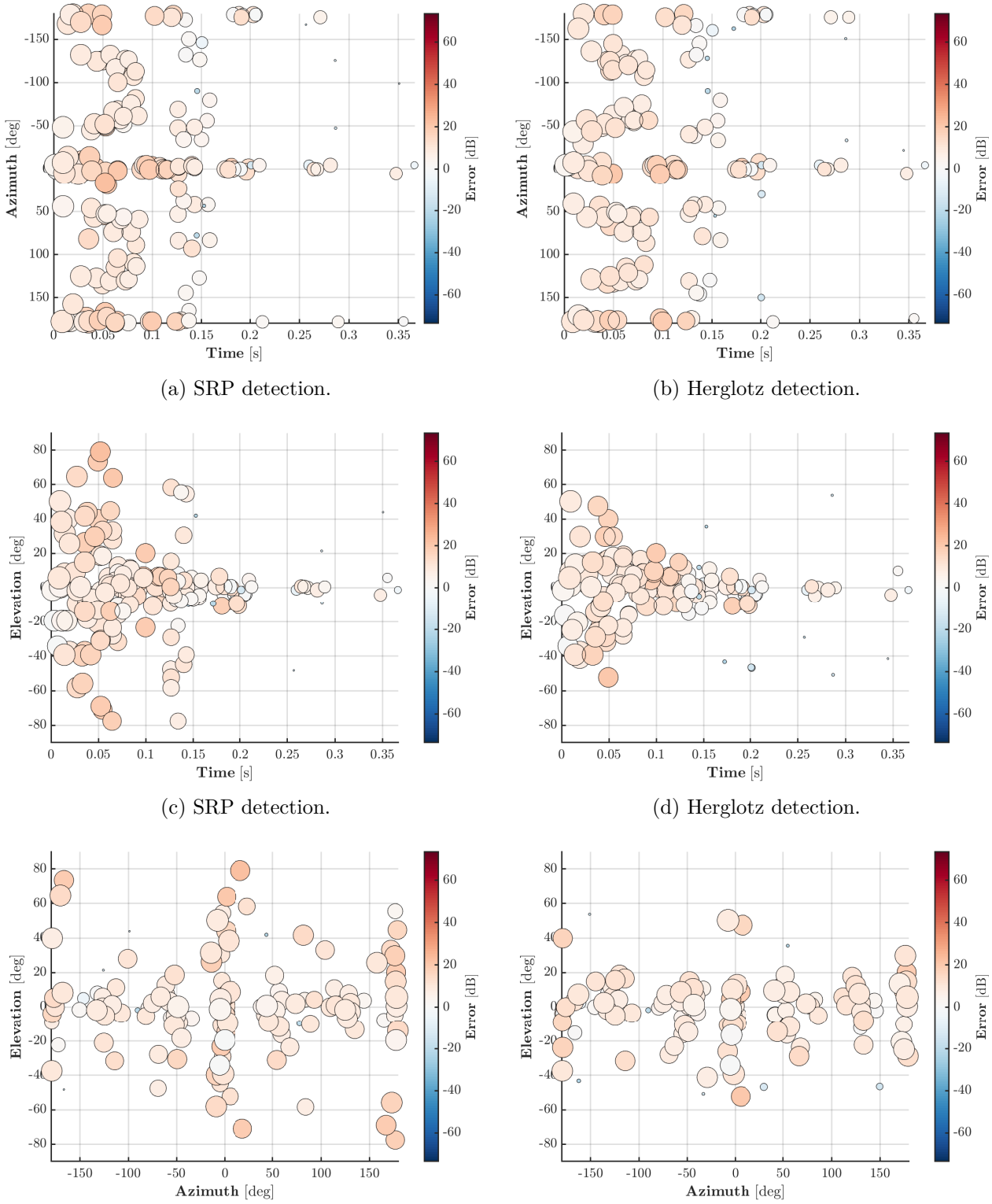
(c) SRP detection.

(d) Herglotz detection.

(e) SRP detection.

(f) Herglotz detection.

Figure B.3: Echo detection angular error maps for the EVERTims "Test Room 1" IS SRIR simulation. Point colours are relative to their angular errors (contained between 0° and 180° by definition), and sizes are again proportional to the estimated reflection energies. For more details on these error calculations, see again Sec. 4.2.1.



(a) SRP detection.

(b) Herglotz detection.

(c) SRP detection.

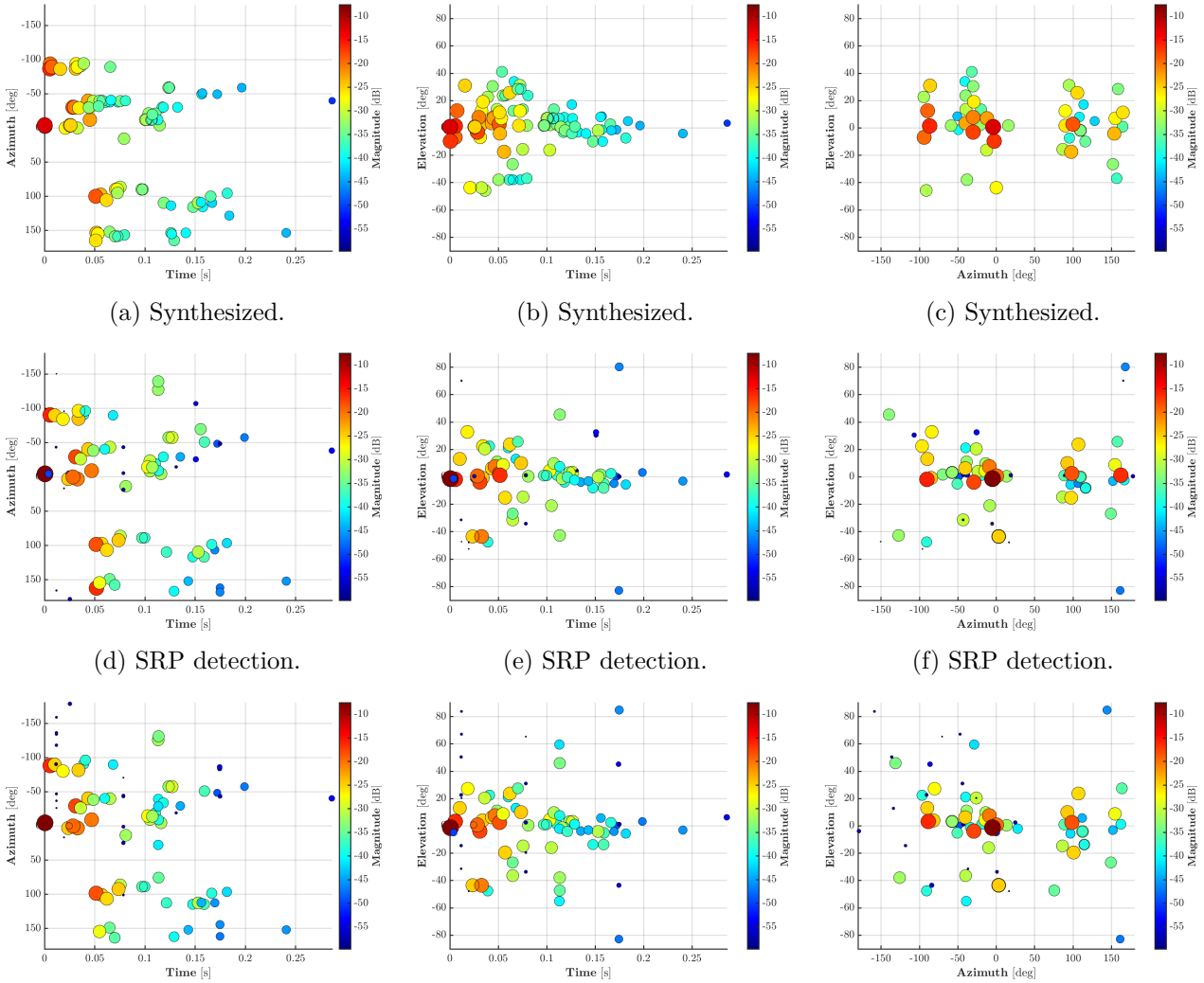
(d) Herglotz detection.

(e) SRP detection.

(f) Herglotz detection.

Figure B.4: Echo detection energy error maps for the EVERTims "Test Room 1" IS SRIR simulation. Point colours are relative to the energy estimation error of the corresponding reflections, and sizes are again proportional to their estimated energies. For more details on these error calculations, see again Sec. 4.2.1.

Test Room 2

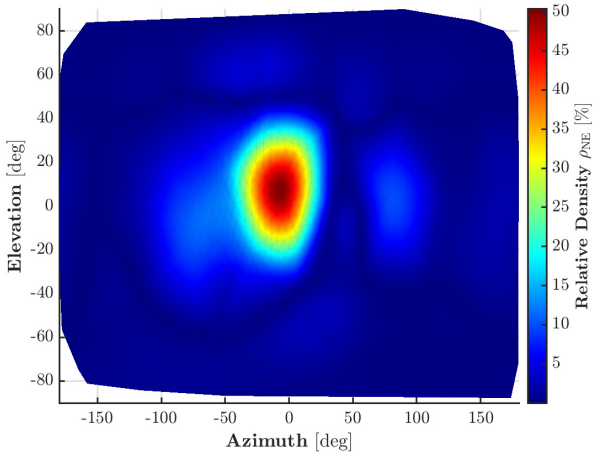


(a) Synthesized. (b) Synthesized. (c) Synthesized.

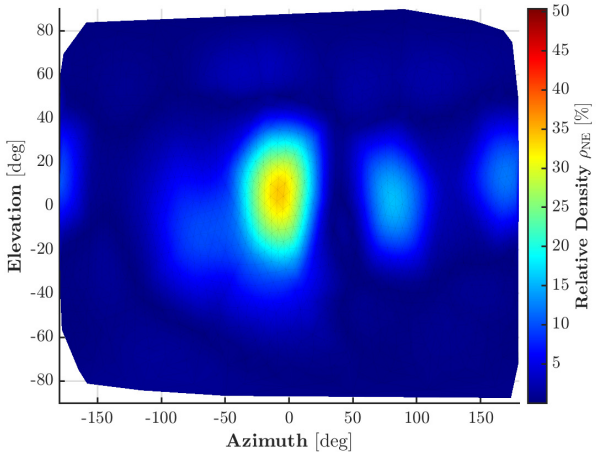
(d) SRP detection. (e) SRP detection. (f) SRP detection.

(g) Herglotz detection. (h) Herglotz detection. (i) Herglotz detection.

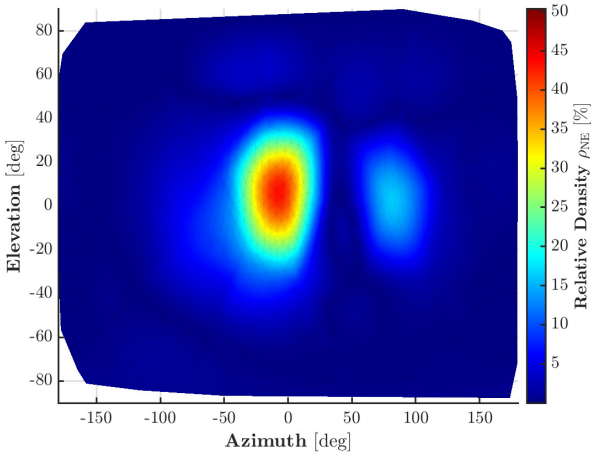
Figure B.5: Synthesized and detected echos for the EVERTims "Test Room 2" IS SRIR simulation. Point colours and sizes are both proportional to the estimated reflection energies. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4. For more details on the synthesis of these simulated SRIRs, see Sec. 4.2.1.



(a) Synthesized.

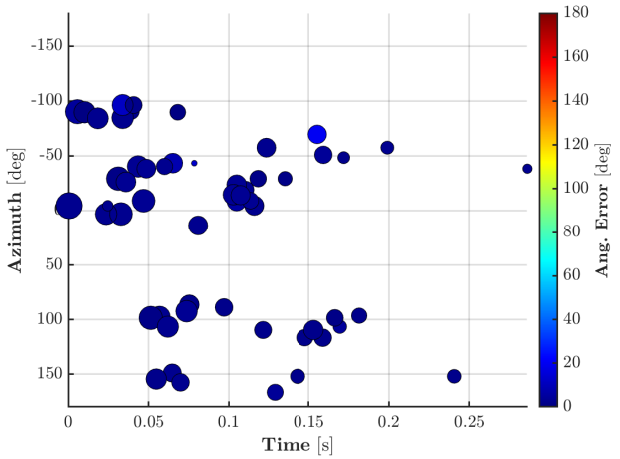


(b) SRP detection.

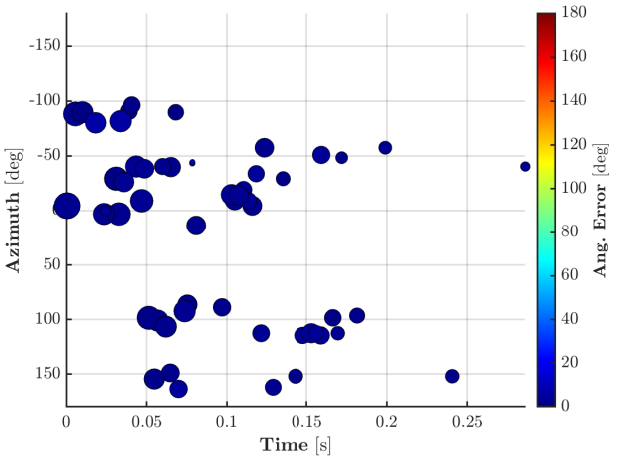


(c) Herglotz detection.

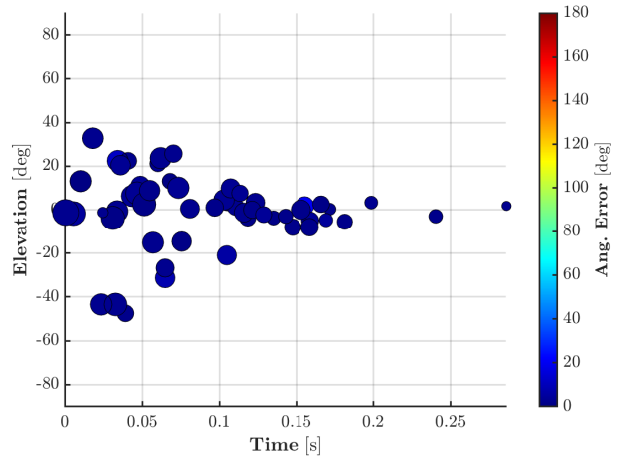
Figure B.6: Directional echo energy density (DEED) maps calculated on synthesized (a), SRP detected (b), and Herglotz detected (c) echoes from the EVERTims "Test Room 2" IS SRIR simulation. For more details on the DEED, see Sec. 2.4.3.



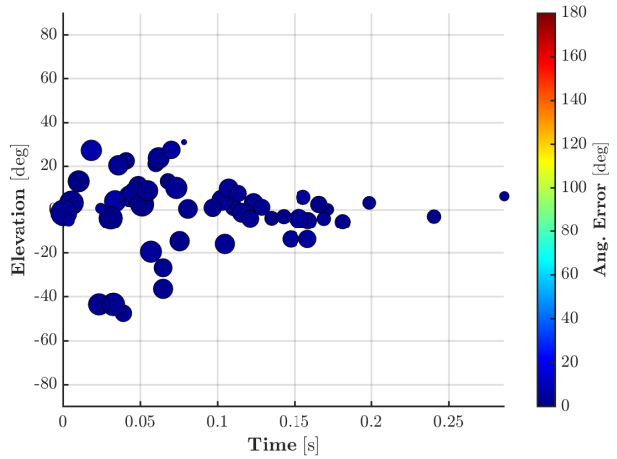
(a) SRP detection error.



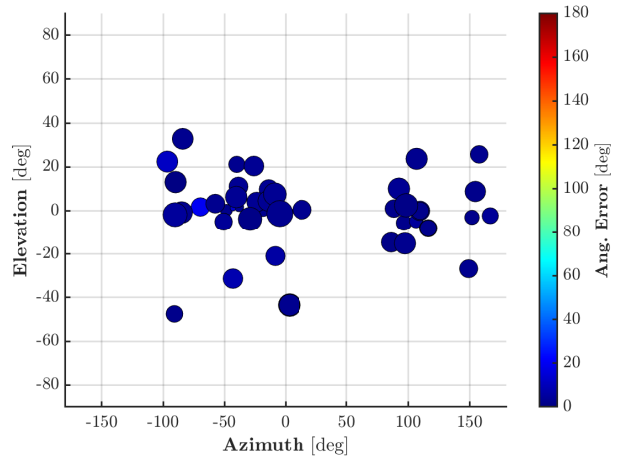
(b) Herglotz detection error.



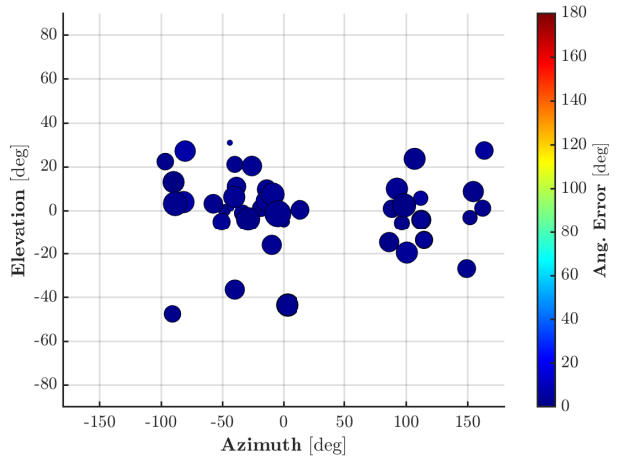
(c) SRP detection error.



(d) Herglotz detection error.

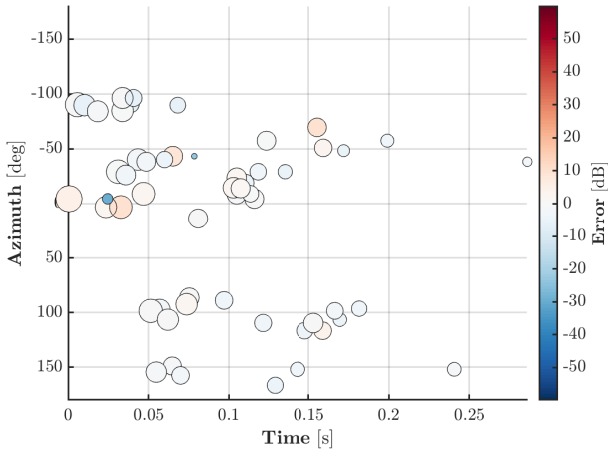


(e) SRP detection error.

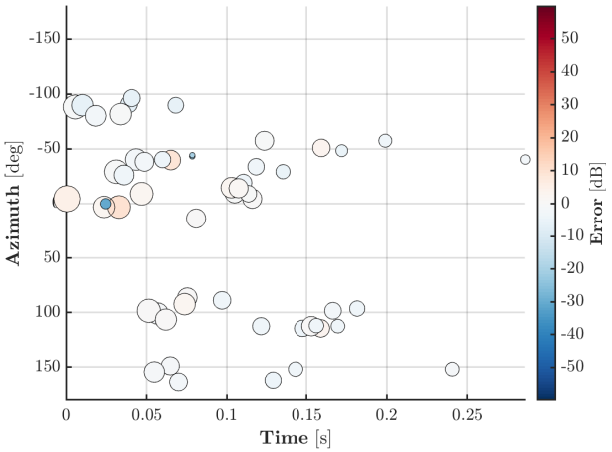


(f) Herglotz detection error.

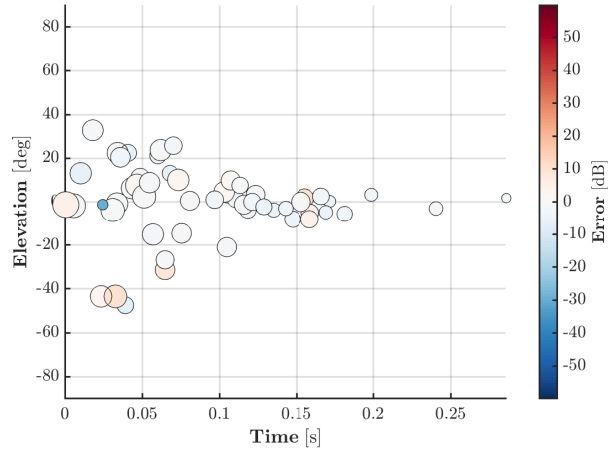
Figure B.7: Echo detection angular error maps for the EVERTims "Test Room 2" IS SRIR simulation. Point colours are relative to their angular errors (contained between 0° and 180° by definition), and sizes are again proportional to the estimated reflection energies. For more details on these error calculations, see again Sec. 4.2.1.



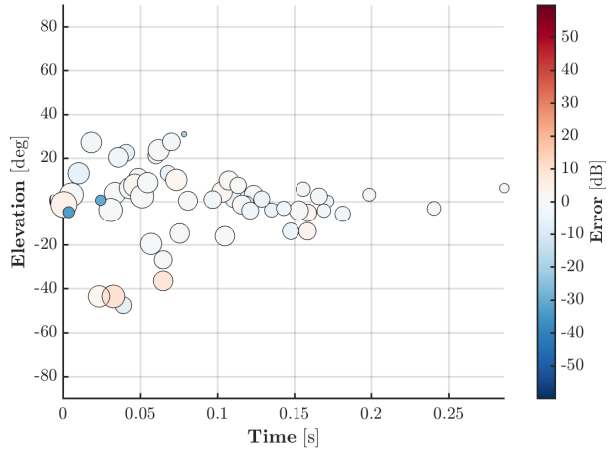
(a) SRP detection.



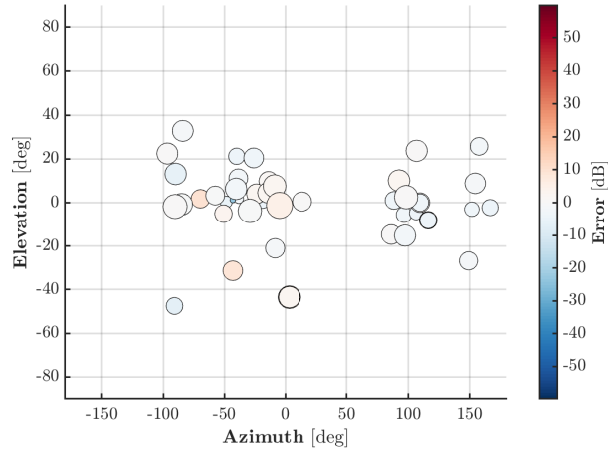
(b) Herglotz detection.



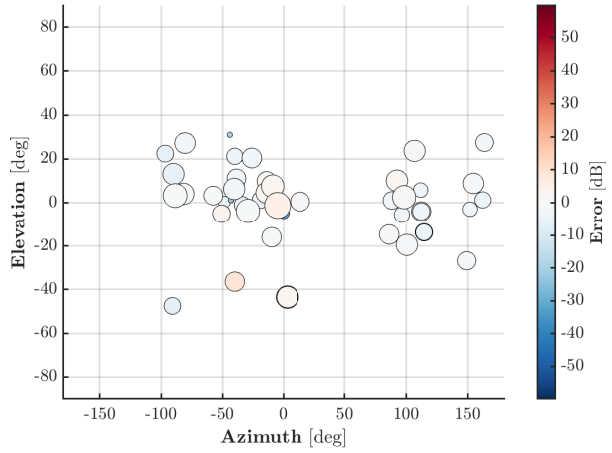
(c) SRP detection.



(d) Herglotz detection.



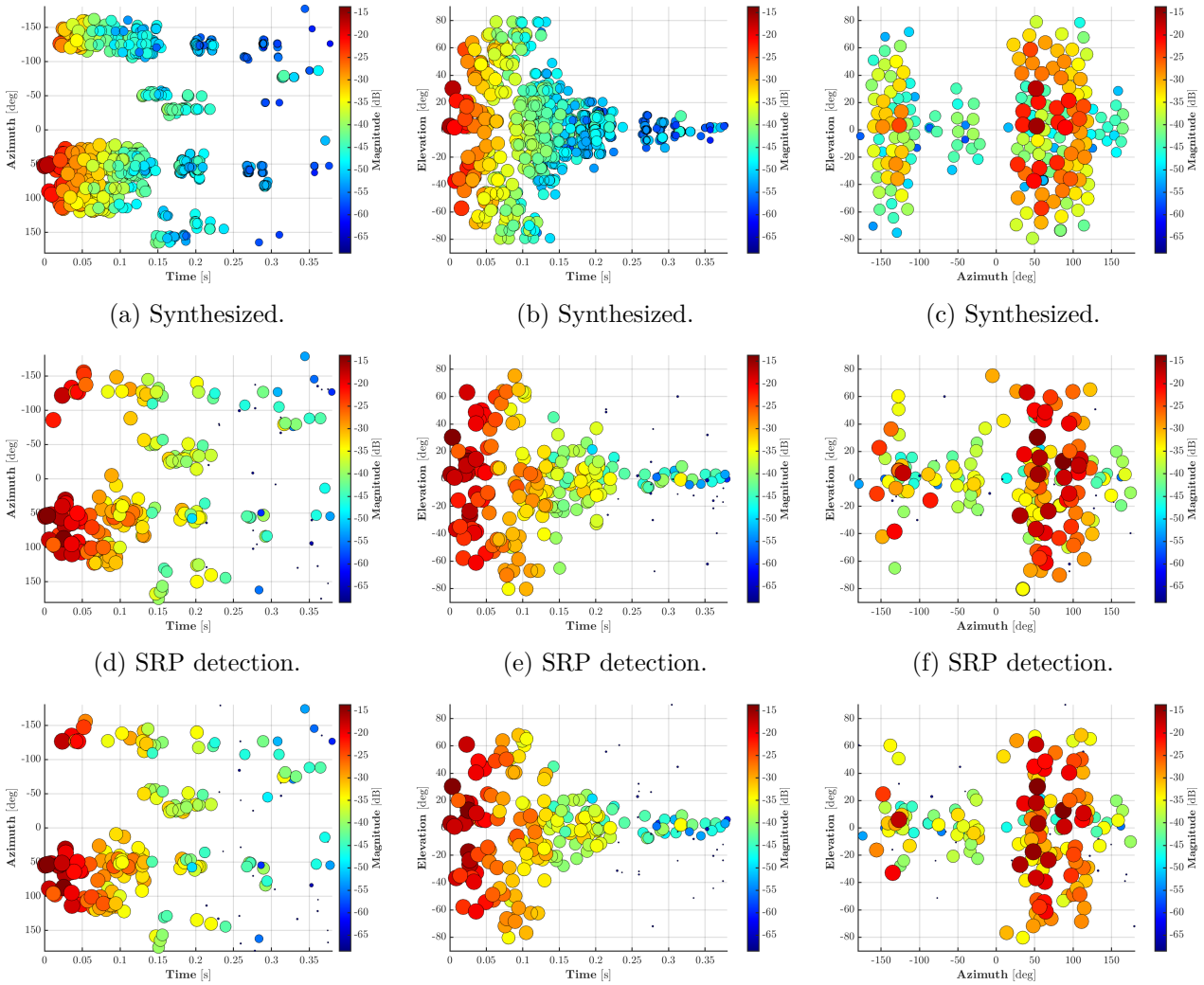
(e) SRP detection.



(f) Herglotz detection.

Figure B.8: Echo detection energy error maps for the EVERTims "Test Room 2" IS SRIR simulation. Point colours are relative to the energy estimation error of the corresponding reflections, and sizes are again proportional to their estimated energies. For more details on these error calculations, see again Sec. 4.2.1.

Test Room 3

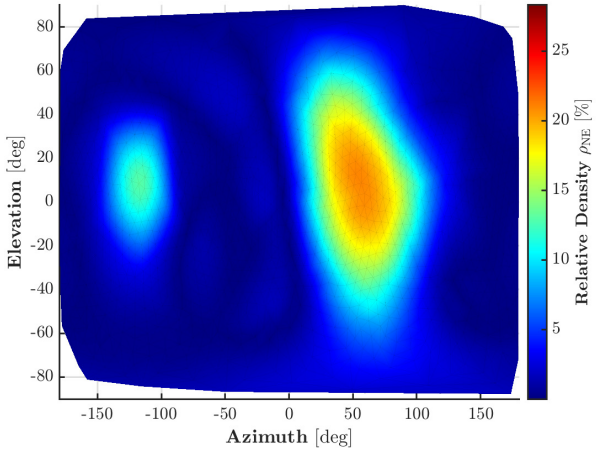


(a) Synthesized. (b) Synthesized. (c) Synthesized.

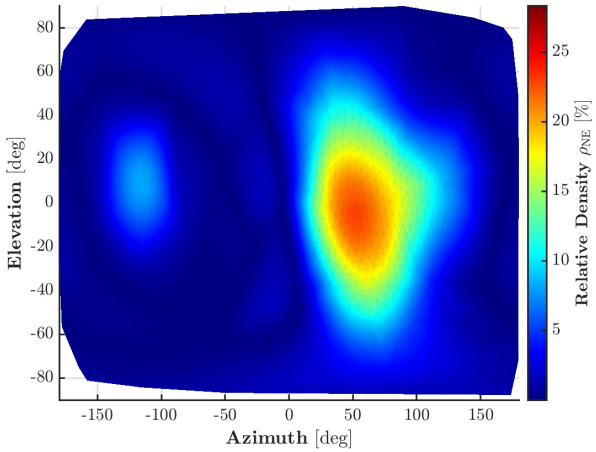
(d) SRP detection. (e) SRP detection. (f) SRP detection.

(g) Herglotz detection. (h) Herglotz detection. (i) Herglotz detection.

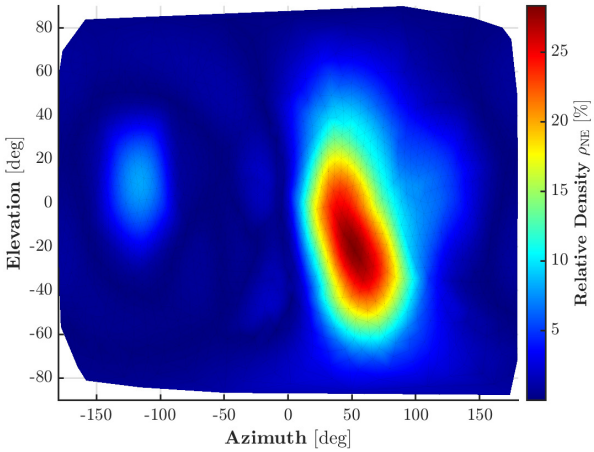
Figure B.9: Synthesized and detected echos for the EVERTims "Test Room 3" IS SRIR simulation. Point colours and sizes are both proportional to the estimated reflection energies. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4. For more details on the synthesis of these simulated SRIRs, see Sec. 4.2.1.



(a) Synthesized.

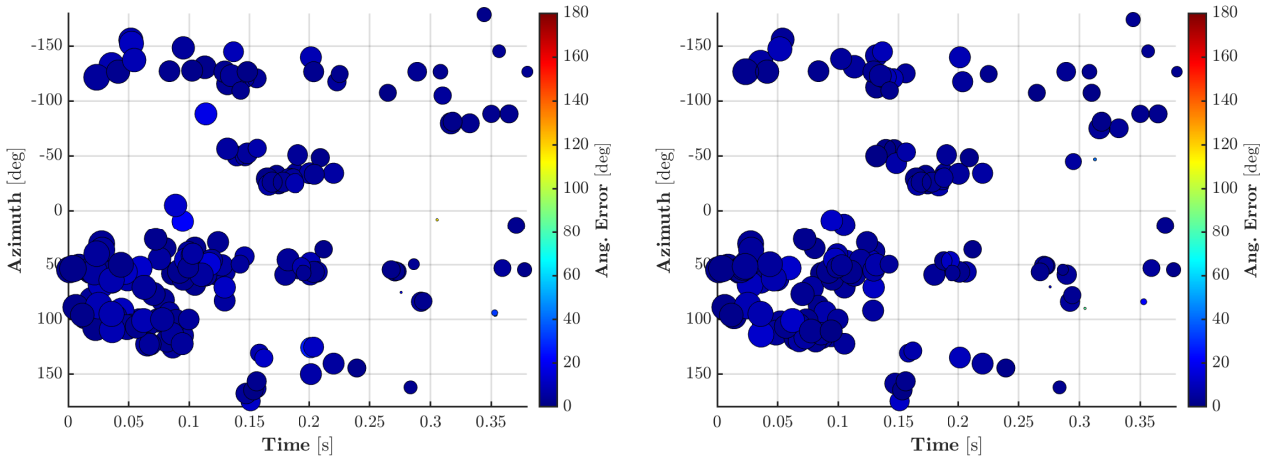


(b) SRP detection.



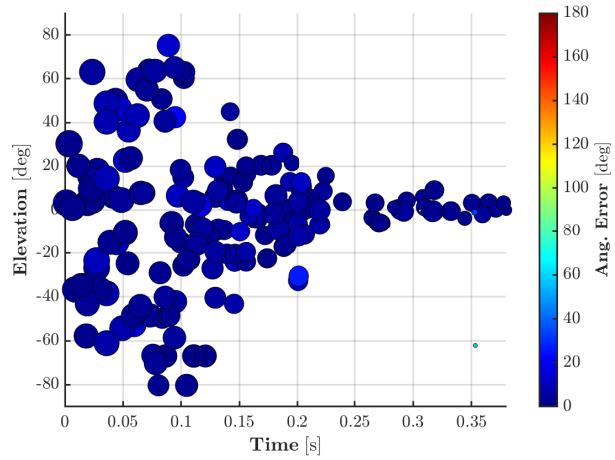
(c) Herglotz detection.

Figure B.10: Directional echo energy density (DEED) maps calculated on synthesized (a), SRP detected (b), and Herglotz detected (c) echoes from the EVERTims "Test Room 3" IS SRIR simulation. For more details on the DEED, see Sec. 2.4.3.

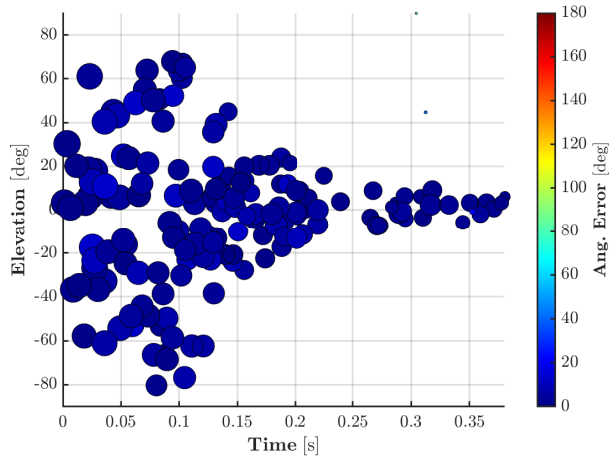


(a) SRP detection error.

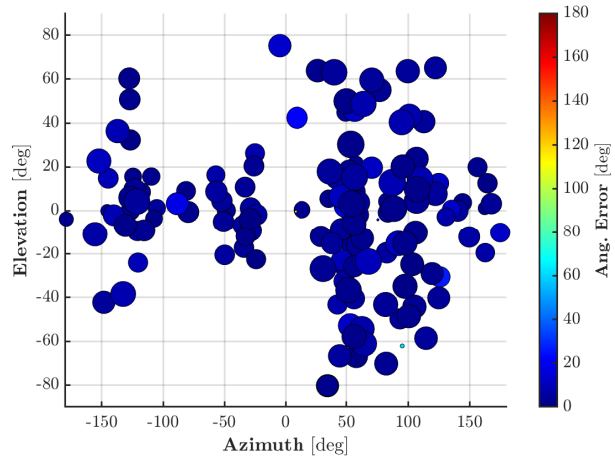
(b) Herglotz detection error.



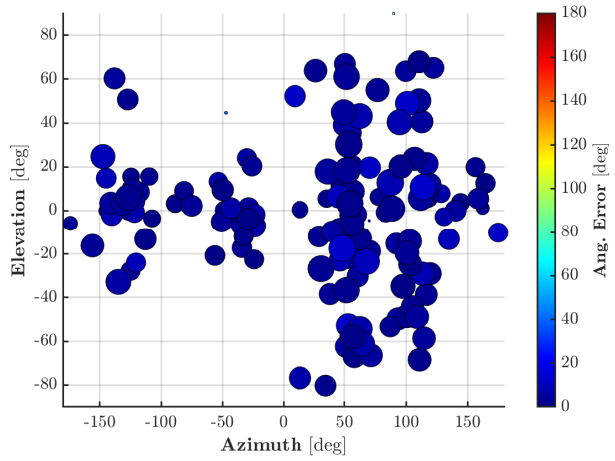
(c) SRP detection error.



(d) Herglotz detection error.

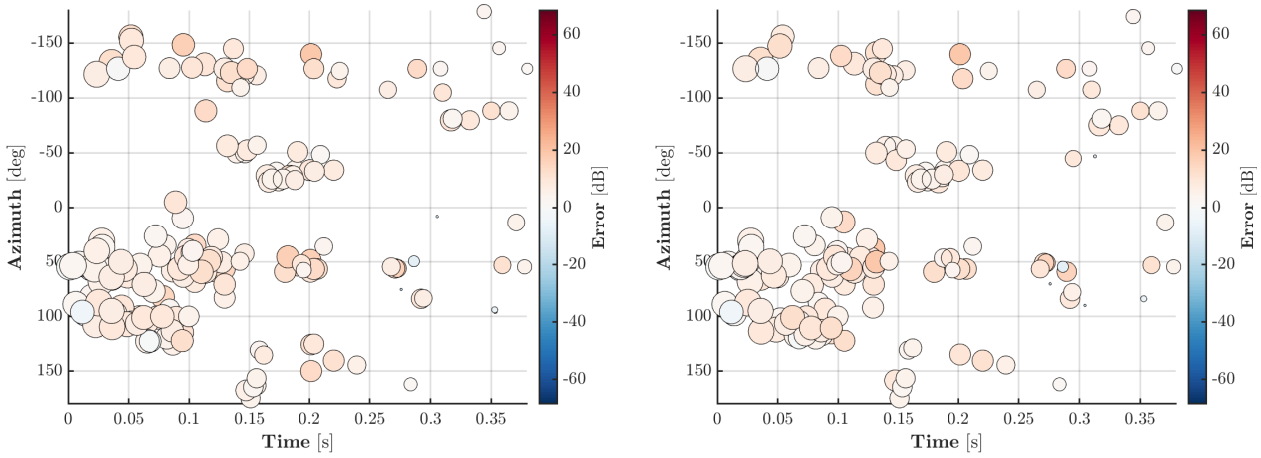


(e) SRP detection error.



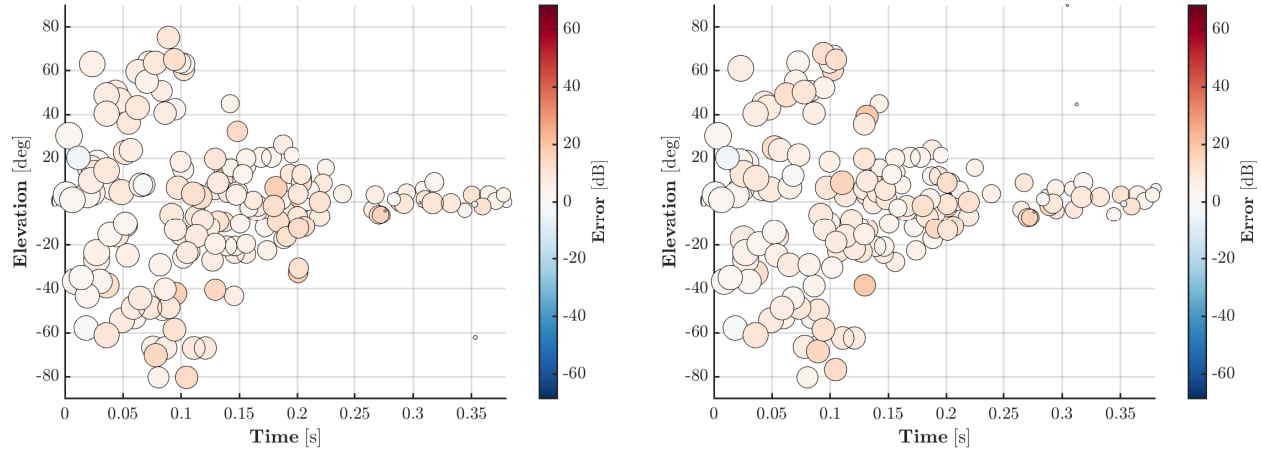
(f) Herglotz detection error.

Figure B.11: Echo detection angular error maps for the EVERTims "Test Room 3" IS SRIR simulation. Point colours are relative to their angular errors (contained between 0° and 180° by definition), and sizes are again proportional to the estimated reflection energies. For more details on these error calculations, see again Sec. 4.2.1.



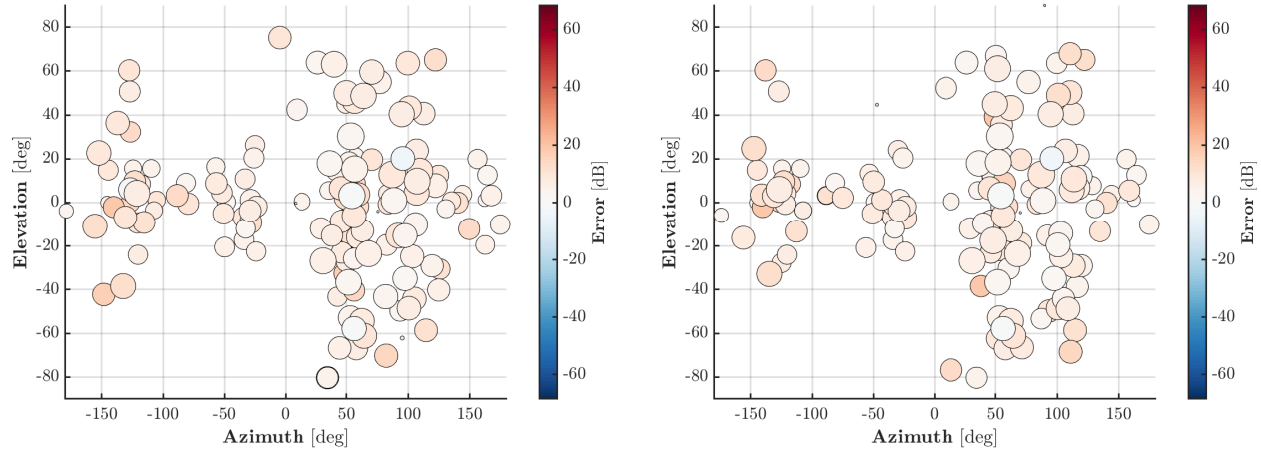
(a) SRP detection.

(b) Herglotz detection.



(c) SRP detection.

(d) Herglotz detection.

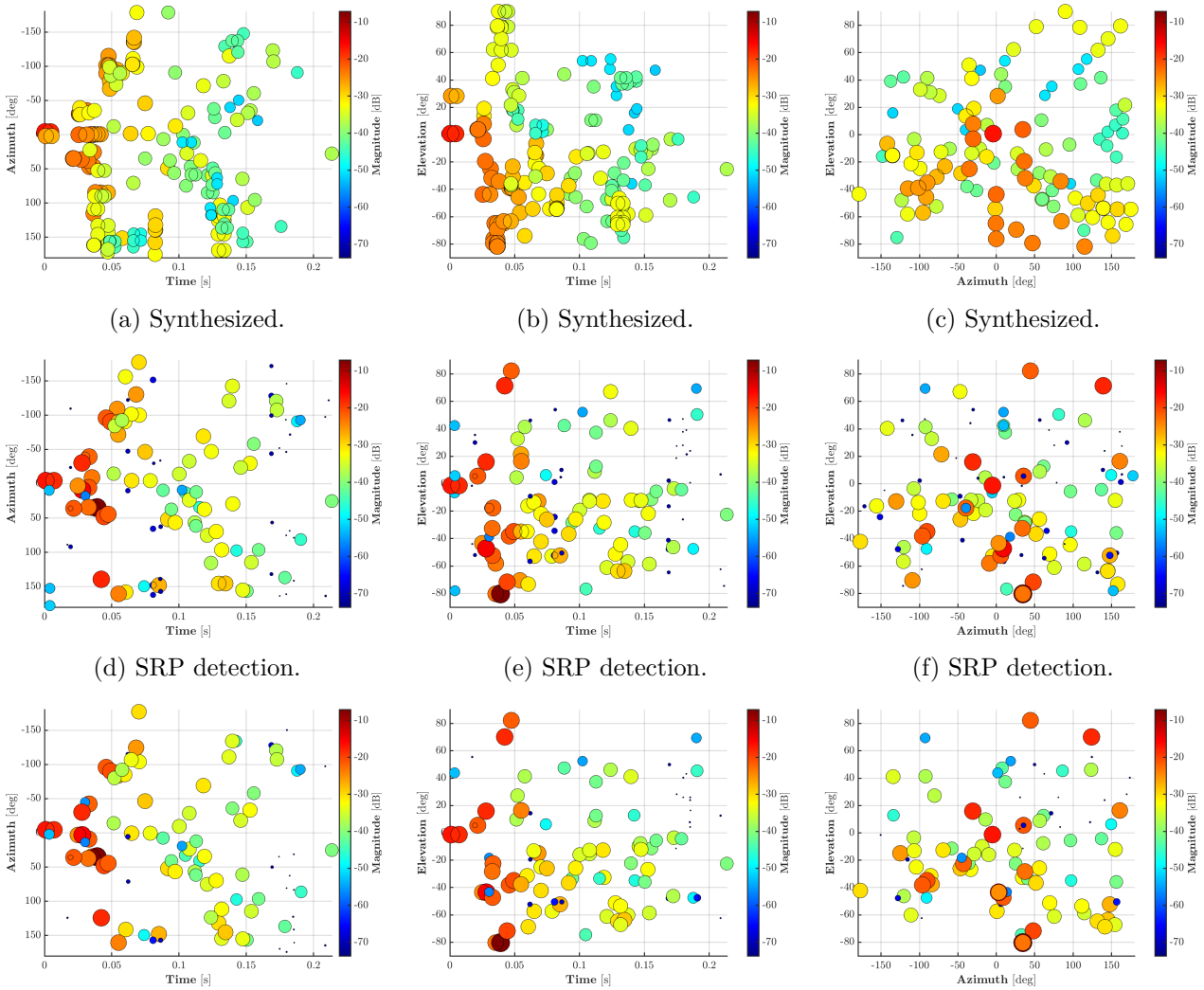


(e) SRP detection.

(f) Herglotz detection.

Figure B.12: Echo detection energy error maps for the EVERTims "Test Room 3" IS SRIR simulation. Point colours are relative to the energy estimation error of the corresponding reflections, and sizes are again proportional to their estimated energies. For more details on these error calculations, see again Sec. 4.2.1.

Fogg

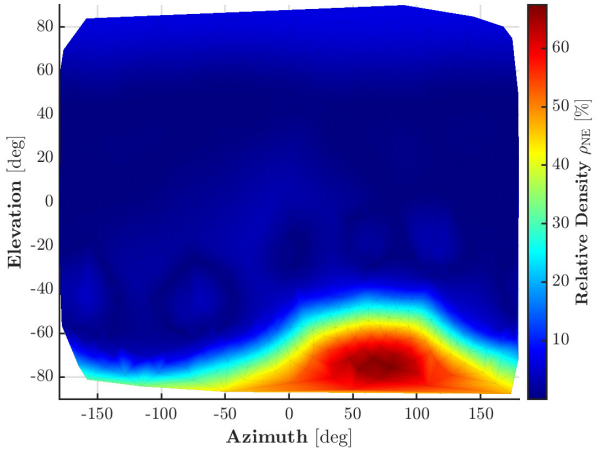


(a) Synthesized. (b) Synthesized. (c) Synthesized.

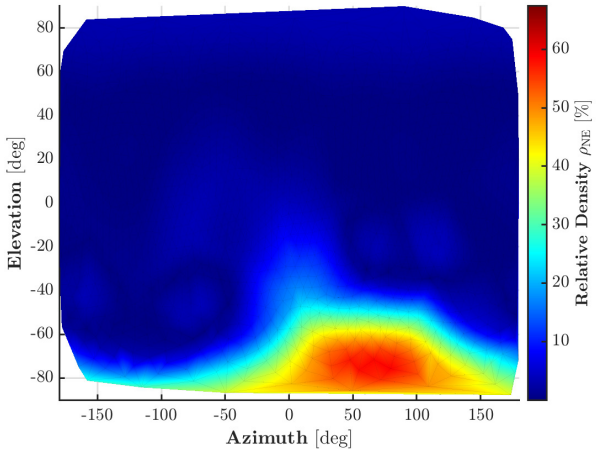
(d) SRP detection. (e) SRP detection. (f) SRP detection.

(g) Herglotz detection. (h) Herglotz detection. (i) Herglotz detection.

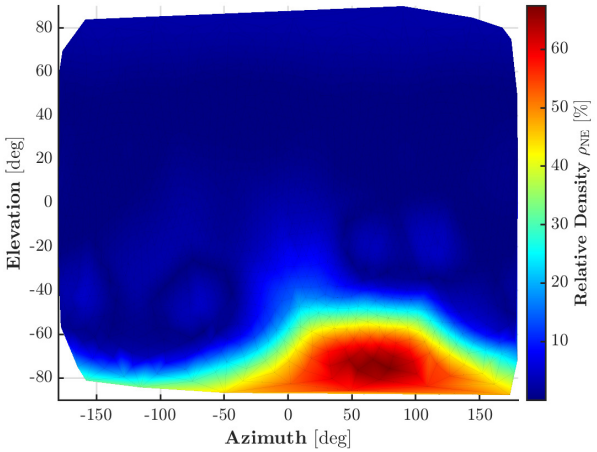
Figure B.13: Synthesized and detected echos for the EVERTims "Fogg" IS SRIR simulation. Point colours and sizes are both proportional to the estimated reflection energies. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4. For more details on the synthesis of these simulated SRIRs, see Sec. 4.2.1.



(a) Synthesized.

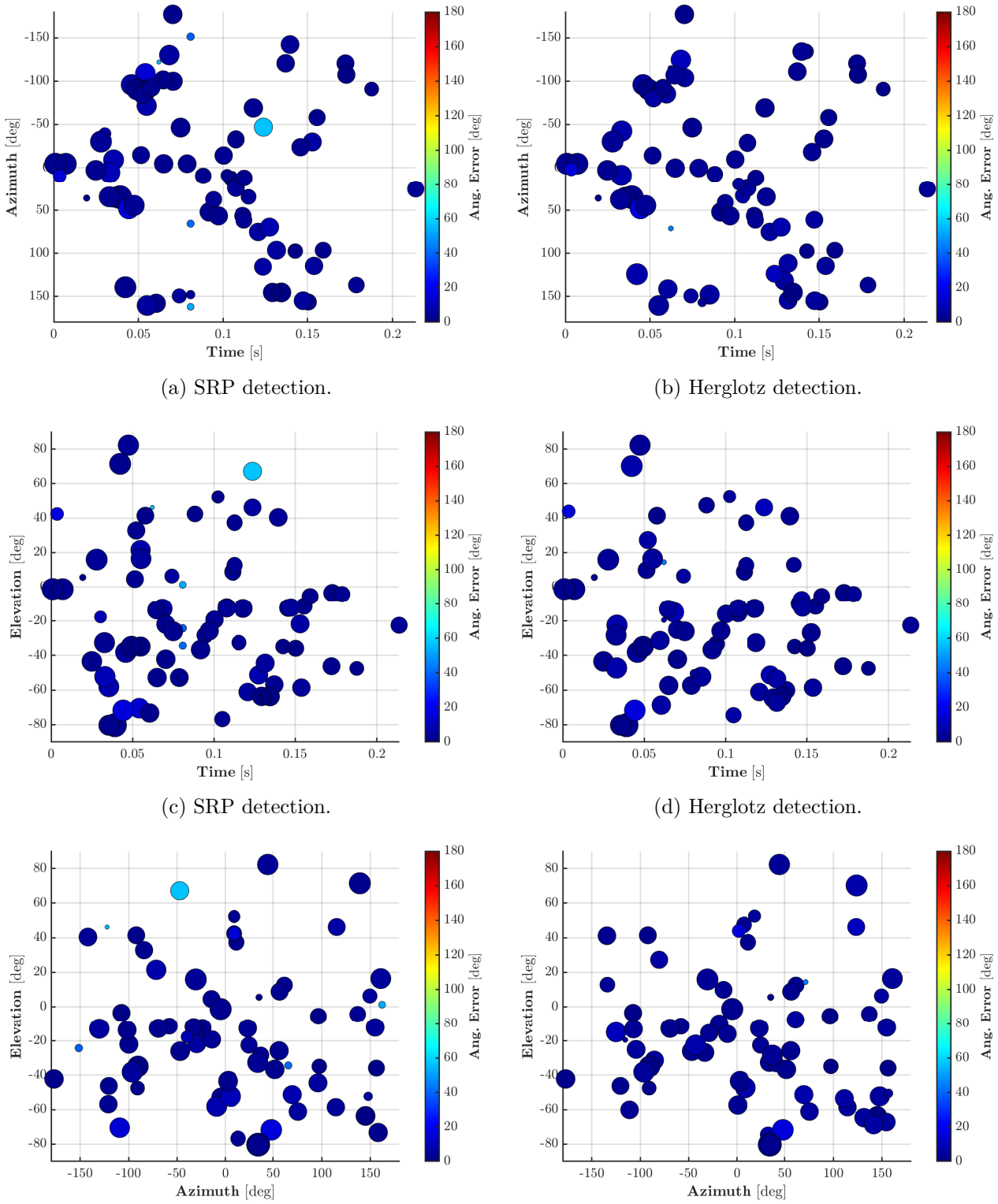


(b) SRP detection.



(c) Herglotz detection.

Figure B.14: Directional echo energy density (DEED) maps calculated on synthesized (a), SRP detected (b), and Herglotz detected (c) echoes from the EVERTims "Fogg" IS SRIR simulation. For more details on the DEED, see Sec. 2.4.3.



(a) SRP detection.

(b) Herglotz detection.

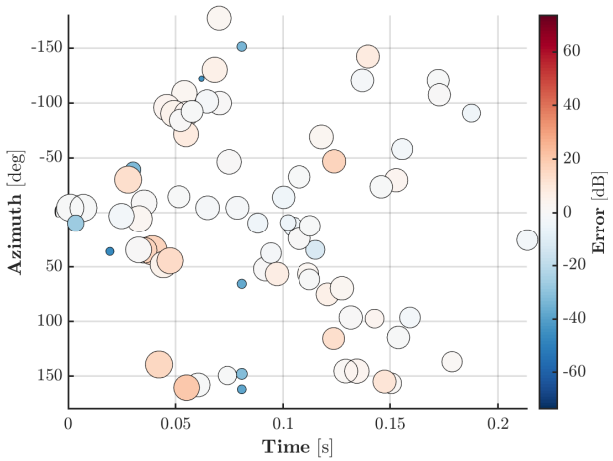
(c) SRP detection.

(d) Herglotz detection.

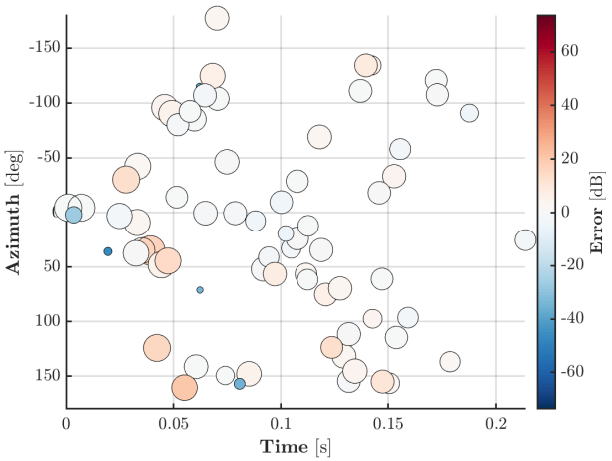
(e) SRP detection.

(f) Herglotz detection.

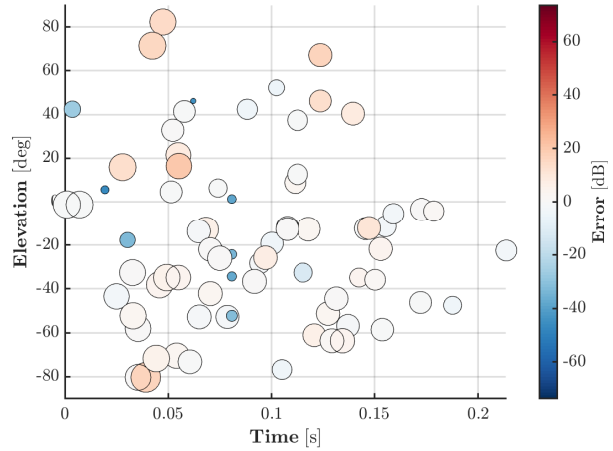
Figure B.15: Echo detection angular error maps for the EVERTims "Fogg" IS SRIR simulation. Point colours are relative to their angular errors (contained between 0° and 180° by definition), and sizes are again proportional to the estimated reflection energies. For more details on these error calculations, see again Sec. 4.2.1.



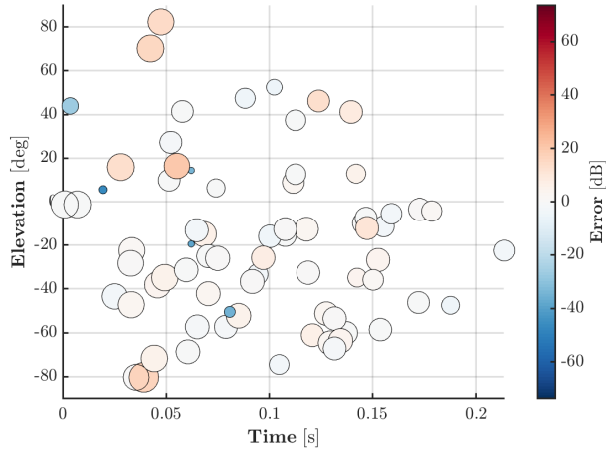
(a) SRP detection.



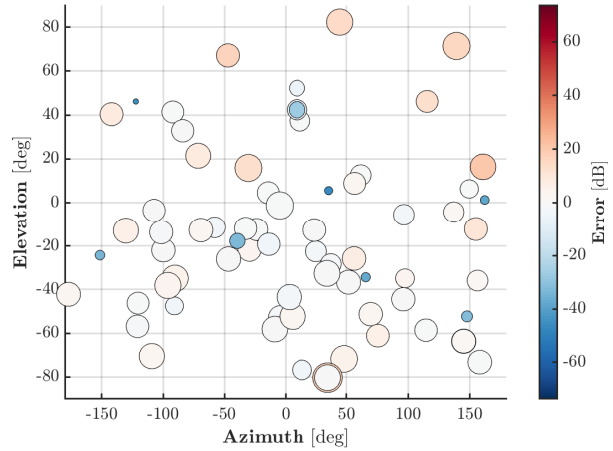
(b) Herglotz detection.



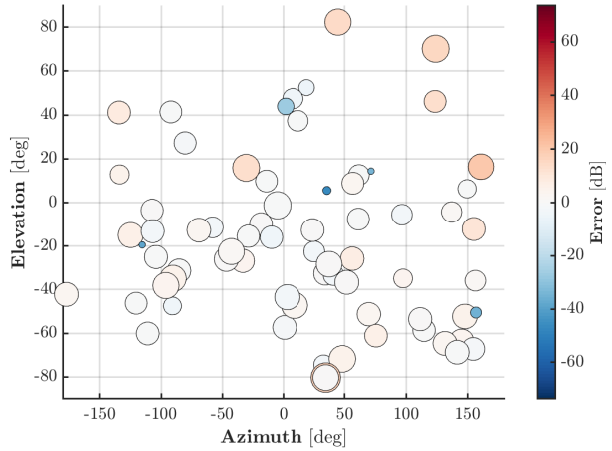
(c) SRP detection.



(d) Herglotz detection.



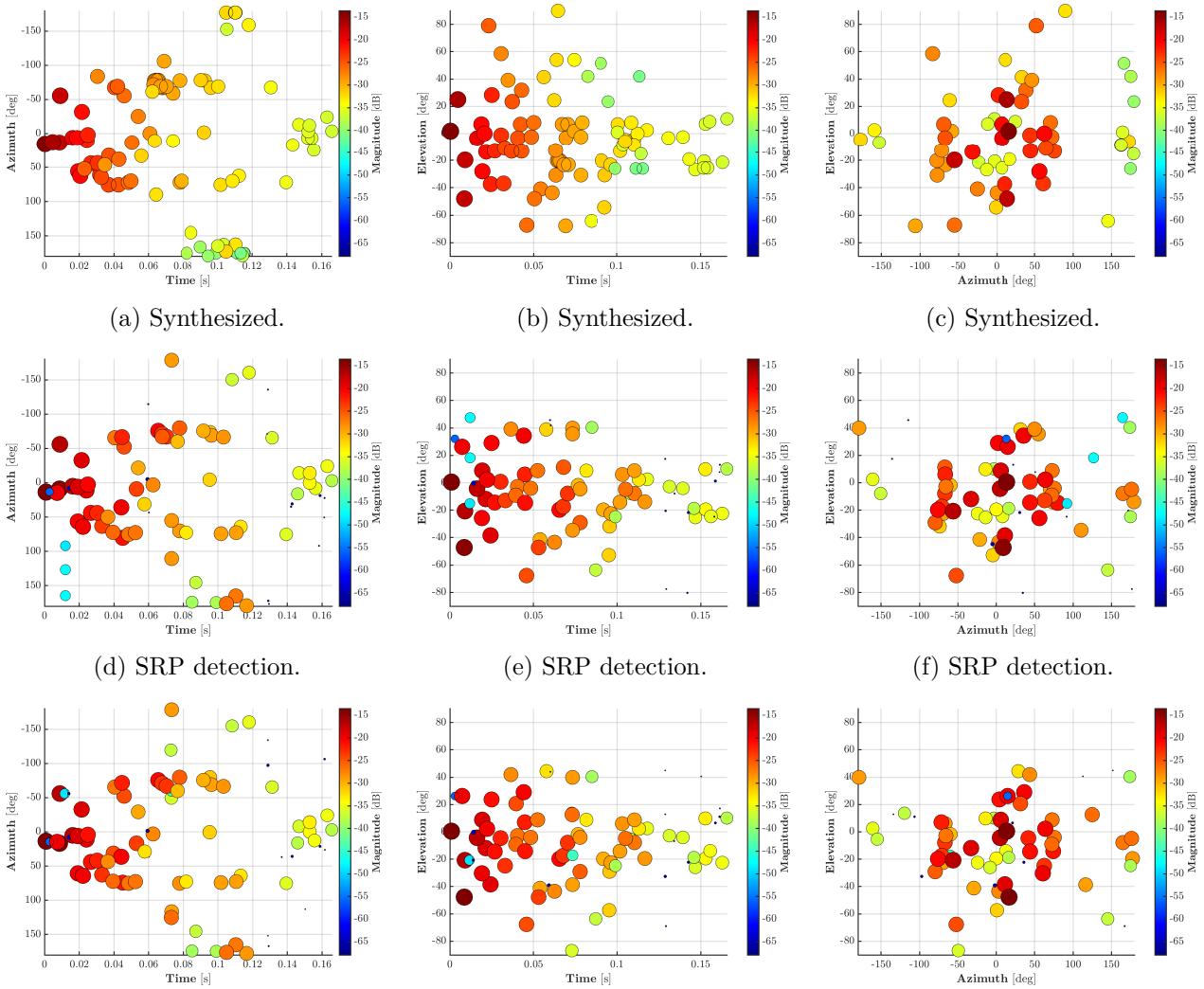
(e) SRP detection.



(f) Herglotz detection.

Figure B.16: Echo detection energy error maps for the EVERTims "Fogg" IS SRIR simulation. Point colours are relative to the energy estimation error of the corresponding reflections, and sizes are again proportional to their estimated energies. For more details on these error calculations, see again Sec. 4.2.1. For more details on these error calculations, see again Sec. 4.2.1.

Morgan

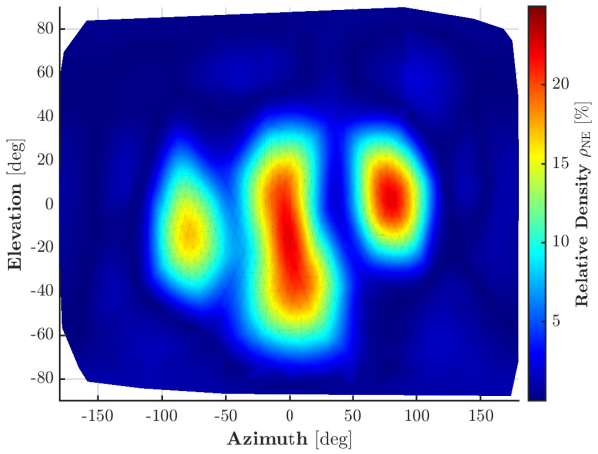


(a) Synthesized. (b) Synthesized. (c) Synthesized.

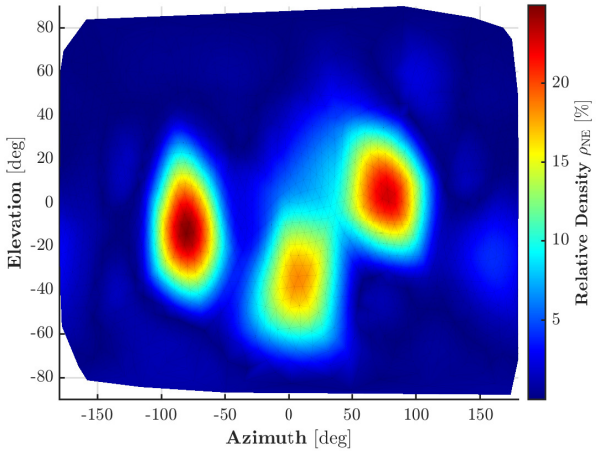
(d) SRP detection. (e) SRP detection. (f) SRP detection.

(g) Herglotz detection. (h) Herglotz detection. (i) Herglotz detection.

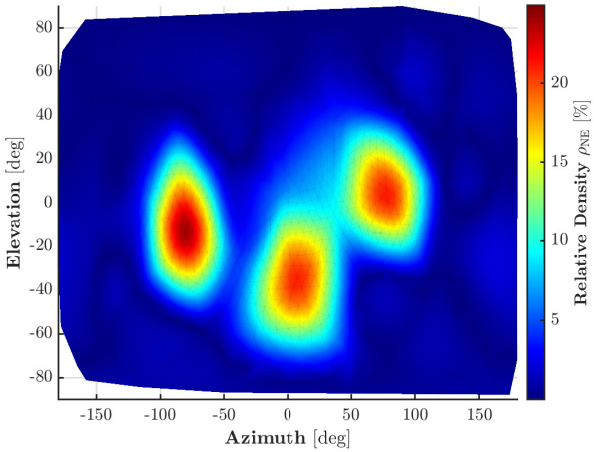
Figure B.17: Synthesized and detected echos for the EVERTims "Morgan" IS SRIR simulation. Point colours and sizes are both proportional to the estimated reflection energies. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4. For more details on the synthesis of these simulated SRIRs, see Sec. 4.2.1.



(a) Synthesized.

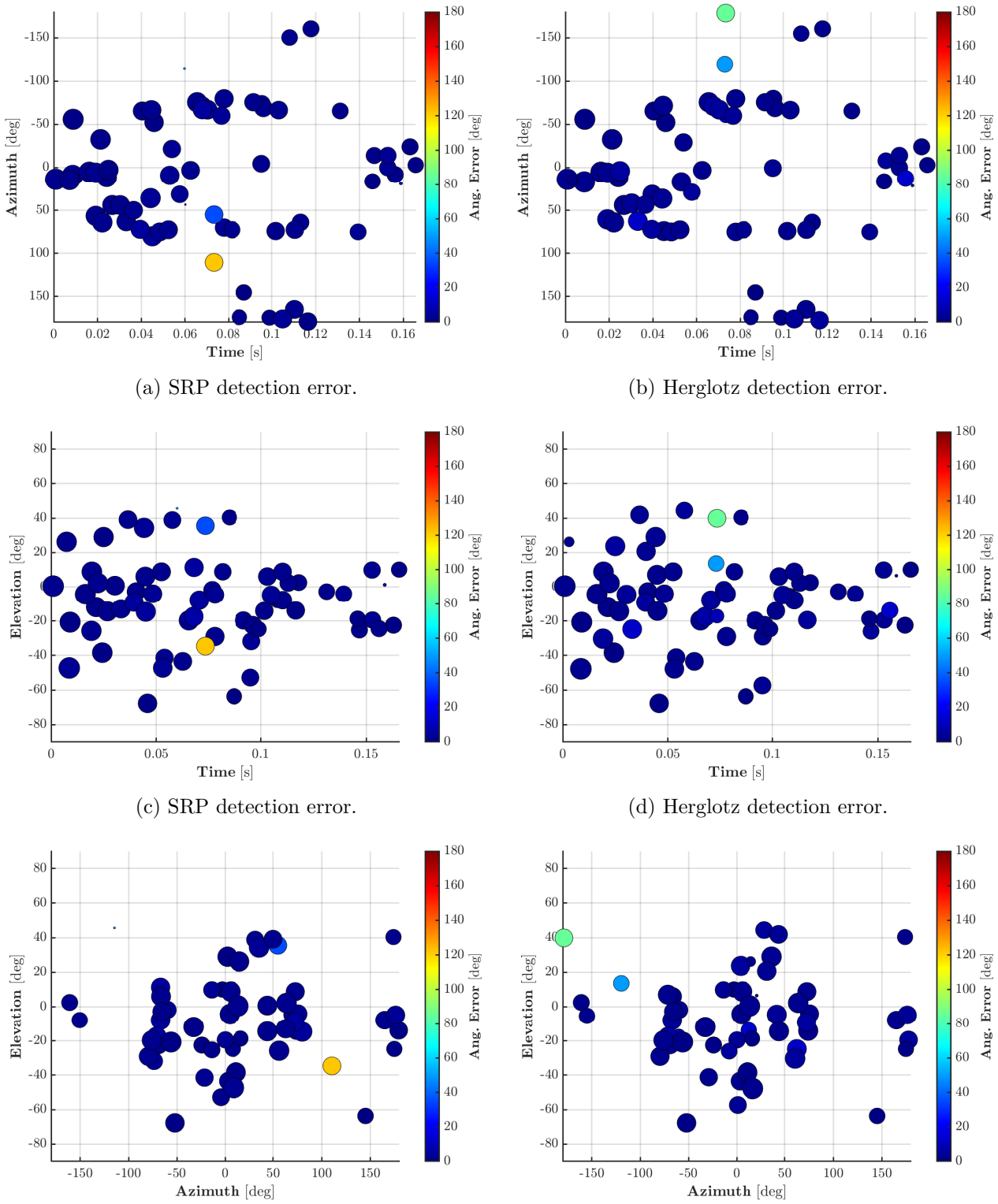


(b) SRP detection.



(c) Herglotz detection.

Figure B.18: Directional echo energy density (DEED) maps calculated on synthesized (a), SRP detected (b), and Herglotz detected (c) echoes from the EVERTims "Test Room 1" IS SRIR simulation. For more details on the DEED, see Sec. 2.4.3.



(a) SRP detection error.

(b) Herglotz detection error.

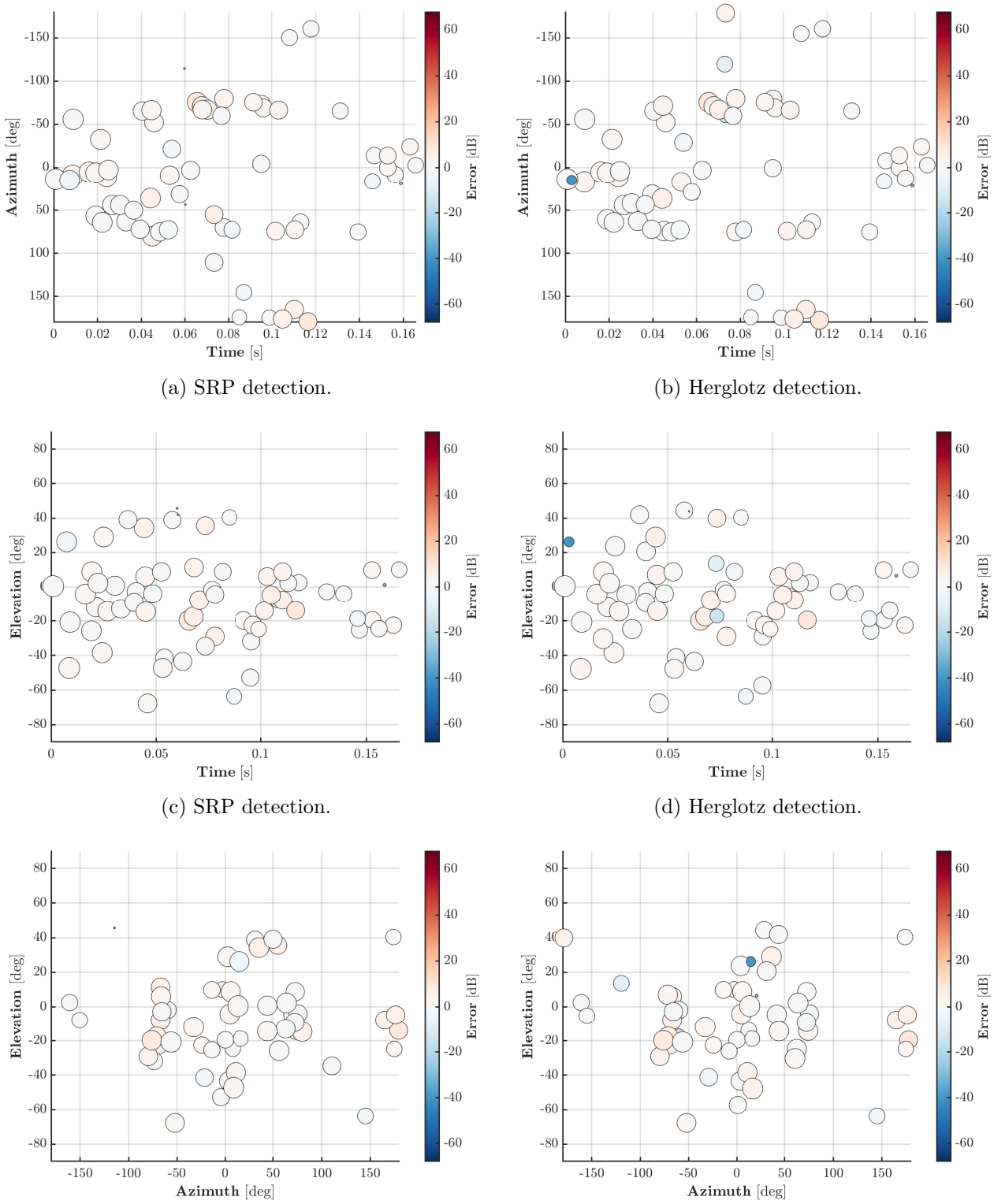
(c) SRP detection error.

(d) Herglotz detection error.

(e) SRP detection error.

(f) Herglotz detection error.

Figure B.19: Echo detection angular error maps for the EVERTims "Morgan" IS SRIR simulation. Point colours are relative to their angular errors (contained between 0° and 180° by definition), and sizes are again proportional to the estimated reflection energies. For more details on these error calculations, see again Sec. 4.2.1.



(a) SRP detection.

(b) Herglotz detection.

(c) SRP detection.

(d) Herglotz detection.

(e) SRP detection.

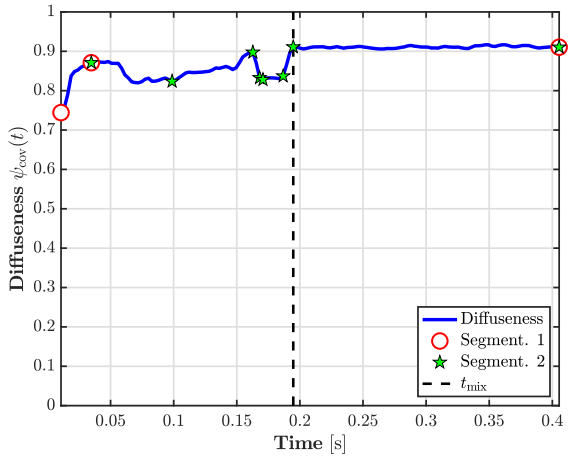
(f) Herglotz detection.

Figure B.20: Echo detection energy error maps for the EVERTims "Morgan" IS SRIR simulation. Point colours are relative to the energy estimation error of the corresponding reflections, and sizes are again proportional to their estimated energies. For more details on these error calculations, see again Sec. 4.2.1.

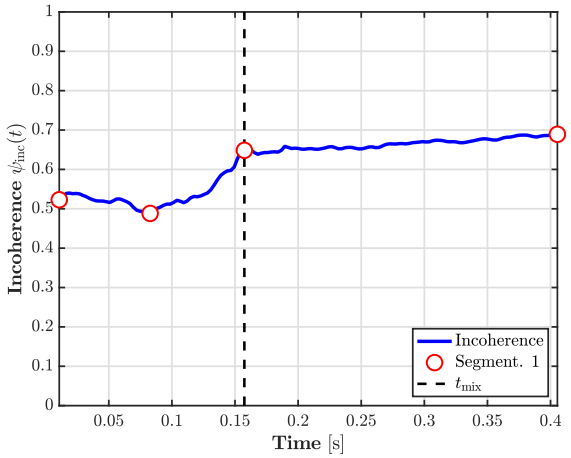
B.2/ Mixing Time Estimation

This appendix contains the complete mixing time estimation incoherence profile plots for the validation test presented in Sec. 4.2.2.

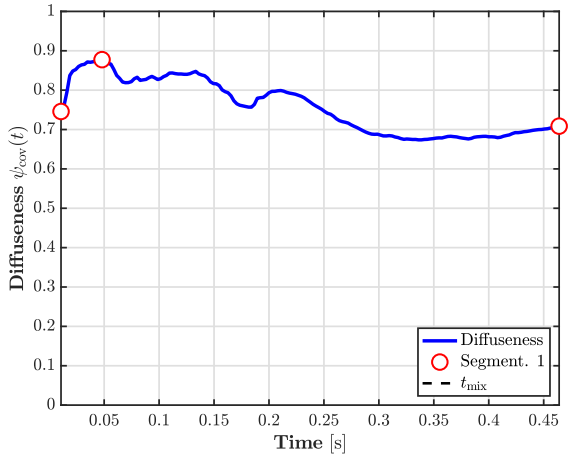
Test Room 1



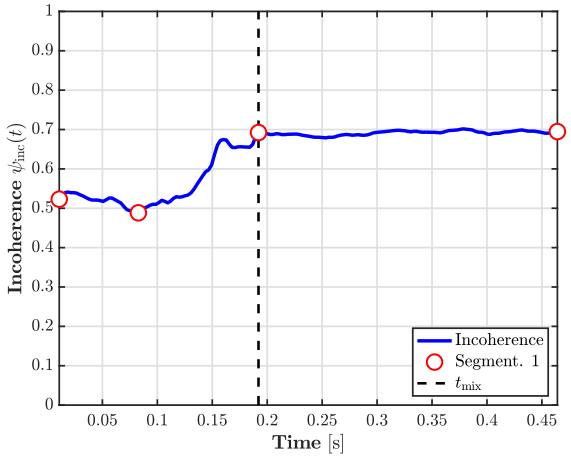
(a) SH-domain CoMEDiE diffuseness, isotropic late reverberation tail.



(b) DRIR spatial incoherence, isotropic late reverberation tail.



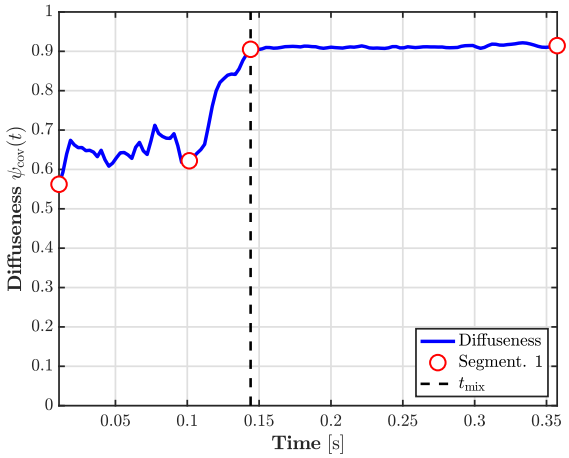
(c) SH-domain CoMEDiE diffuseness, anisotropic late reverberation tail.



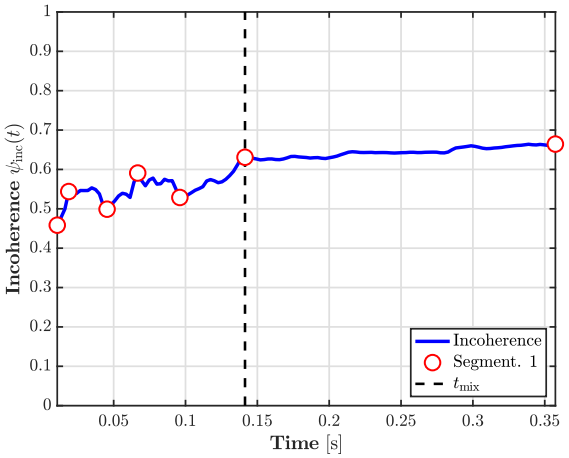
(d) DRIR spatial incoherence, anisotropic late reverberation tail.

Figure B.21: Mixing time estimation on the SH-domain CoMEDiE diffuseness (left, a & c) and DRIR spatial incoherence profiles (right, b & d) for both an isotropic (top, a & b) and an anisotropic (bottom, c & d) reverberation tail fitted to the "Test Room 1" IS SRIR simulation, as described in Sec. 4.2.2. For more details on the mixing time estimation procedure, see Sec. 2.3.

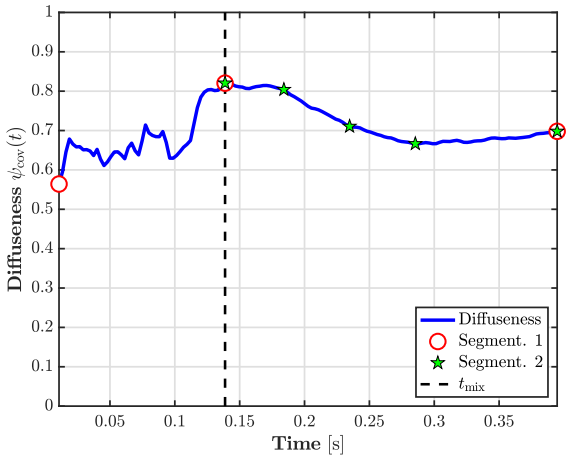
Test Room 2



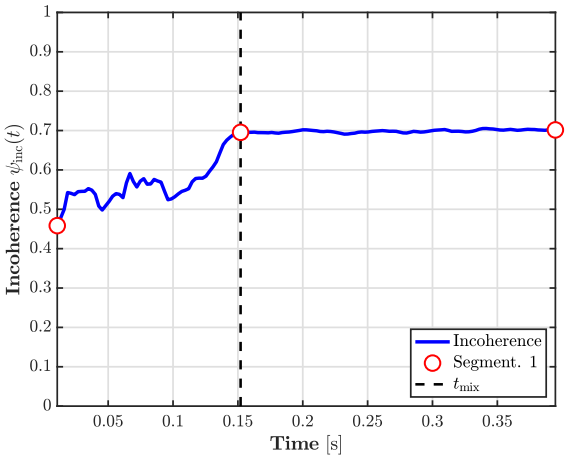
(a) SH-domain CoMEDiE diffuseness, isotropic late reverberation tail.



(b) DRIR spatial incoherence, isotropic late reverberation tail.



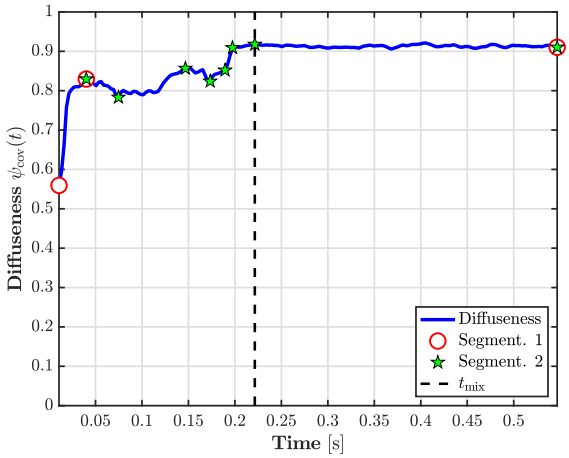
(c) SH-domain CoMEDiE diffuseness, anisotropic late reverberation tail.



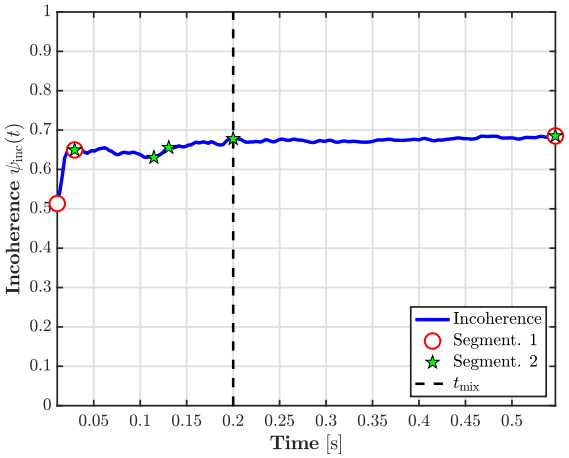
(d) DRIR spatial incoherence, anisotropic late reverberation tail.

Figure B.22: Mixing time estimation on the SH-domain CoMEDiE diffuseness (left, a & c) and DRIR spatial incoherence profiles (right, b & d) for both an isotropic (top, a & b) and an anisotropic (bottom, c & d) reverberation tail fitted to the "Test Room 2" IS SRIR simulation, as described in Sec. 4.2.2. For more details on the mixing time estimation procedure, see Sec. 2.3.

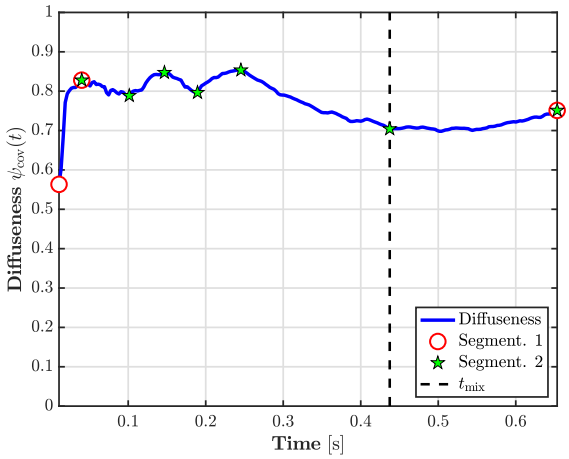
Test Room 3



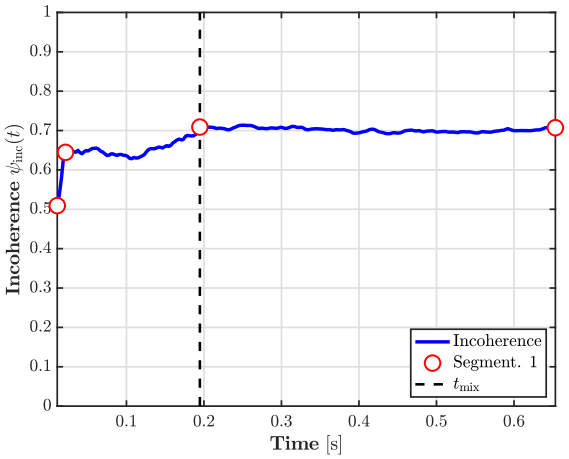
(a) SH-domain CoMEDiE diffuseness, isotropic late reverberation tail.



(b) DRIR spatial incoherence, isotropic late reverberation tail.



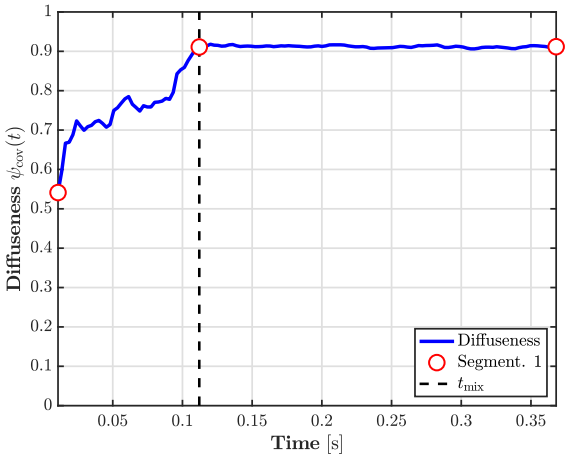
(c) SH-domain CoMEDiE diffuseness, anisotropic late reverberation tail.



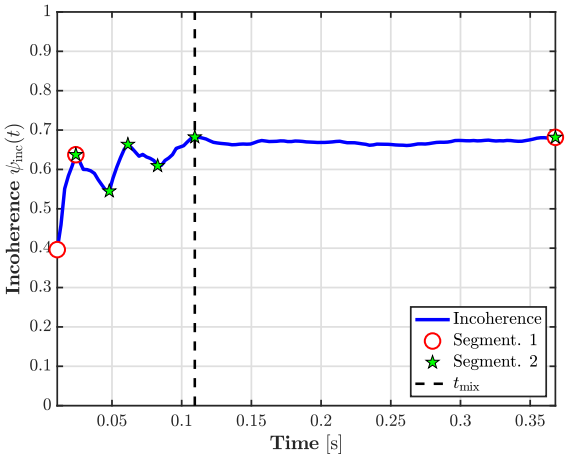
(d) DRIR spatial incoherence, anisotropic late reverberation tail.

Figure B.23: Mixing time estimation on the SH-domain CoMEDiE diffuseness (left, a & c) and DRIR spatial incoherence profiles (right, b & d) for both an isotropic (top, a & b) and an anisotropic (bottom, c & d) reverberation tail fitted to the "Test Room 3" IS SRIR simulation, as described in Sec. 4.2.2. For more details on the mixing time estimation procedure, see Sec. 2.3.

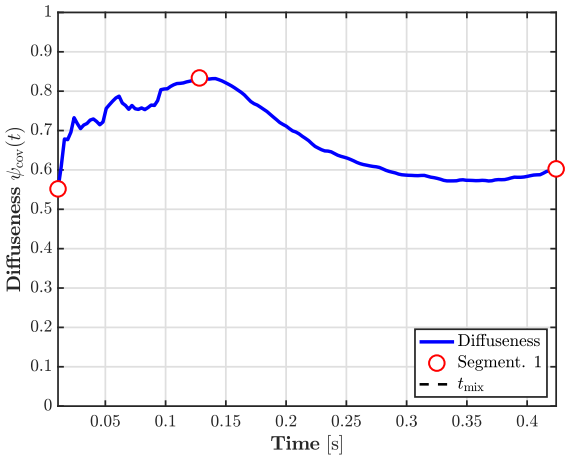
Fogg



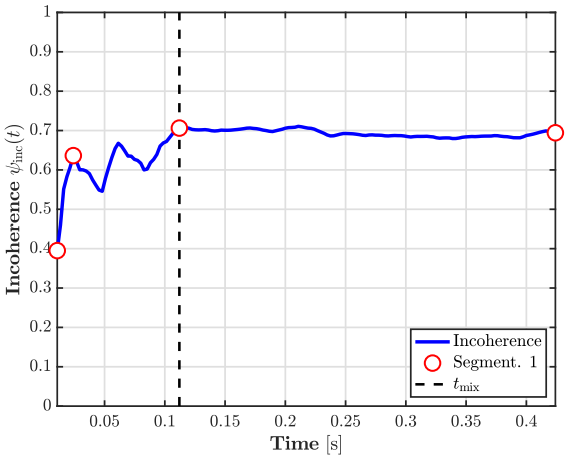
(a) SH-domain CoMEDiE diffuseness, isotropic late reverberation tail.



(b) DRIR spatial incoherence, isotropic late reverberation tail.



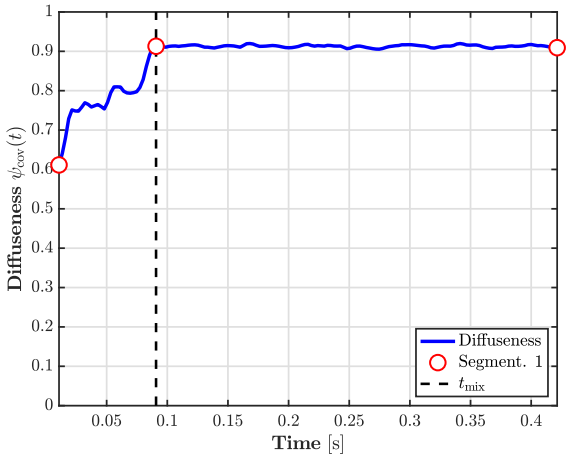
(c) SH-domain CoMEDiE diffuseness, anisotropic late reverberation tail.



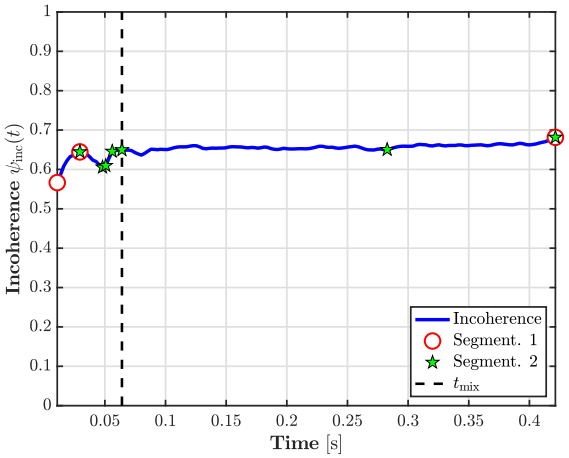
(d) DRIR spatial incoherence, anisotropic late reverberation tail.

Figure B.24: Mixing time estimation on the SH-domain CoMEDiE diffuseness (left, a & c) and DRIR spatial incoherence profiles (right, b & d) for both an isotropic (top, a & b) and an anisotropic (bottom, c & d) reverberation tail fitted to the "Fogg" IS SRIR simulation, as described in Sec. 4.2.2. For more details on the mixing time estimation procedure, see Sec. 2.3.

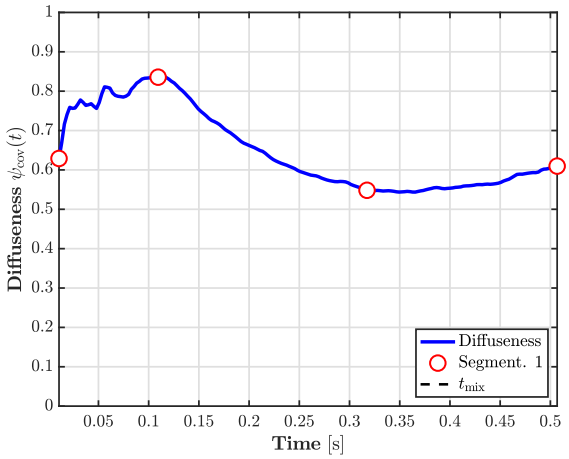
Morgan



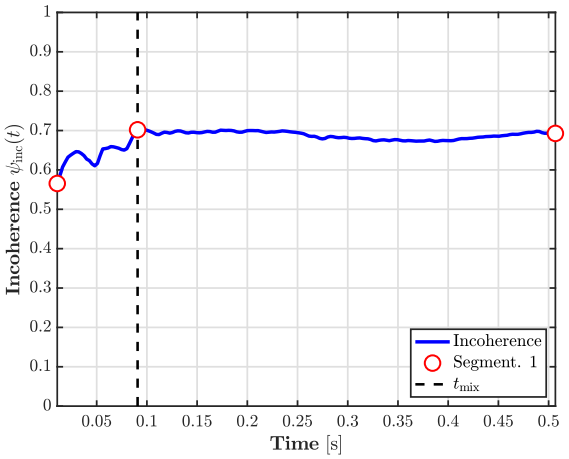
(a) SH-domain CoMEDiE diffuseness, isotropic late reverberation tail.



(b) DRIR spatial incoherence, isotropic late reverberation tail.



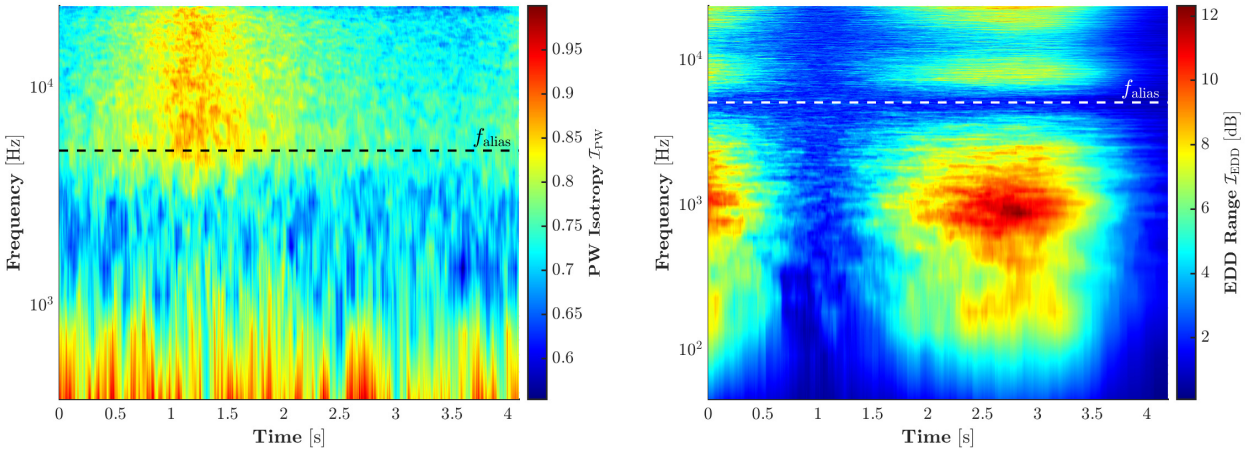
(c) SH-domain CoMEDiE diffuseness, anisotropic late reverberation tail.



(d) DRIR spatial incoherence, anisotropic late reverberation tail.

Figure B.25: Mixing time estimation on the SH-domain CoMEDiE diffuseness (left, a & c) and DRIR spatial incoherence profiles (right, b & d) for both an isotropic (top, a & b) and an anisotropic (bottom, c & d) reverberation tail fitted to the "Morgan" IS SRIR simulation, as described in Sec. 4.2.2. For more details on the mixing time estimation procedure, see Sec. 2.3.

B.3/ Late Tail Isotropy Evaluation

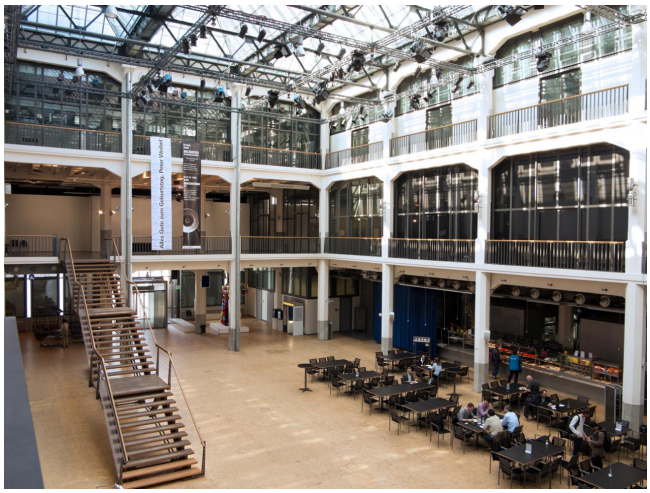


(a) PW isotropy measure (Eq. 1.26, Sec. 1.2.2).

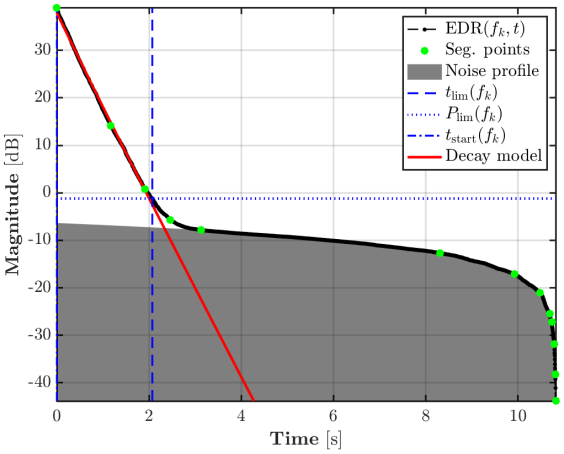
(b) Simulated SMA measurement EDD range (Eq. 2.37, Sec. 2.5.2).

Figure B.26: Frequency-dependent isotropy measures evaluated throughout the length of the anisotropic reverberation tail described and shown in Fig. 4.10. The PW isotropy measure $\mathcal{I}_{\mathrm{PW}}(t)$ (a, left) is compared to the EDD range measure $\mathcal{I}_{\mathrm{EDD}}(t)$ (b, right). Both are calculated on the simulated SMA measurement DRIR described in Sec. 4.2.4. The superimposed black (a, left) and white (b, right) dashed lines represent the spatial aliasing frequency $f_{\mathrm{alias}} = 5192$ Hz for the simulated 4^{th}-order Eigenmike SMA (radius $r_s = 4.2$ cm). Note that the curves in Fig. 4.11a correspond to the frequential averages of these two measures.

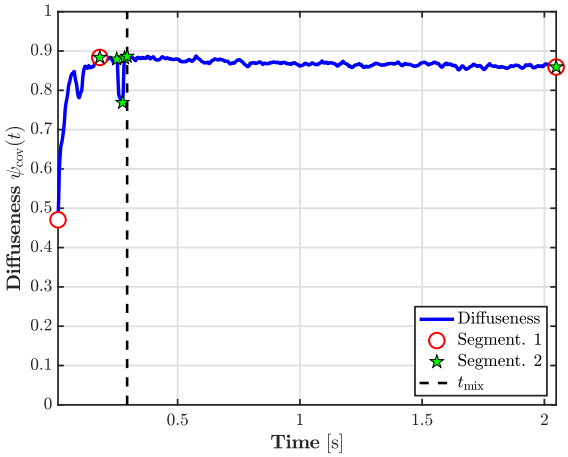
B.4/ Mixing Time Estimation (Measured SRIRs)



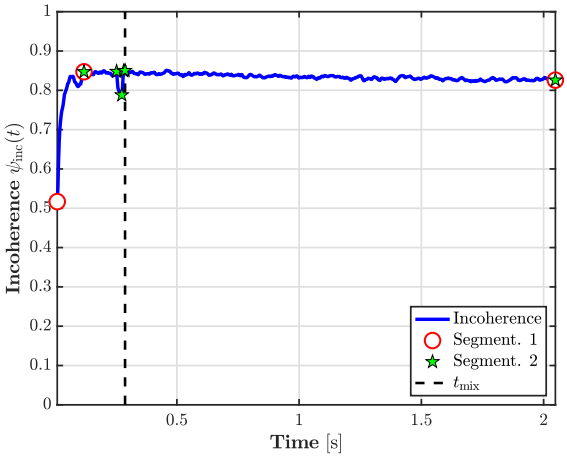
(a) View of the measured foyer at the ZKM Centre for Art and Media, Karlsruhe, Germany.



(b) Broadband $H_{0,0}(f,t)$ decay analysis, $\widetilde{T}_{60} = 2.89$ s.



(c) SH-domain CoMEDiE diffuseness ψ_{cov}, $\tilde{t}^{\text{cov}}_{\text{mix}} = 288$ ms.

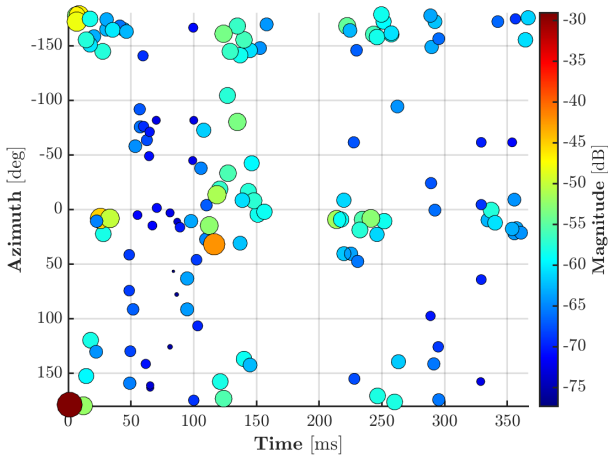


(d) DRIR spatial incoherence ψ_{dir}, $\tilde{t}^{\text{dir}}_{\text{mix}} = 293$ ms.

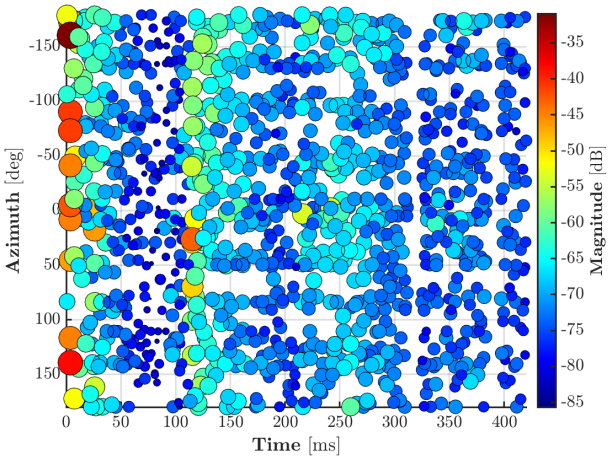
Figure B.27: Mixing time estimation for an SRIR measured in the main foyer at the Zentrum für Kunst und Medien (Centre for Art and Media, ZKM) in Karlsruhe, Germany, with a 32-capsule mh acoustics Eigenmike® SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4$^{\text{th}}$-order HOA. The DRIR representation used in (d) is obtained with a natural PWD beamformer on a 25-point Fliege-Maier [79] look direction layout grid. For more details on the mixing time estimation procedure, see Sec. 2.3.

B.5 / Echo Detection Cartographies (Measured SRIRs)

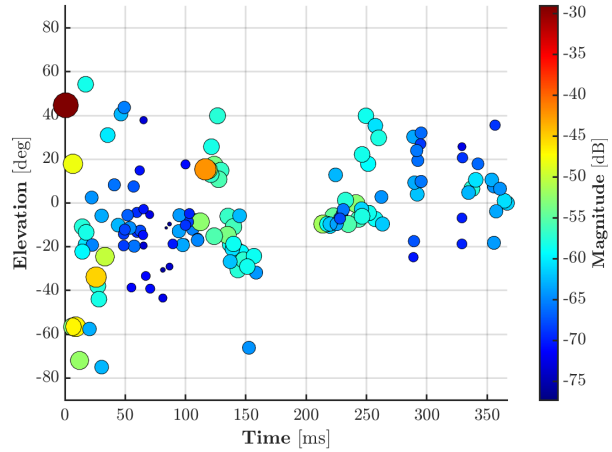
Notre-Dame Cathedral, Créteil, France



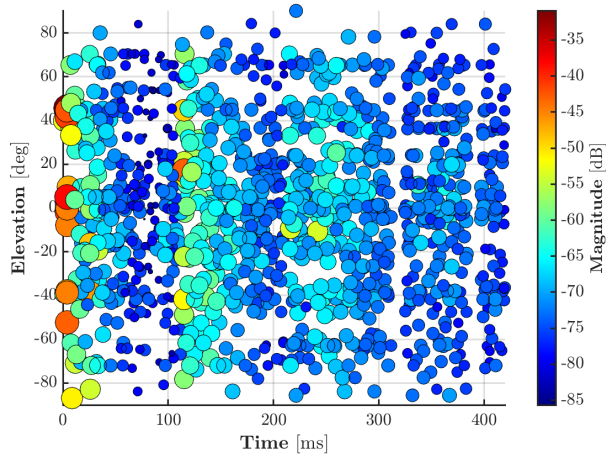
(a) SRP echo detection.



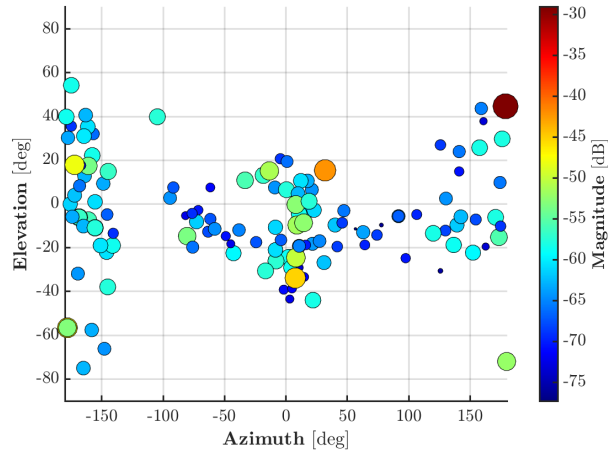
(b) Herglotz inversion echo detection.



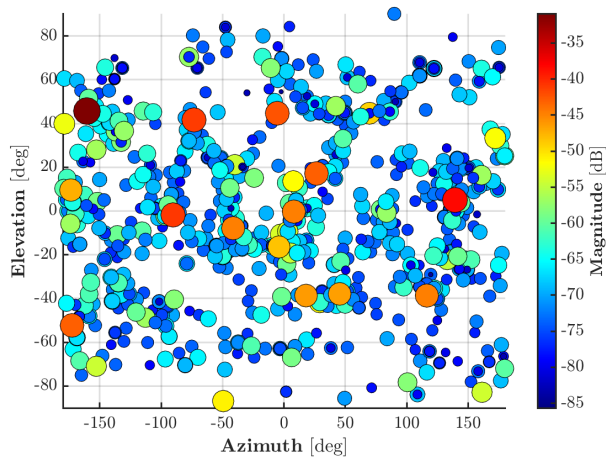
(c) SRP echo detection.



(d) Herglotz inversion echo detection.



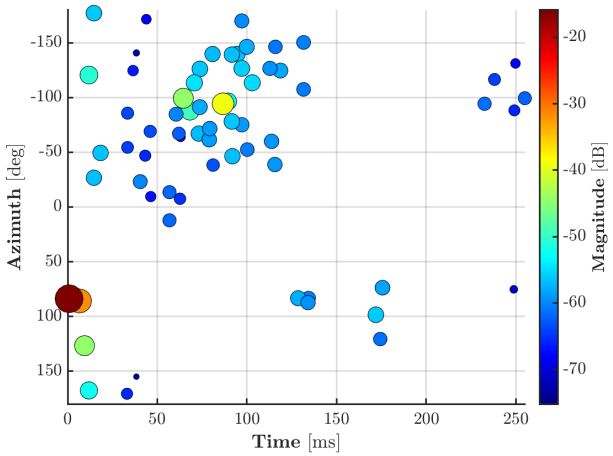
(e) SRP echo detection.



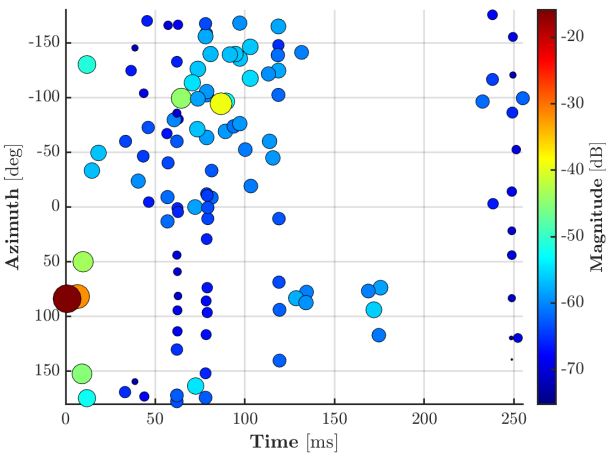
(f) Herglotz inversion echo detection.

Figure B.28: Early reflection cartography for an SRIR measured in the Notre-Dame Cathedral in Créteil, France, with a 32-capsule mh acoustics Eigenmike® SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4$^{\text{th}}$-order HOA. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4.

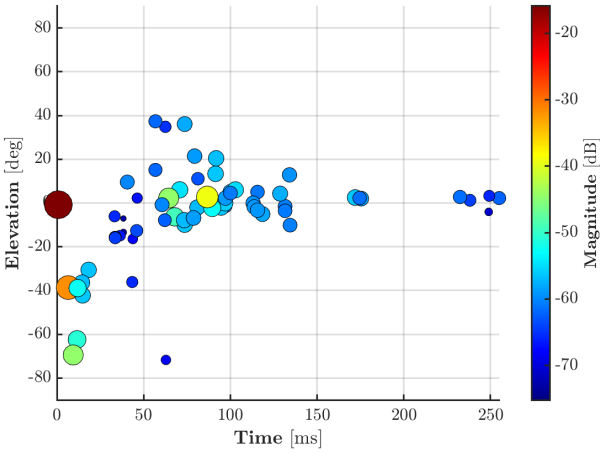
Dominicains de Haute-Alsace convent cloister, Guebwiller, France



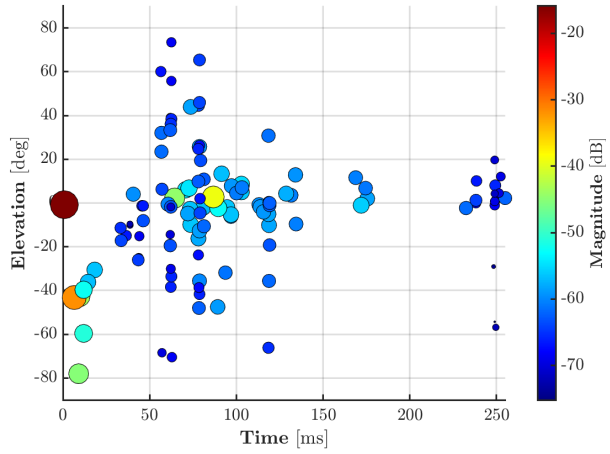
(a) SRP echo detection.



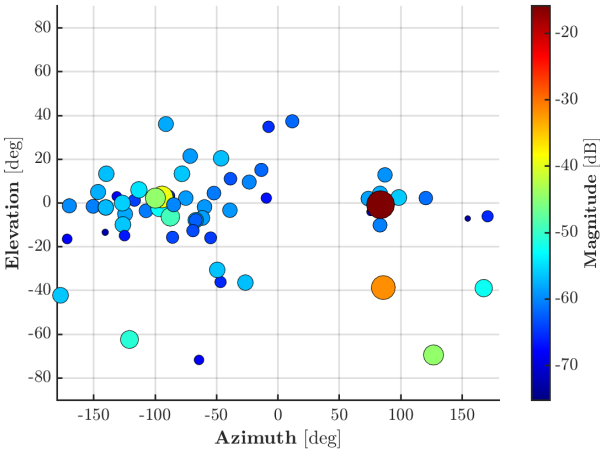
(b) Herglotz inversion echo detection.



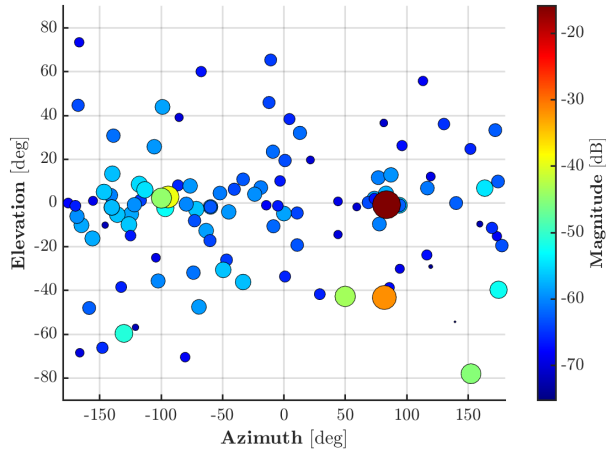
(c) SRP echo detection.



(d) Herglotz inversion echo detection.



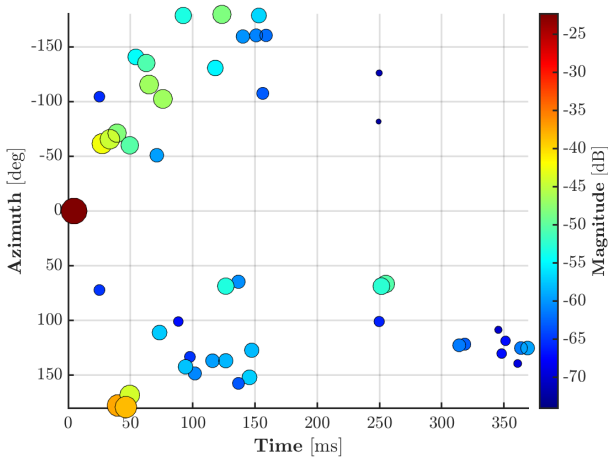
(e) SRP echo detection.



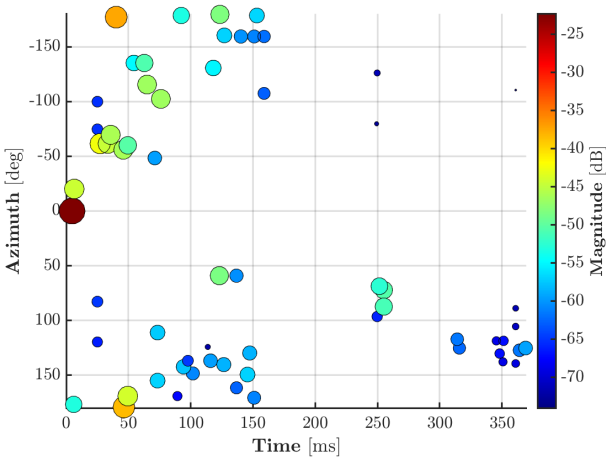
(f) Herglotz inversion echo detection.

Figure B.29: Early reflection cartography for an SRIR measured in the Dominicains de Haute-Alsace convent cloister in Guebwiller, France, with a 32-capsule mh acoustics Eigenmike® SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4^{th}-order HOA. For more details on the SRP and regularized Herglotz inversion methods, see Sec. 2.4.

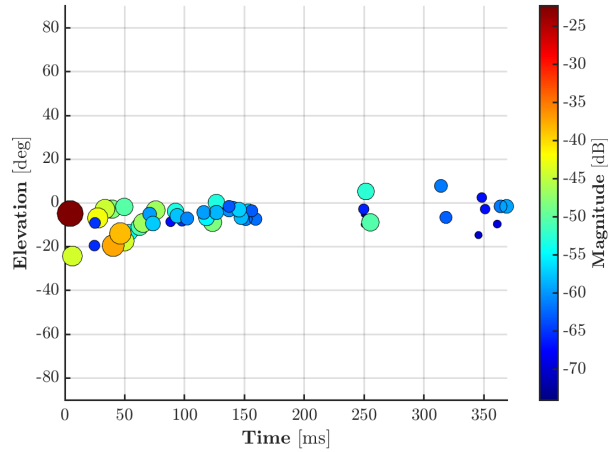
La Défense esplanade, Puteaux, France



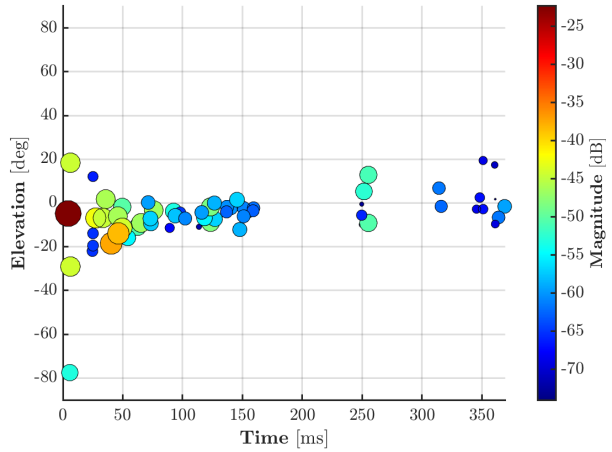
(a) SRP echo detection.



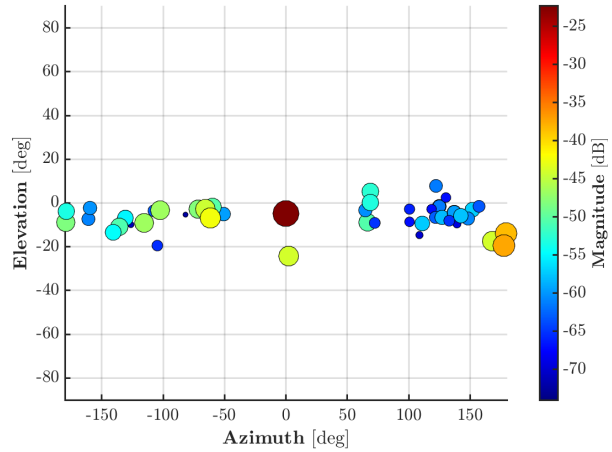
(b) Herglotz inversion echo detection.



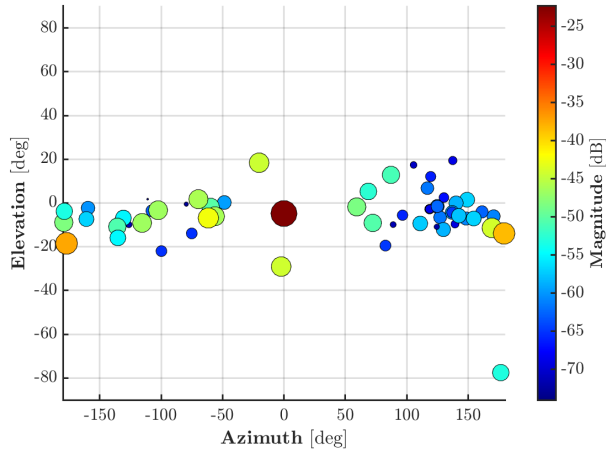
(c) SRP echo detection.



(d) Herglotz inversion echo detection.



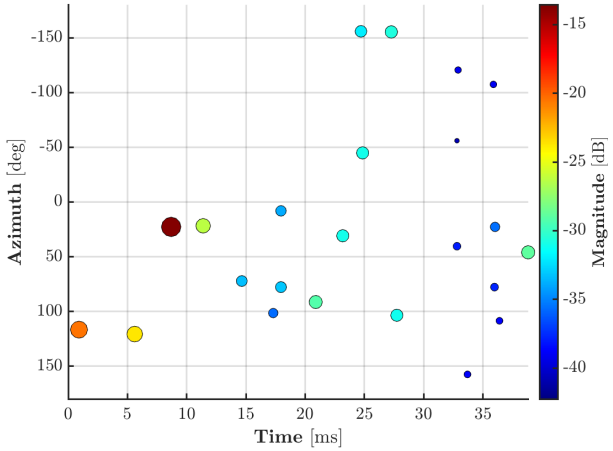
(e) SRP echo detection.



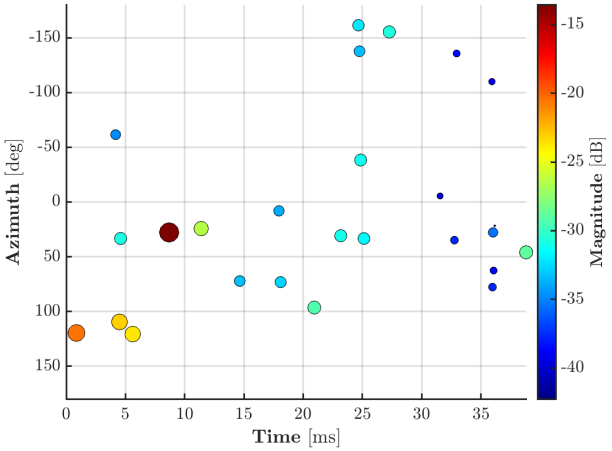
(f) Herglotz inversion echo detection.

Figure B.30: Early reflection cartography for an SRIR measured on the esplanade of La Défense in Puteaux, France, with a 32-capsule mh acoustics Eigenmike® SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4^{th}-order HOA. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4.

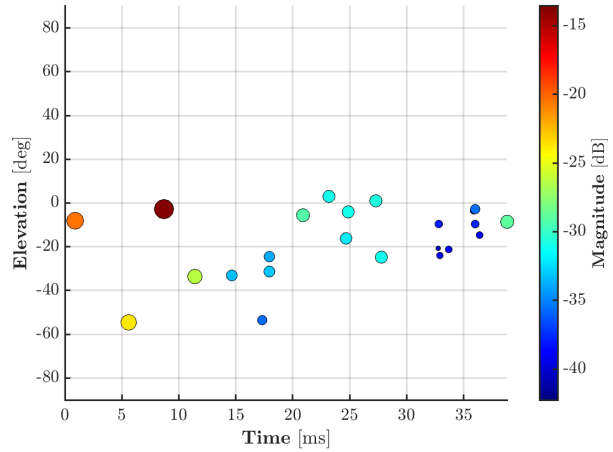
Church of St. Eustache, Paris, France



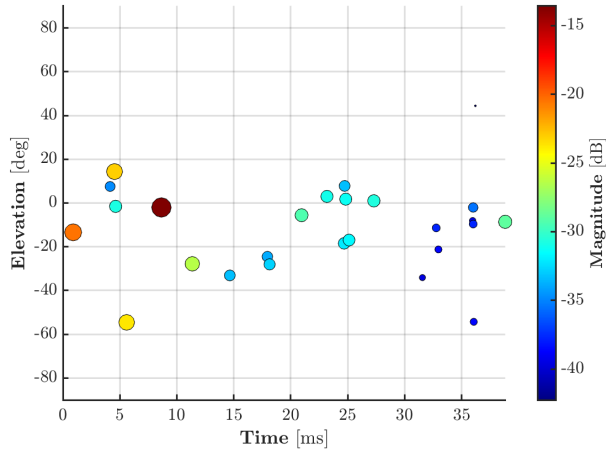
(a) SRP echo detection.



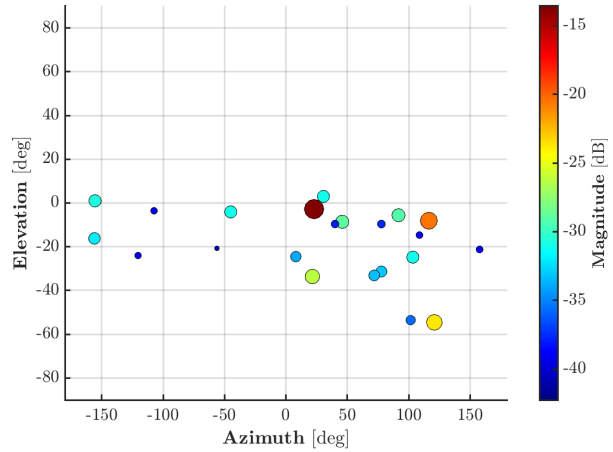
(b) Herglotz inversion echo detection.



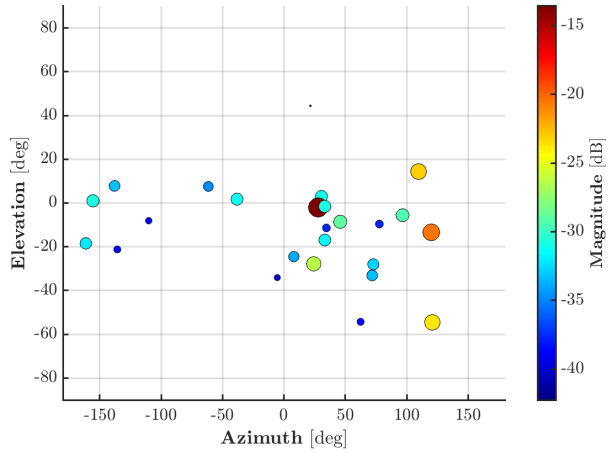
(c) SRP echo detection.



(d) Herglotz inversion echo detection.



(e) SRP echo detection.

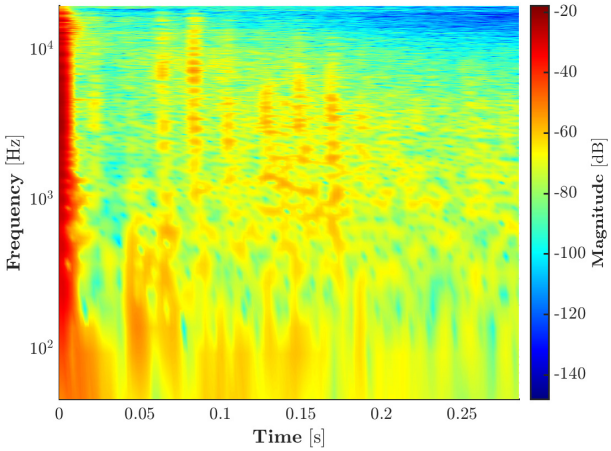


(f) Herglotz inversion echo detection.

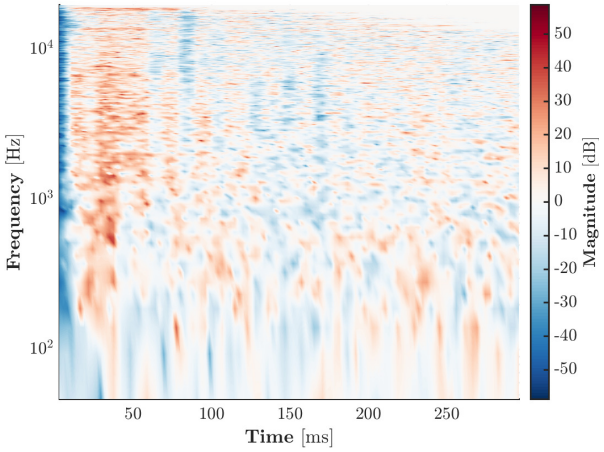
Figure B.31: Early reflection cartography for an SRIR measured in the Church of St. Eustache, Paris, France, with a 32-capsule mh acoustics Eigenmike® SMA. The measured SRIR is pre-treated with the method described in Sec. 3.1 and then encoded to 4^{th}-order HOA. For more details on the SRP and regularized under-determined Herglotz inversion methods, see Sec. 2.4.

B.6/ Early Reflection Salience Manipulation

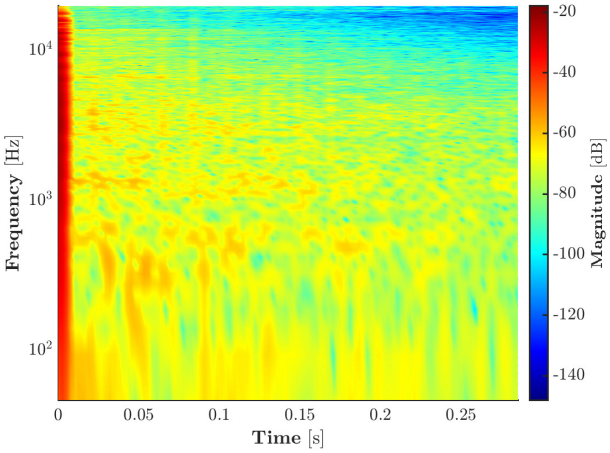
Dominicains de Haute-Alsace convent cloister, Guebwiller, France



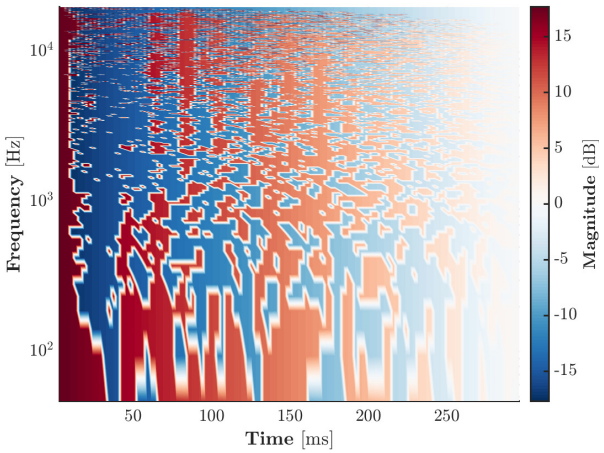
(a) Original early $H(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$.



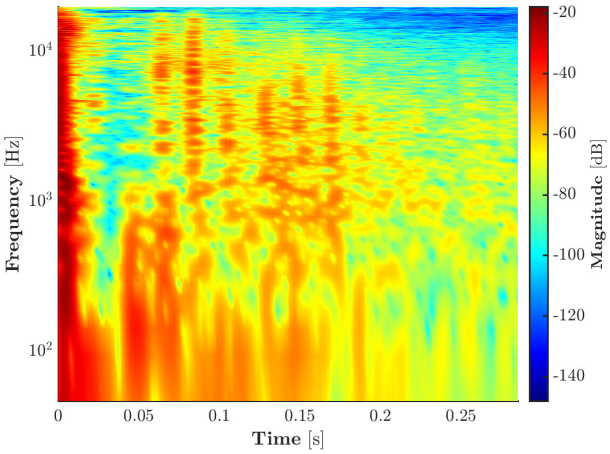
(b) Modification map $M_{\mathrm{exp}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ to constrain echoes to the late energy decay envelope.



(c) Modified $\widetilde{H}_{\mathrm{exp}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ with echoes constrained to the late energy decay envelope.



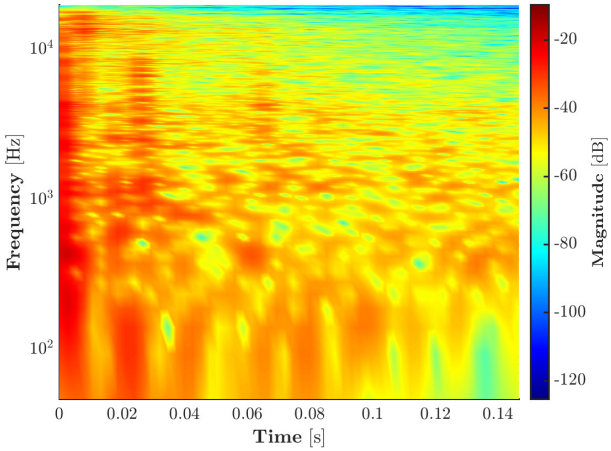
(d) Modification map $M_{\mathrm{acc}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ for accentuating salience with respect to the late energy decay.



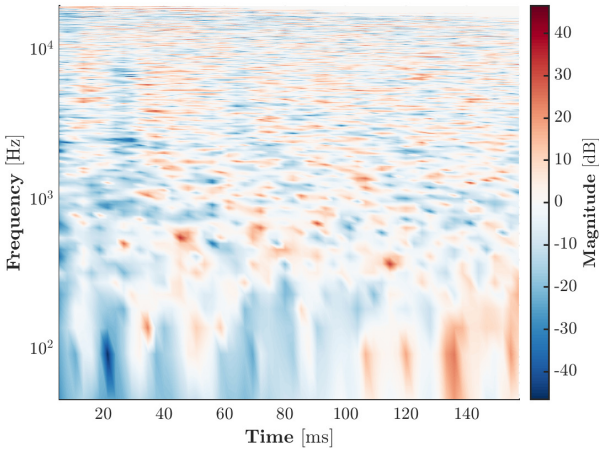
(e) Modified $\widetilde{H}_{\mathrm{acc}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ with increased echo salience.

Figure B.32: Early reflection salience manipulation for an SRIR measured in the Dominicains de Haute-Alsace convent cloister in Guebwiller, France, with a 32-capsule mh acoustics Eigenmike® SMA. For more details on the reflection salience manipulation process, see Secs. 3.2.1, 3.2.2 and 5.3.3.

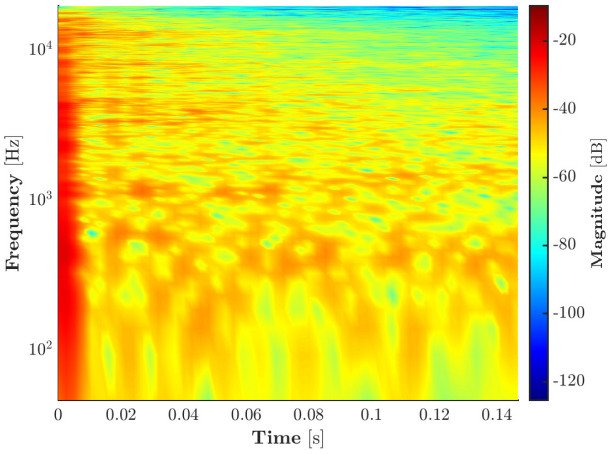
Church of St. Eustache, Paris, France



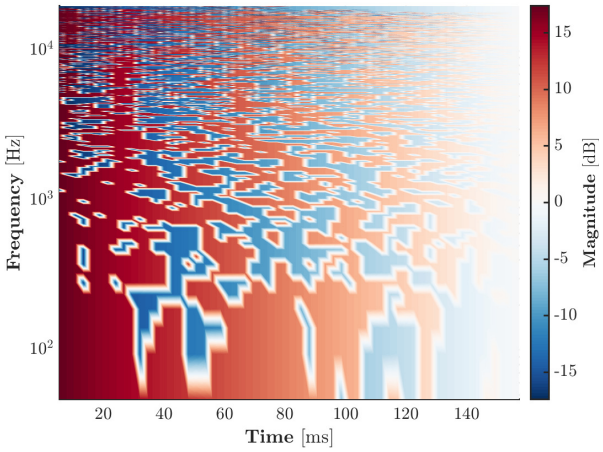
(a) Original early $H(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$.



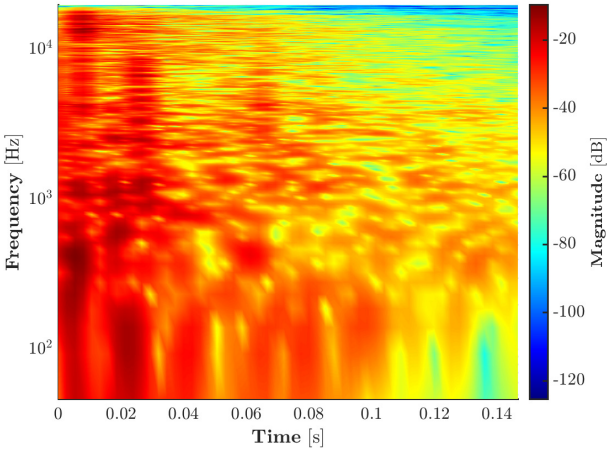
(b) Modification map $M_{\mathrm{exp}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ to constrain echoes to the late energy decay envelope.



(c) Modified $\widetilde{H}_{\mathrm{exp}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ with echoes constrained to the late energy decay envelope.



(d) Modification map $M_{\mathrm{acc}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ for accentuating salience with respect to the late energy decay.

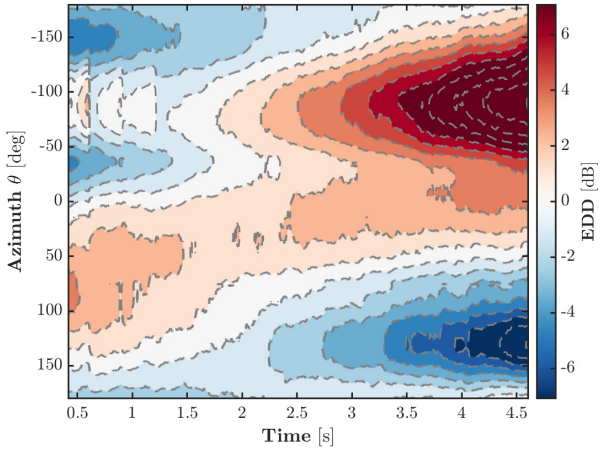


(e) Modified $\widetilde{H}_{\mathrm{acc}}(f, \mathbf{\Omega}_{\mathrm{DS}}, t)$ with increased echo salience.

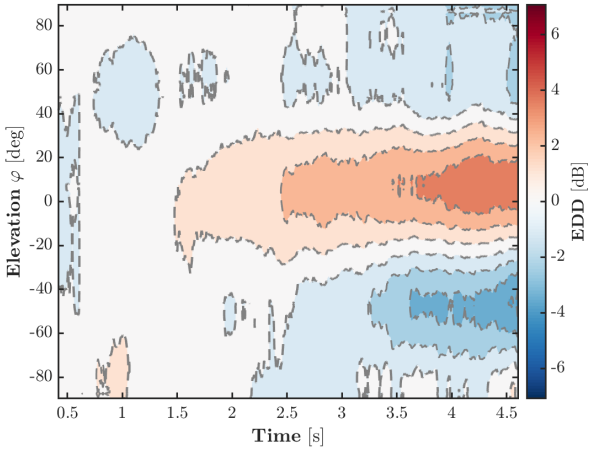
Figure B.33: Early reflection salience manipulation for an SRIR measured in the Church of St. Eustache, Paris, France, with a 32-capsule mh acoustics Eigenmike® SMA. For more details on the reflection salience manipulation process, see Secs. 3.2.1, 3.2.2 and 5.3.3.

B.7/ Late Tail Isotropy Manipulation

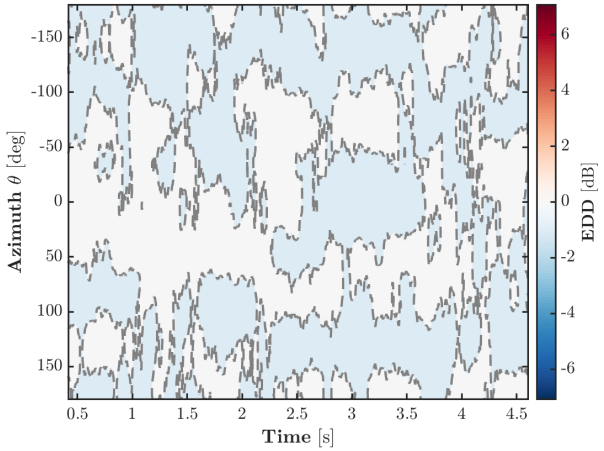
Isotropification



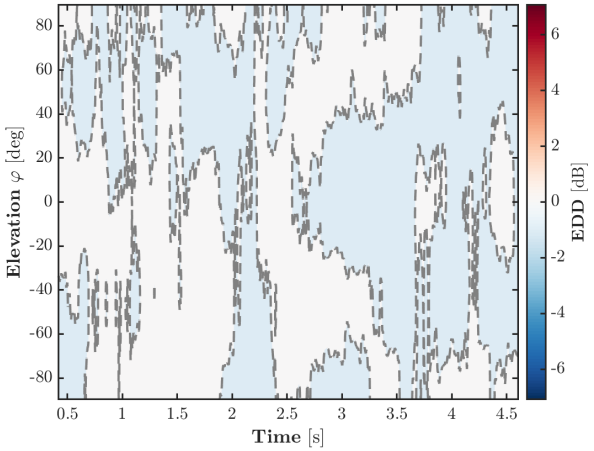
(a) Reference denoised azimuthal EDD.



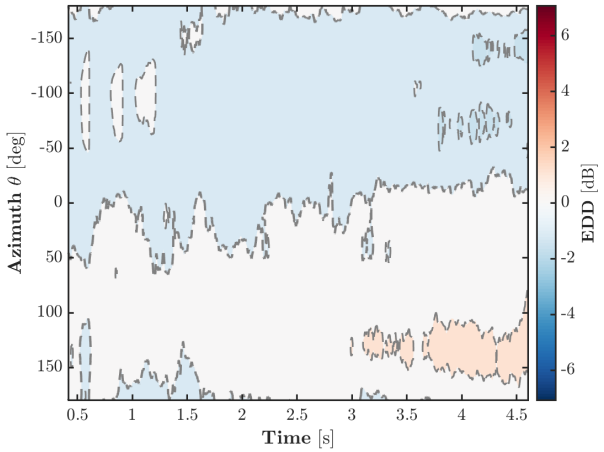
(b) Reference denoised elevational EDD.



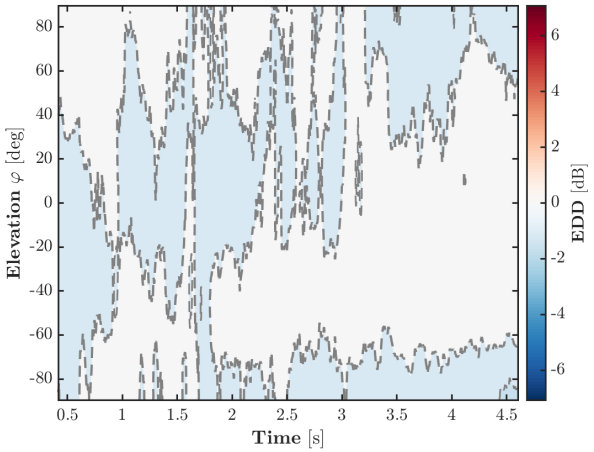
(c) Tail resynthesis azimuthal EDD, $v_{\text{iso}} = 1$.



(d) Tail resynthesis elevational EDD, $v_{\text{iso}} = 1$.



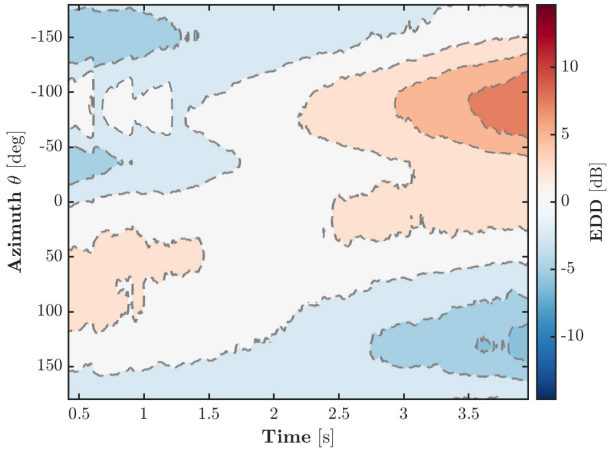
(e) EDD compensation azimuthal EDD, $v_{\text{iso}}^{\text{EDD}} = 1$.



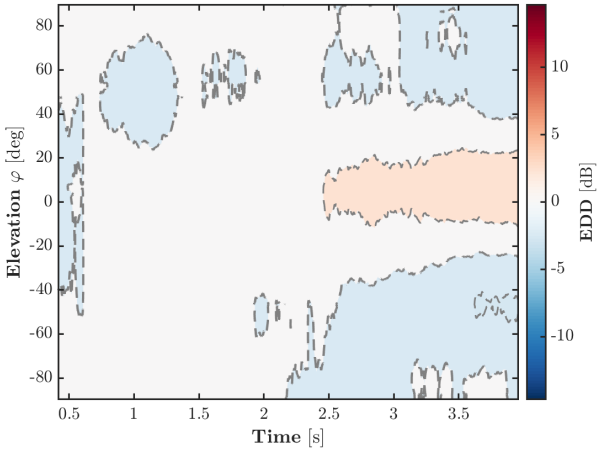
(f) EDD compensation elevational EDD, $v_{\text{iso}}^{\text{EDD}} = 1$.

Figure B.34: Azimuthal and elevational plane EDDs demonstrating the effect of isotropification as applied by (c-d, middle) tail resynthesis and (e-f, bottom) EDD compensation to the SRIR measured at the Grandes Serres de Pantin in Pantin, France. The EDDs of the "original" denoised SRIR is included in (a-b, top) as a reference. The estimated direct sound DoA is $\mathbf{\Omega}_{\text{DS}} \simeq (89.1°, -14.9°)$. For more details on the tail isotropy manipulation methods, see Secs. 3.3.2 and 5.4.4 and Sec. 3.4.

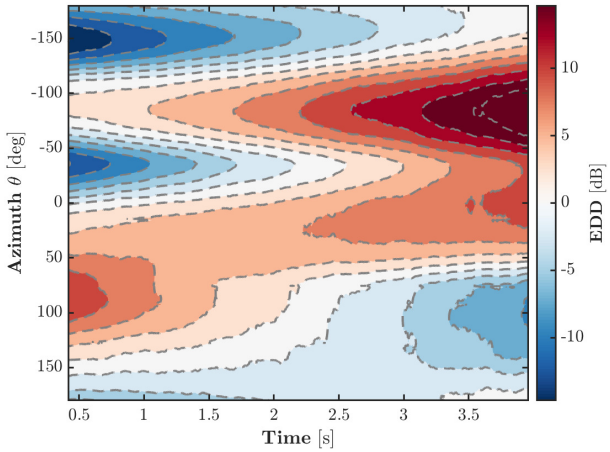
Directification



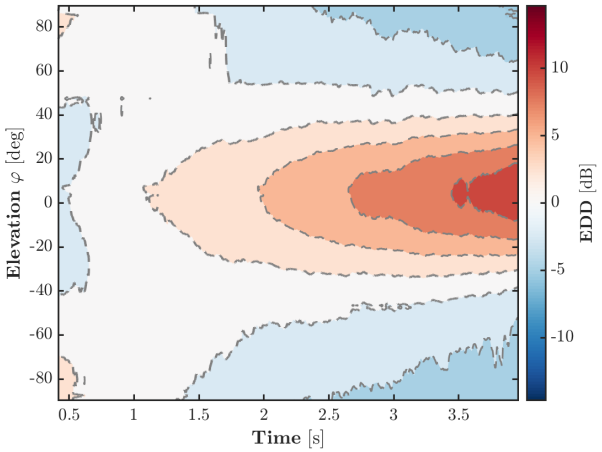
(a) Reference denoised azimuthal EDD.



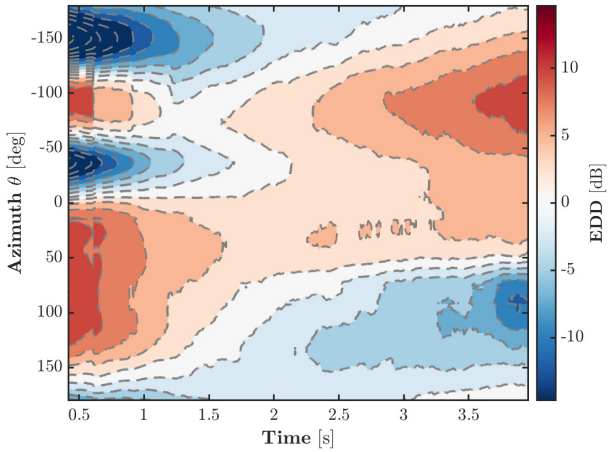
(b) Reference denoised elevational EDD.



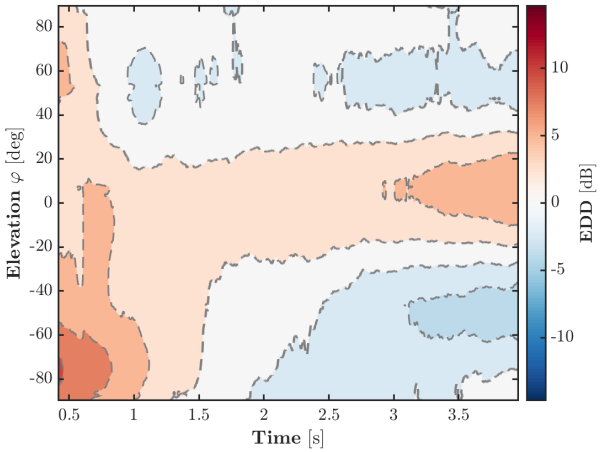
(c) Tail resynthesis azimuthal EDD, $v_{\text{dir}} = 2$.



(d) Tail resynthesis elevational EDD, $v_{\text{dir}} = 2$.



(e) EDD compensation azimuthal EDD, $v_{\text{dir}}^{\text{EDD}} = 18$ dB.



(f) EDD compensation elevational EDD, $v_{\text{dir}}^{\text{EDD}} = 18$ dB.

Figure B.35: Azimuthal and elevational plane EDDs demonstrating the effect of directification as applied by (c-d, middle) tail resynthesis and (e-f, bottom) EDD compensation to the SRIR measured at the Grandes Serres de Pantin in Pantin, France. The EDDs of the "original" denoised SRIR is included in (a-b, top) as a reference. The estimated direct sound DoA is $\mathbf{\Omega}_{\text{DS}} \simeq (89.1°, -14.9°)$. For more details on the tail isotropy manipulation methods, see Secs. 3.3.2 and 5.4.4 and Sec. 3.4.