Optimum-phase primal signal and radiation-filter modelling of musical instruments

Master Thesis

Franck ZAGALA

Supervisor: Dr. Franz ZOTTER
Assessor: O.Univ.Prof. Dr. Robert HÖLDRICH

Institute of Electronic Music and Acoustics
University of Music and Performing Arts, Graz
January 2019

Statutory declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Franck ZAGALA
Graz, January 2019

Acknowledgements

I would like to thank especially Franz Zotter for his outstanding supervision. I could not have imagined a better supervisor than you.

Franz, vielen Dank für alles!

I also thank all the IEM staff for the nice workspace their offered me throughout this work. I really enjoyed your help, support, as well as all the coffee breaks and lunch breaks we shared. Special thanks to Daniel who voluntarily served as a first guinea pig for all the perceptual experiments and made wise suggestions for their improvement.

I also would like to thank all the colleagues I had the pleasure to work with at ÖAW, IRCAM, and SPSC during my studies and who contributed in some way to this thesis.

Thank you also to my parents for their support during my whole studies.

Last but not least, I would like to thank my friends Aurélien, Caroline, Eva, Julian, Margaux, Paulus, Tatjana and all the others, thank you for being there!

Danke und pfiat'di Graz! Es war mir eine Ehre!

Abstract

Capture of musical instruments including directivity by spherically surrounding microphone array often leads to either overly complex radiation patterns or destructive interferences at high frequencies. There, small differences in the distance from the instrument to each of the microphones yield different arrival times in the captured signal and hereby large phase differences. Therefore, be it by linear triangular or spherical harmonics interpolation, frequency-independent directional interpolation of the microphone signals can result either in spectral degradation or in a shift of the signal energy to higher orders; especially affecting high frequencies. In this work, an analysis method is proposed in order to decompose directional signals of any measured instrument into an interference-free primal signal and a directivity filter. Both the directivity filter and the primal source signal utilise only the short-term spectral magnitude of each microphone signal in order to avoid artefacts. The magnitudes are complemented by a simplified phase using phase-retrieval techniques.

Kurzfassung

Die Aufnahme musikalischer Instrumenten mit Richtwirkung anhand einer umhüllenden sphärischen Mikrofon-Anordnung führt häufig entweder zu einer komplizierten Abstrahlcharakteristik oder zu destruktiven Interferenzen bei hohen Frequenzen. Dabei ergeben kleine Distanzunterschiede vom Instrument zu jedem Mikrofon eine unterschiedliche Ankunftszeit und dadurch entstehen große Phasendifferenzen. Aus diesem Grund können frequenzunabhängige Richtungsinterpolationsverfahren für die Mikrofonsignale, sei es durch lineare Dreiecks- oder Kugelflächenfunktionsinterpolation, entweder zu spektralen Verschlechterungen führen oder zu einer Verschiebung der Signalenergie zu höheren Ordnungen. Beides beeinträchtigt das Ergebnis bei hohen Frequenzen. In dieser Arbeit wird eine Analysemethode vorgeschlagen, um Richtungsignale beliebiger Instrumente in ein interferenzfreies Ursignal und in Richtwirkungsfilter zu zerlegen. Die Bestimmung von beidem, Ursignal und Richtwirkungsfilter, basiert ausschließlich auf den Kurzzeitbetragsspektren der Mikrofonsignale, um Artefakten zu vermeiden. Die Betragsspektren werden mit einer vereinfachten Phase aus Phasenrekonstruktionsverfahren ergänzt.

Résumé

La captation d'instruments avec leur directivitée, à l'aide d'un réseau sphérique environnant de microphones entraine souvent, soit une fonction de directivitée excessivement complexe, soit des interférences destructives dans les hautes fréquences. En l'occurence, de légères différences entre les distances source-microphones induisent des différences de temps d'arrivée et amènent, ce faisant, à un déphasage. Pour cette raison, une interpolation indépendante de la fréquence, que ce soit une interpolation triangulaire ou une interpolation se basant sur les harmoniques sphériques, aboutit soit à une dégradation spectrale, soit à un déplacement de l'énergie du signal vers des ordres plus élevés. Dans ce travail, une méthode est proposée afin de décomposer n'importe quel instrument en un signal primaire et un filtre de directivitée tout en garantissant l'absence d'interférences. Le signal primaire, ainsi que le filtre de directivitée sont déduits uniquement à partir des spectres d'amplitudes courts termes des microphones afin d'éviter tout artéfact. La magnitude est complétée par une phase simplifiée en se basant sur différents algorithmes de récupération de phase.

Contents

Chapter 1

Introduction

The study of sound source radiation has been fascinating acousticians for decades, as its studies appear to be beneficial for different domains, such as room acoustics, virtual reality, acoustics simulations or music recordings. In fact, most instruments do not radiate sound omnidirectionally, i.e. the intensity of the radiated signal may depend on the direction of emission, on the frequency but also on the playing techniques or the posture of the player [Mey09]. This inherent complexity of sound sources radiation makes its understanding, virtualisation and restitution a great challenge, which can involve several fields from theoretical physics to psychoacoustic through signal processing. Perceptually, its understanding is capital, as changes in the source radiation characteristics strongly influence the auditory experience [DKS93]. Although the direct signal emitted by the source in the direction of the listener is crucial from a perceptual point of view, the rest of the emitted signal is also particularly of interest, as is may reach the listener through different propagation paths that depend of the surrounding environment and therefore determines how the room ”reacts” to the source.

Among the initiators of the field, one can cite Jürgen Meyer who first investigated the directivity of various orchestral instruments by providing the average directivity for different frequency bands [Mey09]. By this averaging, Meyer's model of sound source radiation remains simplistic, but it delivers knowledge that can be of great interest for audio recording engineers, acousticians, conductors or musicians. Nevertheless, there are still many research questions, as for instance: how can concrete directivity evolve over time? Is there a holographic virtualisation of played instruments?

In this regard, the use of a more advanced mathematical radiation model is desired in order to capture, process and restitute the instruments emission with the greatest perceptual fidelity in a systematic way.

Weinreich and Arnold first proposed the use of spherical expansion from spatially discrete measurements [WA80]. The hardware designed by Weinreich and Arnold made use of 2 microphones mounted on a very elaborated structure, that may not necessarily enable a straightforward virtualisation of the source, however a first link between measurement techniques and mathematical concepts of source modelling were made.

Some of the history can be found in the dissertation of Zotter [Zot09] in Chapter 5.

In 2009, Hohl has constructed a 64-channel spherical microphone array at the Institute of Electronic Music and Acoustics (IEM) in Graz to enable to measurement of radiation up to the 7^{th} order [Hoh09] without need of a turntable, which is quite appropriate for the virtualisation of a musical instrument including its potentially time-varying radiation pattern while played and its musical excitation that is difficult to reproduce by a technical device. However, the virtualisation and auralisation of sources based on such recordings often suffer from different spectral artefacts due to bad centring, great size of the instrument or insufficient spatial sampling.

In recent years, some authors proposed to simplify radiation patterns based on acoustic centring approaches [BHPVR11, SV15, DZ10, DZ11]. Due to the computational complexity required, acoustic centring appears to be unpractical, and alternatives may be more attractive [Süs11, Hol14, Mit16]. Still those alternatives require further developments and investigations.

This work aspires to propose some interesting methods that can be utilised to model an instrument into a primal source signal and a complementary radiation filter by replacing the signals phase by a simplified one. This phase replacement aims to counter spectral artefacts while simplifying the directivity pattern of the instrument.

The work is organised as follows:

Chapter 2 describes some conventional techniques for radiated sound interpolation. This section is also used to introduce the problems caused by the recorded multichannel signal phase and to demonstrate how a spatial phase modification could be beneficial to compensate for the artefacts.

Chapter 3 presents some phase retrieval techniques on the sphere, which may serve as simplification of the radiation pattern based on its spatially sampled magnitude.

Chapter 4 deals with the primal signal and radiation-filter modelling based uniquely on the measured short-term spectral magnitude of surrounding microphones.

Perceptual evaluation of the spatial interpolation techniques, reconstructed primal signal and virtualised instrument based on the primal source signal and radiation-filter decomposition are undertaken in Chapter 5.

Chapter 2

Interpolation of radiated signals from a surrounding spherical array

When auralising multi-channel surrounding spherical array recording of instruments, the spatial interpolation of directivity signals enables to obtain the signal radiated in each specific direction, also for the potentially time-varying directions of an interactive virtual environment. For the room acoustics simulation, geometrical acoustic models, such as image-source, ray-tracing or beam-tracing, where propagation paths have a known radiation direction, can be fed with the interpolated signal [SS15]. This section, discusses some basic interpolations methods suitable for surrounding microphone array signals such as the hyperinterpolation by means of spherical harmonics decomposition as well as Vector Base Amplitude Panning, and a phase-modified version of it.

2.1 Hyperinterpolation with spherical harmonics

The pressure signal emitted by a source enclosed in a sphere with radius smaller than r can be expressed using the elementary solutions of the Helmholtz differential equation in a spherical coordinate system for the external problem [Wil99]

$$x\left(\tfrac{\omega}{c}r, \boldsymbol{\theta}\right) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_l^m(\omega) h_l^{(2)}\left(\tfrac{\omega}{c}r\right) Y_l^m(\boldsymbol{\theta}), \tag{2.1}$$

where $h_l^{(2)}(\frac{\omega}{c}r)$ denotes the spherical Hankel functions of the second kind [1], ω the angular frequency, c the speed of sound, r the distance from the source centre and $\boldsymbol{\theta}$ the normed cartesian vector

$$\boldsymbol{\theta} = \begin{pmatrix} \cos(\varphi)\sin(\vartheta) \\ \sin(\varphi)\sin(\vartheta) \\ \cos(\vartheta) \end{pmatrix}, \tag{2.2}$$

and $Y_l^m(\boldsymbol{\theta})$ the real-valued spherical harmonic functions

$$Y_l^m(\varphi,\vartheta) = \sqrt{\frac{2l+1}{2\pi(1+\delta_m)}\frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}(\cos\vartheta) \begin{cases} \cos(m\varphi), & m \geq 0 \\ \sin(m\varphi), & m < 0 \end{cases}, \tag{2.3}$$

where $P_l^m(\cos\vartheta)$ are the associated Legendre functions.

As implied by (2.1), the emitted sound field can be fully described by the frequency dependent weighting of each spherical harmonic, the spherical wave spectrum denoted $\chi_l^m(\frac{\omega}{c}r) = a_l^m(\omega) \cdot h_l^{(2)}(\frac{\omega}{c}r)$. In the practice, this so-called spherical wave spectrum is approximated based on discrete observations $\mathbf{x}$ on a surrounding sphere with a radius r, this leads to a reduced set of spherical harmonics, namely up to an order $L < \infty$. In vector notation and using n as the discrete time variable, the spherical harmonics signals are obtained by

$$\boldsymbol{\chi}[n] = \mathbf{Y}^\dagger \mathbf{x}[n] \tag{2.4}$$

where

$$\mathbf{Y} = \begin{pmatrix} Y_0^0(\boldsymbol{\theta}_1) & Y_1^{-1}(\boldsymbol{\theta}_1) & Y_1^0(\boldsymbol{\theta}_1) & \cdots & Y_L^L(\boldsymbol{\theta}_1) \\ Y_0^0(\boldsymbol{\theta}_2) & Y_1^{-1}(\boldsymbol{\theta}_2) & Y_1^0(\boldsymbol{\theta}_2) & \cdots & Y_L^L(\boldsymbol{\theta}_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\boldsymbol{\theta}_\Lambda) & Y_1^{-1}(\boldsymbol{\theta}_\Lambda) & Y_1^0(\boldsymbol{\theta}_\Lambda) & \cdots & Y_L^L(\boldsymbol{\theta}_\Lambda) \end{pmatrix} \tag{2.5}$$

and $(\cdot)^\dagger$ is the (Moore-Penrose) pseudo inverse.

From there, the interpolated signal of a virtual microphone $\hat{x}$ at a desired position $\boldsymbol{\theta}$ can be determined by combining the signal of each spherical harmonics channel $\boldsymbol{\chi}$ from (2.4) with the following appropriate weighting [NNZ10]

$$\hat{x}(n,\boldsymbol{\theta}) = \begin{pmatrix} Y_0^0(\boldsymbol{\theta}) & Y_1^{-1}(\boldsymbol{\theta}) & Y_1^0(\boldsymbol{\theta}) & \cdots & Y_L^L(\boldsymbol{\theta}) \end{pmatrix} \boldsymbol{\chi}[n]. \tag{2.6}$$

1. The choice of the Hankel function of the first or second kind is a matter of convention. In [Wil99], Williams uses the spherical Hankel functions of the first kind due to the choice to use the conventional definition for the space variables. The choice here is considered to be more conventional regarding the temporal Fourier transform's definition in signal processing.

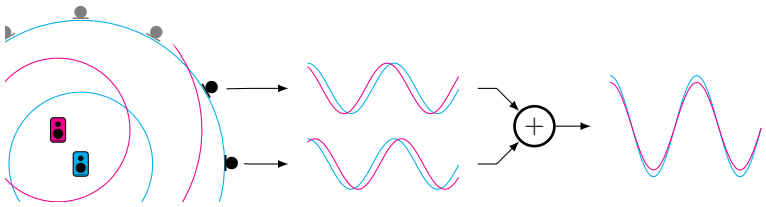
2.1.1 Some considerations on the radiation complexity, source size and positioning

Unfortunately, the direct spherical harmonics expansion of spatially discrete measurements may not always deliver a satisfying representation of the radiated field for the following reasons:

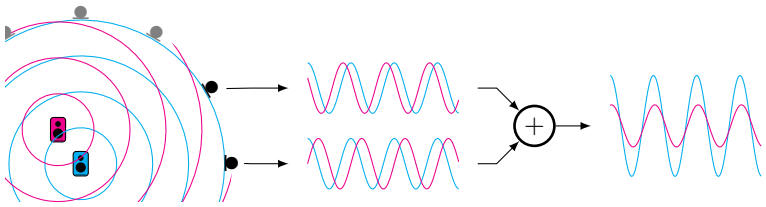
— The sampling of the surrounding sphere does not always permit a high enough spatial resolution to fully represent the directivity of the instrument, especially at high frequencies.

— The poor centring of the instrument leads to a shift of the energy towards higher orders, as described in [Bau11]. From a more intuitive interpretation, the translation of the source induces different times of arrival and therefore different phases between the microphones. As depicted in Fig. 2.1, those induced phase differences can have a huge influence on linear interpolation at high frequencies; Here, an equal weight linear interpolation of two neighbouring microphones leads to a non-negligible signal cancellation at high frequencies.

— Most of the time, the size of an instrument is, regarding the half wave length of the radiated signal, not negligible. For this reason, each frequency exhibits its own different radiation centroid, which would complicate an eventual centring of the source.



(a) Interferences at low frequencies due to small time delay



(b) Interferences at high frequencies due to small time delay

Figure 2.1 – Destructive interferences between 2 microphones through bad source centering

Therefore, a direct hyperinterpolation of the emitted sound field does not appear to be convenient for an interference free directional signal interpolation, in any case.

2.2 Vector Base Amplitude Panning (VBAP)

Similarly to (2.6), where the virtual signal $\hat{x}(\boldsymbol{\theta})$ is obtained by a linear combination of all microphones, an efficient interpolation can also be achieved from a linear interpolation of a reduced set of microphones in the vicinity of the desired position. Initially, Ville Pullki proposed this method in [Pul97] for synthesising a virtual incident sound field from a given direction by distributing a signal to loudspeakers in its vicinity, however it appears to be also convenient for the inverse problem, namely to interpolate between microphone signals in the vicinity of a given direction.

The microphone array layout is decomposed into a convex hull of triangular facets, and for each given direction, a subset of 3 microphones is used for the gain panning on the vertices of the active facet, namely the one containing the point of interest. For each facet ς we define the matrix $\mathbf{L}_\varsigma$, containing the cartesian coordinates of the active three microphones

$$\mathbf{L}_\varsigma = \begin{pmatrix} \boldsymbol{\theta}_{\varsigma,1} & \boldsymbol{\theta}_{\varsigma,2} & \boldsymbol{\theta}_{\varsigma,3} \end{pmatrix}. \tag{2.7}$$

The gains associated to each of the three microphones can be computed as follow

$$\tilde{\mathbf{g}}_\varsigma = \mathbf{L}_\varsigma^{-1}\boldsymbol{\theta} \tag{2.8}$$

and normed in order to guarantee a position independent gain

$$\mathbf{g}_\varsigma = \frac{\tilde{\mathbf{g}}_\varsigma}{\|\tilde{\mathbf{g}}_\varsigma\|}. \tag{2.9}$$

The active facet $\hat{\varsigma}$ is the one whose microphones coefficients all appear non-negative.

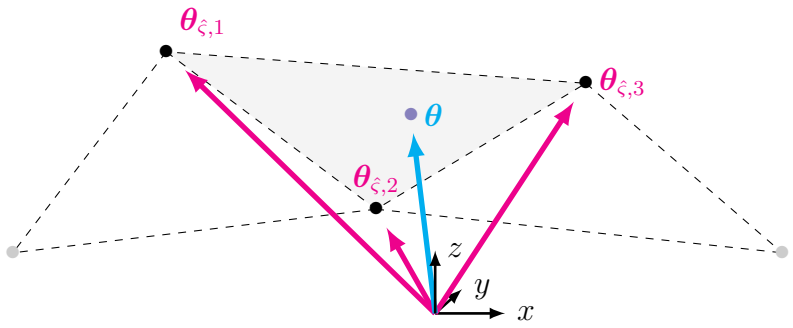


Figure 2.2 – Sample configuration for a virtual microphone at $\boldsymbol{\theta}$

However, as for the hyperinterpolation, the phase difference within the selected subset of microphones may also lead to destructive interferences and, therefore, entail a low pass or comb filter characteristic in the resulting interpolated signal. To overcome this problem, a simple phase selection strategy can be adopted as described in Section 2.3.

2.3 Modified Vector Base Amplitude Panning (MVBAP)

In 2006, Hom *et al.* proposed a strategy for interpolating the binaural signals recorded for different directions (Motion Tracked Binaural, or in short MTB) without energy loss at high frequency [HAD06]. They propose different implementations that are actually similar in the way that the linear interpolation at high frequency is restricted to work on the short-term magnitude spectrum. The phase could either be reconstructed with higher effort using spectrogram inversion as described in Section 4.1, or for a faster and more efficient solution, the phase could be switched to the one observed at the closest measurement angle.

Similarly to MTB, this cost-efficient avoidance of destructive interference is applicable to the VBAP interpolation described in Section 2.2 for the interpolation of surrounding spherical microphone signals. For simplicity, the second option was selected and referred as Modified Vector Base Amplitude Panning (MVBAP), here. The microphone determining the phase of the interpolated signal at high frequencies is the one having the greatest gain (equivalent to the closest microphone). The interpolation procedure with MVBAP is depicted in Fig. 2.3.

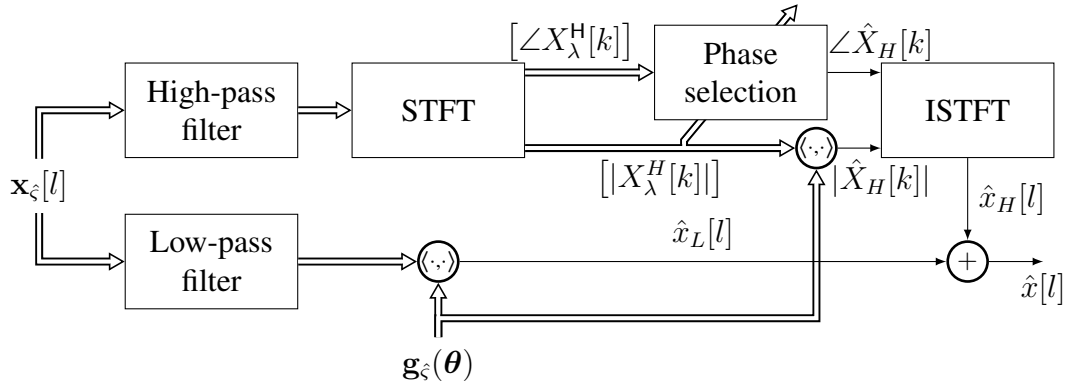


Figure 2.3 – Interpolation procedure from the signals of selected microphones subset with MVBAP.

The phase switching can also be improved in order to avoid recurrent discontinuity when the desired position moves in the vicinity of a transition border by introducing a small hysteresis, this is especially convenient when the desired position is noisy (e.g. sensor based tracking) [ZFZ19].

Chapter 3

Spherical phase retrieval

As seen in Section 2.1, the expansion of the radiated sound field in spherical harmonics is a powerful tool that enables the virtualisation of a measured or synthetic sound source without the need of any information about the decoding paradigm. However, we demonstrated that a poor centring of the instrument, or a large instruments which induces scattering or frequency dependent eccentric centroids may lead to destructive interferences at high frequencies.

For these reason, it may be beneficial to approximate the source radiation with a spherical harmonics expansion of the directivity with a modified phase, which would avoid the low pass or comb-filter behaviour of the signal in the low order channels.

In this sections, the microphone signals are stacked in a complex vector $\mathbf{x} = \mathrm{diag}\{\mathbf{p}\}\mathbf{z}$, where $\mathbf{p} = [|x_\lambda[k]|]$ and $\mathbf{z} = \left[e^{\mathrm{i}\angle x_\lambda}\right]$. The following proposed techniques consist of determining the optimal spatial allpass vector $\mathbf{z}$. The reconstructed directivity can then be expressed in term of spherical harmonics as in (2.4),

$$\boldsymbol{\chi} = \mathbf{Y}^\dagger \,\mathrm{diag}\{\mathbf{p}\}\mathbf{z}. \tag{3.1}$$

Directivity reconstruction based on simulated coefficients To depict the reconstructed directivity based on different methods described in this section, the microphones coefficients are evaluated from a known ideal directivity pattern at some given positions on the surrounding sphere. Hereby, the geometry is based on the spherical 64 microphones array of the IEM [Hoh09] which is shown in Fig. 3.1, the exact coordinates are given in Appendix A.

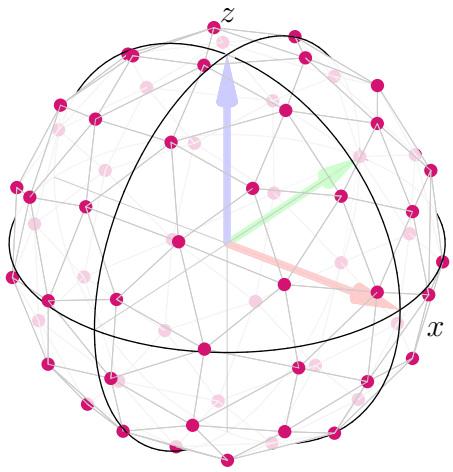


Figure 3.1 – Geometry of the IEM 64 microphones array

3.1 Zero-phase Approximation

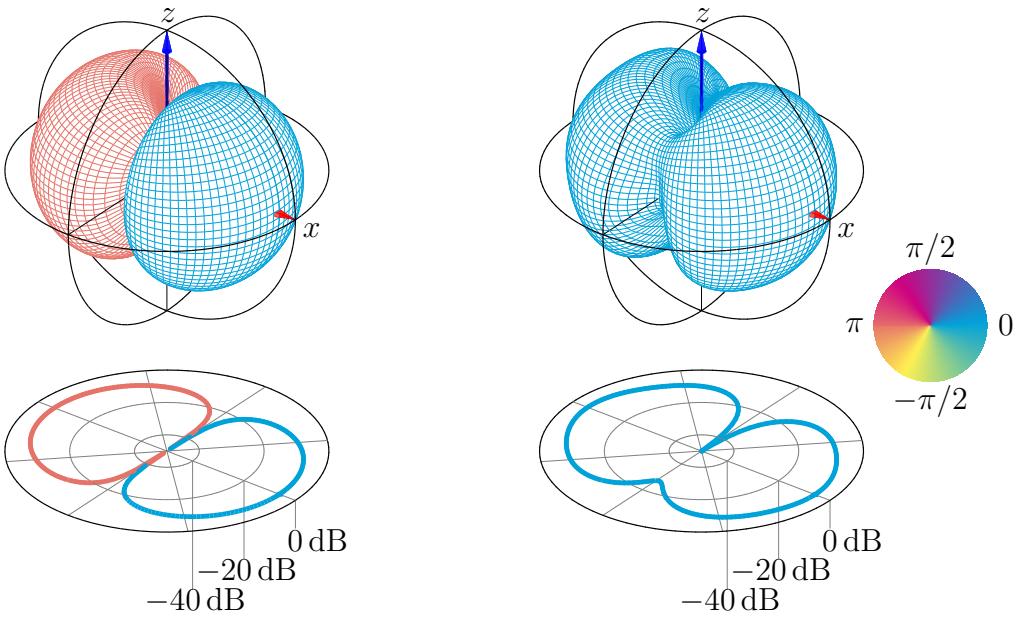
The simplest way to preserve the magnitude of a function while simplifying the phase is to set its phase to zero. The approximated spherical harmonics coefficients are therefore

$$\boldsymbol{\chi} = \mathbf{Y}^{\dagger}\mathbf{p}. \tag{3.2}$$

Alternatively, a weighting can be applied onto the SH coefficients in order to reduce the side lobes,

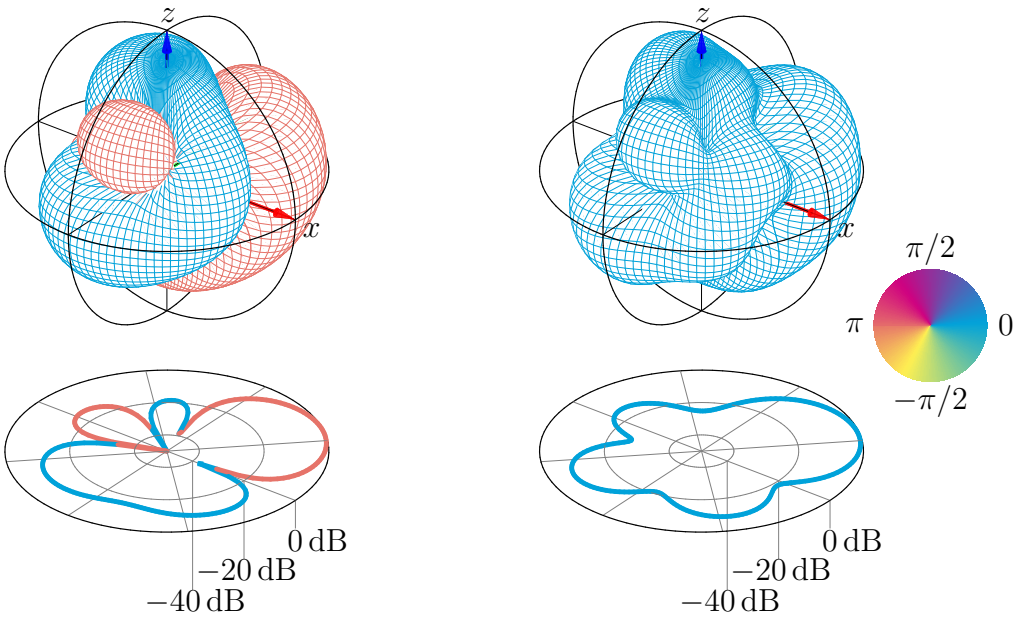
$$\boldsymbol{\chi} = \mathrm{diag}\{\mathring{\mathbf{w}}\}\mathbf{Y}^{\dagger}\mathbf{p}. \tag{3.3}$$

Figures 3.2 and 3.3 shows the reconstructed directivity of a perfect dipole $Y_1^1(\boldsymbol{\theta})$ when applying (3.2) onto the coefficients $\mathbf{x}$ obtained by the microphone array depicted in Fig. 3.1. As expected, it can be observed in Figs. 3.2 and 3.3 that the zeros can not be reconstructed properly, unless a microphone perfectly samples it, but still, the magnitude valleys turn out shallower around zero. However, some negative polarity can still occur in the vicinity of microphones with small gains due to ripple, that will then induce zeros which do not necessarily match the desired one.



(a) original (b) reconstruction

Figure 3.2 – Reconstruction of the dipole directivity $Y_1^1(\boldsymbol{\theta})$ from 64 magnitude points with the zero phase method.



(a) original (b) reconstruction

Figure 3.3 – Reconstruction of a random directivity function ($L = 2$) from 64 magnitude points with the zero phase method.

3.2 Phase reconstruction based on Magnitude Least Squares (MLS)

In his PhD dissertation, Kassakian proposed to apply convex optimisation to the design of radiation filter [Kas06]. The high frequency magnitude preservation can be achieved by modifying the phase of each microphone in a way that minimises the mean square error between the all-pass filtered zero phase signal and the spherical harmonics signal transformed back to the space domain. Thus, the phase recovery problem can be described as

$$\begin{aligned} \underset{\mathbf{z}\in\mathbb{C}^{\Lambda}}{\text{minimise}} \quad & \| \operatorname{diag}\{\mathbf{p}\}\mathbf{z} - \mathbf{Y}\boldsymbol{\chi}\|^2 \\ \text{subject to} \quad & \operatorname{diag}\{\mathbf{z}^*\}\mathbf{z} = \mathbf{1}. \end{aligned} \tag{3.4}$$

where Λ is the number of microphones.

In (3.4), $\boldsymbol{\chi}$ can be approximated by the minimum-mean-square-error solution for known $\operatorname{diag}\{\mathbf{p}\}\mathbf{z}$

$$\boldsymbol{\chi} = \mathbf{Y}^{\dagger} \operatorname{diag}\{\mathbf{p}\}\mathbf{z}, \tag{3.5}$$

enabling to rewrite the minimisation problem as follows

$$\begin{aligned} \underset{\mathbf{z}\in\mathbb{C}^{\Lambda}}{\text{minimise}} \quad & \left\|\operatorname{diag}\{\mathbf{p}\}\mathbf{z} - \mathbf{Y}\mathbf{Y}^{\dagger} \operatorname{diag}\{\mathbf{p}\}\mathbf{z}\right\|^2 \\ \text{subject to} \quad & \operatorname{diag}\{\mathbf{z}^*\}\mathbf{z} = \mathbf{1}. \end{aligned} \tag{3.6}$$

Alternatively, a weighting $\operatorname{diag}\{\mathring{\mathbf{w}}\}$ can be applied on the spherical harmonics coefficients in order to ensure a smooth transition of the reconstructed phase in the space domain. The modified cost function can then be written to a more compact form

$$J = \|(\mathbf{I} - \mathbf{Y} \operatorname{diag}\{\mathring{\mathbf{w}}\}\mathbf{Y}^{\dagger}) \operatorname{diag}\{\mathbf{p}\}\mathbf{z}\|^2 \tag{3.7}$$

$$= \mathbf{z}^{\mathsf{H}} \operatorname{diag}\{\mathbf{p}\}(\mathbf{I} - \mathbf{Y} \operatorname{diag}\{\mathring{\mathbf{w}}\}\mathbf{Y}^{\dagger})^{\mathsf{H}}(\mathbf{I} - \mathbf{Y} \operatorname{diag}\{\mathring{\mathbf{w}}\}\mathbf{Y}^{\dagger}) \operatorname{diag}\{\mathbf{p}\}\mathbf{z} \tag{3.8}$$

$$= \mathbf{z}^{\mathsf{H}}\mathbf{B}^{\mathsf{H}}\mathbf{B}\mathbf{z} \tag{3.9}$$

$$J = \mathbf{z}^{\mathsf{H}}\mathbf{C}\mathbf{z} \tag{3.10}$$

where

$$\mathbf{C} = \mathbf{B}^{\mathsf{H}}\mathbf{B}, \tag{3.11}$$

and

$$\mathbf{B} = (\mathbf{I} - \mathbf{Y} \operatorname{diag}\{\mathring{\mathbf{w}}\}\mathbf{Y}^{\dagger}) \operatorname{diag}\{\mathbf{p}\}. \tag{3.12}$$

The minimisation problem writes

$$\begin{aligned} \underset{\mathbf{z}\in\mathbb{C}^{\Lambda}}{\text{minimise}} \quad & \mathbf{z}^{\mathsf{H}}\mathbf{C}\mathbf{z} \\ \text{subject to} \quad & \operatorname{diag}\{\mathbf{z}^*\}\mathbf{z} = \mathbf{1}, \end{aligned} \tag{3.13}$$

3.2.1 Solving MLS with gradient descent

The minimisation problem (3.13) can be solved by mean of the gradient descent algorithm with an update step $0 \leq \mu \leq 0.5$. Hereby, the gradient of the MLS cost function could be computed w.r.t. $\mathbf{z}$

$$\nabla_{\mathbf{z}} J = 2\mathbf{C}\mathbf{z} \tag{3.14}$$

or directly w.r.t. the angle ϕ after observing that

$$\frac{\mathrm{d}}{\mathrm{d}\phi} z = \frac{\mathrm{d}}{\mathrm{d}\phi}(a + \mathrm{i}b) = \frac{\mathrm{d}a}{\mathrm{d}\phi} + \mathrm{i}\frac{\mathrm{d}b}{\mathrm{d}\phi} = \frac{\mathrm{d}\cos\phi}{\mathrm{d}\phi} + \mathrm{i}\frac{\mathrm{d}\sin\phi}{\mathrm{d}\phi} = -\sin\phi + \mathrm{i}\cos\phi = \mathrm{i}z, \tag{3.15}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\phi} z^* = \cdots = -\mathrm{i}z^*. \tag{3.16}$$

The gradient of the cost function w.r.t. $\boldsymbol{\phi}$ can easily be decomposed to

$$\frac{\partial}{\partial \phi_i} \sum_{k,l} z_k^* c_{kl} z_l = -\mathrm{i}z_i^* \sum_l c_{il} z_l + \mathrm{i}z_i \sum_l c_{li} z_l^*, \tag{3.17}$$

which leads, in the vector notation, to

$$\frac{\partial}{\partial \boldsymbol{\phi}} \mathbf{z}^{\mathsf{H}} \mathbf{C}\mathbf{z} = -\mathrm{i}\,\mathrm{diag}\{\mathbf{z}^*\}\mathbf{C}\mathbf{z} + \mathrm{i}\,\mathrm{diag}\{\mathbf{z}\}\mathbf{C}^{\mathsf{T}}\mathbf{z}^* \tag{3.18}$$

$$= -\mathrm{i}\,\mathrm{diag}\{\mathbf{z}^*\}\mathbf{C}\mathbf{z} + (-\mathrm{i}\,\mathrm{diag}\{\mathbf{z}^*\}\mathbf{C}^{\mathsf{H}}\mathbf{z})^*. \tag{3.19}$$

With a hermitian matrix $\mathbf{C} = \mathbf{C}^{\mathsf{H}}$, this simplifies to

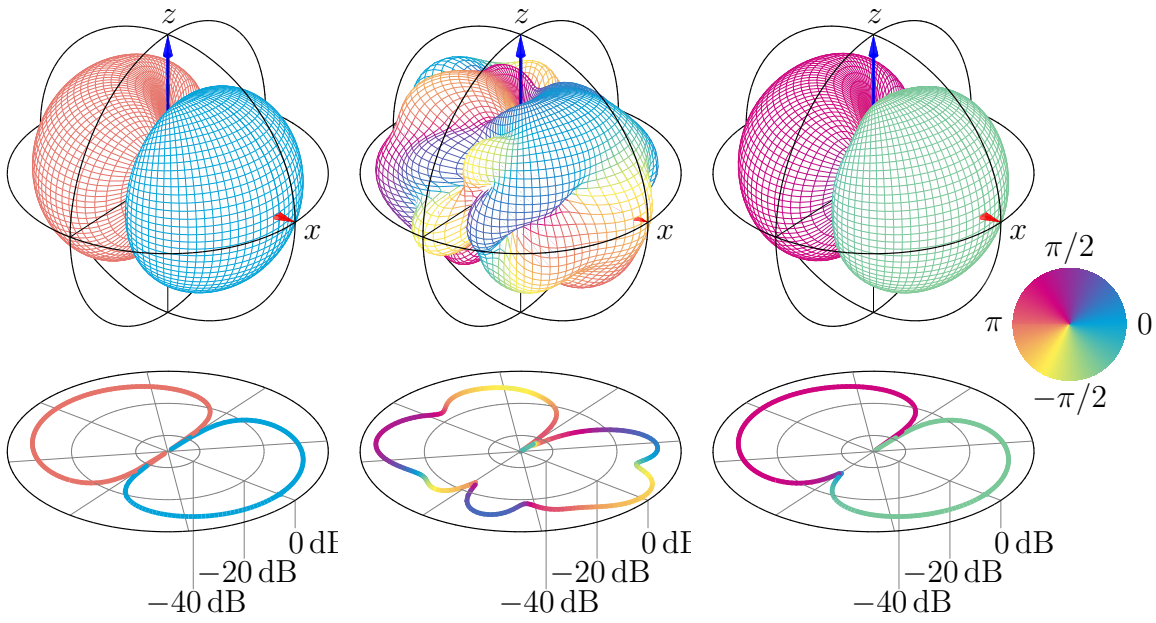
$$\frac{\partial}{\partial \boldsymbol{\phi}} \mathbf{z}^{\mathsf{H}} \mathbf{C}\mathbf{z} = 2\Re\{-\mathrm{i}\,\mathrm{diag}\{\mathbf{z}^*\}\mathbf{C}\mathbf{z}\} = 2\Im\{\mathrm{diag}\{\mathbf{z}^*\}\mathbf{C}\mathbf{z}\}. \tag{3.20}$$

The update rule can be decomposed as following:

1. Initialisation of the phase $\boldsymbol{\phi}^{(0)}$
2. For each iteration step $\iota \in \{1, \ldots, I\}$, the current estimation is moved toward the negative gradient
$$\boldsymbol{\phi}^{(\iota)} = \boldsymbol{\phi}^{(\iota-1)} - 2\mu\Im\{\mathrm{diag}\{\mathbf{z}^*\}\mathbf{C}\mathbf{z}\} \tag{3.21}$$
3. Compute $\mathbf{z}$
$$\mathbf{z} = \left[e^{i\boldsymbol{\phi}^{(I)}}\right]. \tag{3.22}$$

Then, the directivity can be expressed in term of SH as in (3.1).

The reconstructed directivity of the dipole $Y_1^1(\boldsymbol{\theta})$ and a random directivity of maximum order of $L = 2$ are depicted in Figs. 3.4 and 3.5 for different numbers of iterations I with a random initial phase. The maximum order of spherical harmonics used for the Spherical Harmonics Transformation (SHT), the masking of its coefficients and the Inverse Spherical Harmonics transformation (ISHT) implied in $\mathbf{C}$ (see (3.11)) are set to maximum controlable order $\lfloor \sqrt{\Gamma} - 1 \rfloor = 7$. The mask $\mathrm{diag}\{\mathring{\mathbf{w}}\}$ consists of a $\max$-r_E weighting approximated later in (5.3).

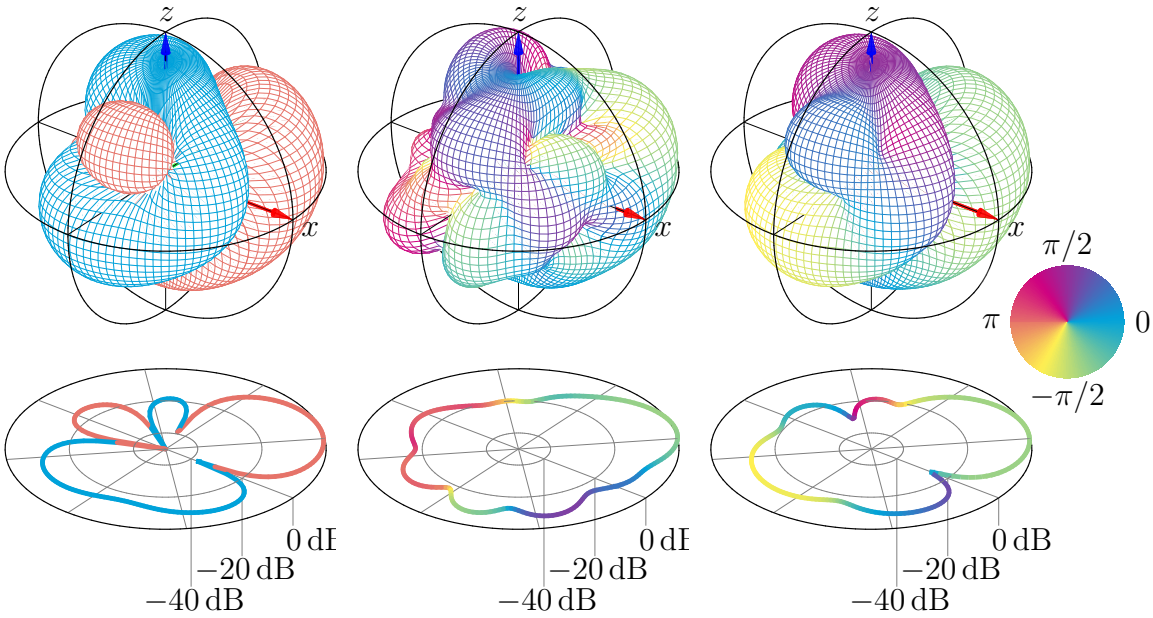


(a) original

(b) reconstruction, $I = 1\mathrm{e}2$

(c) reconstruction, $I = 1\mathrm{e}5$

Figure 3.4 – Reconstruction of the dipole directivity $Y_1^1(\boldsymbol{\theta})$ from 64 magnitude points with gradient descent algorithm for different numbers of iterations I. The initial phase is random.



(a) original

(b) reconstruction, $I = 1e2$

(c) reconstruction, $I = 1e5$

Figure 3.5 – Reconstruction of a random directivity function ($L = 2$) from 64 magnitude points with gradient descent algorithm for different numbers of iterations I. The initial phase is random.

3.2.2 Solving MLS with Semi-Definite Relaxation (SDR)

In contrast to the local optimisation, achieved with e.g. the gradient descent in Section 3.2.1, the Newton or Gauss-Newton method, and that may converge to local minima, semi-definite relaxation (SDR) provides an initialisation to directly converge to global minimum [Kas06, LMS+10, BBE17]. After re-writing $\mathbf{z}^{\mathsf{H}}\mathbf{C}\mathbf{z} = \mathrm{trace}\{\mathbf{z}^{\mathsf{H}}\mathbf{C}\mathbf{z}\} = \mathrm{trace}\{\mathbf{C}\mathbf{z}\mathbf{z}^{\mathsf{H}}\}$ and observing that the matrix $\mathbf{Z} = \mathbf{z}\mathbf{z}^{\mathsf{H}}$ is hermitian positive semidefinite of rank one, the cost function described in (3.10) can be expressed linearly w.r.t $\mathbf{Z}$.

$$\begin{aligned} \underset{\mathbf{Z}\in\mathcal{H}^{\succcurlyeq}}{\text{minimise}} \quad & \mathrm{trace}\{\mathbf{C}\mathbf{Z}\} \\ \text{subject to} \quad & \mathrm{diag}\{\mathbf{Z}\} = \mathbf{1},\ \mathrm{rank}\{\mathbf{Z}\} = 1, \end{aligned} \tag{3.23}$$

with $\mathcal{H}^{\succcurlyeq}$ being the set of hermitian positive semidefinite matrices

$$\mathcal{H}^{\succcurlyeq} = \left\{\mathbf{Z} \in \mathbb{C}^{\Lambda\times\Lambda} | \mathbf{Z}^{\mathsf{H}} = \mathbf{Z}, \mathbf{v}^{\mathsf{H}}\mathbf{Z}\mathbf{v} \geq 0\right\}. \tag{3.24}$$

At first sight, the minimisation problem (3.23) does not differ from (3.13) in terms of

its degrees of freedom, however Eq. (3.23) can also be relaxed by omitting the rank constraint $\mathrm{rank}\{\mathbf{Z}\} = 1$ [WDM15, Kas06, LMS+10], which increases the degrees of freedom from Λ to Λ^2. This leads to the minimisation problem

$$\begin{aligned} \underset{\mathbf{Z}\in\mathcal{H}^{\succcurlyeq}}{\text{minimise}} \quad & \mathrm{trace}\{\mathbf{CZ}\} \\ \text{subject to} \quad & \mathrm{diag}\{\mathbf{Z}\} = \mathbf{1}, \end{aligned} \tag{3.25}$$

which is known as a semi-definite relaxation problem of (3.13) (SDR).

Practically, some freely available programs e.g. the `CVX` MATLAB toolbox [GB08, GB14] can be used for semi-definite programming (SDP) minimisation problems.

According to the (3.11) and (3.12), $\mathbf{C}$ is real-valued, thus the optimal solution $\mathbf{Z}$ to (3.25) is also real-valued. The problem could therefore be simplified by seeking the optimal solution within the set of the symmetric real-valued semi-definite matrices

$$\mathcal{S}^{\succcurlyeq} = \{\mathbf{Z} \in \mathbb{R}^{\Lambda\times\Lambda} | \mathbf{Z}^{\mathsf{T}} = \mathbf{Z}, \mathbf{v}^{\mathsf{T}}\mathbf{Z}\mathbf{v} \geq 0\}. \tag{3.26}$$

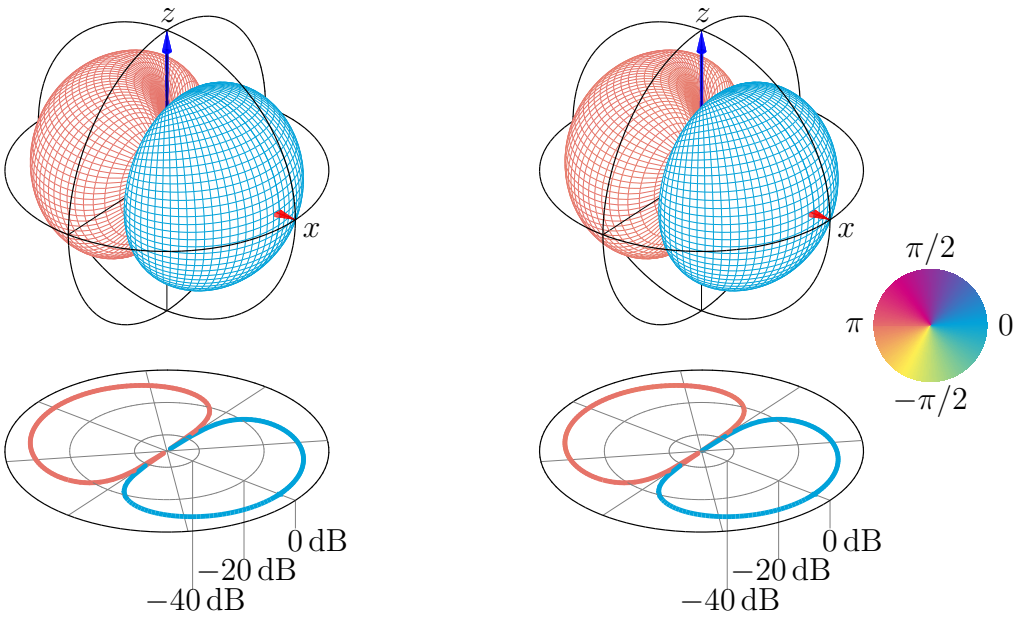
Unfortunately, due to the relaxation, the optimal matrix $\mathbf{Z}$ is not necessarily of rank 1, in particular if the global minimum can not be achieved by a real-valued matrix $\mathbf{Z}$. Therefore, two strategies can be adopted in order to reconstruct the phase:

Main eigenvector Define $\mathbf{z}$ as the main eigenvector of $\mathbf{Z}$ (associated to the greatest eigenvalue). In this case, the solution assign to each microphone a phase of either 0 or π.

Random sampling If the global minimum cannot be reached by a real valued vector $\mathbf{z}$. One may compute the error obtained by different random complex linear combinations of the eigenvectors, wherein the variance of their random coefficients are driven by their respective eigenvalue. $\mathbf{z}$ is set to the normed random vector offering the minimum error when inserted into the cost function J in (3.10). The complex linear combination of the eigenvalues yields a potential complex $\mathbf{z}$ [Kas06].

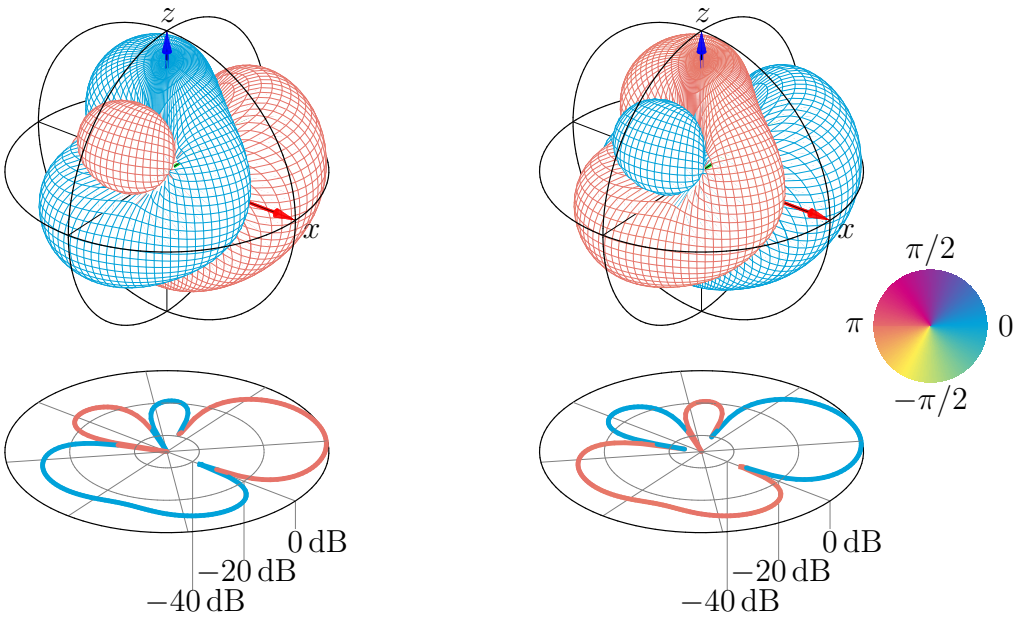
Again the directivity in term of SH is computed using (3.1).

The reconstructed dipole and random directivity using the main eigenvector of $\mathbf{Z}$ as spatial allpass vector $\mathbf{z}$ and without further local optimisations are depicted in Figs. 3.6 and 3.7. Again the maximum order of SH used for the SHT, masking and ISHT is set to $L = 7$.



(a) original (b) reconstruction

Figure 3.6 – Reconstruction of the dipole directivity $Y_1^1(\boldsymbol{\theta})$ from 64 magnitude points with the SDR method using the main eigenvector and without further local optimisation.



(a) original (b) reconstruction

Figure 3.7 – Reconstruction of a random directivity function ($L = 2$) from 64 magnitude points with the SDR method using the main eigenvector and without further local optimisation.

3.3 Phase reconstruction based on Magnitude Squares Least Square (MSLS)

Similarly to the MLS, one can write a minimisation problem that directly search for the optimal SH coefficients without the need of any constraints. For a desired decomposition, the spherical harmonics y_{ij} are real-valued $y_{ij} \in \mathbb{R}$, with the point of observation indexed by i and the harmonic by j. We have the following decomposion problem when given the squared-magnitude measurement $|\chi_i|^2$

$$\underset{\boldsymbol{\chi}\in\mathbb{C}^{\Upsilon}}{\text{minimise}} \quad \left\| |\mathbf{Y}\boldsymbol{\chi}|^{\circ 2} - \mathbf{p}^{\circ 2} \right\|^2. \tag{3.27}$$

where $\Upsilon = (L+1)^2$ is the number of SH for a given maximal order L.

3.3.1 Solving MSLS with Newton descent

The cost function of the minimisation problem (3.27) can be expressed by means of sum notation for better insight

$$J = \sum_{i=1}^{\Lambda} e_i^2 = \sum_{i=1}^{\Lambda} \left[\left|\hat{p}_i\right|^2 - |p_i|^2 \right]^2, \tag{3.28}$$

where

$$\hat{p}_i = \sum_{j=1}^{\Upsilon} y_{ij}\chi_j = \sum_{j=1}^{\Upsilon} y_{ij}(a_j + \mathrm{i}b_j), \tag{3.29}$$

and

$$|\hat{p}_i|^2 = \Big| \sum_{j,j'=1}^{\Upsilon} y_{ij}\chi_j \Big|^2 = \sum_{j,j'=1}^{\Upsilon} y_{ij}\chi_j y_{ij'}^* \chi_{j'}^*. \tag{3.30}$$

The expression $y_{ij}y_{ij'}^*$ is hermitian symmetric, as an index exchange $j \leftrightarrow j'$ yields the conjugate expression $(y_{ij}y_{ij'}^*) = (y_{ij'}y_{ij}^*)^*$. We may use this to get

$$|\hat{p}_i|^2 = \sum_{j,j'=1}^{\Upsilon} \left[\tfrac{1}{2}(y_{ij}y_{ij'}^*) + \tfrac{1}{2}(y_{ij'}y_{ij}^*)^* \right] (\chi_j\chi_{j'}^*) \tag{3.31}$$

$$= \sum_{j,j'=1}^{\Upsilon} \Big[\tfrac{1}{2}(y_{ij}y_{ij'}^*)(\chi_j\chi_{j'}^*) + \underbrace{\tfrac{1}{2}(y_{ij'}^*y_{ij})(\chi_j\chi_{j'}^*)}_{j\leftrightarrow j'} \Big] \tag{3.32}$$

$$= \sum_{j,j'=1}^{\Upsilon} \left[\tfrac{1}{2}(y_{ij}y_{ij'}^*)(\chi_j\chi_{j'}^*) + \tfrac{1}{2}(y_{ij}y_{ij'}^*)(\chi_j\chi_{j'}^*)^*\right] \tag{3.33}$$

$$= \sum_{j,j'=1}^{\Upsilon} y_{ij}y_{ij'}^* \Re\{\chi_j\chi_{j'}^*\} \tag{3.34}$$

$$|\hat{p}_i|^2 = \sum_{j,j'=1}^{\Upsilon} y_{ij}y_{ij'}^*(a_j a_{j'} - b_j b_{j'}) \tag{3.35}$$

with the real and imaginary parts a_j and b_j.

Gradient y_{ij} is real-valued, so that $y_{ij} = y_{ij}^*$. With the above, the complex cost function J can be derived w.r.t. a_k and b_k.

$$\frac{\partial J(\boldsymbol{\chi})}{\partial a_k} = \sum_{i=1}^{\Lambda} \frac{\partial e_i^2}{\partial a_k} = \sum_{i=1}^{\Lambda} 2e_i \frac{\partial |\hat{p}_i|^2}{\partial a_k} \tag{3.36}$$

$$= \sum_{i=1}^{\Lambda} 2e_i \sum_{j,j'=1}^{\Upsilon} y_{ij'}y_{ij}(a_{j'}\delta_{jk} + a_j\delta_{j'k}) \tag{3.37}$$

$$= \sum_{i=1}^{\Lambda} 2e_i \Big[\sum_{j'=1}^{\Upsilon} y_{ij'}y_{ik}a_{j'} + \sum_{j=1}^{\Upsilon} y_{ik}y_{ij}a_j\Big] \tag{3.38}$$

$$= \sum_{i=1}^{\Lambda} 4e_i \sum_{j=1}^{\Upsilon} y_{ik}y_{ij}a_j \tag{3.39}$$

$$\frac{\partial J(\boldsymbol{\chi})}{\partial a_k} = 4\sum_{i=1}^{\Lambda} y_{ik}e_i \Re\{\hat{p}_i\}, \tag{3.40}$$

and

$$\frac{\partial J(\boldsymbol{\chi})}{\partial b_k} = \sum_{i=1}^{\Lambda} \frac{\partial e_i^2}{\partial b_k} = \sum_{i=1}^{\Lambda} 2e_i \frac{\partial |\hat{p}_i|^2}{\partial b_k} \tag{3.41}$$

$$= \sum_{i=1}^{\Lambda} 2e_i \sum_{j,j'=1}^{\Upsilon} y_{ij'}y_{ij}(-b_{j'}\delta_{jk} - b_j\delta_{j'k}) \tag{3.42}$$

$$\vdots \tag{3.43}$$

$$\frac{\partial J(\boldsymbol{\chi})}{\partial b_k} = 4\sum_{i=1}^{\Lambda} y_{ik}e_i \Im\{\hat{p}_i\}. \tag{3.44}$$

Let's write the gradient by concatenating the real gradient and the imaginary gradient as follows

$$\nabla J(\overline{\overline{\boldsymbol{\chi}}}) = \begin{pmatrix} \frac{\partial J(\boldsymbol{\chi})}{\partial a_1} & \cdots & \frac{\partial J(\boldsymbol{\chi})}{\partial a_\Upsilon} & \frac{\partial J(\boldsymbol{\chi})}{\partial b_1} & \cdots & \frac{\partial J(\boldsymbol{\chi})}{\partial b_\Upsilon} \end{pmatrix}^\mathsf{T}, \tag{3.45}$$

$$\nabla J(\overline{\overline{\boldsymbol{\chi}}}) = 4\mathbf{Y}^\mathsf{T} \operatorname{diag}\{\mathbf{e}\} \overline{\overline{\boldsymbol{\chi}}} \tag{3.46}$$

where

$$\overline{\overline{\boldsymbol{\chi}}} = \begin{pmatrix} \Re\{\boldsymbol{\chi}\}^\mathsf{T} & \Im\{\boldsymbol{\chi}\}^\mathsf{T} \end{pmatrix}^\mathsf{T} \tag{3.47}$$

denotes the real/complex stacked vector.

Hessian matrix Similarly to the gradient, the hessian matrix can be computed in order to locally approximate the cost function with a second order polynomial. Thus, the iteration step can be optimised to faster reach the local minimum.

$$\frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_k \partial a_l} = 4\frac{\partial}{\partial a_l} \sum_{i=1}^{\Lambda} \sum_{j=1}^{\Upsilon} e_i y_{ik} y_{ij}\, a_j \tag{3.48}$$

$$= 4\sum_{i=1}^{\Lambda} e_i y_{ik} y_{il} + 8\sum_{i=1}^{\Lambda} \sum_{j,j'=1}^{\Upsilon} y_{il} y_{ij'} y_{ik} y_{ij}\, a_j\, a_{j'} \tag{3.49}$$

$$= 4\sum_{i=1}^{\Lambda} e_i y_{ik} y_{il} + 8\sum_{i=1}^{\Lambda} y_{ik} \Re\{\hat{p}_i\}^2 y_{il} \tag{3.50}$$

$$\frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_k \partial a_l} = 4\sum_{i=1}^{\Lambda} y_{ik}(e_i + 2\Re\{\hat{p}_i\}^2) y_{il} \tag{3.51}$$

$$\frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_k \partial a_l} = 4\frac{\partial}{\partial a_l} \sum_{i=1}^{\Lambda} \sum_{j=1}^{\Upsilon} e_i y_{ik} y_{ij}\, b_j \tag{3.52}$$

$$= 8\sum_{i=1}^{\Lambda} \sum_{j,j'=1}^{\Upsilon} y_{il} y_{ij'} y_{ik} y_{ij}\, b_j\, a_{j'} \tag{3.53}$$

$$\frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_k \partial a_l} = 8\sum_{i=1}^{\Lambda} y_{ik}(\Re\{\hat{p}_i\}\Im\{\hat{p}_i\}) y_{il} \tag{3.54}$$

$$\frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_k \partial b_l} = -4\frac{\partial}{\partial b_l} \sum_{i=1}^{\Lambda} \sum_{j=1}^{\Upsilon} e_i y_{ik} y_{ij}\, b_j \tag{3.55}$$

$$= 4\sum_{i=1}^{\Lambda} e_i\, y_{ik}y_{il} + 8\sum_{i=1}^{\Lambda}\sum_{j,j'=1}^{\Upsilon} y_{il}y_{ij'}y_{ik}y_{ij}\, b_j\, b_{j'} \tag{3.56}$$

$$= 4\sum_{i=1}^{\Lambda} e_i y_{ik}y_{il} + 8\sum_{i=1}^{\Lambda} y_{ik}\Im\{\hat{p}_i\}^2 y_{il} \tag{3.57}$$

$$\frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_k \partial b_l} = 4\sum_{i=1}^{\Lambda} y_{ik}(e_i + 2\Im\{\hat{p}_i\}^2)y_{il} \tag{3.58}$$

Analogously to the stacked real/imaginary expression of the gradient in (3.45), the Hessian matrix can be expressed as follows

$$\mathbf{H}J(\overline{\overline{\boldsymbol{\chi}}}) = \begin{pmatrix} \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_1^2} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_1 \partial a_\Upsilon} & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_1 \partial b_1} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_1 \partial b_\Upsilon} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_\Upsilon \partial a_1} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_\Upsilon^2} & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_\Upsilon \partial b_1} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial a_\Upsilon \partial b_\Upsilon} \\ \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_1 \partial a_1} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_1 \partial a_\Upsilon} & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_1^2} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_1 \partial b_\Upsilon} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_\Upsilon \partial a_1} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_\Upsilon \partial a_\Upsilon} & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_\Upsilon \partial b_1} & \cdots & \frac{\partial^2 J(\boldsymbol{\chi})}{\partial b_\Upsilon^2} \end{pmatrix} \tag{3.59}$$

$$\mathbf{H}J(\overline{\overline{\boldsymbol{\chi}}}) = \begin{pmatrix} 4\mathbf{Y}(\operatorname{diag}\{\mathbf{e}\} + 2\operatorname{diag}\{\Re\{\hat{\mathbf{p}}\}\}^2)\mathbf{Y}^\mathsf{T} & 8\mathbf{Y}\operatorname{diag}\{\Re\{\hat{\mathbf{p}}\}\Im\{\hat{\mathbf{p}}\}\}\mathbf{Y}^\mathsf{T} \\ 8\mathbf{Y}\operatorname{diag}\{\Re\{\hat{\mathbf{p}}\}\Im\{\hat{\mathbf{p}}\}\}\mathbf{Y}^\mathsf{T} & 4\mathbf{Y}(\operatorname{diag}\{\mathbf{e}\} + 2\operatorname{diag}\{\Im\{\hat{\mathbf{p}}\}\}^2)\mathbf{Y}^\mathsf{T} \end{pmatrix} \tag{3.60}$$

From this, the Newton method can be applied in order to iteratively converge toward a stationary point [BV04, Kas06]. The update can be expressed in a real/imaginary stacked form as

$$\Delta\overline{\overline{\boldsymbol{\chi}}} = -\Big(\mathbf{H}J(\overline{\overline{\boldsymbol{\chi}}})\Big)^{-1}\nabla J(\overline{\overline{\boldsymbol{\chi}}}), \tag{3.61}$$

or in its complex form as

$$\Delta\boldsymbol{\chi} = \begin{bmatrix} \mathbf{I}_{\Upsilon\times\Upsilon} & \mathbf{0}_{\Upsilon\times\Upsilon} \end{bmatrix}\Delta\overline{\overline{\boldsymbol{\chi}}} + \mathrm{i}\begin{bmatrix} \mathbf{0}_{\Upsilon\times\Upsilon} & \mathbf{I}_{\Upsilon\times\Upsilon} \end{bmatrix}\Delta\overline{\overline{\boldsymbol{\chi}}}. \tag{3.62}$$

Thus the new vector obtained at each iteration is updated using

$$\boldsymbol{\chi}^{(\iota+1)} = \boldsymbol{\chi}^{(\iota)} + \mu\Delta\boldsymbol{\chi}^{(\iota)}, \tag{3.63}$$

where $0 < \mu < 1$ is an optional damping factor [1].

Some resulting reconstructions are depicted in Figs. 3.8 and 3.9 for a dipole directivity Y_1^1 and a random directivity of maximum order $L = 2$. Different numbers of

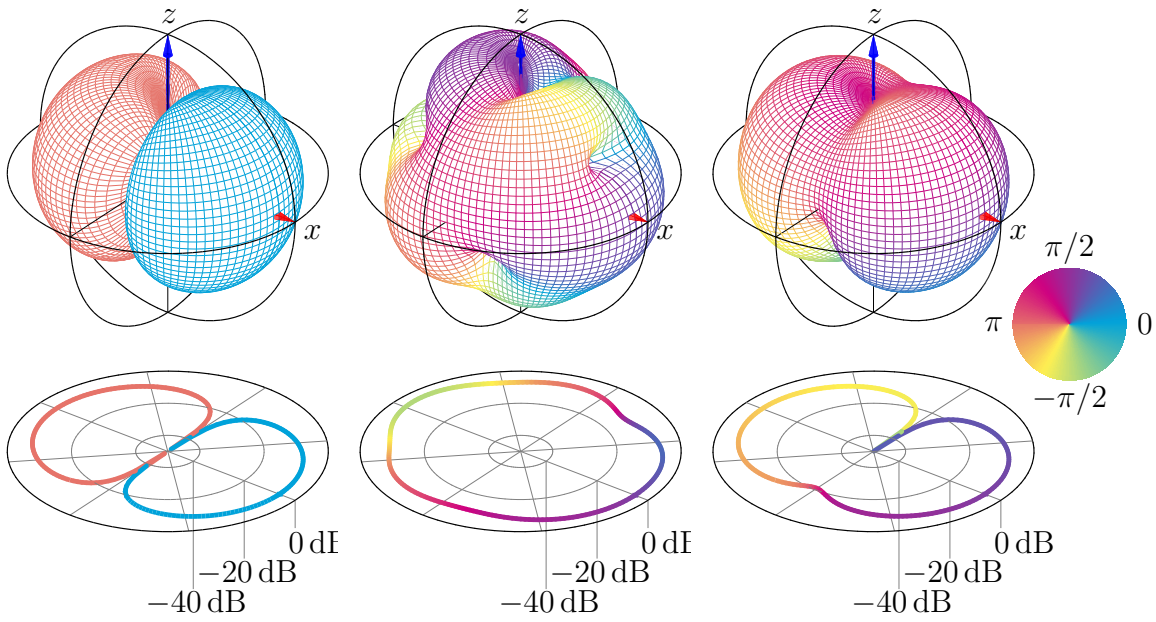
1. Alternatively, a simple line search can be implemented. This has the advantage to avoid converging toward a local maximum and also to increase the convergence speed.

iterations are depicted in order to give a good impression of the convergence speed starting from a random initialisation of the spherical wave spectrum $\boldsymbol{\chi}$. Hereby the minimisation is achieved with a damping factor of $\mu = 0.25$ and the inversion of the hessian matrix includes a regularisation of the form

$$\mathbf{H}_{\text{reg}}^{-1} = (\mathbf{H} + \varepsilon \|\mathbf{H}\|_2 \mathbf{I})^{-1}, \tag{3.64}$$

where $\|\cdot\|_2$ denotes the maximum singular value of the matrix. The matrix Y is set to a maximum order of $L = 3$ without any further SH weighting.

As it can be observed, the minimisation of MSLS does not necessarily converge toward the global minimum such as in Fig. 3.8c. However, after a rapid testing, the convergence speed appears way greater than the one obtained with gradient descent on MLS as in Fig. 3.4.

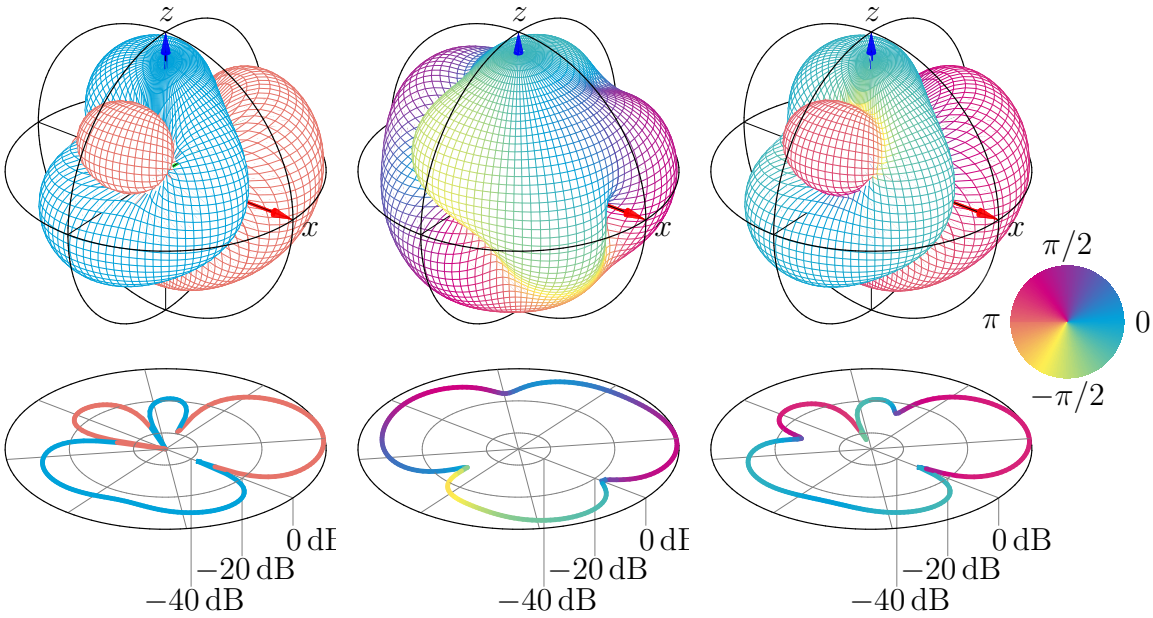


(a) original

(b) reconstruction, $I = 20$

(c) reconstruction, $I = 100$

Figure 3.8 – Reconstruction of the dipole directivity $Y_1^1(\boldsymbol{\theta})$ from 64 magnitude points by solving the MSLS problem with the Newton method.



(a) original (b) reconstruction, $I = 20$ (c) reconstruction, $I = 100$

Figure 3.9 – Reconstruction of a random directivity function ($L = 2$) from 64 magnitude points by solving the MSLS problem with the Newton method.

3.4 Dimensionality reduction by peak-regions grouping

Assuming that neighbouring microphones contained in the same lobe of the directivity function share the same phase, the dimensionality of the different minimisation problems presented above could be reduced by optimising the phase of $\Gamma < \Lambda$ regions, instead of the Λ single microphones.

As a example the matrix $\mathbf{C}$ in the MLS minimisation problem written in (3.13) would take the form

$$\mathbf{C} = \mathbf{B}^{\mathsf{H}}\mathbf{B} \tag{3.65}$$

where

$$\mathbf{B} = (\mathbf{I} - \mathbf{Y}\,\mathrm{diag}\{\mathring{\mathbf{w}}\}\mathbf{Y}^{\dagger})\,\mathrm{diag}\{\mathbf{p}\}\mathbf{H} \tag{3.66}$$

and $\mathbf{H}$ is the $[\Lambda \times \Gamma]$ microphones grouping matrix, e.g.

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \end{pmatrix} \tag{3.67}$$

The grouping must be quite conservative, in a sense that it is preferred to obtained more groups than necessary instead of too few, therefore a simple heap-based maximum search with an iterative incorporation of neighbouring microphones with monotone decreasing magnitude does not appear appropriate. A more ideal peak-region segmentation could take the following form [2]:

1. Create a set $\mathcal{S}_1 = \{1, \cdots, \Lambda\}$ containing the index of all microphones, this will represent the microphones that are not assigned to a region yet. A second set $\mathcal{S}_2$ is also initialised in the same way, this will represent the pool of points in which the current reference microphone is picked up during each loop. Note that $\mathcal{S}_1 \subset \mathcal{S}_2$ is always valid.
2. In $\mathcal{S}_2$, the microphone with the maximum magnitude is set as the current reference microphone m_c.
3. If $m_c \in \mathcal{S}_1$, then a new region is created. In the case where neighbouring microphones build up full triangular facets, then the microphones from the facet with the maximum average magnitude are added to the region. In the case where there is no "free" facet, we only add to the region the microphone m_c plus its neighbouring microphone with the greatest gain, if it is available.

   If $m_c \notin \mathcal{S}_1$, then the available microphones in $\mathcal{S}_1$, that are in the neighbouring of m_c and another microphone from the same region are added to the region if their magnitude are at most the value of one of the two microphones and at most 1.25 $\times$ the value of the other microphone.
4. $\mathcal{S}_1$ is updated by removing the points that just have been assigned to a region.
5. $\mathcal{S}_2$ is updated by removing the points that do not have any free neighbour or removing m_c in the case where no new point has been added to a region during this loop.
6. Repeat from step 2, until $\mathcal{S}_1$ is emptied.

An example of microphones grouping from a random directivity of maximum order $L = 2$ based on the above algorithm is depicted in Fig. 3.10.

2. There, the term "neighbouring" means the microphones that share a common triangular facet from a convex hull

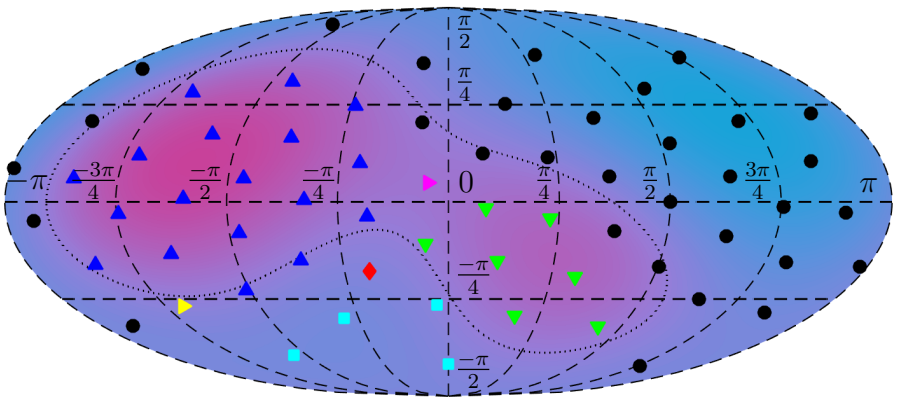


Figure 3.10 – Example of a peak region grouping for a random directivity pattern of maximum order $L = 2$. The dotted line represents the polarity change.

Chapter 4

Primal signal/radiation-filter decomposition

The sound emission of a natural source can be seen a single-input-multiple-output (SIMO) system, whose input consists of the so called primal signal, which contains all emitted spectral components and the radiation filter, which describes how the primal signal propagates in each direction depending on the frequency. The far-field spherical wave spectrum $\chi_n^m(\omega) = \left[a_n^m(\omega)h_n^{(2)}(\omega)\right]$ that drives the spherical harmonics in (2.1) can be represented as a vector

$$\boldsymbol{\chi}[k] = U[k]\boldsymbol{\Psi}[k], \tag{4.1}$$

where $U[k]$ represents the single channel primal source [1] and $\boldsymbol{\Psi}[k]$ the radiation filter.

Such a model based on natural source measurements should :

— consider the time variation, of the system, this includes movements/rotations of the instrument, changes of the playing techniques and currently played note (e.g. a given frequency propagates differently depending on the combination of opened tone holes on wind instruments, ...).

— avoid the potential destructive interferences at high frequency discussed in Chapter 2, even in the case of a poor source centring within the spherical microphone array.

— yield a natural-sounding primal source signal in terms of timbral and temporal properties (no echo, time spreading and relatively probable spectral cues, that differ much from known spectral properties of the instrument).

— yield a directivity robust against eccentric positioning of the instrument within the spherical microphone array.

1. The letter u has been chosen after the german name *Ursignal* that means primal signal

— yield a directivity filter ideally independent of the signal gain (the directivity pattern of each frequency should be arbitrarily at 0dB for perfect omnidirectional source).
— permit independent manipulation of both the primal source signal and directivity filter.
— allow to reduce the directivity filter order without much spectral distortion or magnitude directivity error.

This chapter proposes a straightforward framework, in which the spectrogram of each microphone serves as a basis for both the primal source estimation (as proposed in [Süs11, Hol14]) and the directivity-filter estimation. For both, the measured phase component of the Short-Time Fourier Transform (STFT) of the microphones signals is ignored and replaced by an optimum phase to avoid phase-related interference errors.

Primal signal estimation To guarantee the presence of all the spectral components within the estimated primal signal, the desired interference free primal source spectrogram is defined as the l^p-norm of the microphones spectrogram

$$|U_\tau[k]| = \left(\sum_{\lambda=1}^{\Lambda} |X_{\tau,\lambda}[k]|^p \right)^{\frac{1}{p}}, \tag{4.2}$$

where $U_\tau[k]$ is the desired primal source signal spectrogram, $X_{\tau,\lambda}[k]$ the spectrograms of the microphones $\lambda \in \{1, \ldots, \Lambda\}$.

The phase of the primal signal STFT is then estimated from the spectrogram using spectrogram inversion methods as discussed in Section 4.1. For convenience, the primal source signal as well as its derived signals (e.g. time frames signals, STFT, ...) are referred by the letters x instead of u in Section 4.1.

Directivity filter estimation The original idea of keeping the phase relations between microphones and primal signal gave overly complicated filter, therefore the radial filter design is elaborated from a given magnitude, then an optimum phase is applied. The gain of the directivity filter in the space domain $P_{\tau,\lambda}[k]$ can be obtained by dividing the microphones spectrogram by the desired primal source signal spectrogram

$$|P_{\tau,\lambda}[k]| = \frac{|X_{\tau,\lambda}[k]|}{|U_\tau[k]|}. \tag{4.3}$$

For convenience, the directivity filter is then decomposed into the spherical harmonics as introduced in Section 2.1 , for which the coefficients are denoted by the

vector $\mathbf{\Psi}_\tau[k]$. Similarly to the primal source signal estimation, the phase still has to be reconstructed, however, this time, not only the time and frequency dimensions have to be taken into account, but also the space dimensions, namely across the 2-sphere $\mathbb{S}^2$ representing the set of all radiation directions. Such phase reconstruction problems have already been presented in Chapter 3 and are applicable for a single time-frequency element. In Section 4.2, we propose a method for applying spherical phase retrieval algorithm that is powerful enough to obtain a time varying filter as a model.

The whole SIMO analysis framework is depicted in Fig. 4.1.

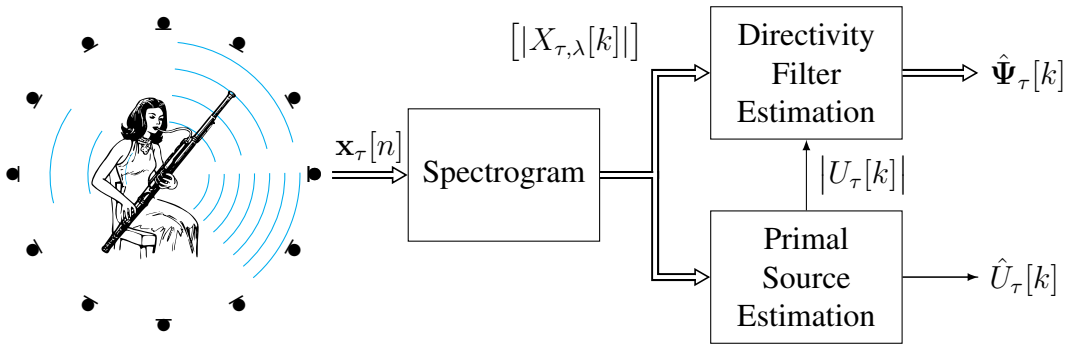


Figure 4.1 – Primal Source – Directivity Filter analysis.

4.1 Primal signal reconstruction with Spectrogram Inversion (SI)

The Short Time Fourier Transform (STFT) analysis of the time serie $x[n]$ is defined as

$$X_\tau[k] = \sum_{n=0}^{M-1} x[n+\tau R] w_a[n] e^{\frac{-\mathrm{i}2\pi kn}{M}}, \tag{4.4}$$

where τ is time frame index, k the frequency bin index, n the time sample index, R the hopsize in samples, M the fft-length and $w_a[n]$ the analysis window. In this work it is assumed that w_a is real and finite with a range of summation $\{0, \ldots, M-1\}$.

A perfect reconstruction from the STFT can be achieved with the overlap and add method

$$x[n] = \sum_{\tau=-\infty}^{\infty} x_\tau[n - \tau R], \tag{4.5}$$

where $x_\tau[n]$ is the time serie of the frame τ on which the suitable synthesis window has been applied

$$x_\tau[n] = w_s \sum_{k=0}^{M-1} X_\tau[k] e^{\frac{\mathrm{i}2\pi kn}{M}}, \qquad \forall n \in \{0, \dots, M-1\}. \tag{4.6}$$

In order to guarantee a perfect reconstruction, the synthesis window w_s has to fulfill

$$\sum_{\tau \in \mathbb{Z}} w_a[n - \tau R] w_s[n - \tau R] = \frac{1}{M}. \tag{4.7}$$

A corrected synthesis window obtained from a given window w_a can be defined as

$$w_s[n] = \frac{1}{M} \frac{w_a[n]}{\sum_{\tau \in \mathbb{Z}} w_a[n - \tau R]^2}, \qquad \forall n \in \{0, \dots, M-1\}. \tag{4.8}$$

Note that the summation range $\mathbb{Z}$ used in (4.8) can be reduced to $\left\{-\left\lceil \frac{M}{R} \right\rceil + 1, \dots, \left\lceil \frac{M}{R} \right\rceil - 1\right\}$ in our case, as w_a and w_s are fully supported by $\{0, \dots, M-1\}$ and adjacent time frames are distant from R.

When omitting the phase, we obtain the spectrogram $|X_\tau[n]|$. Our goal is to find the original signal $x[n]$ without knowing the phase, by exploiting the redundancy hidden in the overlap of adjacent time frames. As requirement, the overlapping factor $\frac{M}{R}$ must be greater than 2, in order to guarantee redundancy between adjacent time frames.

The signal estimate is denoted $\hat{x}[n]$. For this, many methods have been proposed, some are based on iterative procedures [GL84, BZW05, ZBW06, GS08, GS10], non-iterative [PBS17, PS16] and hybrid [PR17]. In this work, all time signals are assumed to be real valued, therefore the processing of frequency bins with indices $k = \left\{0, \dots, \left\lfloor \frac{M}{2} \right\rfloor + 1\right\}$ are sufficient in all following methods.

4.1.1 Iterative methods

Griffin & Lim's algorithm (G&L) In 1984, Griffin and Lim [GL84] proposed a simple iterative method in order to reconstruct a signal from a spectrogram.

The algorithm is based on the observation that not every STFT $\tilde{X}_\tau[k]$ is valid, in a sense that there might not exist any time signal $\tilde{x}[n]$ having the given STFT with phase. From this, it has been proposed, that for any "invalid" STFT, we can estimate a time signal $\hat{x}[n]$ which minimises the mean squared error (MSE) between the invalid STFT $\tilde{X}_\tau[k]$ and the feasible STFT of the estimated signal $\hat{X}_\tau[k]$. The distance

criterion to be minimised is

$$D(\hat{x}[n], \tilde{X}_\tau[k]) = \sum_{j\in\mathbb{Z}} \frac{1}{M} \sum_{k=0}^{M-1} \left|\hat{X}_\tau[k] - \tilde{X}_\tau[k]\right|^2, \tag{4.9}$$

which, according to the Parseval's theorem, can be expressed in the short time domain,

$$D(\hat{x}[n], \tilde{X}_\tau[k]) = \sum_{\tau\in\mathbb{Z}} \sum_{n=0}^{M-1} \left[\hat{x}_\tau[n] - \tilde{x}_\tau[n]\right]^2, \tag{4.10}$$

after setting its gradient with respect to $\hat{x}[n]$ to zero, it can be shown that the optimal signal takes the form

$$\hat{x}[n] = \frac{\sum_{\tau\in\mathbb{Z}} w_a[l-\tau R]\tilde{x}_\tau[l-\tau R]}{\sum_{\tau\in\mathbb{Z}} w_a[l-\tau R]^2}. \tag{4.11}$$

Therefore, an initial – potentially invalid – STFT $\tilde{X}_\tau^{(0)}[k]$ could be constructed by applying a random or an arbitrary phase to a spectrogram of interest $|X_\tau[k]|$ and compute the time signal $\hat{x}_\tau^{(0)}[l]$ minimising the MSE. Then, for each iteration $\iota \in \{1, 2, \ldots, I\}$, the phase $\hat{\phi}_\tau^{(\iota)}[k]$ is extracted from the STFT $\hat{X}_\tau^{(\iota)}[k]$ of the new estimated signal and applied to the original spectrogram of interest $|X_\tau[k]|$. It appears that $D(\hat{x}^{(\iota)}[n], \tilde{X}_\tau^{(\iota)}[k])$ decreases at each iteration and converges toward 0, although it has not been mathematically proven yet.

Unfortunately in its original form, this algorithm is not suitable for a real-time implementation as the whole time signal is required, furthermore without any adequate initialisation of the phase, an impractical large number of iterations may be required to ensure perceptually satisfying results, and hereby renders the algorithm slow.

Real-Time Iterative Spectrogram Inversion (RTISI) Despite the fact, that G&L is not designed for real-time processing, it can be easily adapted for a real-time use on streamed audio, as proposed by Beauregard *et al.* in 2005 [BZW05].

According to RTISI, the ι^{th} G&L-based iteration may estimate the phase of the current time frame $\angle X_{\tau_c}^{(\iota)}[k]$ by only using the $\left\lceil \frac{M}{R} \right\rceil - 1$ times frames past, and the phase values previously found for them

The algorithm can be decomposed as following for each time frame:

1. A partial reconstruction of the signal is computed with the help of the previous $\left\lceil \frac{M}{R} \right\rceil - 1$ overlapping time frames

$$\hat{x}_{part,\tau_c}[n] = \sum_{\tau=\tau_c-\left\lceil \frac{M}{R} \right\rceil+1}^{\tau_c-1} \hat{x}_\tau[n - \tau R] \tag{4.12}$$

   where $\hat{x}_\tau[n]$ is computed using (4.6). The partial reconstruction of those previous frames remains unchanged in each iteration for the iterative estimation in the current frame and therefore does only need to be computed once before starting to iterate.

2. A initial estimate is made by setting $\hat{x}_{\tau_c}^{(0)}[n] = 0$.

3. The estimate $\hat{x}_{\tau_c}^{(\iota)}[n]$ from the previous iteration is overlapped and added to the partially estimated signal $\hat{x}_{part,\tau_c}^{(\iota)}[n]$

$$\hat{x}_{sum,\tau_c}^{(\iota)}[n] = \hat{x}_{part,\tau_c}[n] + \hat{x}_{\tau_c}^{(\iota)}[n - \tau_c R]. \tag{4.13}$$

4. An analysis window is applied on the reconstruction $\hat{x}_{sum,\tau_c}[n]$ at the position of time frame x_{τ_c}

$$\tilde{x}_{\tau_c}^{(\iota)}[n] = w_a[n]\hat{x}_{sum,\tau_c}^{(\iota)}[n - \tau_c R] \tag{4.14}$$

5. The new estimate $\hat{x}_{\tau_c}^{(\iota+1)}[n]$ is computed by applying the phase of $\tilde{X}_{\tau_c}^{(\iota)}[k]$ onto the desired magnitude $|X_{\tau_c}[k]|$

$$\hat{x}_{\tau_c}^{(\iota+1)}[n] = w_s[n] \sum_{\tau \in \mathbb{Z}} \tilde{X}_{\tau_c}^{(\iota)}[k] \frac{|X_{\tau_c}|}{|\tilde{X}_{\tau_c}^{(\iota)}[k]|} e^{\frac{\mathrm{i}2\pi kn}{M}}, \tag{4.15}$$

   where

$$\tilde{X}_{\tau_c}^{(\iota)}[k] = \sum_{\tau=0}^{M-1} \tilde{x}_{\tau_c}^{(\iota)}[n] e^{\frac{-\mathrm{i}2\pi kn}{M}}. \tag{4.16}$$

6. Repeat steps 3-5 until the maximum iteration number I is reached, the final estimate of time frame x_{τ_c} is

$$\hat{x}_{\tau_c}[n] = \hat{x}_{\tau_c}^{(I)}[n]. \tag{4.17}$$

7. Proceed to the next time frame and start with step 1.

The whole process for the phase estimation of a certain time frame x_{τ_c} is depicted in Fig. 4.2.

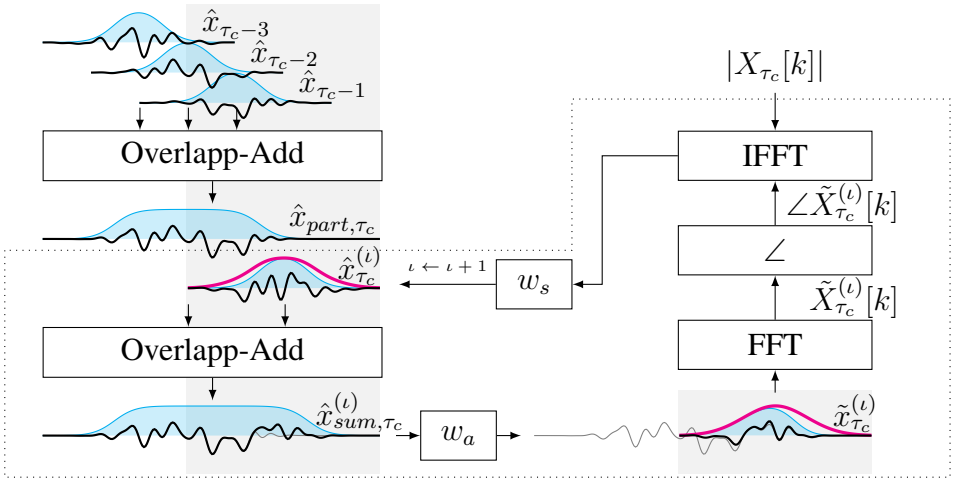


Figure 4.2 – Iterative phase estimation of time frame around τ_c by mean of RTISI. The cyan surfaces depict the effective windowing of the signal. The magenta lines depict the window applied by the previous processing block (w_a or w_s). The dotted frame highlights the part altered by each iteration. The gray patchs show the region of the time-frame to be committed.

Gnann & Spiertz's Real-Time Iterative Spectrogram Inversion with Look-Ahead (GSRTISI-LA) From RTISI, many improvements can be made in order to enhance the convergence speed with a lower computational complexity, e.g. as proposed in [ZBW06, GS08, GS10]. Hereby we will focus on two improvements that seem particularly interesting, namely

— The use of Look-Ahead time frames, introduced by Zhu *et al.*. Therefore, not only the information from previous time frames is used, but also the information from future time frames.

— The use of window compensation on partially reconstructed signal as advised by Gnann & Spiertz.

The following procedure includes both improvements and is referred in this work as the Gnann & Spiertz's Real-Time Iterative Spectrogram Inversion with Look-Ahead (GSRTISI-LA). N_{LA} denotes the number of Look-Ahead time frames and may be arbitrarily set to any non-negative integer smaller than or equal to the number of future overlapping time frames $\left\lceil \frac{M}{R} \right\rceil - 1$. The algorithm can be decomposed into the following steps:

1. Similarly to RTISI, the partial reconstruction $\hat{x}_{part,\tau_c}$ is computed based on the previously estimated time frames using (4.12). A variable p is set to N_{LA} , corresponding to the relative index of the currently processed time frame ($p = 0$ corresponds to the time frame to be committed, $p = N_{LA}$ indicates the latest

look-ahead time frame).

2. The current and the N_{LA} future time frames are overlapped and added to $\hat{x}_{part,\tau_c}$. It results

$$\hat{x}^{(\iota)}_{sum,\tau_c}[n] = \hat{x}_{part,\tau_c}[n] + \sum_{\tau=\tau_c}^{\tau_c+N_LA} \hat{x}^{(\iota)}_{\tau}[n-\tau R]. \tag{4.18}$$

   Hereby, we assume that at least the $N_{LA}-1$ first look-ahead time frames have been initially estimated.

3. The summed signal $\hat{x}^{(\iota)}_{sum,\tau_c}$ is segmented at the position of the $\tau_c + p$ time frame. Unlike the standard RTSI, we do not apply a simple analysis window w_a but a modified one ($w_{\tilde{a},p}$) in order to avoid inconsistency between the time and frequency representation. In fact, the overlap and add procedure with a finite set of time frames leads to a amplitude decay at the beginning and ending of the reconstruction $\tilde{x}^{(\iota)}_{sum,\tau_c}$ as depicted in Fig. 4.2. An appropriate windowing would enable to maintain a constant gain and therefore, the effective window on $\hat{x}^{(\iota)}_{\tau_c}$ will be equal to the original analysis window w_a.

$$w_{\tilde{a},p}[n] = M\frac{w_a[n]}{w_{sum}[n+pR]}, \tag{4.19}$$

   where

$$w_{sum}[n] = \sum_{\tau=0}^{N_{LA}} w_a[n-\tau R]w_s[n-\tau R]. \tag{4.20}$$

   This leads to

$$\tilde{x}^{(\iota)}_{\tau_c,p}[n] = w_{\tilde{a},p}[n]\hat{x}^{(\iota)}_{sum,\tau_c}[n-\tau_c R] \tag{4.21}$$

4. Similarly to RTISI, the new estimate $\hat{x}^{(\iota)}_{\tau_c+p}$ is computed by applying the phase of $\tilde{X}^{(\iota)}_{\tau_c+p}[k]$ onto the desired magnitude $|X_{\tau_c+p}[k]|$ using (4.15) and (4.16). Note that, this time, this iteration index is not incremented.
5. Steps 2-4 are repeated from $p = N_{LA}$ and decreasing to 0.
6. Steps 2-5 are repeated until the number of desired iterations is reached.
7. The final estimate $\hat{x}_{\tau_c}$ is committed, as in (4.17).

The whole algorithm is depicted in Fig. 4.3.

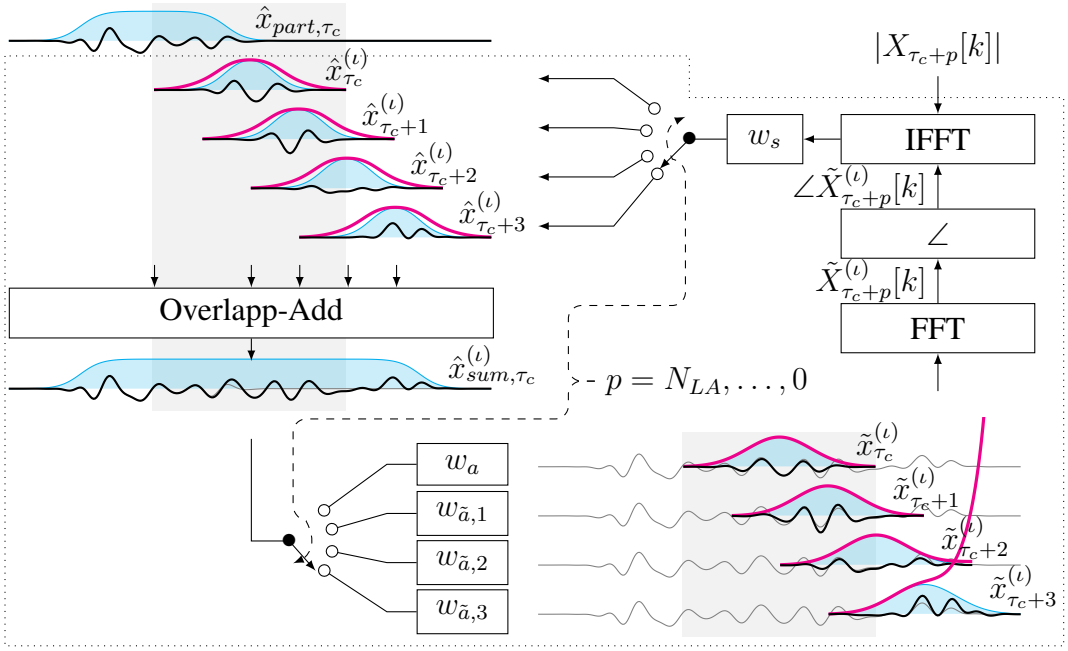


Figure 4.3 – Iterative phase estimation of time frame around τ_c by mean of GSRTISI-LA. The cyan surfaces depict the effective windowing of the signal. The magenta lines depict the window applied by the previous processing block (w_a, $w_{\tilde{a},p}$ or w_s). The dotted frame highlights the part altered by each iteration. Within each iteration, p changes decreasingly from N_{LA} to 0. The gray patchs show the region of the time-frame to be committed.

4.1.2 Non-iterative methods

Phase Gradient Heap Integration In 2017 Průša *et al.* proposed an efficient non-iterative algorithm for STFT reconstruction [PBS17]. The main idea consists in taking advantage of the relationship between the gradient of the log-magnitude Gabor transform to the gradient of its phase.

Using a truncated Gaussian window, where $h \in]0, 1[$ denotes the relative height at its truncation, we can approximate the entries of the phase gradient $\nabla\phi = (\frac{\partial\phi}{\partial k}, \frac{\partial\phi}{\partial\tau})$ using the centred differences over the time frame and frequency bin indices τ and k

$$\frac{\partial\phi_\tau[k]}{\partial k} = \frac{\gamma}{2RM}\left(\log(|X_{\tau+1}[k]|) - \log(|X_{\tau-1}[k]|)\right), \tag{4.22}$$

$$\frac{\partial\phi_\tau[k]}{\partial\tau} = \frac{RM}{2\gamma}\left(\log(|X_\tau[k+1]|) - \log(|X_\tau[k-1]|)\right) + \frac{2\pi Rk}{M}, \tag{4.23}$$

where γ denotes the "time-frequency ratio" of the Gaussian window [2] defined as

$$\gamma = \frac{-\pi}{4}\frac{M^2}{\log(h)}. \tag{4.24}$$

From the gradient $\nabla\phi$, the phase can be reconstructed over the whole spectrogram by using the trapezoidal rule and cumulative sum over a certain integration path. In order to get some phase consistency between neighbouring (τ, k)-cells, the integration path obeys to a simple heap-based rule.

Firstly, a random phase may be applied for all (τ, k)-cells beyond a certain relative tolerance tol, in order to reduce the computational effort. The coordinates (τ, k) of the remaining cells of the incomplete STFT are placed in a set $\mathcal{I}$. An empty heap is created, which will permit to order the (τ, k) coordinates of cells according to their magnitude.

Until $\mathcal{I}$ empties (i.e. the phase of the whole STFT is estimated), the following procedure is repeated: If the heap is empty, the cell with the greatest magnitude is identified, then its coordinates (τ, k) are removed from $\mathcal{I}$, is added to the heap and an arbitrary phase $\hat{\phi}_\tau[k] = 0$ is applied to the incomplete STFT

$$\hat{X}_\tau[k] = |X_\tau[k]| \cdot e^{i\hat{\phi}_\tau[k]}. \tag{4.25}$$

Recursively, the neighbouring STFT values of the heap first cell with coordinates (τ_h, k_h), are given the following phase

$$\hat{\phi}_{\tau_h}[k_h+1] = \hat{\phi}_{\tau_h}[k_h] + \frac{1}{2}\left(\frac{\partial\phi_{\tau_h}[k_h]}{\partial k} + \frac{\partial\phi_{\tau_h}[k_h+1]}{\partial k}\right) + \pi, \quad \text{if } (\tau_h, k_h+1) \in \mathcal{I} \tag{4.26}$$

$$\hat{\phi}_{\tau_h}[k_h-1] = \hat{\phi}_{\tau_h}[k_h] - \frac{1}{2}\left(\frac{\partial\phi_{\tau_h}[k_h]}{\partial k} + \frac{\partial\phi_{\tau_h}[k_h-1]}{\partial k}\right) + \pi, \quad \text{if } (\tau_h, k_h-1) \in \mathcal{I} \tag{4.27}$$

$$\hat{\phi}_{\tau_h+1}[k_h] = \hat{\phi}_{\tau_h}[k_h] + \frac{1}{2}\left(\frac{\partial\phi_{\tau_h}[k_h]}{\partial k} + \frac{\partial\phi_{\tau_h+1}[k_h]}{\partial k}\right), \quad \text{if } (\tau_h+1, k_h) \in \mathcal{I} \tag{4.28}$$

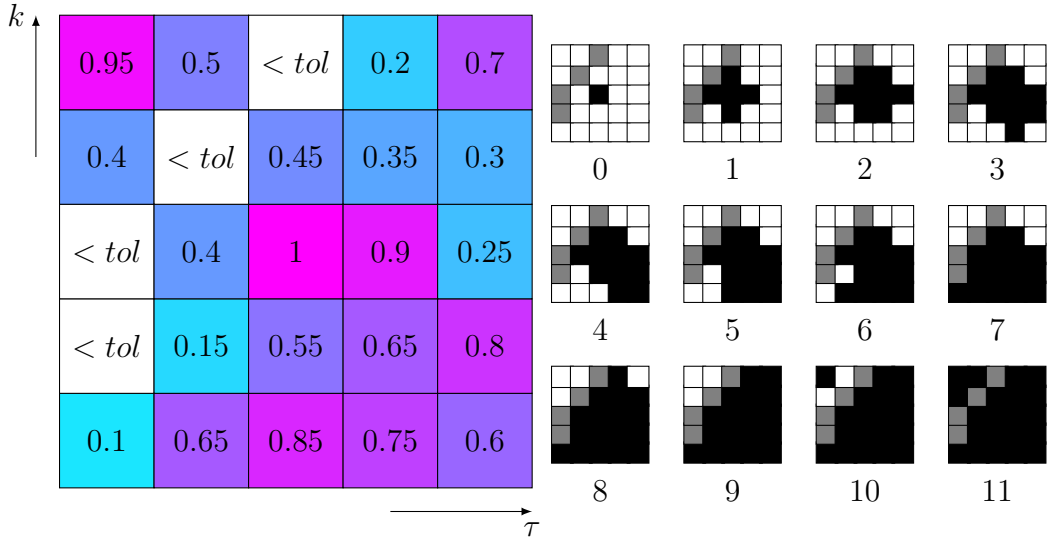
$$\hat{\phi}_{\tau_h-1}[k_h] = \hat{\phi}_{\tau_h}[k_h] - \frac{1}{2}\left(\frac{\partial\phi_{\tau_h}[k_h]}{\partial k} + \frac{\partial\phi_{\tau_h-1}[k_h]}{\partial k}\right), \quad \text{if } (\tau_h-1, k_h) \in \mathcal{I} \tag{4.29}$$

The new cells are then added to the heap and their coordinates are removed from $\mathcal{I}$.

When the heap and $\mathcal{I}$ are both empty, the STFT is complete and the time signal can be reconstructed using (4.5) and (4.6).

An example of an integration path for a given spectrogram is shown in Fig. 4.4.

2. According to [PBS17, PS16], this algorithm can even be applied on non-Gabor transform, i.e.



(a) Sample spectrogram.

(b) Integration sequence. Gray cells depict below-threshold magnitude, black and white cells show bins with estimated or missing phase respectively.

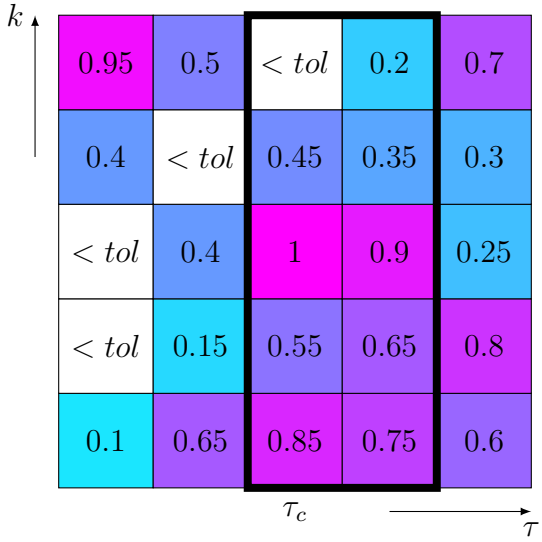
Figure 4.4 – Integration path for PGHI.

Real-Time Phase Gradient Heap Integration The real-time implementation of the PGHI (RTPGHI) is very similar to the off-line one, with the only difference that the integration path only enables the estimation of the phase within a single time frame at the time [PS16]. Contrary to the standard PGHI, as the phase of the previous time frame is already estimated and won't be modified anymore, all their values are directly copied into the heap before the integration as shown in Fig. 4.5.

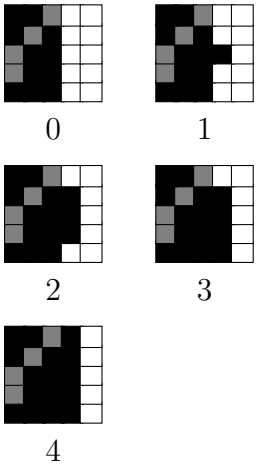
It should be noticed that according to (4.22), a look-ahead time frame is needed [3], which will introduce a delay to fulfil the causality constraint.

when using other types of window functions, the approximate "time-frequency" ratio can then be computed by taking the one of the closest Gaussian window (e.g. which minimises the mean squared error).

3. Note that a further approximation of (4.22) requiring no look-ahead time frame is also possible, as proposed in [PS16], however it is not discussed in this work.



(a) Sample spectrogram.



(b) Integration sequence. Gray cells depict below-threshold magnitude, black and white cells show bins with estimated or missing phase respectively.

Figure 4.5 – Integration path for RTPGHI.

4.1.3 Hybrid methods

As discussed in Section 4.1.1, all iterative algorithms require the use of an initial phase estimation. Both of the non-iterative methods described above could be used to compute an initial phase with little computational complexity, then a further convergence of the phase could be achieved by using any desired iterative method.

In [PR17], Průša *et al.* proposed a high quality real-time spectrogram inversion by initialising the phase with RTPGHI iteratively improving the phase estimate by mean of GSRTISI-LA [4]. This algorithm has been implemented in `C++`. Its source code or a Matlab Executable File (MEX) compilation of it can be obtained from the author on request.

An off-line implementation can also be obtained, e.g. by initialising the phase with PGHI and then use the GL algorithm.

4. There, the latest look-ahead time frame used by GSRTISI-LA also needs a further look-ahead time frame for its initial estimation.

4.2 Directivity-filter estimation

4.2.1 Directivity gain simplification in space domain

This magnitude spectrum division in (4.3) used to determine the microphone dependent directivity gain $|P_{\tau,\lambda}[k]|$ may be quite problematic under certain circumstances, even though the primal signal possesses all spectral components, as demonstrated by the noisiness of the cyan line in Fig. 4.7c. For this reason a regularisation appears to be necessary. However, in the case of a sparse spectrum, such a regularisation yield vanishing radiation filter between the relevant frequency bins. Therefore, a magnitude spectrum smoothing based on the high-energy spectral components may be beneficial.

Practically, a simple two-way non-linear filter is applied on the frequency domain of the microphone spectrum and the primal source before computing the relative gain. This filter is referred as a *Skirt*-filter and represented by the operator $Sk\{\cdot\}$.

The skirt filter applies a decaying slope onto the maxima of the magnitude spectrum with a decay rate specified by $\alpha[k] \in]0, 1[$. The block-diagram of its one-way version is depicted in Fig. 4.6, the two-way version used here is obtained running the spectrum through the skirt operator in reverse order [5].

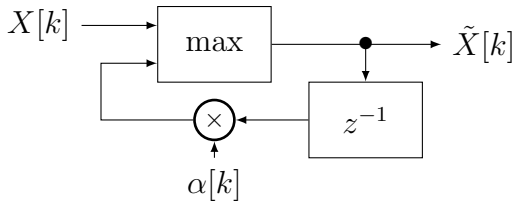


Figure 4.6 – One-way skirt filter with frequency dependent parameter $\alpha[k]$

The frequency dependent parameter $\alpha[k]$ is set to achieve a constant decaying slope, therefore the filter parametrisation can be reduced to a single frequency independent parameter β

$$\alpha^k = e^{\frac{-1}{\beta}}. \tag{4.30}$$

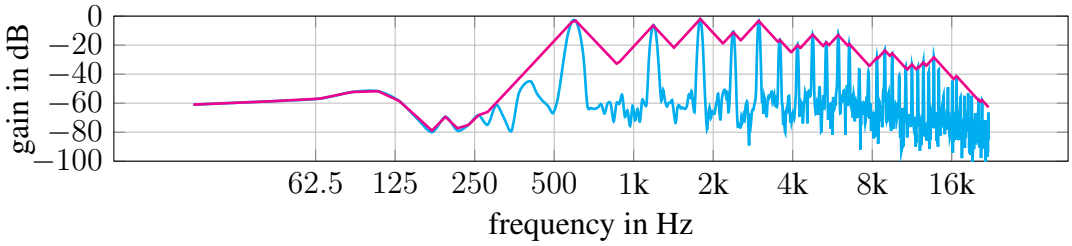
The resulting magnitude spectrum of the radiation filter for each microphone direction can be expressed as follows

$$|P_{\tau,\lambda}[k]| = \frac{Sk_\beta\{|X_{\tau,\lambda}[k]|\}}{Sk_\beta\{|U_\tau[k]|\}}. \tag{4.31}$$

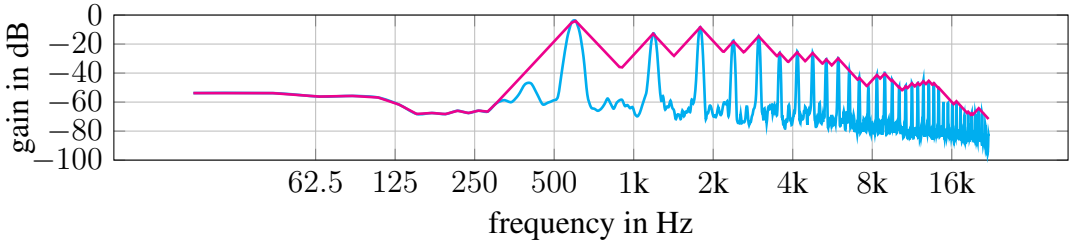
5. In Fig. 4.6, the z^{-1} element depicts the unit saving the value of the previous frequency bin and not the value of the previous time instant, as it is commonly used in the literature.

As an example, Fig. 4.7 presents the original and simplified magnitude spectrum of a microphone and the corresponding primal source as well as the their relative gains. It can be observed that the resulting gain curve is quite smooth and appears to be a great candidate for the radiation-filter design.

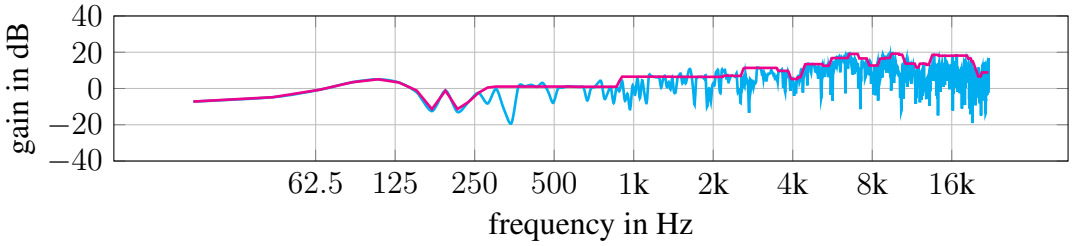
In contrast to the traditional filter bank (e.g. third octave filter bank) with fix bandwidth, the proposed method has the benefit to be signal dependent and to segment the short term magnitude spectrum into regions adapted to the frequency content.



(a) Magnitude spectrum of microphone 1



(b) Magnitude spectrum of primal source



(c) Magnitude spectrum of radiation filter at the position of microphone 1

Figure 4.7 – Magnitude spectrum of microphone 1, primal signal and their corresponding radiation filter. — Original gain — Simplified gain by applying a Skirt filter with factor $\beta = 0.1$

4.2.2 Phase estimation for frequency regions

The simplified magnitude spectrum of the primal source helps to segment the spectrum of the radiation filter into regions of common directivity patterns. The spatial spherical phase of this regions that represent the directivity of several frequency bins gathered can be determined separately. For those regions, the local minima of the primal source magnitude spectrum with skirt delimit the frequency regions, while the magnitude values of the radiation filter are obtained at the local maxima. Those values are used for the spherical phase retrieval algorithm.

As the retrieved absolute phase is not necessarily consistent with the previous time frame, it may be desired to rotate the phase of the microphones in a given frequency region ν and current time frame τ_c in order to best match the phase retrieved in the previous time frame τ_{c-1}. Hereby, the use of the main eigenvector from the phase matrix $\mathbf{Z}$ obtained with the SDR method which deliver a binary phase (0 or π) appears relevant, as a simple appropriate polarity inversion is sufficient. To achieve this, the following scalar product can be computed

$$\rho_\nu = \Re\big\{\mathbf{z}_{\tau_{c-1}}[k_{\nu,\max}]\big\}^{\mathsf{T}} \Re\{\mathbf{z}_{\tau_c}[k_{\nu,\max}]\} \tag{4.32}$$

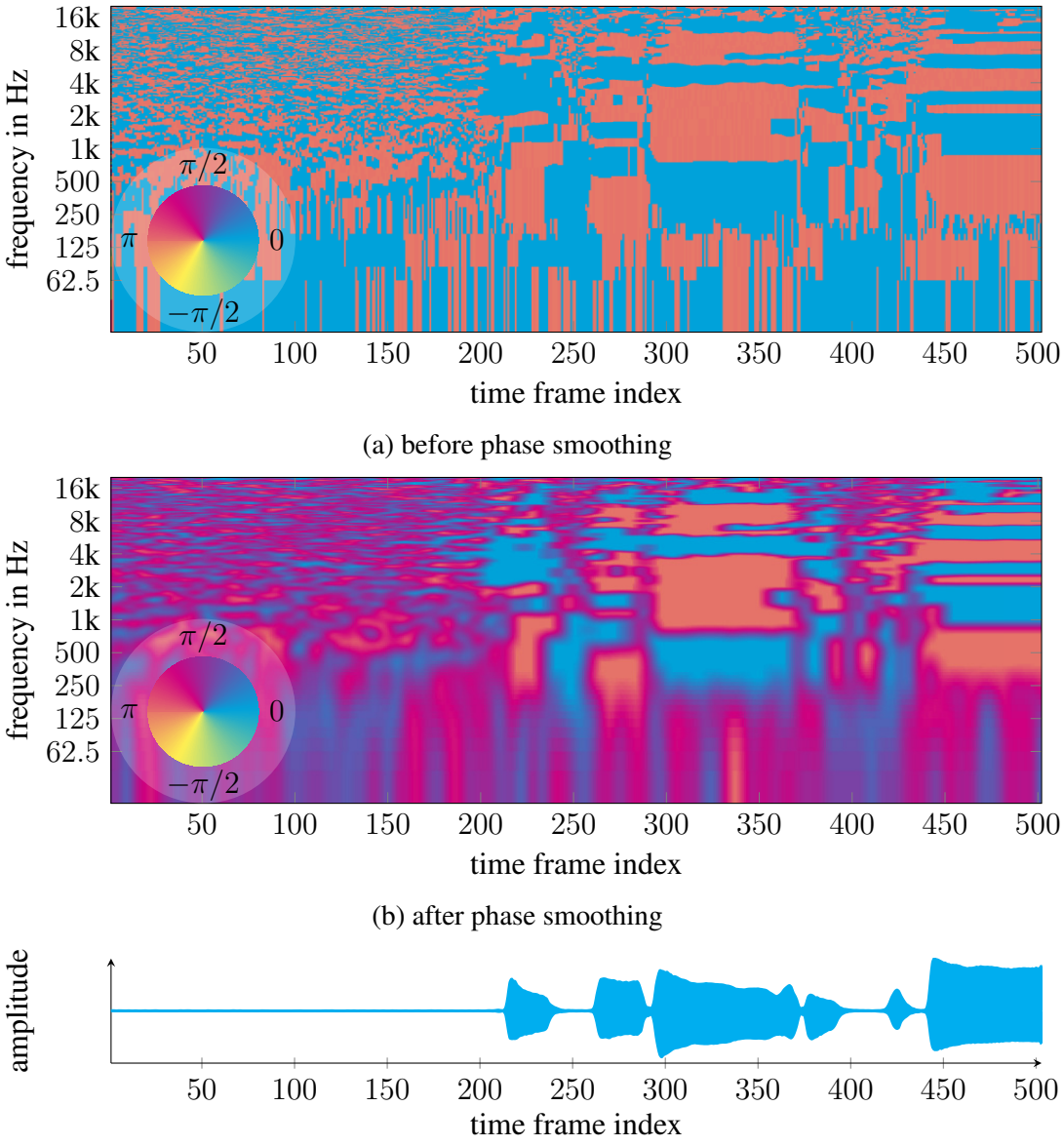
where $k_{\max}$ is the frequency bin of the peak of the frequency region ν and current time frame τ_c. The sign of ρ_ν determines wether the polarity of the microphones has to be inverted for the frequency region ν. Thus, a phase of 0 or π is assigned to each microphone for each frequency region and time frame. Figure 4.8a depicts such a retrieved phase for microphone 1.

The second step is to "smooth" the phase in order to avoid discontinuity in the signal that may lead to a spectral whitening or other non-linear artefacts. By taking into account that the phase is either 0 or π, a simple 2 dimensional Gaussian filter can be applied into the real part of the complex phasor $\mathbf{z}_\tau[k]$ of each microphone, then the allpass properties of $\mathbf{z}$ can be retrieved by applied the $\arccos$ function onto the smoothed real part

$$\hat{\phi}_\tau[k] = \arccos\left(\Re\{\mathbf{z}_\tau[k]\} * * \mathcal{G}\right), \tag{4.33}$$

where $\mathcal{G}$ denotes a discrete Gaussian kernel of arbitrary variance. The resulting phase for microphone 1 is shown in Fig. 4.8b. It can be observed, that due to the non-injectivness of the $\cos$ function, the image of the $\arccos$ function is restricted to $[0, \pi]$, therefore the retrieved phase also lies within $[0, \pi]$.

Apart from the cancellation of spectral artefacts, the smoothing also offers a steadiness of the directivity that is appreciable for visualisation, especially in the absence of a radiated signal where the radiation filter tends toward a monopole instead of a noisy directivity pattern.



(a) before phase smoothing

(b) after phase smoothing

(c) enveloppe of primal source signal

Figure 4.8 – Retrieved phase for microphone 1 from an alto saxophone recording by selecting the main eigenvalue of the SDR solution.

4.2.3 Some considerations on the impulse response compactness

The energy of the resulting SH impulse responses (IR) obtained from the phase optimisation process described above appears to be concentrated against the first and last sample. As this would lead to a strong echo, the IR is simply shifted in a circular way (i.e. complex linear modulation in the frequency domain) in order to concentrate the energy in the centre. The price to be paid is an additional delay corresponding to the half length of the IR. A compaction of the filter into a "minimum-phase"-like SIMO-radiation-filter would be a great help, however such a solution could not be found during this work. Though, as it will be shown in Sections 5.3.1 and 5.3.2, the proposed definition yields perceptually very satisfactory results.

Chapter 5

Perceptual evaluation

5.1 Perceptual evaluation of interpolation methods on a sampled sphere

The different directional interpolation methods described in Chapter 2 are investigated with the help of a short listening experiment. This experiment stands as a demonstration of the interference problems described in Section 2.1.1 and therefore illustrates the problematics that motivate this work.

The experiment consists of a multi stimuli test, where the subject has to rate the similarity of interpolated signals obtained with different algorithms with a reference. Ratings were given by means of some sliders on a GUI implemented in MATLAB.

The different interpolation methods are described in Chapter 2 and consist of VBAP, MVBAP, hyperinterpolation with SH of order $L = 3$ (referred as HI-$L3$) and hyperinterpolation with SH of order $L = 7$ (referred as HI-$L7$). All of them are based on simulated microphone signals that are obtained by applying an appropriate delay and gain depending on their relative position to the virtual source and its directivity (assuming an amplitude decay of $1/d$ and a speed of sound of $c = 343\,\mathrm{ms}^{-1}$). The geometry of the array is given in Appendix A and depicted in Fig. 3.1 while the virtual source consists of an ideal dipole, oriented toward the x-direction and shifted from the microphone array centre by $(0.1\,\mathrm{m}, 0.2\,\mathrm{m}, 0\,\mathrm{m})$.

All stimuli played by the source consisted of a looped $3\,\mathrm{s}$ audio signal of either pink noise, speech, guitar and tablas recording.

The virtual microphone which determines the signal to be rated was rotated on the transverse plane, at an arbitrary distance of $1\,\mathrm{m}$ from the centre of the microphone array, starting a $\varphi_{start} = \frac{\pi}{2}$ and rotating anti-clock-wisely at a frequency of $0.5\,\mathrm{Hz}$

and stopping therefore at $\varphi_{end} = \frac{-\pi}{2}$.

The investigated signals were obtained by applying an interpolation block-wisely with a half-$\cos$ window of length $M = 512$ and an hopsize of $R = 256$, while the reference signal was obtained by directly weighting the stimulus according to the directivity of the ideal source for each sample. Both low and high frequency channels of the MVBAP algorithm were obtained with a second order Linkwitz-Riley filter pair with a cross frequency of $2\,\mathrm{kHz}$ [LV83].

Seven expert listeners from the IEM took part on the test and had to rate the (4 methods + 1 reference) $\times$ 4 stimulus types $\times$ 2 repetitions $= 40$ samples, which lasted about $15\,\mathrm{min}$ per subject. This leads to 56 values for each interpolation method.

The experiment results, grouped according to stimulus type, are depicted in the form of confidence intervals for the median in Fig. 5.1. This representation aspires to give a better overview of the subject answers.

The null-hypothesis that the mean-rank of each interpolation method is equal, is tested with the help of a Wilcoxon signed-rank test. The resulting p-values are shown in Table 5.1 and lead to a rejection of the null-hypothesis for each pair of method. Thus the average rank gives us a good indication of the quality of the restitution.

Regarding the hyperinterpolation, it seems that the quality of the restitution benefits from an increase of the SH order. In fact, an increase of the order leads to a narrower weighting beam; the microphones far from the desired direction – which potentially have a greater relative phase difference – contribute less to the interpolated signal, thus leading in a reduction of the destructive interference. However, according to Fig. 5.1, a 7^{th} order hyperinterpolation might not be sufficient to obtain a satisfactory restitution of the source.

On the other hand, the VBAP methods has the benefit to reduce even more the contribution of distant microphones (with potentially destructing potential) by limiting itself to the 3 closest microphones. Therefore the destructive interferences can be avoided up to a higher frequency, thus increasing the spectral quality. However, with steady broadband signals such as pink noise, this method has the disadvantage to create strong audible artefacts when rapidly moving the interpolation direction, as the destructive interferences are strongly direction depedent.

Finally MVBAP outperforms all other methods. This confirms that a microphone signal interpolation employing a clever phase modification is promising for a more transparent virtualisation of the source.

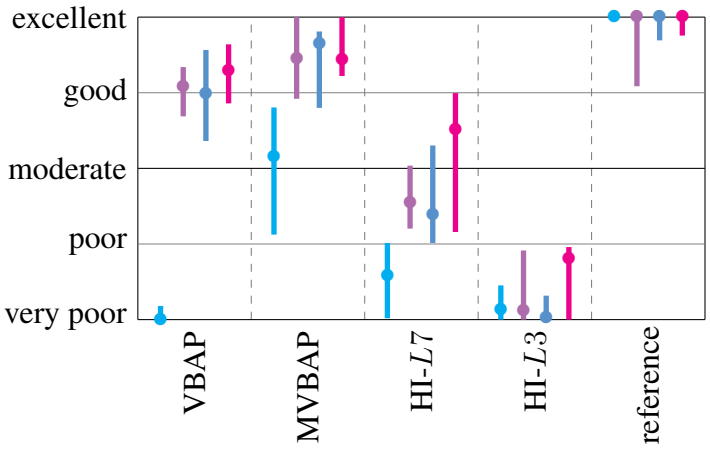


Figure 5.1 – Perceived quality for different spatial interpolation methods. The large points depict the estimated median and the whiskers depict 95% confidence intervals of the median. Pink noise, Speech, Guitar, Tablas.

5.2 Perceptual evaluation of reconstructed primal source signal using RTPGHI + GSRTISI-LA

In order to evaluate the perceptual quality of a reconstructed primal source signal using the spectrogram inversion discussed in Sections 4.1.2 and 4.1.3, two simple listening experiments were conducted.

The first focuses on the perceived quality of a monophonic signal reconstructed from its spectrogram using RTPGHI+GSRTISI-LA to find the best spectrogram parameters and for the tested sounds. The second experiment is conducted in order to evaluate the quality of a reconstructed primal signal from a multichannel recording by the means of spectrogram inversion and other strategies.

5.2.1 Optimal/required spectrogram parameters

In order to determine the optimal parameters to be used in RTPGHI+GSRTISI-LA to reconstruct a time signal with an acceptable computational complexity, different parameters were perceptualy evaluated in different scenarios. While some parameters are kept identical to those used in [PS16, PR17] as the usage of Gaussian window truncated at 1% of the height, the relative tolerance was arbitrary set to $tol = 10^{-3}$.

Method	VBAP	MVBAP	HI-$L7$	HI-$L3$	rank
VBAP					3.11
MVBAP	2.11e^{-6}				2.10
HI-$L7$	1.32e^{-6}	5.97e^{-10}			3.74
HI-$L3$	2.31e^{-9}	7.94e^{-11}	3.48e^{-10}		4.78
reference	5.23e^{-10}	4.02e^{-7}	8.40e^{-11}	5.92e^{-11}	1.28

Table 5.1 – Average rank of each interpolation method and p-values obtained with a Wilcoxon signed-rank test when mixing all types of stimuli together. Gray cells depict the p-values which lead to a rejection of the null-hypothesis with a significance level of $\alpha = 0.05$. Redundant p-values are not shown.

Apart from these fixed parameters, there remains many degree of freedoms to be analysed as the length of the analysis window M, the overlapping factor $Ov = \frac{M}{R}$ and the number of iterations I [1].

A MUlti Stimuli with Hidden Reference and Anchor (MUSHRA)-like test was designed with a Graphical User Interface (GUI) in MATLAB.

The ranges of the investigated independent variables were selected from an informal comparative listening test. The different variables consist of all combinations of window length $M \in \{512, 1024, 2048, 4096, 8192\}$ and overlapping factor $Ov \in \{4, 8\}$ plus hidden reference and anchor.

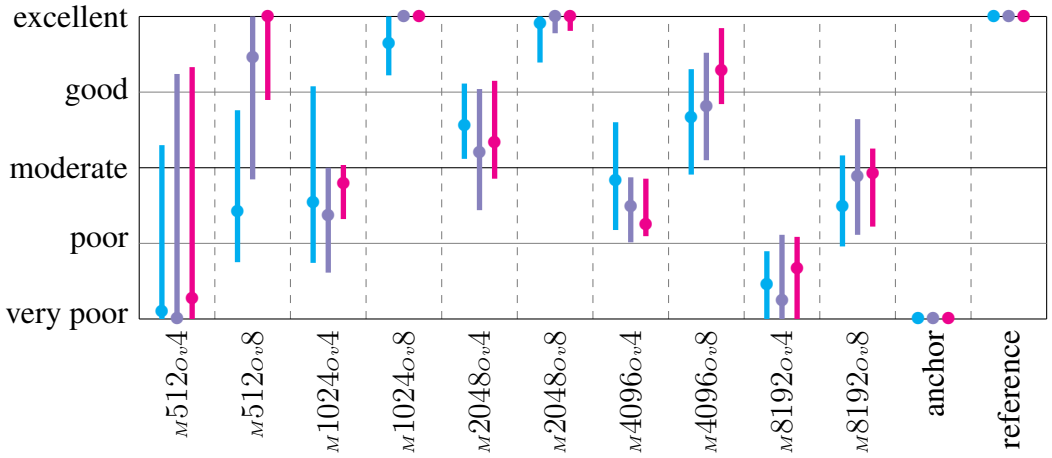
Different treatments were used in a full-factorial design, combining the numbers of iterations $I \in \{0, 3, 10\}$ and the type of stimulus, among a speech signal, a clean electric guitar recording and a tablas recording.

For each type of stimulus, the anchor consisted of a reconstruction without iteration. The window length used for the anchor signals was set to 512 samples for the speech and guitar and 8192 samples for the tablas.

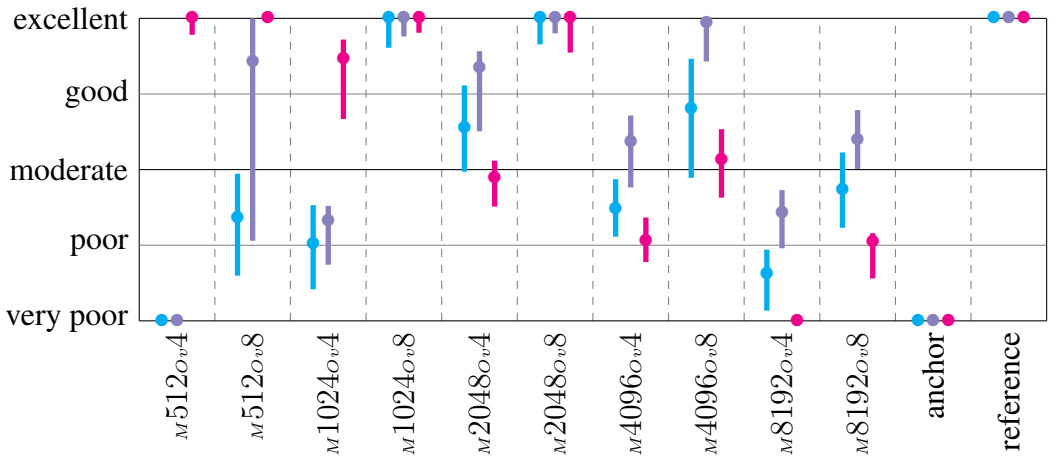
Nine expert subjects took part in the listening test and had to rate $(5 \times 2 + 2 = 12$ blocks) $\times (3 \times 3 = 9$ treatments) = 108 samples once.

All results are grouped for each number of iteration and type of stimuli and depicted separately in Fig. 5.2 in order to avoid plotting all treatment combinations. The choice to depict the medians and their 95% confidence intervals over the mean is motivated by the strong floor and ceiling effect observed. Furthermore it offers a robust confidence interval – although quite conservative due to the small amount of dataset – , which is insensitive against outliers.

1. Note that RTPGHI + GSRTISI-LA with $I = 0$ corresponds to RTPGHI



(a) Perceived quality for different numbers of iterations, all stimuli together. $I = 0$, $I = 3$, $I = 10$.



(b) Perceived quality for different stimuli, all iteration numbers together. Speech, Guitar, Tablas.

Figure 5.2 – Perceived quality for different STFT parameters. The large points depict the estimated median and the whiskers depict 95% confidence intervals of the median.

After a rapid overview, 2 parameters combinations, namely $(M = 1024, Ov = 8)$ and $(M = 2048, Ov = 8)$ (referred respectively as $_M1024_{Ov}8$ and $_M2048_{Ov}8$) appear to outperforms the others groups. This is confirmed by testing the null-hypothesis that their mean ranks are both equal to those of the others groups.

A Wilcoxon signed-ranked test is used to test this hypothesis on all paired formed by both the two groups $_M1024_{Ov}8$ and $_M2048_{Ov}8$ and the other groups (21 pairs testing) when all treatments are considered.

From Table 5.2, it can be confirmed that both the groups $_M1024_{Ov}8$ and $_M2048_{Ov}8$ deliver the best general results compared to other (M, Ov)-combinations when mixing all treatments together (I, type of stimuli) as their average ranking is better and significantly different from the other groups.

M	Ov	1024 8	2048 8	rank
512	4	$3.62e^{-12}$	$8.47e^{-12}$	8.62
	8	$1.34e^{-8}$	$5.02e^{-8}$	5.50
1024	4	$9.32e^{-14}$	$1.27e^{-13}$	7.78
	8			3.00
2048	4	$2.24e^{-11}$	$4.73e^{-13}$	6.01
	8	$8.75e^{-1}$		2.93
4096	4	$3.38e^{-14}$	$8.94e^{-14}$	7.79
	8	$1.38e^{-8}$	$2.27e^{-10}$	5.08
8192	4	$5.73e^{-15}$	$6.49e^{-15}$	10.10
	8	$1.93e^{-14}$	$6.48e^{-14}$	7.45
anchor		$1.79e^{-15}$	$2.24e^{-15}$	11.35
reference		$4.67e^{-3}$	$7.08e^{-3}$	2.40

Table 5.2 – Average rank of each (M, Ov)-group and p-values obtained with a Wilcoxon signed-rank test when mixing all treatments (I, type of stimuli) together. Gray cells depict the p-values which lead to a rejection of the null-hypothesis with a significance level of $\alpha = 0.05$. Redundant p-values are not shown.

Further investigations on the influence of the number of iterations on both these groups are undertaken with 15 additional Wilcoxon signed-rank tests. As shown in Table 5.3, groups with an iteration greater than 0 score significantly better than groups without any iteration with respect to their ranking.

However, despite the significant improvement of the quality by increasing the number of iterations, the results obtained without any iterations appear very satisfying,

especially for non-expert listeners (e.g. the 95% confidence range for the median of the perceived quality of $({}_M1024_{Ov}8\ , I = 0)$ is $[0.84, 1]$ and 77% of its values are greater than 0.8). Therefore, regarding the high computational cost of each iteration, it was decided to discard the iterative part of the algorithm in the following experiment.

Furthermore, despite the non-significant difference between $({}_M1024_{Ov}8\ , I = 0)$ and $({}_M2048_{Ov}8\ , I = 0)$, the window length $M = 2048$ was preferred for the following experiments.

M, Ov		1024, 8			2048, 8		rank
	i	0	3	10	0	3	
1024, 8	0						4.50
	3	3.18e^{-3}					3.17
	10	2.29e^{-3}	7.33e^{-1}				2.78
2048, 8	0	2.79e^{-1}	1.49e^{-1}	1.23e^{-2}			4.20
	3	7.90e^{-3}	1.27e^{-1}	9.77e^{-2}	2.35e^{-1}		3.37
	10	3.76e^{-3}	8.55e^{-1}	6.07e^{-1}	7.40e^{-3}	1.23e^{-1}	2.98

Table 5.3 – Average rank of each (M, Ov, I)-group and p-values obtained with a Wilcoxon signed-rank test when mixing all types of stimuli together. Gray cells depict the p-values which lead to a rejection of the null-hypothesis with a significance level of $\alpha = 0.05$. Redundant p-values are not shown.

5.2.2 Perceived quality of the reconstructed primal signal

Section 5.2 demonstrated that, when RTPGHI(+GSRTISI-LA) is well parametrised, the signal reconstructed from its spectrogram is perceptualy extremely similar to the original one. The question remains wether this can be used to accomplish a satisfactory primal signal estimation from a modified spectrogram under physical situations, which could be obtained by combining the individual spectrogram of each microphone together.

Again, the evaluation consists of a MUSHRA-like test where the perceived quality of primal signals estimated from different methods is assessed.

Hereby a simulation is designed to compute the signal of a circular microphone array for a source emitting a reference signal with controlled position, orientation, movement and directivity.

Simulated microphone signals for testing The geometry of the array is depicted in Fig. 5.3 and consists of 15 microphones uniformly positioned on a circle of 1.2m radius. According to the source characteristics (position, orientation, movement, directivity), each microphone signal is computed by applying the following instructions:

— The distance taken into account is computed by applying a $1/d$ linear gain, where d is the source-microphone distance and the right delay according to a speed of sound of $343\,\mathrm{m\,s^{-1}}$,

— The Doppler shift, induced by the movement of the source is modeled by an adequate frequency modulation of the signal, this is implemented by a simple signal interpolation (e.g. using the `interp1` function in MATLAB),

— The directivity of the source is modelled by applying a gain corresponding to the direction of the microphone relatively from the local coordinate system of the source described by its position and orientation. Assuming the directivity pattern of the source $h(\boldsymbol{\theta})$ being rotationally symmetric, the gain at microphone λ associated to the directivity of the source can be described as

$$g_\lambda = h(\phi) = h\left(\arccos\left(\frac{\langle(\mathbf{m}_\lambda - \mathbf{s}), \mathbf{d}_s\rangle}{\|\mathbf{m}_\lambda - \mathbf{s}\| \cdot \|\mathbf{d}_s\|}\right)\right) \tag{5.1}$$

where $\mathbf{m}_\lambda$ is the position vector of the λ^{th} microphone, $\mathbf{s}$ the position vector of the source, $\mathbf{d}_s$ the vector representing the orientation of the source.

Architecture of the experiment Different alternatives for the determination of the primal signal are investigated:

— The signal of two microphones have been selected (mic 1, mic2, drawn in cyan in Fig. 5.3),

— An estimation of the primal signal by use of the mean of all microphones spectrograms (referred as SI l^1),

— An estimation of the primal signal by use of the generalised mean with exponent 10 of all microphones spectrograms (referred as SI l^{10}),

— The mean signal obtained by superposing the microphone signals in the time-domain (referred as time mix).

Different conditions are modelled, the source is either fix at the position ($0.2\,\mathrm{m}$, $0.1\,\mathrm{m}$) and pointing toward the x-direction (referred as *off-centre*) or following the movement/rotation described in Fig. 5.3 (referred as *moving*). The source directivity is either *omnidirectional* or *directional*. Hereby, *directional* is modelled by combining the microphone signal obtained with an omnidirectional source below $800\,\mathrm{Hz}$ and the microphone signal obtained with a source of $L_b = 5^{\text{th}}$ order $\max$-r_E beam

directivity above $800\,\mathrm{Hz}$. This rotational symmetric directivity is described as

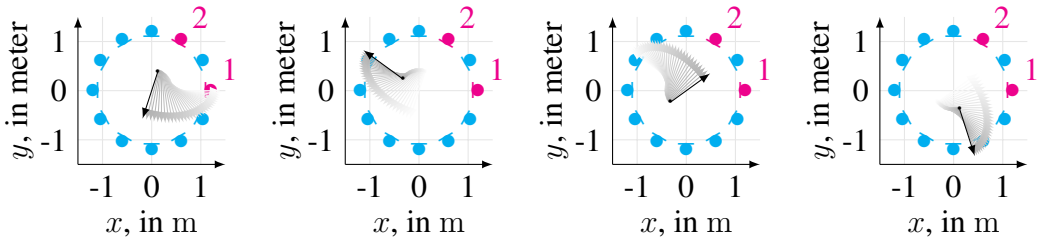
$$h(\phi) = \sum_{l=0}^{L_b} \frac{2l+1}{4\pi} a_l P_l(\phi), \tag{5.2}$$

where the coefficients a_l can be approximated as following [ZF12]

$$a_l \approx P_l\left(\cos\left(\frac{137.9^\circ}{N_b + 1.51}\right)\right). \tag{5.3}$$

The junction between the low frequency and high frequency directivity is made by mean of a second order Linkwitz-Riley filter pair to ensure a smooth transition [LV83].

The anchor signal for each scenario corresponds to the linearly mixed time signal, when the source is moving rapidly ($2\,\mathrm{Hz}$ anti-clockwise rotation on a circle of $0.4\,\mathrm{m}$ radius centred on ($0.2\,\mathrm{m}$, $0.1\,\mathrm{m}$). The source orientation rotates in the clock-wise direction with a frequency of $0.6\,\mathrm{Hz}$).



(a) from 0 to 1 s (b) from 1 to 2 s (c) from 2 to 3 s (d) from 3 to 4 s

Figure 5.3 – Source movement used in the simulation. The source position defined a by $0.2\,\mathrm{Hz}$ anti-clockwise rotation on a centred circle of $0.4\,\mathrm{m}$ radius. The source orientation rotates in the clock-wise direction with a frequency of $0.3\,\mathrm{Hz}$.

All stimuli consist of the same guitar loop used in the previous experiments.

The 7 groups (5 strategies + 1 anchor + 1 reference) for all 4 conditions (2 movements $\times 2$ directivity) with 2 repetitions leads to 56 samples per test person.

Seven expert listeners took part in the test, their results are depicted in Fig. 5.4.

Both groups SI l^1 and SI l^{10} are compared to others with a Wilcoxon signed rank test by testing the null-hypothesis that their average ranking are identical. The results are depicted in Table 5.4.

SI l^{10} appears to obtain the best results with mic 1 together according to their average ranking, however the use of an individual microphone appears not to be an optimal

solution especially for a directional source moving or pointing away from the microphone, since spectral elements may not all reach the microphone with an equal gain. The better results of SI l^{10} over SI l^1 for a directional source can be explained by the fact, that the generalised mean of the spectrograms with a high exponent enables a better conservation of the spectral elements that are less well represented in the space domain, on average.

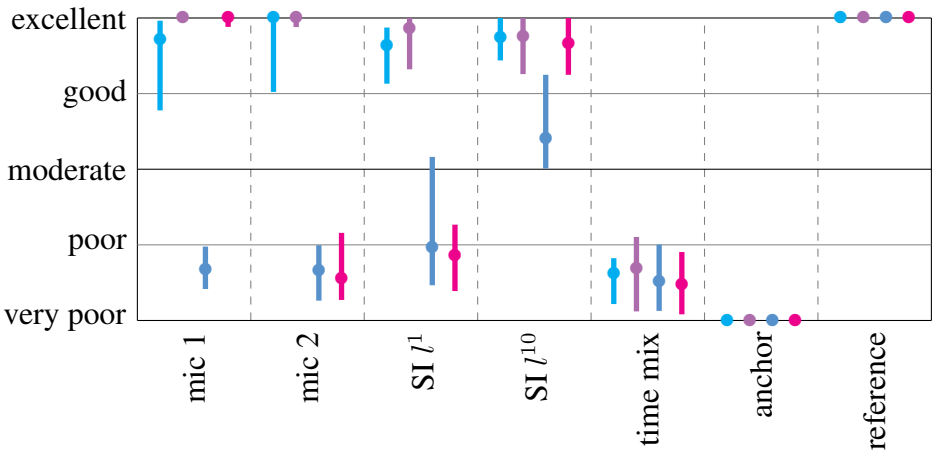


Figure 5.4 – Perceived quality of the estimated primal signal for different methods. The large points depict the estimated median and the whiskers depict 95% confidence intervals of the median. omnidirectional moving, omnidirectional off-centre, directional moving, directional off-centre.

5.3 Perceptual evaluation of primal source/radiation-filter decomposition methods

After demonstrating the great potential of the primal signal estimation by means of phase retrieving algorithm introduced in Section 4.1, the perceived quality of the whole primal signal/directivity-filter estimation algorithm described on Chapter 4, is evaluated in this section. To this end, two perceptual tests are conducted in a virtual free field environment and in a virtual reverberant environment respectively. Again, both experiments are based on simulated microphone signals as described in Sections 5.1 and 5.2.2 which has the benefit to provide a reference signal for comparison. The virtual array consists of the IEM 64 microphones array whose geometry is given in Appendix A and Fig. 3.1.

	SI l-1	SI l-10	rank
mic 1	5.51e^{-3}	1.28e^{-1}	3.13
mic 2	2.61e^{-1}	3.79e^{-5}	3.72
SI l-1			3.78
SI l-10	1.77e^{-6}		3.00
mix time	8.55e^{-10}	7.96e^{-11}	5.71
anchor	7.49e^{-11}	6.94e^{-11}	6.92
reference	1.02e^{-9}	7.87e^{-8}	1.73

Table 5.4 – Average rank of each group and p-values obtained with a Wilcoxon signed-rank test when mixing all treatments (source movement and directivity) together. Gray cells depict the p-values which lead to a rejection of the null-hypothesis with a significance level of $\alpha = 0.05$. Redundant p-values are not shown.

5.3.1 Perception under free-field conditions

In this experiment, the test persons were asked to rate different virtual source modelling algorithms similarly to the experiment described in Section 5.1 with some variations. Contrary to the experiment on the spatial interpolation in Section 5.1, the test persons were free to move the virtual microphone around the source with the help of a GUI implemented in Pure data. The global quality of the restitution of each source model, compared to the reference, was rated with sliders on a GUI implemented in MATLAB. The source properties used of the microphone signals simulation was identical to Section 5.1, i.e. positioning at $(0.1\,\text{m},\ 0.2\,\text{m},\ 0\,\text{m})$ and dipole directivity corresponding to the SH Y_1^1.

The auralisation techniques consist of:

— A direct SHT of the microphone signals with order 3. The resulting SH signal of the source after re-synthesis is weighted according to the $\max$-r_E approximation given in (5.3). The interpolation is done linearly with an update every $20\,\text{ms}$ between consecutive directions. This model is referred as HI-$L3$.

— A direct SHT of the microphone signals with order 7 and $\max$-r_E weighting. This model is referred as HI-$L7$.

— A zero phase variant of the primal signal/radiation-filter modelling of order 3. The primal signal was obtained with RTPGHI+GSRTISI-LA and its parameters $M = 2048$, $R = 256$, gaussian window with truncation at $1\,\%$, $tol = 1e-3$, $I = 20$ on a l^2 normed spectrogram. The gain of the radiation-filter was obtained with a skirt filter with parameters $\beta = 0.1$ and a $\max$-r_E weighting was applied on the resulting synthesised SH signal. The interpolation is done linearly with

an update every $20\,\mathrm{ms}$ between consecutive directions. This model is referred as ZP-$L3$.

— A second zero phase variant, this time with a maximal order of 7. This model is referred as ZP-$L7$.

— A 3^{rd} order modelling using the same primal signal as the zero phase variant and a radiation filter obtained with semi definite relaxation as described in Chapter 4. The phase of each frequency region is obtained from the main eigenvector of the optimal $\mathbf{Z}$ matrix and smoothed with a Gaussian kernel with a standard deviation of 5 along the time frames and 3 along the frequency bins (see Section 4.2.2). A $\max$-r_E weighting is also applied on the resulting SH signal. The interpolation is done linearly with an update every $20\,\mathrm{ms}$ between consecutive directions. This model is referred as SDR-$L3$.

— the MVBAP method, which scored best in Section 5.1 is also present for comparison. The interpolation is undertaken on 128 samples long half-$\cos$ windowed time-frames with an overlap of 64 samples. The cut-off frequency was set to $4\,\mathrm{kHz}$.

A total of 6 expert listeners from the IEM took part on the experiment. Each one of them had to rate (6 methods + 1 hidden reference) $\times$ 4 stimulus types $\times$ 2 repetitions = 56 samples, the experiment lasted about $15\,\mathrm{min}$.

The results are presented in Fig. 5.5 in the form of confidence intervals for the median for each stimuli type separately to give the reader a good overview of the ratings. The null-hypothesis that the average rank of each method is equal was tested by using the Wilcoxon signed-rank test, whose results are presented in Table 5.5.

Despite the non-significant difference between ZP-$L3$ and MVBAP or HI-$L7$, ZP-$L7$ seems to deliver an overall better virtualisation of the source than the methods implying a direct hyperinterpolation of microphone signal (HI-$L3$ and HI-$L7$ as well a MVBAP).

The phase retrieving primal signal/radiation-filter model SDR-$L3$ appears to outperform all other models significantly, which indicates an overall better spectral and directivity experience.

5.3.2 Perception in a reverberant field

One of the motivations to model natural sound sources is to include them in a virtual environment that can be auralised (e.g. for virtual reality purposes or artistic works). For this reason, the perceived quality of such models might be particularly interesting to investigate in such environments. In this experiment, the virtualised sources from the previous experiment Section 5.3.1 (i.e. HI-$L3$, HI-$L7$, MVBAP, ZP-$L3$, ZP-$L7$ and SDR-$L3$) are inserted in a virtual shoe-box room by means of an image source

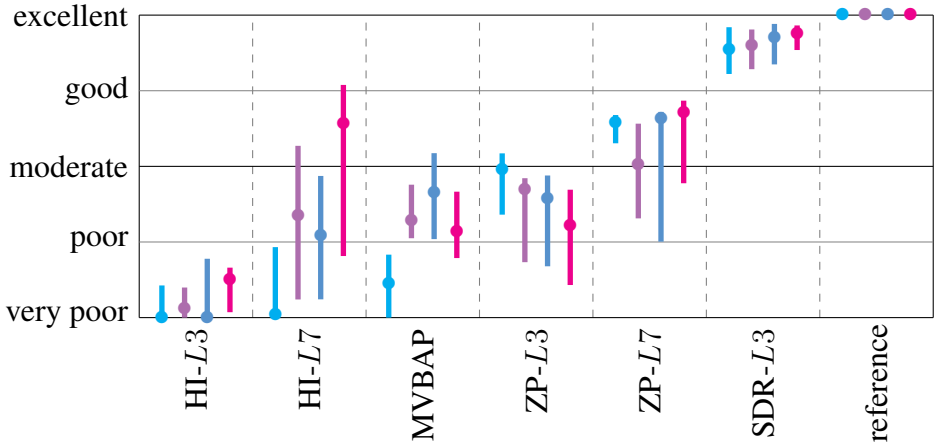


Figure 5.5 – Perceived quality of the radiated signal in free field for different source auralisation methods. The large points depict the estimated median and the whiskers depict 95% confidence interval of the median. Pink noise, Speech, Guitar, Tablas.

Method	HI-$L3$	HI-$L7$	MVBAP	ZP-$L3$	ZP-$L7$	SDR-$L3$	rank
HI-$L3$							6.34
HI-$L7$	1.53e^{-6}						4.97
MVBAP	1.28e^{-4}	8.57e^{-1}					5.05
ZP-$L3$	4.66e^{-5}	3.12e^{-1}	1.82e^{-1}				4.77
ZP-$L7$	1.69e^{-8}	2.22e^{-4}	5.32e^{-5}	5.26e^{-6}			3.71
SDR-$L3$	1.63e^{-9}	2.40e^{-9}	1.63e^{-9}	2.24e^{-9}	8.72e^{-9}		2.07
reference	1.11e^{-9}	1.57e^{-9}	1.63e^{-9}	1.62e^{-9}	2.24e^{-9}	1.03e^{-7}	1.08

Table 5.5 – Average rank of each source auralisation method in a free field environment and p-values obtained with a Wilcoxon signed-rank test when mixing all types of stimuli together. Gray cells depict the p-values which lead to a rejection of the null-hypothesis with a significance level of $\alpha = 0.05$. Redundant p-values are not shown.

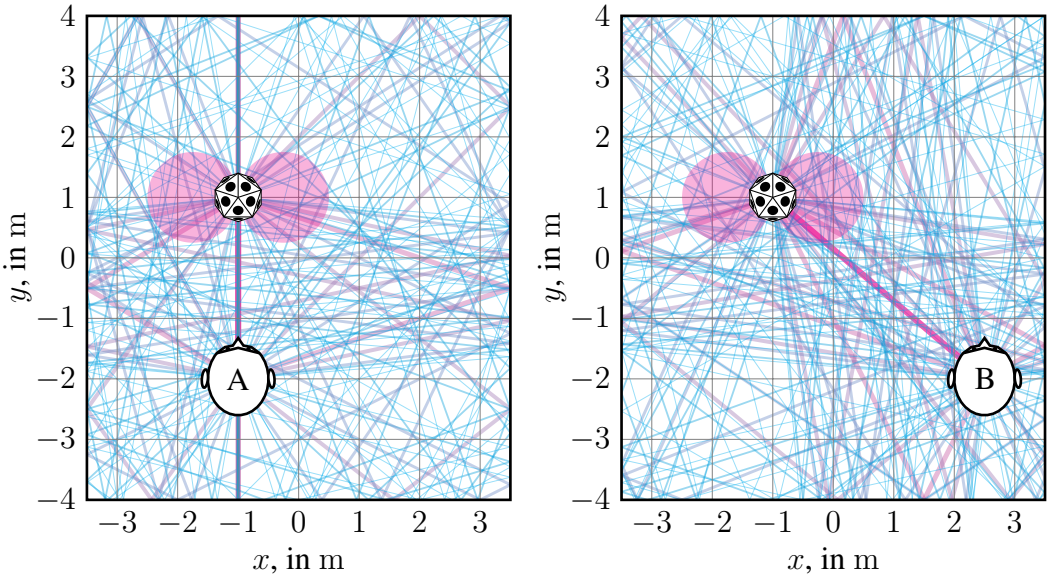
model. Additionally to those models, the primal signal as an omnidirectional source is also provided for comparison to all other techniques and is referred as Omni.

Two positions are investigated, namely directly in the zero's direction of the dipole with a distance of $3\,\text{m}$ (position A) and another one at a distance of about $4.6\,\text{m}$ and $40°$ from the dipole direction (position B). Both virtual scene are depicted in Fig. 5.6 and the rays up the 5^{th} order are represented with a width and color depending of the reflexion order. The resulting signal at the listener is computed with the help of the

room encoder from the IEM Plug-in Suite. The room has the dimension $7\,\mathrm{m} \times 8\,\mathrm{m} \times 6\,\mathrm{m}$, the first 100 reflections were computed and each reflections lead to a gain reduction of $1\,\mathrm{dB}$ and a bass and high frequency attenuation modelled by two first order shelving filters of $-2.5\,\mathrm{dB}$ @ $99\,\mathrm{Hz}$ and $-5\,\mathrm{dB}$ @ $8\,\mathrm{kHz}$ respectively.

The scenes are restituted binaurally in a static way using the Neumann KU100 Head Related Transfer functions (HRTF) based on the decoding strategy proposed in [SZH18]. Practically, this was computed using the *binaural decoder* VST plug-in, also from the IEM Plug-in Suite. As headphones a pair of Sennheiser HD650 was used.

The ratings are based on the overall listening experience, therefore the test person could assess either the spectral properties, temporal properties, the perceived localisation or any other spatial cues that seem important to him.



(a) Listener at position A

(b) Listener at position B

Figure 5.6 – Geometry of the virtual environment. The reflections are represented up to the 5^{th} order.

Six experts listeners from the IEM took part on the experiment has to rate (6 methods + 1 reference) $\times$ 4 stimulus types $\times$ 2 positions $\times$ 2 repetitions = 112 samples that resulted in a test of about $25\,\mathrm{min}$.

The confidence intervals for the median are depicted for all 4 stimulus type and positions in Figs. 5.7 and 5.8 respectively. The results of the Wilcoxon signed rank test, testing the hypothesis that all methods share the same average rank are presented in Table 5.6.

All groups appear to be significantly different, and the method proposed in this work (i.e. SDR-$L3$) deliver the best rank.

Interestingly, the increase of the order used for the direct SHT of the microphones (HI-$L3$ and HI-$L7$) did not lead to an amelioration of the restitution quality, despite the fact that the Sections 5.1 and 5.3.1 showed the opposite under free-field conditions.

An other interesting observation is that the omnidirectional restitution of the primal signal lead to a better rating from the participant when taking all positions into consideration. Actually, after a short informal comparison of the stimuli, even if the spatial cues of the omnidirectional source might not represent well the reference source, the accuracy of its spectral properties are preferred against the distorted spectrum of SH expansion of the microphone signals.

Both the sources with a zero-phase radiation filter (ZP-$L3$ and ZP-$L7$) offer a significantly better listening experience than Omni, HI-$L3$ and HI-$L7$, despite less accurate than SDR-$L3$.

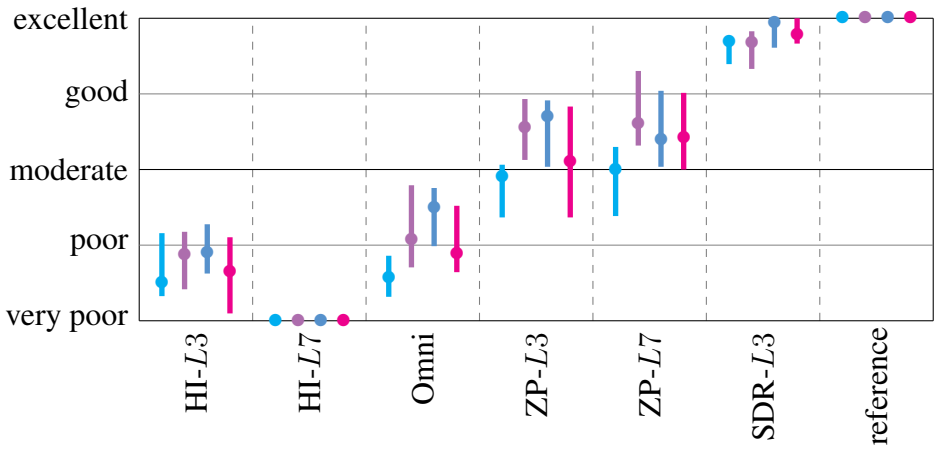


Figure 5.7 – Perceived quality of the radiated signal in reverberant environment for different source auralisation methods. The large points depict the estimated median and the whiskers depict 95% confidence intervals of the median. Pink noise, Speech, Guitar, Tablas.

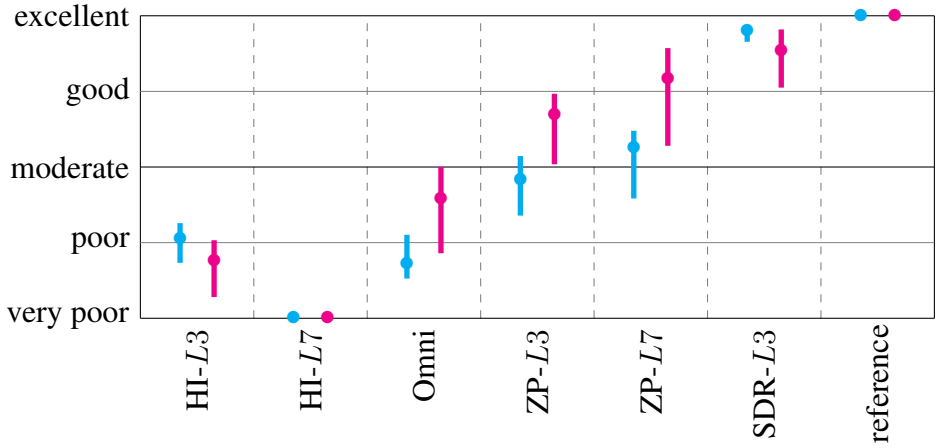


Figure 5.8 – Perceived quality of the radiated signal in reverberant environment for different positions. The large points depict the estimated median and the whiskers depict 95% confidence intervals of the median. Position A Position B.

Method	HI-$L3$	HI-$L7$	Omni	ZP-$L3$	ZP-$L7$	SDR-$L3$	rank
HI-$L3$							5.62
HI-$L7$	$5.55e^{-13}$						6.86
Omni	$4.21e^{-3}$	$4.13e^{-17}$					5.12
ZP-$L3$	$2.29e^{-16}$	$1.78e^{-17}$	$5.94e^{-12}$				3.78
ZP-$L7$	$8.87e^{-17}$	$1.78e^{-17}$	$7.04e^{-15}$	$4.65e^{-3}$			3.46
SDR-$L3$	$1.78e^{-17}$	$1.64e^{-17}$	$2.15e^{-17}$	$9.34e^{-17}$	$6.76e^{-15}$		2.01
reference	$1.73e^{-17}$	$1.74e^{-20}$	$1.78e^{-17}$	$2.60e^{-17}$	$3.81e^{-17}$	$5.45e^{-13}$	1.15

Table 5.6 – Average rank of each source auralisation method in a virtual reverberant environment and p-values obtained with a Wilcoxon signed-rank test when mixing all types of stimuli together. Gray cells depict the p-values which lead to a rejection of the null-hypothesis with a significance level of $\alpha = 0.05$. Redundant p-values are not shown.

Chapter 6

Conclusion and outlook

In this work, I proposed a new signal-processing framework in order to virtualise musical instruments by means of a surrounding spherical microphone array, by decomposing the source signal into a monophonic primal signal and a radiation filter. Perceptually, this method has proved to enable a highly qualitative virtualisation.

Hereby complicated centring algorithm can be avoided by simply replacing the signals phase by a more simplistic one. On the one hand, the primal signal can be estimated by retrieving the phase of a generalised mean of the microphones spectrograms, thus avoiding destructive interferences by using non-greedy algorithms such as Real-Time Phase Gradient Heap Integration. On the other hand, the radiation filter can be estimated by retrieving a simplified phase for each microphone and signal dependent-frequency regions using different strategies.

While other spatial interpolation algorithms such as the Modified Vector Base Amplitude Panning already appear convenient for avoiding spectral artefacts due to complicated phase patterns, the proposed global interpolation method has the benefit to offer a compact radiation signal in the form of a spherical wave spectrum, which can be useful for some auralisation techniques (e.g. with a variable directivity loudspeaker).

A simplification of my global method, requiring zero-phase radiation filter also yields perceptually satisfactory results, even though such a virtualisation require higher orders than the original source for a perceptually comparable result.

In a listening experiment, I have shown that the proposed method (primal signal estimated by means of RTPGHI + GSRTISI-LA and a 3^{rd} order radiation filter estimated region-wisely by solving an MLS problem with SDP) clearly outperformed other methods such as those involving the 3^{rd} and 7^{th} order zero-phase radiation-filter approximation or the 3^{rd} and 7^{th} order hyperinterpolation of the microphone signals

under both free field and reverberant conditions. In the free field case, the result are also significantly better than those obtained using MVBAP.

Of course the full potential of the proposed approach has not been completely explored, however this method offers a very promising alternative to other overly complicated acoustic centring-based algorithms.

Some improvement could consist of achieving a faster convergence for the spherical phase retrieval algorithm. The method providing time consistency of the radiation filter could be extended to complex spectral values, as it is now only implemented for binary phase (0 or π). Furthermore the compactness of radiation-filter IR could be optimised by a suitable phase simplification over frequency ("minimum-phase"). The spherical phase optimisation has not been tested for peak-region grouping; here some improvements could be undertaken in order to avoid erroneous microphone grouping.

Bibliography

[Bau11] R. Baumgartner, "Time domain fast-multipole translation for ambisonics," Master's thesis, Institute for Electronic Music and Acoustics (IEM), Graz, Austria, June 2011.

[BBE17] T. Bendory, R. Beinert, and Y. C. Eldar, "Fourier phase retrieval: Uniqueness and algorithms," 2017.

[BHPVR11] I. Ben Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic centering of sources measured by surrounding spherical microphone arrays," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2003–2015, Oct 2011. [Online]. Available: http://dx.doi.org/10.1121/1.3624825

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[BZW05] G. Beauregard, W. Zhu, and L. Wyse, "An efficient algorithm for real-time spectrogram inversion," in *8th Int. Conference on Digital Audio Effects (DAFX-05)*, Madrid, Spain, September 2005.

[DKS93] B.-I. Dalenbäck, M. Kleiner, and U. Svensson, "Audibility of changes in geometric shape, source directivity, and absorptive treatment - experiments in auralization," *AES: Journal of the Audio Engineering Society*, vol. 41, pp. 905–913, 11 1993.

[DZ10] D. Deboy and F. Zotter, "Acoustic center and orientation analysis of sound-radiation recorded with a surrounding spherical microphone array," in *Proceedings of the Int. Symp. on Ambisonics and Spherical Acoustics (Proceedings of the Int. Symp. on Ambisonics and Spherical Acoustics)*, M. Noisternig, Ed., Paris (Frankreich), 09 2010, procedure: peer-reviewed.

[DZ11] ——, "Tangential intensity algorithm for acoustic centering," in *Fortschritte der Akustik - DAGA 2011 (DAGA 2011)*, Düsseldorf (Deutschland), 03 2011, procedure: without peer reviewing.

[GB08] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

[GB14] ——, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, March 2014.

[GL84] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time fourier transform," *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*, vol. ASSP-32, no. 2, April 1984.

[GS08] V. Gnann and M. Spiertz, "Comb-filter free audio mixing using STFT magnitude spectra and phase estimation," in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, 2008.

[GS10] ——, "Improving RTISI phase estimation with energy order and phase unwrapping," in *13th Int. Conference on Digital Audio Effects (DAFx-10)*, Graz, September 2010.

[HAD06] R.-M. Hom, V. Algazi, and R. Duda, "High-frequency interpolation for motion-tracked binaural sound," *Audio Engineering Society - 121st Convention Papers 2006*, vol. 3, pp. 1166–1178, 01 2006.

[Hoh09] F. Hohl, "Kugelmikrofonarray zur Abstrahlungsvermessung von Musikinstrumenten," Master's thesis, Institute for Electronic Music and Acoustics (IEM), November 2009.

[Hol14] C. Hollomey, "Real time spectrogram inversion," Master's thesis, Institute for Electronic Music and Acoustics, University of Music and dramatic Arts Graz, March 2014.

[Kas06] P. W. Kassakian, "Convex approximation and optimization with applications in magnitude filter design and radiation pattern synthesis," Ph.D. dissertation, Electrical Engineering and Computer Sciences University of California at Berkeley, May 2006.

[LMS$^+$10] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[LV83] S. P. Lipshitz and J. Vanderkooy, "In-phase crossover network design," in *Audio Engineering Society Convention 74*, October 1983.

[Mey09] J. Meyer, *Acoustics and the Performance of Music*. Springer, 2009.

[Mit16] R. Mittmannsgruber, "Adaptive MISO-Filter zur möglichst konstruktiven Mehrkanalsignalüberlagerung," Master's thesis, Institute for Electronic Music and Acoustics (IEM), 2016.

[NNZ10] C. Nachbar, G. Nistelberger, and F. Zotter, "Listening to the direct sound of musical instruments in freely adjustable surrounding directions," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, France, Mai 2010.

[PBS17] Z. Pruša, P. Balazs, and P. L. Søndergaard, "A noniterative method for reconstruction of phase from STFT magnitude," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 1154–1164, May 2017.

[PR17] Z. Pruša and P. Rajmic, "Towards high quality real-time signal reconstruction from STFT magnitude," *IEEE SIGNAL PROCESSING LETTERS*, 2017.

[PS16] Z. Pruša and P. L. Søndergaard, "Real-time spectrogram inversion using phase gradient heap integration," in *Proc. Int. Conf. Digital Audio Effects (DAFx-16)*, 2016, pp. 17–21.

[Pul97] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

[SS15] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, 2015.

[Süs11] S. Süss, "Aufinden von "Ursignalen" aus Aufnahmen umgebender kugelförmiger Mikrofonanordnungen," Toningenieur Projekt, June 2011.

[SV15] N. R. Shabtai and M. Vorländer, "Acoustic centering of sources with high-order radiation patterns," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 1947–1961, Apr 2015.

[SZH18] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *DAGA 2018 München*, 2018.

[WA80] G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," *The Journal of the Acoustical Society of America*, vol. 68, no. 2, pp. 404–411, 1980. [Online]. Available: https://doi.org/10.1121/1.384751

[WDM15] I. Waldspurger, A. D'Aspremont, and S. Mallat, "Phase recovery, MAXCUT and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1, pp. 47–81, February 2015.

[Wil99] G. Williams, *Fourier Acoustics: Sound radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[ZBW06] W. Zhu, G. Beauregard, and L. Wyse, "Real-time iterative spectrum inversion with look-ahead," *2006 IEEE International Conference on Multimedia and Expo*, pp. 229–232, 2006.

[ZF12] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *Journal of Audio Engineering Society*, vol. 60, no. 10, p. 807, October 2012.

[ZFZ19] M. Zaunschirm, M. Frank, and F. Zotter, "Perceptual evaluation of variable-orientation binaural room impulse response rendering," in *AES International Conference on Immersive and Interactive Audio (to be published)*, York, 2019.

[Zot09] F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays," Ph.D. dissertation, Institute of Electronic Music and Acoustics, 2009.

Chapter A

IEM microphone array geometry

λ	φ	ϑ
1	0.00	165.00
2	90.03	89.88
3	78.08	147.79
4	−167.16	28.44
5	164.49	49.17
6	14.28	67.70
7	176.93	120.13
8	−107.65	88.46
9	68.12	103.55
10	65.14	51.09
11	137.41	166.12
12	136.43	23.02
13	96.90	37.66
14	−98.43	161.01
15	−52.34	143.80
16	−101.75	58.86
17	−169.14	98.77
18	−179.07	74.28
19	55.55	124.80
20	−43.09	45.64
21	154.73	141.37
22	−11.67	53.12
23	32.88	142.88
24	−58.48	89.15
25	143.99	117.91
26	−62.76	117.23
27	−153.16	78.95
28	−85.88	104.31
29	150.47	71.12
30	−14.16	25.69
31	−151.23	119.03
32	161.65	95.11
33	−33.21	96.74
34	−34.32	122.12
35	−36.49	71.92
36	−83.61	79.09
37	−116.75	114.24
38	−9.42	109.50
39	−105.14	7.70
40	20.91	117.53
41	114.6	105.90
42	−134.27	95.61
43	115.16	78.25
44	−91.78	131.07
45	26.52	44.55
46	−79.63	34.32
47	−129.24	68.39
48	90.34	120.00
49	−124.43	39.43
50	−126.76	138.3
51	128.95	52.70
52	−158.41	52.65
53	15.32	93.09
54	65.72	78.34
55	41.24	69.42
56	−8.35	81.27
57	93.99	62.89
58	−5.55	137.63
59	41.35	97.80
60	52.29	21.81
61	136.13	92.25
62	−67.33	60.37
63	−164.35	147.58
64	116.94	135.04

Table A.1 – Geometry of the IEM 64 microphones array in spherical coordinates, the angles are given in degrees