Stefan Riedel

Distance-coded Ambisonics Formats and their Reproduction on Headphones and Loudspeaker Arrays

Project Thesis

University of Technology Graz
University of Music and Performing Arts Graz

Institute of Electronic Music and Acoustics
Head: O.Univ.-Prof. Mag.art. Dipl.-Ing. Dr.techn. Robert Höldrich

Supervisor: M.Sc. Julien De Muynke, Eurecat Barcelona

Barcelona, June 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.

Graz, ____________________ ____________________

Date Signature

Eidesstattliche Erklärung[1]

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ____________________ ____________________

Datum Unterschrift

[1]Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008; Genehmigung des Senates am 1.12.2008

Abstract

This work introduces distance-coded Ambisonics formats and their reproduction on headphones and loudspeaker arrays.

The first simple and practically motivated format proposes two Ambisonic signals, a far-field and a near-field signal, to which sounds are distributed according to a distance parameter at encoding stage. In binaural decoding this enables the application of near-field HRTFs with inherent binaural cues which cannot be applied at encoding stage, for example the frequency-dependent increase in interaural level differences compared to far-field HRTFs. Blending between two Ambisonic reverberation patterns (modeled or measured DRIR) is combined with a physically meaningful level attenuation to achieve a plausible distance effect that includes a change in the direct-to-reverberant sound energy ratio. Compatibility with loudspeaker arrays is given by summation of the two Ambisonic signals after introducing level differences and the two reverberation patterns to retain a relative distance effect.

An efficient and more accurate way to render distance is to restrict the effect to the horizontal plane. Therefore, a second format that interprets negative elevation as the distance of a horizontal source is proposed. In binaural reproduction, this format allows for a high spatial resolution in the precomputation of distance-dependent HRTFs and early reflections, applied at decoding stage. Moreover, this format could motivate future research on loudspeaker systems that employ horizontal sound field synthesis (rendering of near-field sources) combined with AllRAD for elevated sources.

Contents

1 Introduction

The Ambisonics technology is evolving as the de facto standard for spatial audio and music due to its flexibility in terms of reproduction both on non-uniform loudspeaker arrays and on headphones. Especially the the rise of 360° video streaming and AR/VR applications required a sound scene representation that allows for an efficient rotation corresponding to the user's head movements.

In practice, the distance of sound sources is often considered to be either inherent in Ambisonic microphone recordings, or to be achieved by adjusting the level or reverberation of a source at encoding stage in a studio production. While the assumption about recordings seems justified, automating level and reverb effect for each source of a studio production can become quite tedious. This already motivates a more systematic way of encoding a distance intention for a sound source. Moreover, distance encoding obviously allows for new decoding strategies that can now include near-field effects, especially in the binaural reproduction, where applications such as movies, video games and VR/AR experiences profit from a realistic rendering of nearby sound sources.

In fact, the idea of introducing distance-coding and near-field compensation to Ambisonics was pioneered by Jerome Daniel, that proposed compensation filters at encoding and decoding stage (NFC-HOA) [6]. However, a physical sound field synthesis to accurately render a curved wavefront for a near-field source and a plane wave for a far-field source requires unpractically high amounts of loudspeakers and ambisonic channels to be stored and transmitted, the latter being a problem even in the case of binaural reproduction.

Moreover, filtering sound at the encoding stage has the obvious drawback of imprinting the compensation filter characteristics to the initial sound source and apart from that requires further adaptation to the actual loudspeaker array radius at decoding stage, which might not be ensured in any reproduction situation. Furthermore, the proposed compensation assumes a uniform-radius

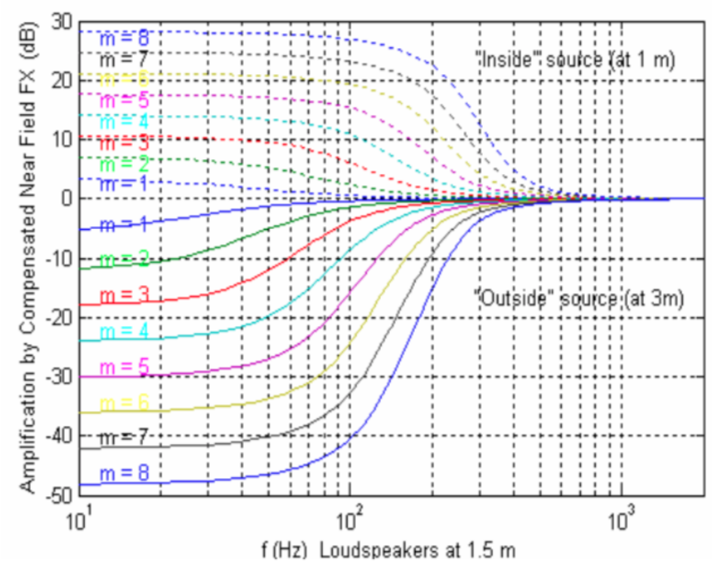


Figure 1.1: J. Daniel's proposed Near-Field Compensation Filters that equalize the reproduction array radius in relation to the intended source distance. [6]

loudspeaker array, which is often not the case in reproduction rooms. Additionally acoustic room modes can corrupt the low-frequency corrections that initially assumed free-field conditions. NFC-HOA was lately revisited and combined with Vector-Base Amplitude Panning (VBAP) to achieve Near-Field Compensation for angularly non-uniform layouts [12, 9].

The current ambisonic production workflow (see e.g. IEM Plug-In suite) is based on frequency independent encoding and decoding at reasonable orders (N≤7) achieving comparable results on arbitrary loudspeaker arrays.

Therefore this project thesis explores the potential of new distance-coded Ambisonics formats that allow to encode the intention of a source distance in addition to the intended direction of a sound, postponing the actual rendering of distances to the stage of the Ambisonic scene representation.

The first format presented in chapter 2 introduces the concept of two Ambisonic buses (near-field and far-field), to which Ambisonic signals (*azimuth,elevation*) are distributed according to an additional *distance* parameter of the encoder (*DistanceEncoder*, see Fig. 2.4). Section 2.2 discusses the binaural reproduction of this full-sphere distance format. Apart from monaraul distance cues (level change, direct-to-reverberant sound ratio) this format now allows to account for additional frequency-dependent interaural level differences (ILDs) of near-field HRTFs included in the binaural decoding. For loudspeaker arrays a downmix (addition) of the two streams can be applied, after introducing dynamics in level and direct-to-reverberant sound by applying different DRIRs (Directional Room Impulse Responses) on the two streams. These basic psychoacoustic distance effects (level and room effect) are termed All-round Ambisonic Distance effects. The resulting ambisonic stream is compliant with the current standard and may be distributed to existing loudspeaker decoders as well as far-field binaural decoders, achieving a relative distance effect based on psychoacoustic cues.

In an alternative format, see chapter 3, a continous set of horizontal distances is available by interpreting negative elevation as the distance of a horizontal source, which is meaningful in the sense of a horizontal wave-field synthesis loudspeaker array reproduction combined with All-round Ambisonic Decoding (AllRAD) [15] for elevated sound sources.

This format has also been used to create a prototype binaural decoder that renders realistic near-field effects based on a set of near-field BRIRs mapped to the negative hemisphere. The decoders (*mcfx convolver-config* files and filters) are downloadable[1] and ready to be tested with standard AmbiX-format encoder plug-ins. For a quick listening experience the interested reader finds rendered binaural demos on a *soundcloud* page[2].

[1] `https://github.com/stefanriedel/NearField-BinauralDecoder`

[2] `https://soundcloud.com/stefan-mario-riedel-1/sets/ambisonic-binaural-near-field-demos`

2 Full-Sphere Distance: Near-Field and Far-Field Ambisonic Signals

2.1 Format Definition and Distance Encoder Workflow

A full-sphere distance format in Ambisonics can be defined by multiple (theoretically n) Ambisonic signals representing different distances.

In practice, it shows that the minimum case of $n = 2$ Ambisonic signals is enough to efficiently render distance effects in binaural and loudspeaker array reproduction. The first Ambisonic bus shall be defined as the near-field bus ($r = 0$) and the second bus as the far-field bus ($r = 1$), with Ambisonic signals being distributed continously according to a distance parameter at encoding stage ($0 \leq r \leq 1$, see Fig. 2.1).

A first prototype of a distance encoder was implemented in Max/MSP (Fig. 2.3). It encodes the source direction based on azimuth and elevation angles and blends the Ambisonic signal between a near-field and a far-field bus, according to a relative distance parameter. Subsequently, distance processing is done at the scene stage as Allround Ambisonic Distance effects, as seen in Fig. 2.2. Additionally, specific near-field processing (near-field HRTFs in binaural or near-field compensation filters for array reproduction) can be applied. After the processing of the two Ambisonic signals, they can be summed to a single Ambisonic signal for standard decoding to loudspeaker arrays or for storage in a multichannel WAV-File.

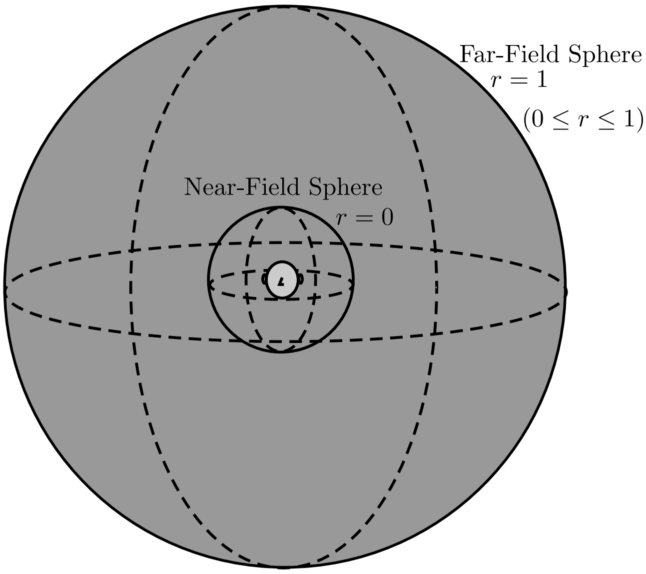


Figure 2.1: Rendering distance by blending between a near-field and far-field Ambisonic signal.

2.2 Binaural Reproduction with Near-Field and Far-Field HRIRs

The binaural synthesis of spatial audio uses so-called Head-Related Impulse Responses (HRIRs) and their frequency domain representation Head-Related Transfer Functions (HRTFs) that contain perceptual cues like Interaural Level Differences (ILDs) and Interaural Time Differences (ITDs) and other spectral cues that depend on the incoming direction and distance of sound [3].

2.2.1 Near-Field vs. Far-Field HRIR Properties

Near-Field HRTFs contain inherently different properties compared to Far-Field HRTFs [5] which are necesarry to realistically auralize nearby sound sources. While recent studies suggest that source distance estimation is not significantly affected by near-field cues measured under free-field conditions [1], older studies by Brungart et. al. [4] suggest the contrary.

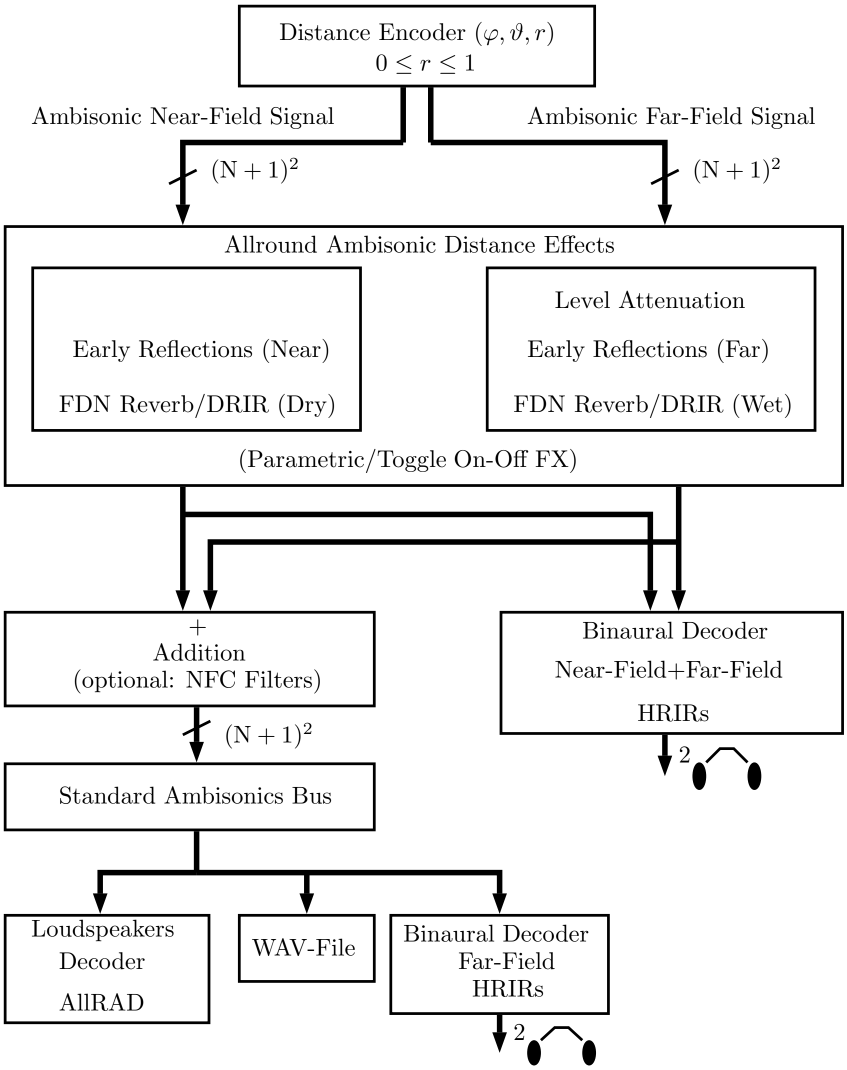


Figure 2.2: The signal flow for the Near-Field/Far-Field Ambisonic Signals approach.

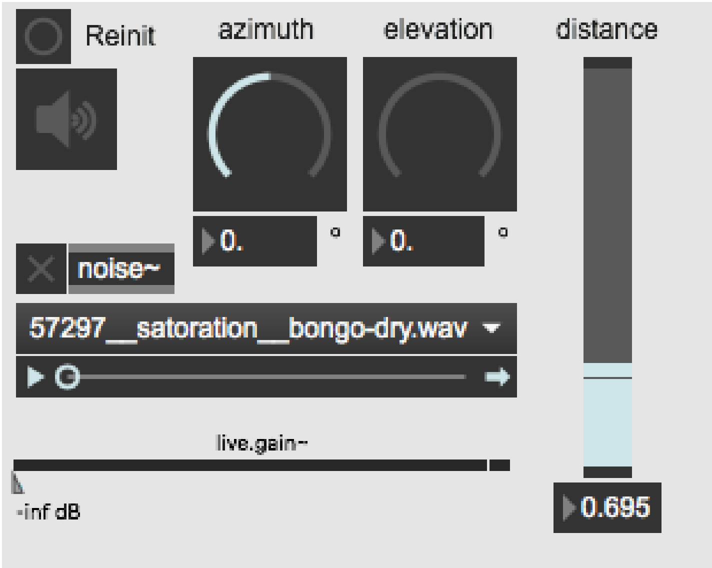


Figure 2.3: Prototype of the DistanceEncoder implemented in Max/MSP. It outputs two ambisonic signals and the source is blended between them according to the distance parameter on the GUI interface.

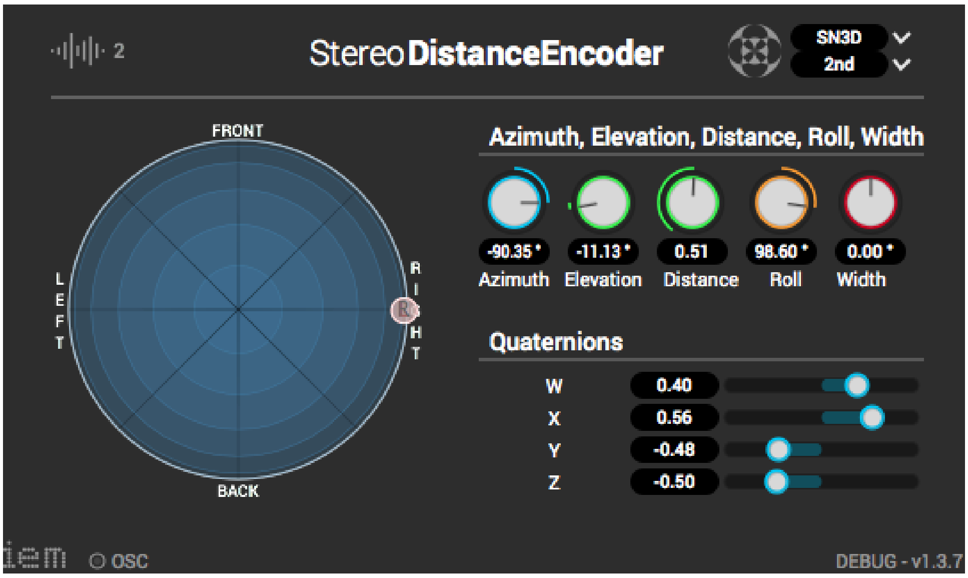


Figure 2.4: *StereoDistanceEncoder* VST plug-in based on the *IEM StereoEncoder* plug-in by Daniel Rudrich. It outputs two ambisonic signals and the source is blended between them according to the distance parameter on the GUI interface.

In practice, we find examples ranging from dummy-head recordings to near-field BRIR[1]-based rendering which prove that realistic proximity effects can be achieved with binaural reproduction techniques.

Monaural Cues

A monaural cue is a distance cue that affects both ears equally, such as the generally increased volume of a nearby sound source.

Binaural Cues

Binaural cues are based on differences between the two ear signals that become more and more pronounced as the sound source gets close to the listener's head. Studying a database of free-field measured circular near-field HRTFs [1, 8], we find an increase in the overall interaural level difference, when averaging over a range of frequencies. This increase is frequency dependent and especially pronounced at high-frequencies due to the acoustic shadow of the human head, see Figs. 2.5 to 2.9.

The dataset of Near-Field HRTFs contains a minimum distance of 0.5 meters for a full-spherical 2702-node Lebedev grid. Modeling near-field cues and extrapolating to closer distances has been demonstrated [2] and may be desired in some use-cases. Extrapolation based on a simplified head model (sphere model) seems appropriate to exaggerate near-field cues without further emphasizing spectral notches and ripples that would hinder seamless blending between a far-field and near-field HRTF set in the specific rendering approach of this work.

2.2.2 Room Model for Early Reflections

It has been shown that externalisation is significantly improved in binaural decoding by simulating the early reflections caused by the walls of a room [10]. With the near-field/far-field signal approach, we precompute two ambisonic

[1] Binaural Room Impulse Response, see Sec. 3.2

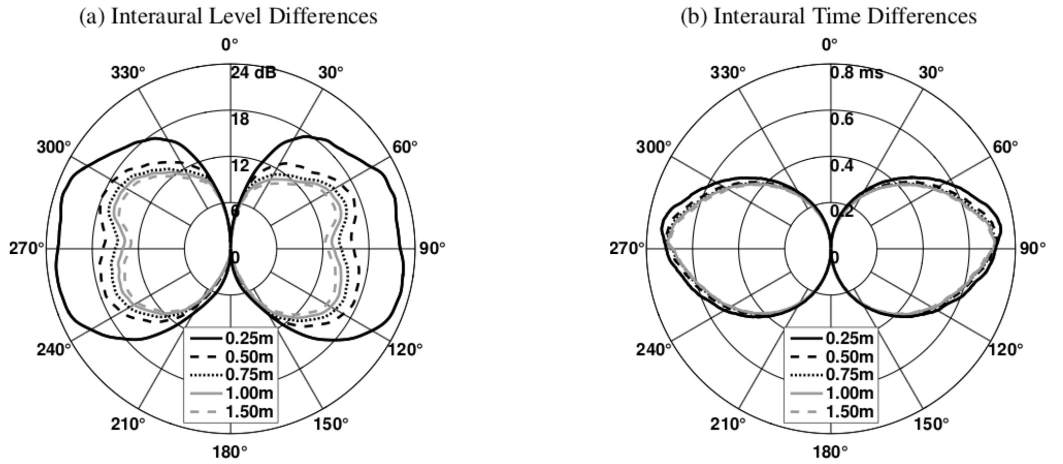


Figure 2.5: Interaural Level Differences escalate at distances of 25 cm and below. The change in Interaural Time Differences seems to be comparatively small, although the so called 'paralax effect' is often mentioned in literature. These plots suggest that it is justified to blend a set of far-field HRTFs with a set of near-field HRTFs and thereby create intermediate distances (as proposed with the near-field/far-field signal approach). Plots taken from (cite)

directional room impulse responses (DRIRs) corresponding to two distances. To avoid comb filtering when simply blending these two responses in our approach, we neglect the time of arrival of the direct sound (zero delay direct sound) and simply blend the reflection patterns that differ in their time of arrival, which is an important cue for distance perception.

2.2.3 Magnitude Least-Squares Decoding

When transforming high resolution free-field HRIR sets into the Ambisonics domain, one typically faces a severe loss of high frequencies in the frontal zone. This is due to rapid phase changes and would require Ambisonic orders of up to N = 35. Recently proposed methods mitigate this effect by simplifying the phase above a frequency f_c , where localization is achieved by the HRTF magnitude only [11].

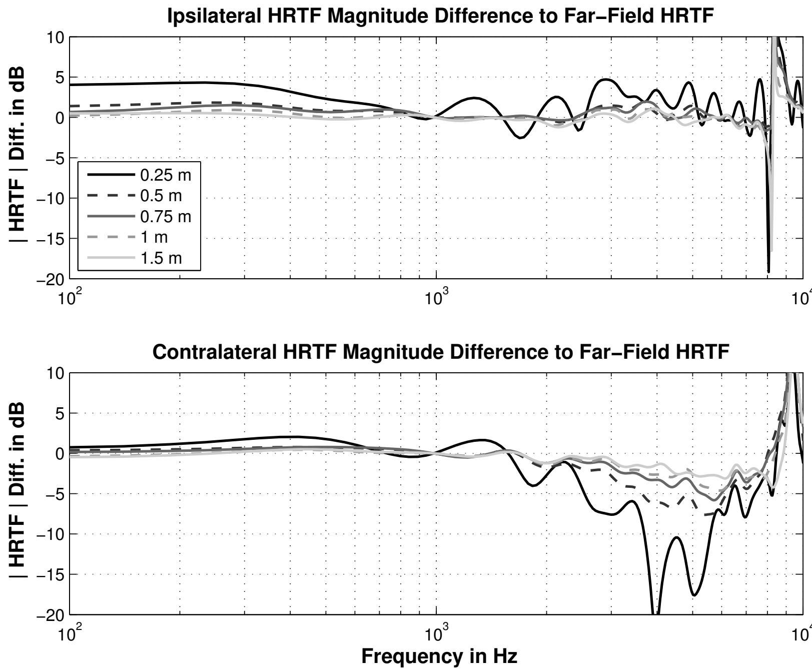


Figure 2.6: The increase in ILD for close lateral sources is especially pronounced above frequencies of 2 kHz, when the head diameter is larger than the wave length (acoustic shadowing). The HRTFs are normalized at a frequency of 1 kHz to expose the frequency dependent behaviour. The source direction is azimuth $\varphi = 90°$, elevation $\vartheta = 0°$ (lateral source).

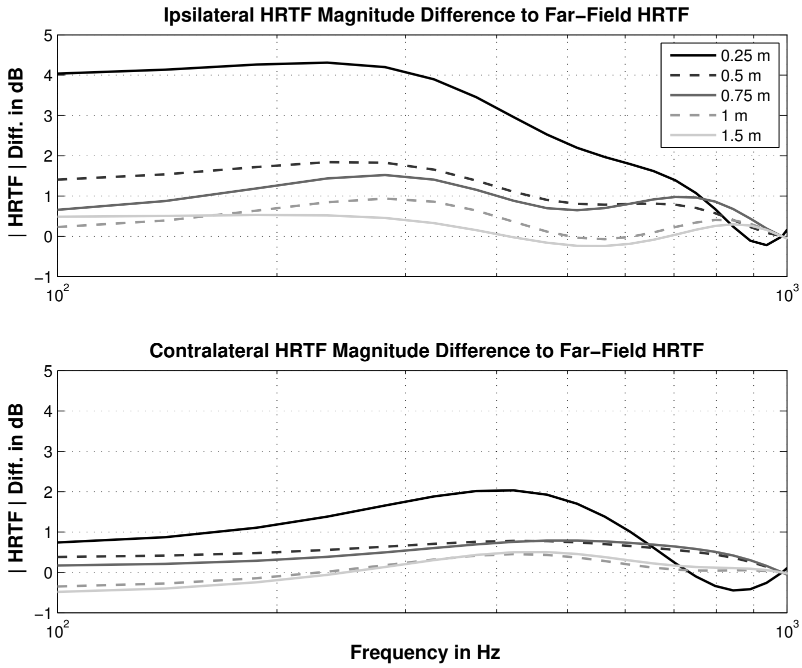


Figure 2.7: Due to the curvature of the wavefront a significant boost of low frequencies occurs for near-field sources. This bass boost must be considered a binaural cue, as it is more pronounced on the ipsilateral ear. The HRTFs are normalized at a frequency of 1 kHz to expose the frequency dependent behaviour. The source direction is azimuth $\varphi = 90°$, elevation $\vartheta = 0°$ (lateral source).

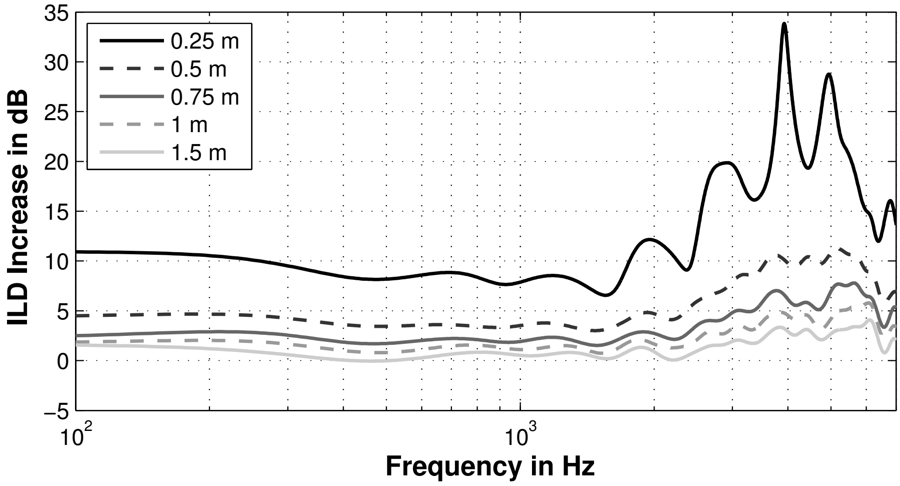


Figure 2.8: This plot shows the frequency dependent increase in Interaural Level Differences for nearby sound sources. The source direction is azimuth $\varphi = 90°$, elevation $\vartheta = 0°$ (lateral source).

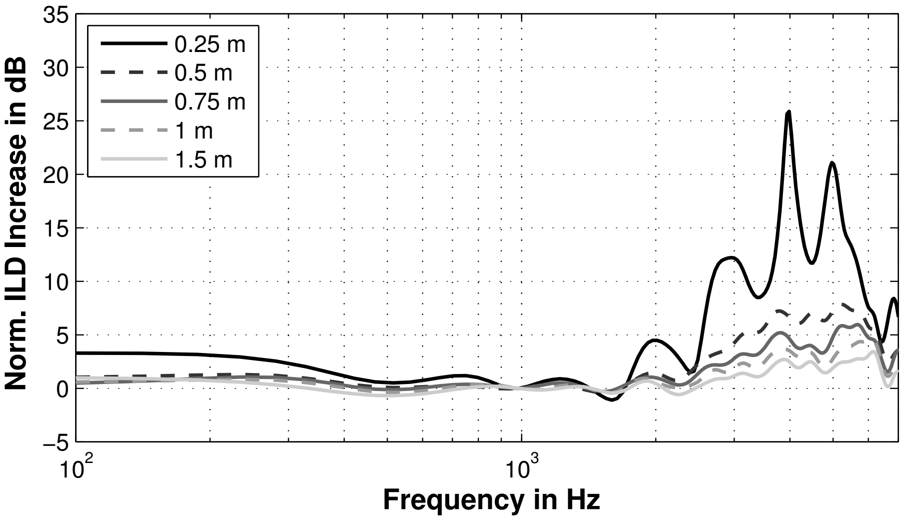


Figure 2.9: ILD increase curves normalized at a frequency of 1 kHz. The source direction is azimuth $\varphi = 90°$, elevation $\vartheta = 0°$ (lateral source).

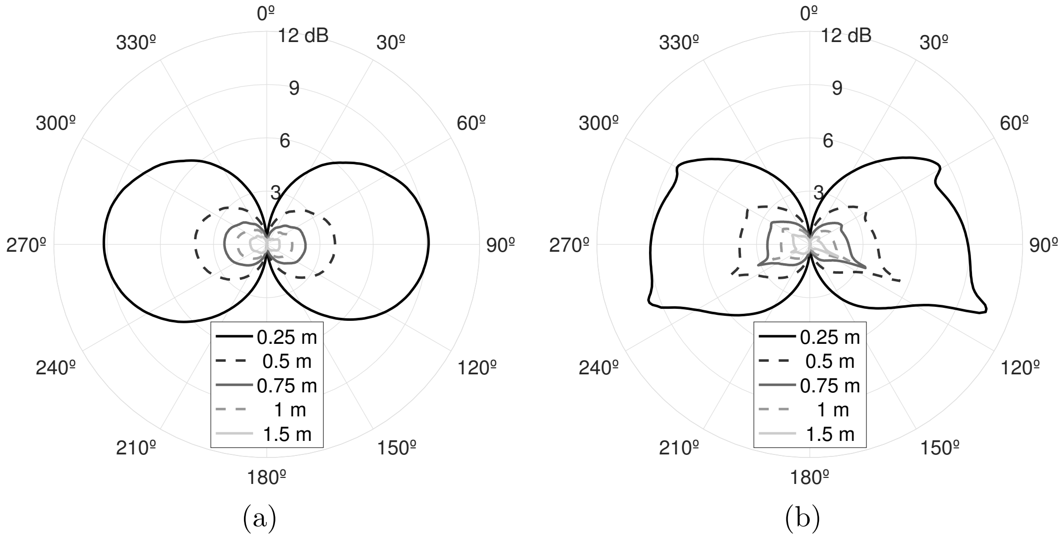


Figure 2.10: Increase in Interaural Level Differences (far-field HRTF as reference). The ILD increase in dB is averaged over all frequency bins up to 1 kHz (a) and 2 kHz (b). Up to frequencies of 1 kHz the head can be approximated by a rigid sphere and the polar diagram is symmetric.

For frequencies $f < f_c = 2$ kHz:

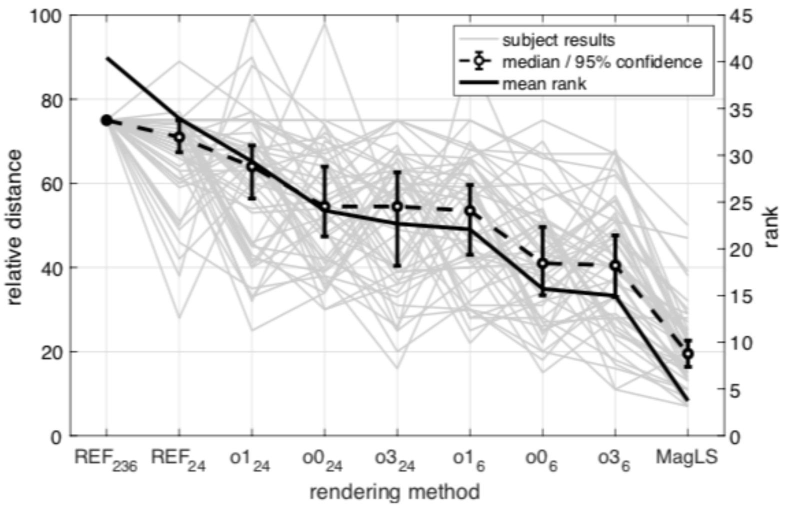
$$\hat{\boldsymbol{h}}_{\mathrm{SH},k} = (\boldsymbol{Y}_{\mathcal{M}}^{\mathrm{T}} \boldsymbol{Y}_{\mathcal{M}})^{-1} \boldsymbol{Y}_{\mathcal{M}}^{\mathrm{T}} \, \boldsymbol{h}_{\mathcal{M},k} \tag{2.1}$$

For frequencies $f \geq f_c = 2$ kHz:

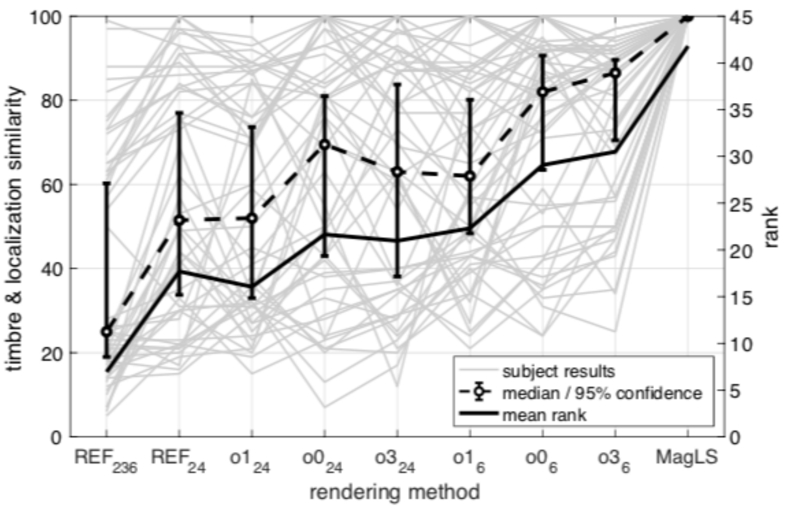
$$\hat{\boldsymbol{\phi}}_{\mathcal{M},k-1} = \angle(\boldsymbol{Y}_{\mathcal{M}}^{\mathrm{T}} \, \hat{\boldsymbol{h}}_{\mathrm{SH},k-1}) \tag{2.2}$$

$$\hat{\boldsymbol{h}}_{\mathrm{SH},k} = (\boldsymbol{Y}_{\mathcal{M}}^{\mathrm{T}} \boldsymbol{Y}_{\mathcal{M}})^{-1} \boldsymbol{Y}_{\mathcal{M}}^{\mathrm{T}} \big[\, |\boldsymbol{h}_{\mathcal{M},k}| \, e^{\mathrm{i}\hat{\boldsymbol{\phi}}_{\mathcal{M},k-1}} \,\big] \tag{2.3}$$

, where $\boldsymbol{h}_{\mathcal{M},k} = [H(k,\Omega)]_{1\ldots\mathcal{M}}^{\mathrm{L,R}}$ denotes the k-th FFT-bin of the original HRTF set of $\mathcal{M} = 2702$ nodes and $\boldsymbol{Y}_{\mathcal{M}}$ is the matrix of spherical harmonics sampled at these node points. By discarding the original phase and replacing it by a phase retrieved from the Ambisonic subspace representation (e.g. order $\mathrm{N} = 3$), it is possible to retain the magnitude response of the original HRTF set (see Figure 2.13). As a measure of similarity the composite loudness level error



(a) distance; 0 *very close*, 100 *further away*, 75 *same as reference*



(b) timbre/localization; 0: *very different*, 100 *identical* (reference)

Figure 2.11: Listening Experiments by Rudrich/Frank[10] show that early reflections give a sense of distance and externalization, especially high numbers of simulated reflections (method oN_i: N-th order ambisonic directivity of reflections, i simulated early reflections). The directivity order seems not to effect the localization accuracy. In this work here, apart from the far-field reflections, a near-field reflection pattern is computed that perceptually achieves a closer distance than a dry MagLS decoding and expands the distance range available.

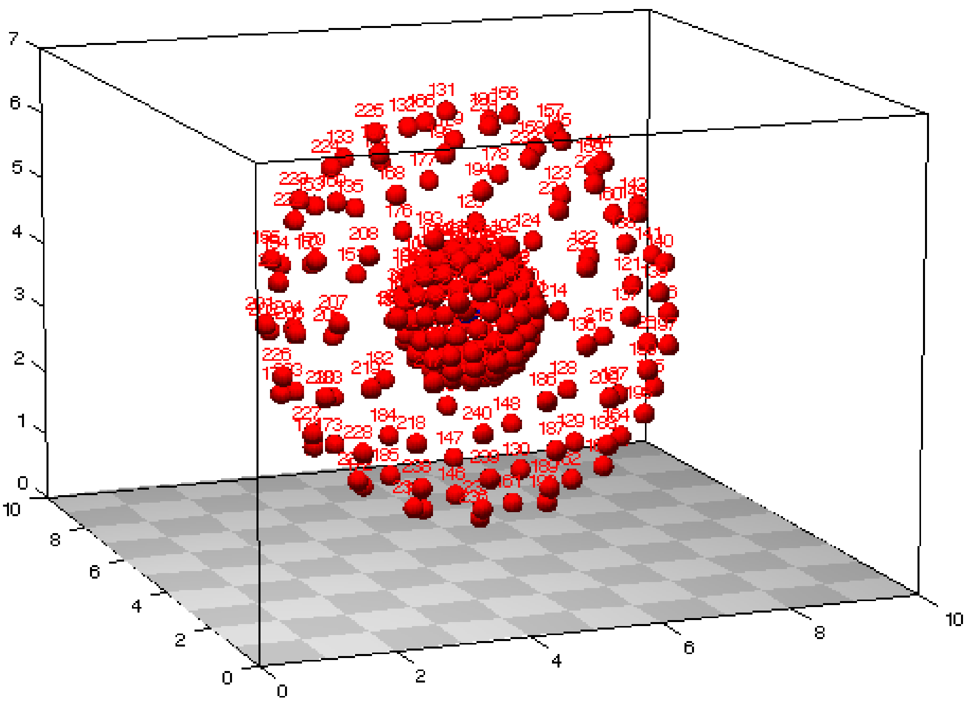


Figure 2.12: Early Reflections are precomputed for two spheres (near- and far-field) of 120 t-design points and applied at decoding stage as Ambisonic DRIRs. Plot created with the McRoomSim toolbox. Early reflection DRIRs were recorded from an OSC-controlled IEM RoomEncoder plug-in. A reflection attenuation of -6 dB decreases the sound colouration and the direct path was set to 'zero delay' to avoid strong comb filtering when blending the two patterns.

(CLL error e_{CLL}) is defined as follows:

$$\mathrm{CLL}(H(k,\Omega)) = 10 \cdot \log\left(|H(k,\Omega)|^2 + |H(k,\Omega')|^2\right) \tag{2.4}$$

$$e_{\mathrm{CLL}}(k,\Omega) = \mathrm{CLL}\big(\hat{H}(k,\Omega)\big) - \mathrm{CLL}\big(H(k,\Omega)\big) \tag{2.5}$$

, with the mirrored direction angles $\Omega' = (-\varphi, \theta)$, assuming symmetric HRTFs.

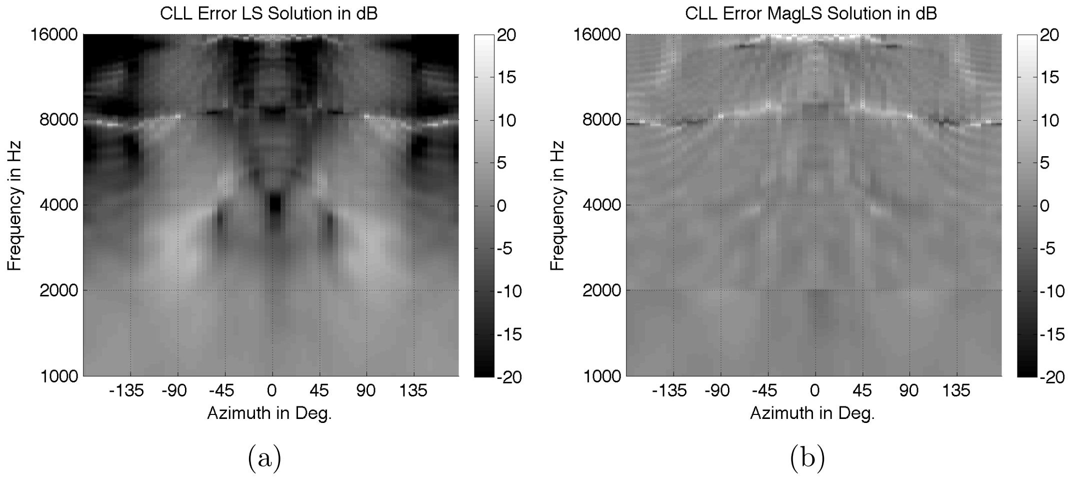


Figure 2.13: The CLL error e_{CLL} for the linear Least-Squares (LS) Solution (a) and a two-band solution that applies a Magnitude-Least-Squares (MagLS) above $f_c = 2$ kHz (b), both for an ambisonic order N=3 when transforming the TH Koeln HRIR Set with L=2702 measurement points into the ambisonic domain. The MagLS prevents the loss of high frequencies by simplifying the phase, especially for the frontal zone ($\varphi = -45° \ldots 45°$). Computed for horizontal source directions ($\varphi = -180° \ldots 180°$, $\theta = 0°$).

Additionally a diffuse-field covariance constraint can be defined that results in a 2-by-2 filter matrix and equalizes a lack of diffuseness when decoding with low orders N< 3. [16, 14].

2.3 Loudspeaker Array Reproduction through Allround Ambisonic Distance Effects

When working with the Ambisonics technology, compatibility between loudspeaker and headphones playback is essential and a central idea of the spatial

Figure 2.14: The production studio at Eurecat Barcelona was used to test the Allround Ambisonic Distance effect. Reproduction rooms can be 'virtually expanded' by dynamic Ambisonic reverberation simulating distant sources in much larger rooms than the actual studio or reproduction venue. The same idea applies to free-field reproduction at music festivals with multichannel loudspeaker systems.

audio format. The de facto standard of Allround Ambisonic Decoding (AllRAD) [15] as robust Ambisonic amplitude panning motivates to achieve at least a relative distance effect when rendering the full-sphere distance-coded format into a single Ambisonic signal for loudspeaker playback. The basic idea is illustrated in Fig. 2.2. The two Ambisonic signals are processed with different amplitude factors to simulate the sound pressure decay and additionally dynamic Ambisonic reverberation can be applied by means of two 'FDN-Reverb' plug-ins on the two buses. This simulates the increased diffuseness of a source moving away from the listener in a room (see Fig. 2.15).

The amount of these Allround Ambisonic Distance effects can be adjusted before adding the two signals in the final Ambisonics bus that is used to feed loudspeakers by AllRAD. Obviously, also a headphones reproduction that uses far-field HRTFs only is perfectly viable and will achieve a distance effect.

Apart from these basic psychoacoustically-motivated distance effects, near-field compensation filters could be applied to the two signals, before blending them together [12]. This would ideally allow to reproduce plane waves for sources that are encoded into the far-field Ambisonic signal. At least, it could serve as a spectral equalization of the bass-boost effect experienced by listeners in the

Figure 2.15: A simple way to create dynamic Ambisonic reverberation is to use two 'FDNReverb' plug-ins with different 'Dry/Wet' settings to simulate increased diffuseness when moving a source away from the listener. The room settings 'Room Size' or 'Rev. Time' can be set equal to achieve a coherent scene-based room effect. Additionally the two buses should exhibit level differences before being mixed on the final Ambisonics bus.

near-field of the reproduction loudspeakers.

3 Horizontal Distance: The Projection-based Approach

3.1 Format Definition and Workflow

This format originates from the idea to combine horizontal wave field synthesis (WFS) with AllRAD for elevated sources. By interpreting negative elevation as the distance of a horizontal source, one is able to render a continous set of distances in the horizontal plane.

3.2 Binaural Reproduction based on Near-Field and Far-Field BRIRs

3.2.1 Measurement Setup and Data Sets

A set of near-field BRIRs was measured in Eurecat's production studio (see Fig. 2.14). The measurement positions are depicted in Fig. 3.3. The horizontal near-field measurements were complemented with measurements of the loudspeaker cupula for far-field and elevated sources.

3.2.2 Virtual Loudspeaker Decoding using the Projection-based Format

As a decoding strategy for the BRIRs we use the traditional virtual loudspeaker approach, however interpolating the measurements to a denser set of a t-design

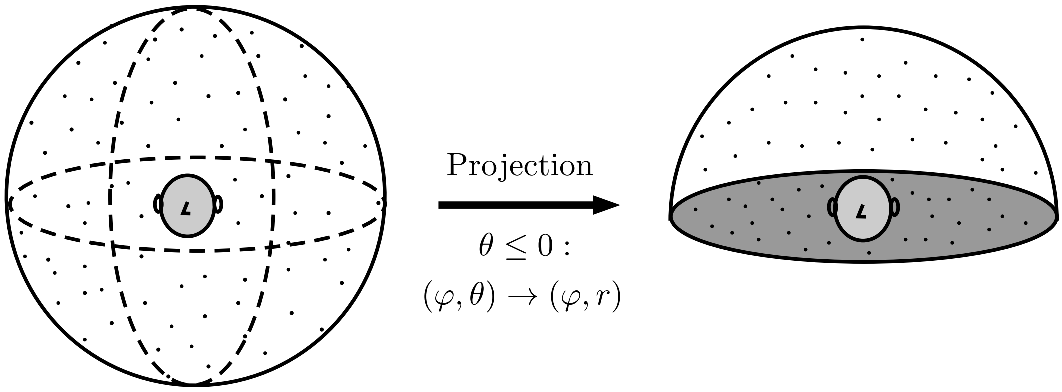


Figure 3.1: Rendering horizontal distance by a projection of the negative hemisphere into the horizontal plane.

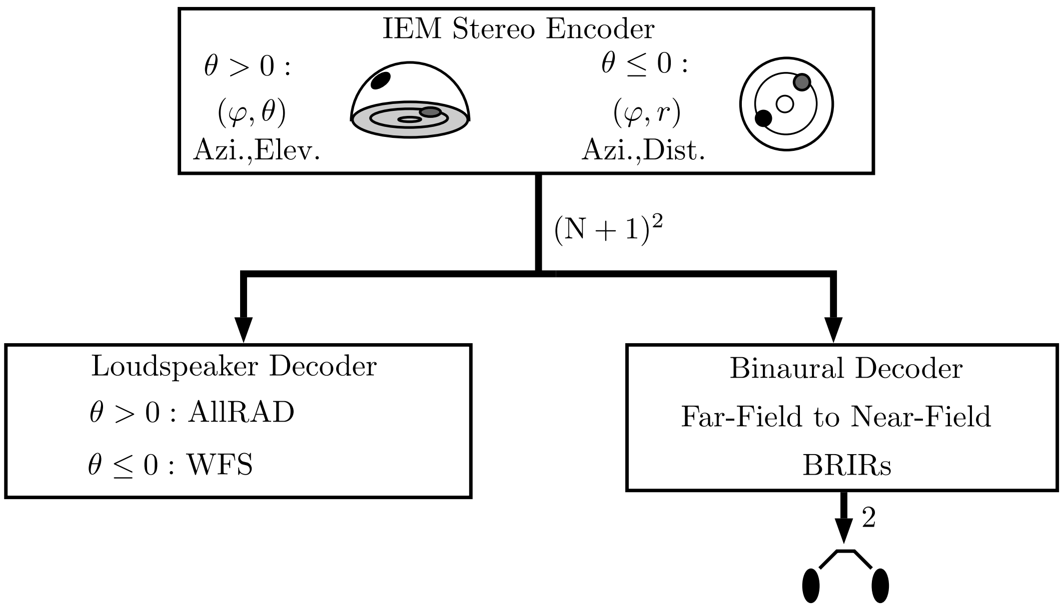


Figure 3.2: The signal flow for the Projection-approach.

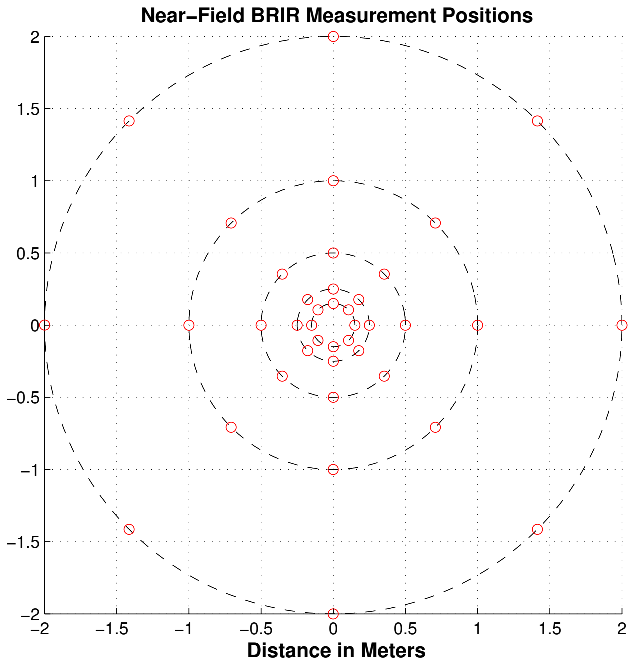


Figure 3.3: Measurement positions for the near-field BRIRs with an angular resolution of 45 degrees.

Figure 3.4: The Neumann KU100 dummy head was used to measure a set of near-field BRIRs. Measured at [2 m, 1 m, 50 cm, 25 cm, 15 cm, 12 cm] from the head center. The loudspeaker is a Genelec 8020, where we positioned the tweeter slightly above the ear level to capture head-shadowing effects at high frequencies.

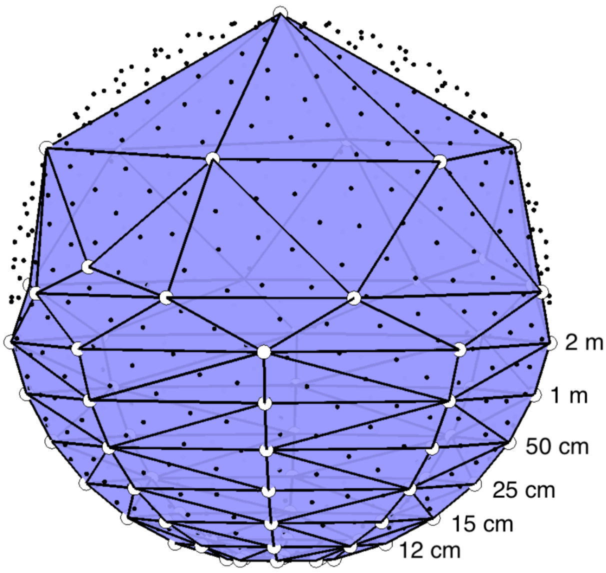


Figure 3.5: Final triangulation of the virtual loudspeaker layout for the projection-based approach with the horizontal near-field BRIRs in the negative hemisphere. BRIRs are interpolated to t-design points (small black dots) which gives a spatial upsampling and a well-conditioned spherical harmonics matrix for ambisonic decoding.

with, e.g. 120 points on the sphere. The final Ambisonics interpolation of the BRIRs blends different distances seamlessly.

The decoders are computed from SOFA files that contain the near-field measurements in the negative hemisphere (Projection-based approach). This virtual loudspeaker layout is triangulated to compute interpolated BRIRs at the t-design positions via a VBAP-type procedure (after time-alignment of the BRIRs, see Fig. 3.6). The final filters are shortened to a length of 4096 samples which reduces the DRIR to the early relfections of the Eurecat studio. This allows to minimize sound colouration while ensuring good externalization through real and diffuse early reflection patterns.

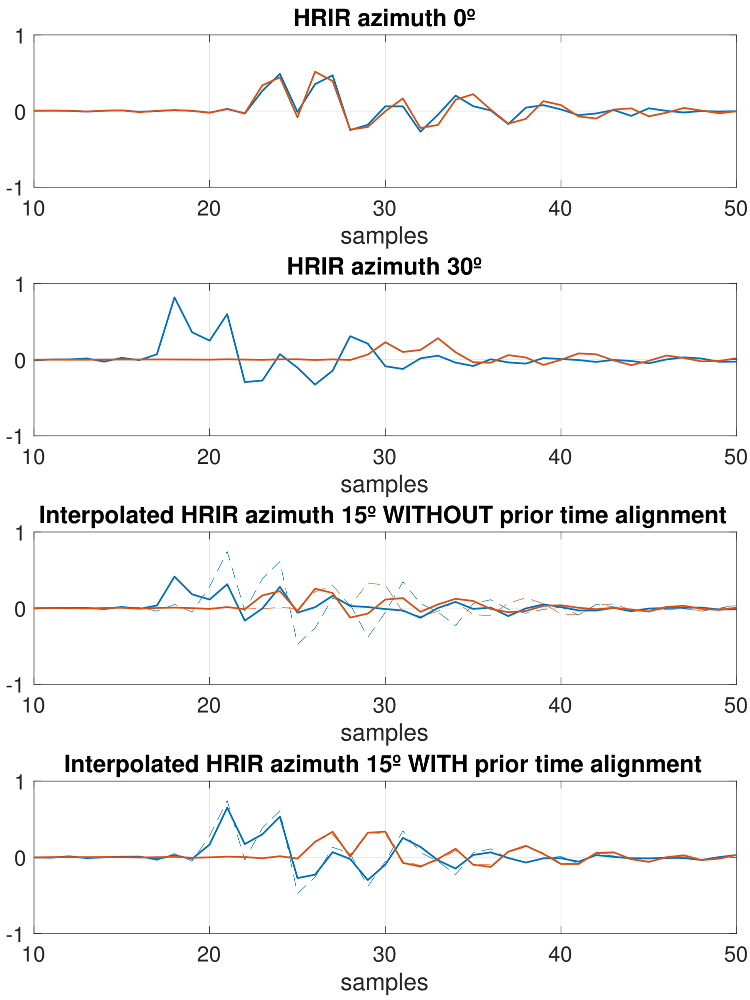


Figure 3.6: Interpolation of sparse BRIR/HRIR sets requires prior time-alignment. The plot shows the ipsilateral HRIR (blue) and the contralateral HRIR (orange) for the azimuth angles of 0° and 30° and subsequently HRIRs retrieved by direct interpolation without prior time alignment. An interpolation with prior time alignment (based on peak detection) matches the reference HRIRs (dashed lines) after reintroducing the ITD based on a spherical head model.

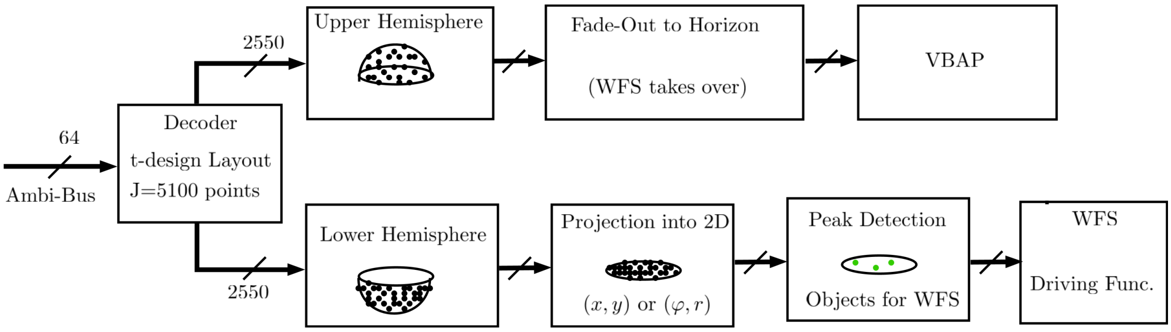


Figure 3.7: Block diagram for a rendering system that combines WFS for horizontal sources and AllRAD for elevated sound sources.

3.3 Loudspeaker Array Reproduction combining Wave Field Synthesis and AllRAD

Combining Wave Field Synthesis (WFS) and All-round Ambisonic Decoding (AllRAD) seems like an interesting approach and would allow to render near-field sources with a dense array of horizontal loudspeakers complemented by a sparse array of elevated loudspeakers to render elevated sources via AllRAD. While the two systems could obviously run separately, it is more desirable to drive them from a single format. The proposed projection-based format seems appropriate to render all sources from the Ambisonics domain.

The critical part is the quality of the horizontal wave field synthesis, which is investigated by simulations based on the Sound Field Synthesis (SFS) Toolbox [13]. Regarding the underlying theory and the SFS toolbox documentation, the interested reader is referred to the corresponding webpage[1].

To simulate an ambisonically driven wave field synthesis, a source position is encoded into Ambisonics and decoded onto a dense grid of t-design points [7]. The t-design points in the negative hemisphere represent different distances in the horizontal plane. The drawback of an ambisoncally driven WFS is the finite spatial resolution of the source that is now more or less distributed, depending on the ambisonics order. This affects the quality of the reproduced sound field significantly.

[1] http://sfstoolbox.org

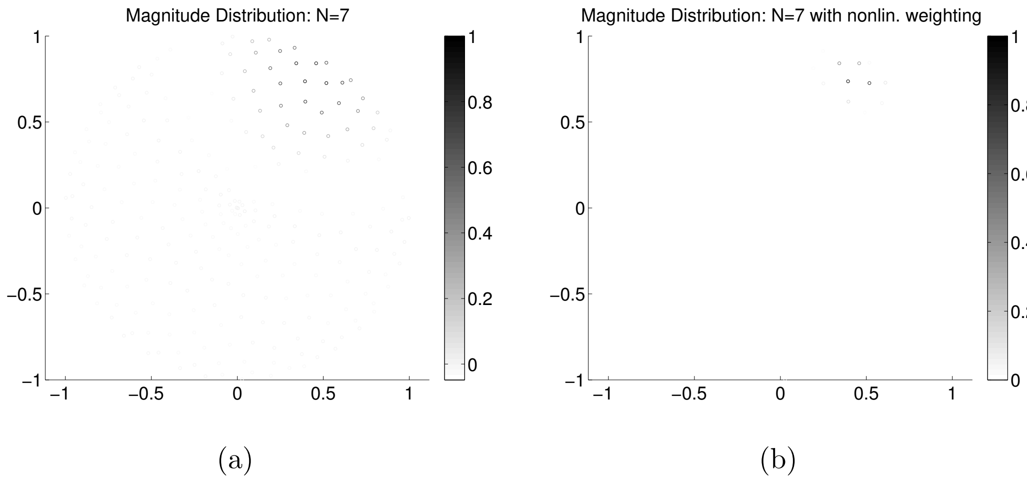
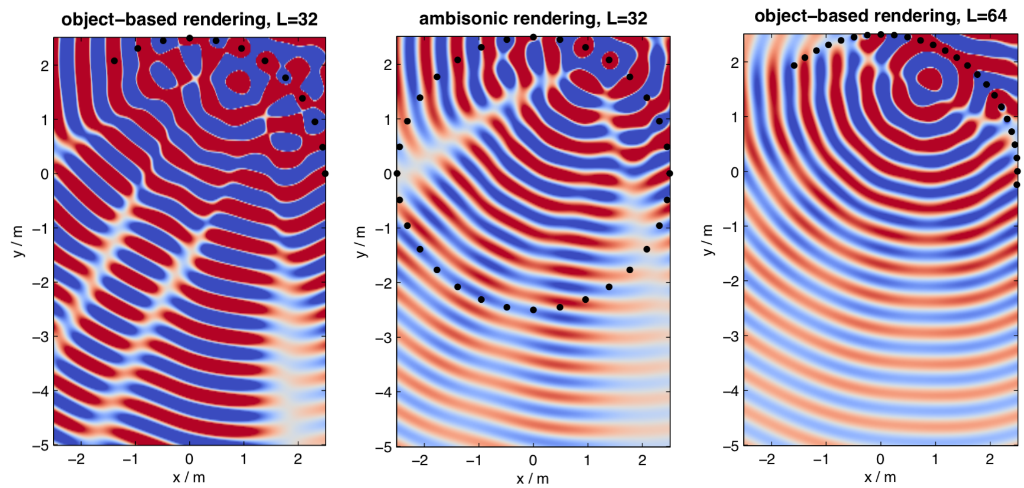


Figure 3.8: Linear magnitude distribution of a projected 7-th order Ambisonics directivity function (a) and subsequent nonlinear processing of the weights with $f(w) = w^{10}$ as spatial dynamic expansion (b). All dots will be rendered as a focused source in the simulations. A decrease in the spread is necesarry to obtain a meaningful soundfield synthesis.

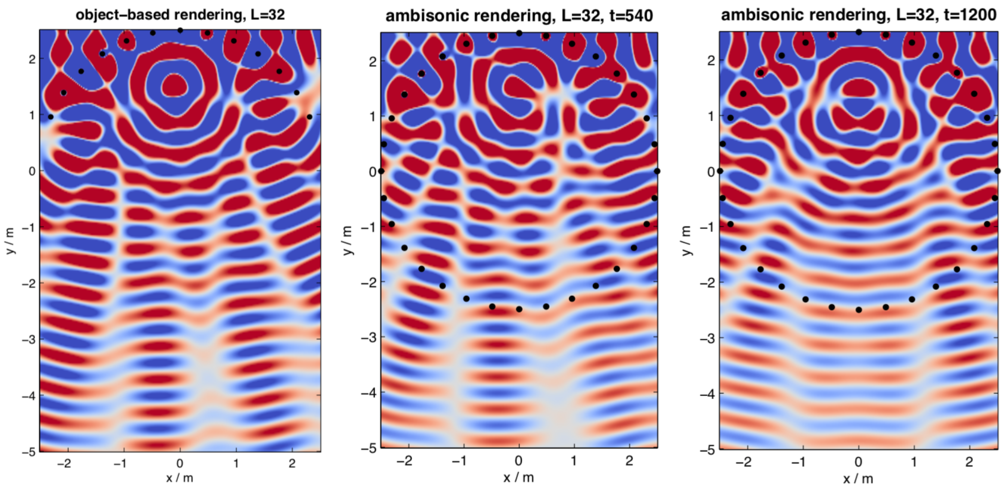
Ways to mitigate this effect could be found in nonlinear processing applied to the t-design weights, as a spatial dynamic expansion, minimizing the spatial spread by e.g. a squaring of the weight distribution (see Fig. 3.8). The obvious drawback of nonlinear weighting in the spatial domain is that small weights (due to low volume encoded signals) are supressed and such an algorithm can effectively only render one source at a time. Considering this single rendered source, ambisonic WFS shows a regularization effect by giving up precise wave front reconstruction around the focused source, allowing to minimize aliasing effects in a wider area of the sound field, see Figs. 3.9 to 3.11.

More sophisticated approaches could try to retrieve object positions by a 'spatial peak detection' method, which would then allow the standard object-based WFS rendering (block diagram in Fig. 3.7).



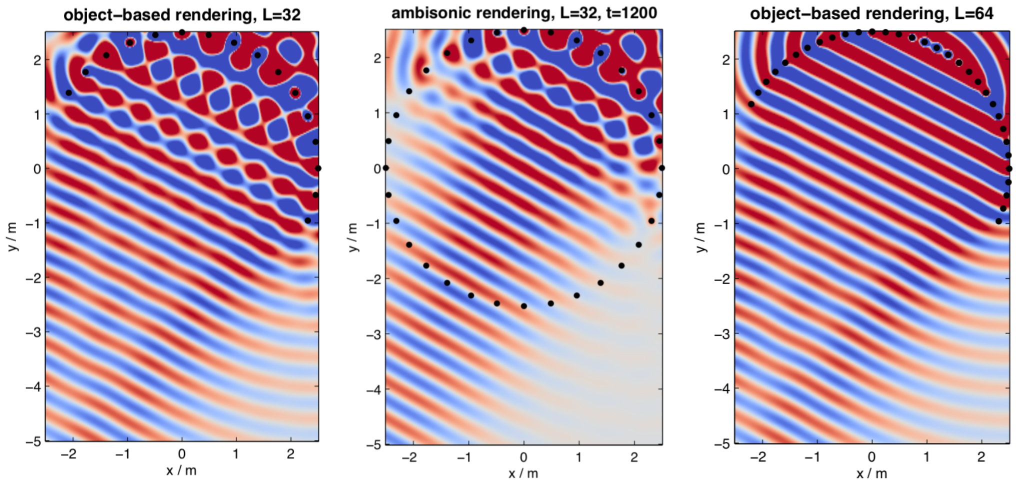
(a) (b) (c)

Figure 3.9: Comparison of sound field synthesis of a focused source at f=600 Hz. Ambisonic rendering involves decoding to a dense t-design followed by a nonlinear processing of the t-design weights (here: $f(w) = w^{10}$) that minimizes the ambisonic spread. Theoretically, all L=32 secondary sources are involved in the ambisonically driven WFS, that renders t/2=270 t-design points within the array diameter simultaneously.



(a) (b) (c)

Figure 3.10: Comparison of sound field synthesis of a focused source at f=600 Hz. The density of the t-design influences the quality of the wave field, as asymmetries and spread for sparse t-designs as well as interference between the points are detrimental especially at frequencies around and above aliasing. Around the aliasing frequency (here at approx. 680 Hz), ambisonic WFS achieves a wider area of reduced aliasing, while neglecting precise wave front synthesis around the focal point.



(a) (b) (c)

Figure 3.11: Comparison of sound field synthesis of a plane wave at f=600 Hz. Ambisonic rendering involves decoding to a dense t-design followed by a nonlinear processing of the t-design weights (here: $f(w) = w^{10}$) that minimizes the ambisonic spread. Theoretically, all L=32 secondary sources are involved in the ambisonically driven WFS, that renders t/2=600 t-design points within the array diameter simultaneously.

4 Conclusion

This work presented a flexible full-sphere distance format for Ambisonics that is compatible with the existing Ambisonics production workflow. It simplifies the creation of depth and distance in Ambisonics audio productions, which previously required a tedious automation of volume and room effects per encoded source to achieve a coherent distance impression. Compatibility between loudspeaker and headphones reproduction has been tested and is ensured. Besides basic distance cues such as level changes and direct-to-reverberant sound ratio, the format allows for near-field processing and compensation.

A second format was defined and is motivated by a combination of wave field synthesis and All-round Ambisonic Decoding. Simulations have been carried out that showed promising effects of rendering WFS from the Ambisonic domain. Moreover, a binaural decoder based on BRIRs (near-field and far-field) serves as a reference for near-field effects in Ambisonics over headphones and is fully compatible with existing encoder plug-ins.

Future work could investigate the distance perception of the proposed rendering methods through formal listening experiments. A general comparison between HRIR-based and BRIR-based rendering concerning externalization seems highly interesting. Finally, the Ambisonic rendering of wave field synthesis ('Ambisonic WFS') could be further investigated and needs to be explored from a theoretical viewpoint.

Bibliography

[1] Johannes M Arend, Annika Neidhardt, and Christoph Pörschmann. "Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set (Messung und perzeptive Evaluierung eines sphärischen Satzes von Nahfeld-HRTFs)." In: *Tonmeistertagung TMT* (2016).

[2] Johannes M Arend and Christoph Pörschmann. "Synthesis of Near-Field HRTFs by Directional Equalization of Far-Field Datasets." In: *Fortschritte der Akustik DAGA* (2019).

[3] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization.* MIT press, 1997.

[4] Douglas S Brungart, Nathaniel I Durlach, and William M Rabinowitz. "Auditory localization of nearby sources. II. Localization of a broadband source." In: *The Journal of the Acoustical Society of America* 106.4 (1999), pp. 1956–1968.

[5] Douglas Scott Brungart. "Near-field auditory localization." PhD thesis. Massachusetts Institute of Technology, 1998.

[6] Jérôme Daniel. "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format." In: *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction.* Audio Engineering Society. 2003.

[7] M. Graef. URL: `https://homepage.univie.ac.at/manuel.graef/quadrature.php`.

[8] Christoph Pörschmann, Johannes Arend, and Annika Neidhardt. "A Spherical Near-Field HRTF Set for Auralization and Psychoacoustic Research." In: *Audio Engineering Society Convention 142.* Audio Engineering Society. 2017.

[9] Ville Pulkki. “Virtual sound source positioning using vector base amplitude panning.” In: *Journal of the Audio Eng. Soc.* 45.6 (1997), pp. 456–466.

[10] Daniel Rudrich and Matthias Frank. “Improving Externalization in Ambisonic Binaural Decoding.” In: *Fortschritte der Akustik*. DAGA. 2018.

[11] Christian Schörkhuber, Markus Zaunschirm, and Robert Höldrich. “Binaural Rendering of Ambisonic Signals via Magnitude Least Squares.” In: *Fortschritte der Akustik*. DAGA. 2018.

[12] Tong Wei et al. “Near-Field Compensated Higher-Order Ambisonics Using a Virtual Source Panning Method.” In: *Audio Engineering Society Convention 145*. Audio Engineering Society. 2018.

[13] Hagen Wierstorf and Sascha Spors. “Sound field synthesis toolbox.” In: *Audio Engineering Society Convention 132*. Audio Engineering Society. 2012.

[14] Markus Zaunschirm, Christian Schörkhuber, and Robert Höldrich. “Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint.” In: *The Journal of the Acoustical Society of America* 143.6 (2018), pp. 3616–3627.

[15] F. Zotter and M. Frank. “All-Round Ambisonic Panning and Decoding.” In: *Journal of the Audio Eng. Soc.* 60.10 (2012), pp. 801–820.

[16] Franz Zotter and Matthias Frank. *Ambisonics*. Springer Open Access, 2019, pp. 90–91.