TOWARDS CHARACTERIZING IRREGULAR LOUDSPEAKER ARRAYS IN DIVERSE ACOUSTIC ENVIRONMENTS

Georgios Papadimitriou[1, 2], Gardar Edvaldsson[2], Nils Meyer-Kahlen[3, 4], Matthias Kohler[2], Rama Gottfried[1]

[1] *Institute for Computer Music and Sound Technology, Zurich University of the Arts, Switzerland*
[2] *Gramazio Kohler Research, ETH Zurich, Switzerland*
[3] *Dpt. of Information and Communications Engineering, Aalto University, Finland*
[4] *Dyson School of Design Engineering, Imperial College London, UK*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

The expansion of immersive audio into multipurpose venues, such as concert halls, music studios, and cinemas, poses a challenge to the reproducibility of spatial audio content. Unlike standardized listening rooms, these environments exhibit unique geometric and acoustic constraints. As a highly transportable and effective method for describing and reproducing sound fields, Higher-Order Ambisonics is often preferred for immersive audio applications. However, multipurpose loudspeaker setups often feature non-uniform speaker distributions and vary in room acoustics. Ambisonic decoders for irregular arrays exist, but the spatial quality may be suboptimal. Also, they typically do not account for the room acoustics, and standard system calibration relies primarily on single-channel metrics. Consequently, there is a methodological gap in diagnosing and predicting how room acoustics distort spatial audio rendering. This paper is a comparative case study of six diverse venues in Switzerland. First, we compare vector measures based on an All-Round Ambisonic Decoder calculated for the selected loudspeaker layouts. Then, we analyze spatial room impulse responses measured on-site using the Spatial Decomposition Method and demonstrate how early reflections may create spatial distortions. This contribution is a foundational step toward an open-access dataset of multipurpose loudspeaker arrays. These measurements will drive future listening tests to assess the performance of the physical arrays and the reproducibility of spatial reproduction algorithms.

Corresponding author: *papadimitriou@arch.ethz.ch.*

Keywords: *spatial audio, ambisonics, loudspeaker arrays, room acoustics*

1. Introduction

The development of immersive audio technologies has extended the reproduction environment beyond standardized listening rooms to multipurpose venues. For ideal spatial reproduction, spherical arrangements of near-uniformly distributed, equidistant loudspeakers are desirable [1]. However, many real-world multipurpose environments impose geometric constraints, leading to non-uniform layouts with unequal radial distances.

In such irregular arrays, standard loudspeaker calibration, which addresses frequency response, level matching, and time-of-arrival (delay) alignment, cannot fully account for the system's spatial performance, even when done precisely. Moreover, although digital signal processing (DSP) equalization effectively corrects the timbral dimension of a system in a specific listening position, the spatial qualities of the interaction of the reproduced sound field with the room remain uncorrected. Spatial distortions and room effects may hinder translation of artistic intent between venues [2], compromise the validity of presented architectural acoustic auralizations, and impact experiments in hearing science and psychoacoustic research, which require strictly controlled, predictable sound fields to ensure the validity of experimental results [3]. Commercial venues use various methods, such as object-based renderers based on Vector Base Amplitude Panning (VBAP) [4], for example. For decoding scene-based, Ambisonic content, the most common approach is All-Round Ambisonic Decoding (AllRAD) [5], which allows for mapping continuous sound fields to irregular layouts. AllRAD, like most decoders, computes decoder weights under the assumption of a free-field environment. In the field of Ambisonics, a set of vector-based measures has been established that characterizes the spatial properties of a given loudspeaker layout and decoder [5, Ch. 2]. We employ these measures as a first approximation to spatial perception induced by practical loudspeaker layouts, knowing that recent research on spatial audio rendering systems has shown that purely mathematical measures are not always reliable predictors of human perceptual quality or localization [6].

Yet, optimizing cross-venue reproducibility also requires understanding the impact of perceived room acoustic characteristics, such as envelopment and spatial clarity, that drive listener preference [7]. Therefore, an analysis of the physical sound field, which accounts for the acoustics of the listening room, is required.

Due to the diversity of multipurpose spaces, it is necessary to collect extensive measured data across different venues to support the analytical and perceptual evaluation of their quality. Here, we present a first visual analysis of large loudspeaker arrays from a recent measurement campaign conducted across six venues in Switzerland. By visually comparing the vector measures computed from the loudspeaker layouts and measured spatial room impulse responses (SRIRs) of the loudspeakers, we analyze the spatial consistency of the rendered sound field. This paper serves as the foundational contribution to a larger research initiative that will: first, provide a dataset with all the measurements and characteristics of loudspeaker arrays and the rooms in which they are built; second, assess the perceptual qualities of these listening spaces through binaural listening tests.

2. Background & Measures

The characterization of immersive sound systems in non-standard environments necessitates distinguishing between the individual loudspeaker performance, the array's theoretical capabilities due to its layout, and the room acoustic conditions. Here, we first address the loudspeaker calibration, then use vector measures computed for an AllRAD decoder, and finally employ a parametric SRIR methods to analyze and visualize the data.

2.1. Loudspeaker Array Calibration

Before rendering algorithms are applied, the physical loudspeaker arrays are typically calibrated to a central geometric reference point (sweet spot). The process requires delay compensation to ensure simultaneous wavefront arrival times from loudspeakers at varying distances, alongside gain adjustments to achieve equal Sound Pressure Level (SPL) at the origin from all sources. Furthermore, DSP filters are often applied to flatten the frequency response of individual transducers and mitigate the acoustic influence of room modes, ensuring a timbrally neutral and cohesive reproduction system [8]. No loudspeaker tuning was performed during our measurements, instead we evaluated the reproduction systems as already configured.

2.2. Vector models & Ambisonics

To evaluate an array's spatial properties, we assume that users would like to apply a continuous, sound-field-based framework such as Ambisonics. Arguably, the most common decoder for irregular layouts is AllRAD [9], which first decodes the scene to a high-density quasi-uniform grid and then uses VBAP [4] to internally map the resulting signals onto irregular layouts.
The spatial fidelity of an Ambisonic reproduction system can be based on mathematical prediction of the perceived trajectory and source width of a virtual phantom source [5]. Following Gerzon's localization theory [10], low-frequency wavefront reconstruction is evaluated via the velocity vector ($\boldsymbol{r}_V$), whereas mid- and high-frequency localization is assumed to be approximated by the energy vector ($\boldsymbol{r}_E$). Here, we focus on the $\boldsymbol{r}_E$ vector.
When panning a source to different directions and decoding, the panning gains g_i for each loudspeaker located in the direction of unit vector $\boldsymbol{u}_i$ are obtained. The total energy of the source is defined as the sum of the squared gains: $E = \sum_i g_i^2$, and the energy vector $\boldsymbol{r}_E$ is the energy-weighted sum of the loudspeaker directions, normalized by the total energy: $\boldsymbol{r}_E = (\sum_i g_i^2 \boldsymbol{u}_i)/E$.
Without considering specific room characteristics, the vector's properties provide a theoretical baseline for spatial fidelity. The direction of the energy vector ($\boldsymbol{r}_E/\|\boldsymbol{r}_E\|$) indicates the perceived direction of the phantom source. The *Panning Error* is quantified as the angular deviation between the intended target panning direction and the resulting energy vector direction. The magnitude of the energy vector ($r_E = \|\boldsymbol{r}_E\|$) describes the spatial concentration of the sound field, approaching 1 for a maximally concentrated phantom source and 0 for a wide and unlocalizable auditory event. Geometrically, r_E corresponds to the cosine of the half-angle of the synthesized energy cone. Therefore, the theoretical *Angular Spread* (source width or localization blur) is mathematically approximated as $2\arccos(r_E)$ [5]. By translating the vector length into a physical angle, this metric shows that a shorter r_E widens the aperture of the energy distribution, resulting in a wider, more blurred auditory image.
Based on the spatial coordinates of each venue's loudspeakers, a decoder was created using the *AllRADecoder* plugin from the *IEM Plug-In Suite*[1]. The subwoofers were excluded from the decoders to avoid skewing the described sound field. Theoretical measures — *Energy Distribution*, *Panning Error*, and *Angular Spread* — were mapped onto a spherical grid to illustrate the array's expected spatial performance, without considering the room acoustics yet.

2.3. In-situ Spatial Analysis

In real-world venues, the theoretical E measure and $\boldsymbol{r}_E$ vector is impacted by the acoustic environment. Specular reflections and reverberation introduce com-

[1]https://plugins.iem.at/

peting energy that arrives shortly after the direct sound from the loudspeakers. To capture this spatial representation of the sound field, measured first-order Ambisonic RIRs of each loudspeaker were analyzed using the Spatial Decomposition Method (SDM) [11]. In existing literature, SDM is used for applications ranging from the visualization of room reflections to the Ambisonic upmixing of measured impulse responses for room auralization [12], [5], [13]. Furthermore, recent studies have utilized SDM to capture, analyze, and reproduce SRIRs to compare the sound fields of different reproduction rooms, such as studios and cinemas [2].

Based on these principles, this study captures the directional characteristics of loudspeaker arrays and uses SDM as an objective visual diagnostic tool. SDM assumes that a single sound event occurs at any given sample; we use the pseudo-intensity vector (PIV) to estimate the Direction-of-Arrival (DOA), which is then projected onto a 2D equal-area Hammer-Aitoff spherical map, in order to visually inspect the venue's spatial acoustic behavior. To ensure visual clarity, an energy-threshold decimation was applied to the vectors, rendering discrete markers scaled by their relative broadband pressure within a dynamic range of 60 dB.

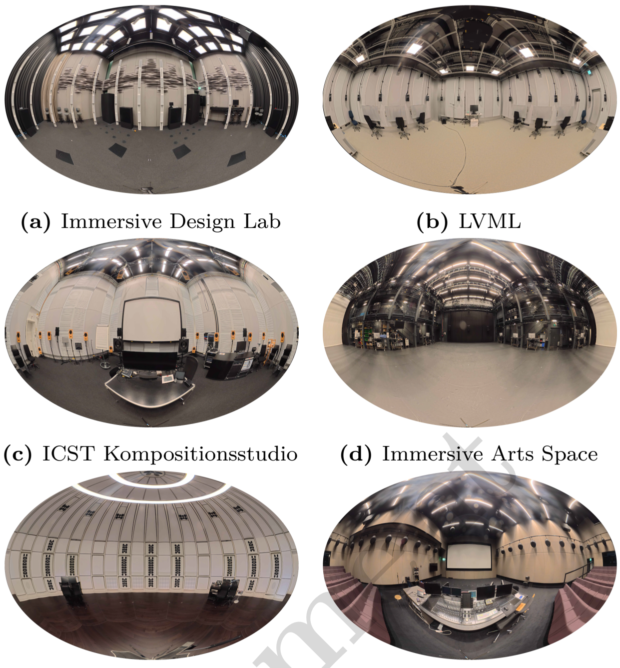
This mapping approach allows the application of custom temporal windows to isolate specific acoustic phenomena, such as direct sound, early reflections, and the late reverberant tail. It is important to note that SDM's parameterization is most accurate for discrete events, such as direct sound and early reflections. In the denser late tail, where multiple reflections overlap within a single analysis window, the PIV estimator returns a weighted spatial average rather than discrete directions. However, in the late reverb, PIV estimates approximate possible anisotropic energy to at least some extent [14]. Segmenting the impulse response allows both the discrete early specular reflections and the overall directional envelope of the late reverberation to be diagnosed.

3. Venues & Measurements

To begin building an open-access dataset, we conducted a measurement campaign in six multipurpose venues in Switzerland. The selection is intended to encompass a wide range of acoustic environments, from nearly symmetric domes to highly irregular architectures. It should be noted that while not all of these venues were explicitly designed for Ambisonic reproduction, most notably the Dolby Studio, Ambisonic theory is leveraged to analyze their theoretical performance, as it provides a robust, format-independent framework for evaluating spatial sound field reconstruction across diverse loudspeaker arrays.

3.1. Venue Specifications

The six measured sites (Fig. 1) represent various geometric topologies and acoustic conditions. The **Immersive Design Lab [IDL]** (ETH Zürich) ($V = 460\,\mathrm{m}^3$) features a dense 75.4 layout, specifically designed for Ambisonic reproduction. The **Large-scale Virtualization and Modeling Lab [LVML]** (ETH Zürich) ($V = 306.6\,\mathrm{m}^3$) uses a sparser 20.1 hemispherical design. The **ICST Kompositionsstudio [ICST Studio]** (ZHdK) ($V = 130.5\,\mathrm{m}^3$) is a 17.1 studio environment that features mixing desks and a speaker layout suited to commercial music production. The **Immersive Arts Space [IA Space]** (ZHdK) ($V = 2354\,\mathrm{m}^3$) is a flexible 32.1 black-box theater designed to accommodate a variety of productions. The **SE Musiclab** ($V = 170\,\mathrm{m}^3$) provides a highly symmetric 26.4 dome structure optimized for critical listening of all commercial music formats. Finally, the **Dolby Studio** (ZHdK) ($V = 591\,\mathrm{m}^3$) serves as a commercial theatrical baseline, featuring a Dolby-certified 44.1 rectangular "shoebox" geometry. Figure 2 shows the frequency-dependent reverberation times (Fig. 2a), as well as the overall Definition for speech (D50) and music (D80), and the Speech Transmission Index (STI) (Fig. 2b) at the listening position.



(a) Immersive Design Lab **(b)** LVML

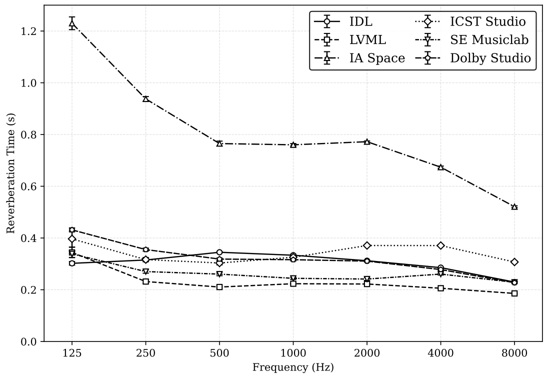
(c) ICST Kompositionsstudio **(d)** Immersive Arts Space

(e) SE Musiclab **(f)** Dolby Studio

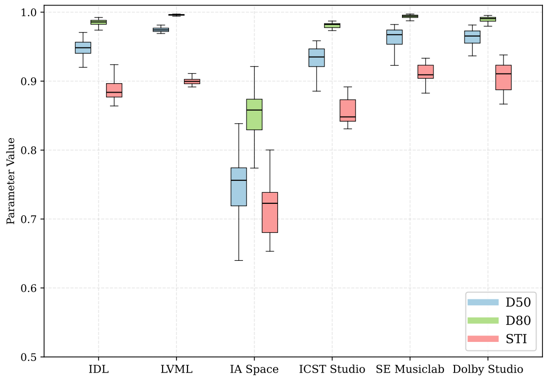
Figure 1: 360° images of the loudspeaker arrays.

3.2. Measurement Protocol

To capture a complete acoustic fingerprint, we measured RIRs at the geometric sweet spot and two off-axis positions (30% and 60% of the array radius at 225°azimuth). Exponential sine sweeps (ESS) [15] were recorded (48kHz/24bits) at each position using three microphones: an NTi M2230 omnidirectional microphone (1/2" Class 1) to extract precision metrics for Time-of-Arrival (ToA), gain, and frequency response; a Sennheiser Ambeo VR Mic (first-order Ambisonics) to visualize directional energy and reflections; and a Neumann KU100 dummy head that



(a) Reverberation times in seconds.



(b) Definition for speech (D50) and music (D80), and Speech Transmission Index (STI), all evaluated on a 0-1 scale. Boxplots show the median and interquartile range across loudspeakers per venue.

Figure 2: Room acoustic parameters of each loudspeaker array environment.

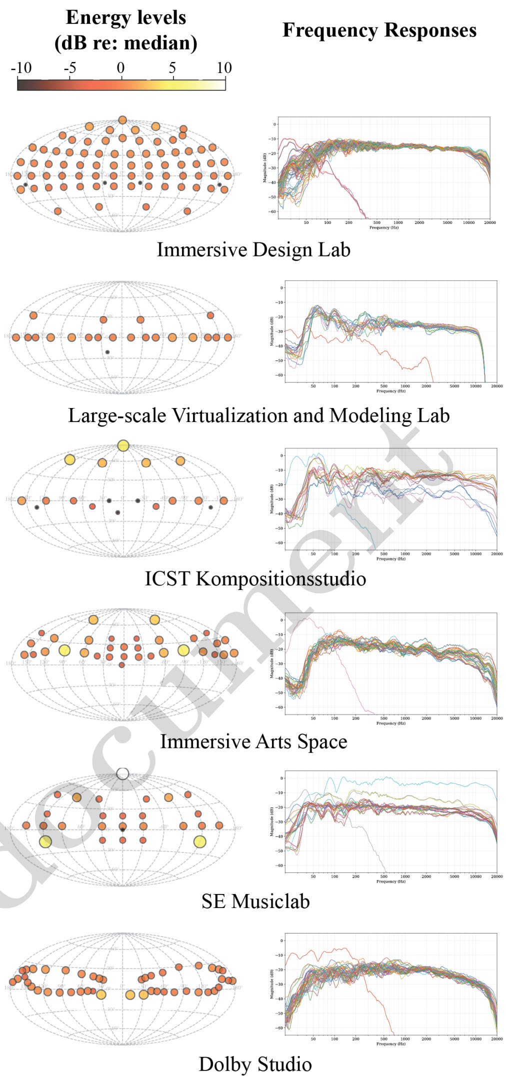


Figure 3: Level deviations in dB relative to the median (left) and frequency responses (right) of all loudspeakers per venue.

captured binaural room impulse responses for future perceptual experiments. Signal processing operations (deconvolution, windowing, and acoustic parameter extraction) were implemented in Python, while raw A-Format signals were encoded to B-Format in AmbiX format (ACN/SN3D), using the *SPARTA array2sh* VST plugin [16] in Cockos Reaper.

4. Results & Discussion

The presented results are limited to the RIRs at the listening position (sweet spot) of the arrays, and their discussion is divided into three stages, progressing from the loudspeakers, to the theoretical capabilities of the geometric layouts, to the measured acoustic reality.

Starting with the electroacoustic capabilities of the loudspeaker arrays, Fig. 3 details the per-channel characteristics of the systems. The energy level plots display the relative amplitude variance across the arrays, and the frequency responses show the frequency compensation at the sweet spot, indicating the effectiveness of the initial gain and frequency alignment. Some arrays, such as IDL and LVML, are precisely tuned to emit nearly equal sound levels and flat spectra at the sweet spot, whereas others, such as ICST Studio, exhibit large discrepancies in both metrics. It is important to note that users of such arrays often apply correction filters to their signal chain to account for discrepancies caused by loudspeaker tuning. However, the scope of this study was to measure the baseline condition of these reproduction systems without any additional compensation.

Figure 4 illustrates the theoretical spatial performance of the arrays as predicted by the vector measures when applying the AllRAD decoder. Observing the energy distribution highlights how loudness is more evenly distributed in dense, symmetrical arrays, such as the IDL. In this dense loudspeaker configuration, the panning error remains below 10° in most spatial regions. On the contrary, arrays with large geometric gaps, such as LVML and ICST Studio, demonstrate energy peaks and dips, significantly larger panning errors exceeding 15°, and wider angular spread up to 60°. These discrepancies occur because regions with low angular

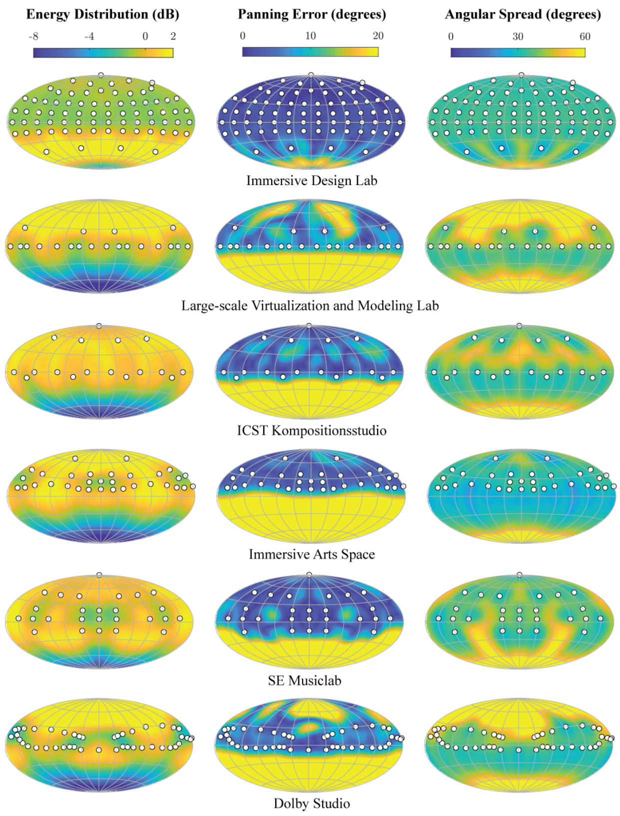


Figure 4: AllRAD decoder measures per venue: Energy Distribution in dB (left), Panning Error in degrees (middle), and Angular Spread in degrees (right).

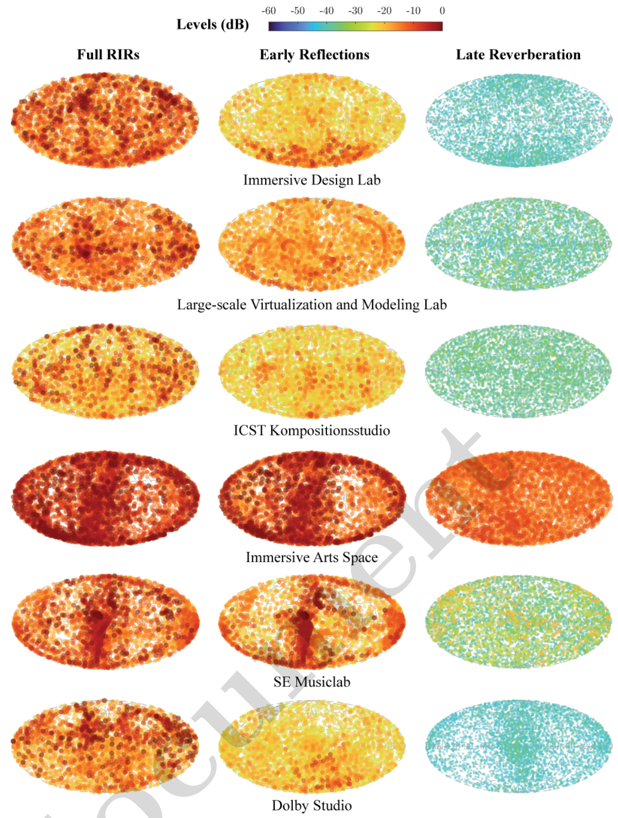


Figure 5: SDM visualizations across all loudspeakers per venue, showing relative energy levels normalized to the peak energy within a 60 dB dynamic range for the Full RIRs (left), Early Reflections up to 100ms (middle), and Late Reverberation after 100ms (right).

density (spatial holes) force the AllRAD decoder's underlying VBAP to form large projection facets on the physical loudspeaker hull. Panning a source in these sparse regions distributes energy across distant loudspeakers, reducing the energy vector length, and misaligning its direction even before room acoustics are considered.

Finally, in Fig. 5, the SDM-analyzed SRIRs from each array's loudspeakers are superimposed per venue to map the temporal evolution of the measured spatial sound field and capture the transition from direct sound to the late reverberation. By segmenting these visualizations into Full RIRs, Early Reflections, and Late Reverberation and scaling them within a relative 60 dB dynamic range normalized to the RIRs' peak energy, we can directly compare the measured in-situ reality against the theoretical geometric baseline. Looking at the early reflections, the maps reveal exactly how boundary interactions inject directional energy into the geometric spatial holes predicted by the AllRAD model. For instance, at IDL and SE Musiclab, reflected energy from the rigid floors is clearly visible, while in the ICST Studio, scattered reflections emerge from the studio equipment. Contrasting this behavior with the theoretical energy distribution in Fig. 4 reveals that room reflections introduce reverberant energy in geometric spatial holes, particularly at lower elevations lacking loudspeaker coverage, where the decoder predicts large energy dips. In late reverberation, the spatial energy distribution becomes uniformly distributed in some venues while retaining directionality in others. For instance, the IA Space, which exhibits longer reverberation times as seen in Fig. 2a, demonstrates strong early and late energy arriving from many directions. The impact of this reflected energy is likely twofold: while excessively strong reflections blur the reproduced spatial image, a moderate, uniformly distributed diffuse field can decorrelate the loudspeaker signals, helping mitigate comb-filtering effects at the sweet spot and broadening the usable listening area.

5. Summary and Future Work

This study introduced a visual diagnostic workflow to characterize non-standard immersive audio environments. By showing theoretical Ambisonic capabilities and SDM analysis of SRIRs measured across six venues, we demonstrate the limitations of free-field rendering assumptions for irregular layouts. Visualizations reveal that geometric gaps in loudspeaker placement degrade the energy vector. However, room reflections may fill spatial holes, while potentially sustaining directional biases. Consequently, achieving cross-venue reproducibility requires evaluating the ar-

ray and room as a joint system.
This comparative framework is a first step toward understanding the physical limits of spatial audio reproduction in multipurpose loudspeaker arrays. Factors such as the perceptual thresholds at which reflections shift from localizable artifacts to beneficial decorrelation cannot be definitively determined only by visualization. Consequently, future work will leverage the binaural impulse responses from these measurements to conduct listening tests, evaluate the robustness of reproduction algorithms, and correlate these objective spatial maps with human perception. Finally, upon completion of ongoing measurement campaigns, the complete dataset, including impulse responses and geometric characteristics from these and additional venues, will be published in an open-access repository, providing the community with resources for modeling, auralization, and algorithm development.

Acknowledgments

This work was funded by the Swiss National Science Foundation (grant no. 10000803).

References

[1] *ITU-R BS.2051-3* - Advanced sound system for programme production*, 2022.

[2] J. Riionheimo and T. Lokki, "Spatial Sound, Part 1: Exploring the Impact of Listening Room on Perceptual Immersion," *Journal of the Audio Engineering Society*, vol. 72, no. 12, pp. 913–929, Dec. 2024. DOI: 10.17743/jaes.2022.0179.

[3] S. E. Favrot and J. Buchholz, "Reproduction of nearby sound sources using higher-order ambisonics with practical loudspeaker arrays," *Acta Acustica United with Acustica*, vol. 98, no. 1, pp. 48–60, 2012. DOI: 10.3813/AAA.918491.

[4] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Audio Engineering Society*, 1997.

[5] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Cham: Springer International Publishing, 2019, vol. 19. DOI: 10.1007/978-3-030-17207-7.

[6] M. Gerken, V. Hohmann, and G. Grimm, "Comparison of 2D and 3D multichannel audio rendering methods for hearing research applications using technical and perceptual measures," *Acta Acustica*, vol. 8, p. 17, 2024. DOI: 10.1051/aacus/2024009.

[7] J. Francombe, T. Brookes, and R. Mason, "Evaluation of Spatial Audio Reproduction Methods (Part 1): Elicitation of Perceptual Differences," en, *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 198–211, Mar. 2017. DOI: 10.17743/jaes.2016.0070.

[8] F. Toole, *Sound reproduction: the acoustics and psychoacoustics of loudspeakers and rooms*, Nachdr. Amsterdam: Elsevier, Focal Press, 2009, ISBN: 978-0-240-52009-4.

[9] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 807–820, Nov. 2012.

[10] M. A. Gerzon, "General Metatheory of Auditory Localisation," *Journal of the Audio Engineering Society*, no. 3306, Mar. 1992.

[11] S. Tervo, J. Pätynen, and A. Kuusinen, "Spatial Decomposition Method for Room Impulse Responses," en, *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, 2013.

[12] J. Pätynen, S. Tervo, and T. Lokki, "Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses," *The Journal of the Acoustical Society of America*, vol. 133, no. 2, pp. 842–857, Jan. 2013. DOI: 10.1121/1.4770260.

[13] N. Meyer-Kahlen, S. V. A. Garí, and T. Lokki, "What the spatial decomposition method can and cannot do," in *ICA 2022 proceedings*, Gyeongju, Korea: Acoustical Society of Korea (ASK), 2022.

[14] N. Meyer-Kahlen and S. J. Schlecht, "Directional distribution of the pseudo intensity vector in anisotropic late reverberation," en, *The Journal of the Acoustical Society of America*, vol. 155, no. 2, pp. 1515–1526, Feb. 2024. DOI: 10.1121/10.0024960.

[15] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *AES 108th Convention*, 2000.

[16] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinardi, and V. Pulkki, "Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis," in *144th Audio Engineering Society Convention 2018*, Audio Engineering Society, Jan. 2018, pp. 294–303.