DOI: 10.61782/fa.2025.0411

FORUM ACUSTICUM
EURONOISE 2025

FREQUENCY AND SPATIAL LIMITATIONS OF SOUND FIELD RECONSTRUCTION METHODS FOR NAVIGABLE REPRODUCTION

Antonio Figueroa-Duran[1*] **Nils Meyer-Kahlen**[2]
Samuel A. Verburg[1] **Tapio Lokki**[2] **Efren Fernandez-Grande**[3,1]
[1] Acoustic Technology, DTU Electro, Technical University of Denmark, Kgs. Lyngby, DK
[2] Acoustics Lab, Dept. Information and Communications Eng., Aalto University, Espoo, FI
[3] Dept. Audiovisual and Comunications Eng., Technical University of Madrid, Madrid, SP

ABSTRACT

Sound field reconstruction methods enable the prediction of acoustic field quantities based on a limited number of measurements, including locations where direct measurements are unavailable. Their predictive capabilities over large spatial domains make them a promising tool for navigable reproduction and '6-degrees-of-freedom' applications. Despite their potential, the methods are band-limited due to spatial sampling, and the predictions can deteriorate far away from the measurement positions. In this study, we quantitatively examine the frequency and spatial limitations of sound field reconstruction methods using psychoacoustic metrics, including binaural auditory models to assess perceived source localisation and spectral colouration. The evaluated reconstruction methods include different wave-based approaches, such as plane waves and point sources. The analysis is conducted using an experimental dataset of spatial room impulse responses, measured sequentially along a line of high spatial resolution in a variable-acoustic room. The findings provide insights into sound field reconstruction techniques for navigable reproduction and the extent to which physical models contribute to practical, navigable auditory experiences.

Keywords: *sound field reconstruction, 6DOF, plane wave decomposition, ambisonics*

**Corresponding author*: *anfig@dtu.dk.*

1. INTRODUCTION

The advent of immersive audio technology has largely influenced the acoustic literature in the recent years. Methods for '6-degrees-of-freedom' (6DOF) reproduction have been developed, where the aim is to enable seamless spatial navigation within a desired sound field. When these fields encompass large spatial domains, such as entire rooms or auditoria, experimental characterisation becomes increasingly challenging due to the large data and processing requirements. To address this, data acquisition is often supported by interpolation and extrapolation techniques, where different theoretical principles are adopted depending on the specific application.

In this context, audio reproduction methods often rely on pragmatic assumptions, prioritising perceptual goals and full-band rendering [1–7]. On the other hand, sound field reconstruction methods target a complete prediction of acoustic field quantities, adopting more explicit considerations about the physical structure, and enabling active manipulation and reproduction of sound fields. These kind of approaches were traditionally developed to provide physically-accurate characterisation in sound radiation problems [8, 9], and were later extended to diverse applications including active noise control [10, 11], sound field analysis in rooms [12–14], and others. Recent studies have extended their applicability to large-scale domains for sound field analysis [15, 16], making them a potential tool for 6DOF reproduction using a more physically accurate perspective.

Naturally, different applications entail using different error metrics. Whilst audio reproduction is evaluated taking perceptual considerations into account, where psychoacoustic metrics and listening tests are common mech-

anisms for assessing reproduction performance [17, 18], sound field reconstruction methods are frequently evaluated using physical metrics.

This paper aims to provide insight into the frequency and spatial limitations of extrapolation methods when different theoretical principles are adopted, assessed using both physical and perceptual metrics. To this end, we compare four simple methods that include plane wave decomposition, spherical propagation and uniform late reverberation to extrapolate from a single array position. The perceptual metrics focus on spectral colouration and source localisation, and their comparison with physical metrics is intended to reveal how these two evaluation approaches are mutually informative.

2. SOUND FIELD MODELS

The sound field in reverberant environments can be characterised by a collection of room impulse responses (RIRs). The RIRs refer to the transfer path between source and receiver, and exhibit a distinctive general structure where two parts can be identified: the early part, comprised of the direct sound and the early reflections, and the late reverberation [8]. Both the early and the late part of the RIR can be described as a superposition of elementary waves [8, 19]. This is, the acoustic pressure $p(\mathbf{r}, \omega)$ at a given angular frequency ω can be written as

$$p(\mathbf{r}, \omega) = \sum_{l=1}^{L} w_l(\omega)\phi_l(\mathbf{r}, \omega) + n(\mathbf{r}, \omega), \tag{1}$$

where ϕ_l is the l^{th} wave, and n is the measurement noise. The wave coefficients w_l can be inferred from the pressure data at the measurement positions and used to reconstruct the sound field elsewhere via

$$\hat{p}(\mathbf{r}_\bullet, \omega) = \sum_{l=1}^{L} \hat{w}_l(\omega)\phi_l(\mathbf{r}_\bullet, \omega), \tag{2}$$

where $_\bullet$ denotes prediction and $\hat{w}_l$ are the estimated coefficients.

Sound field reconstruction methods typically choose the elementary waves to be solutions to the Helmholtz equation. Depending on the geometry of the problem and the distance to sound sources, common choices for sound field models are plane waves, defined as

$$\phi_l(\mathbf{r}, \omega) = \mathrm{e}^{\mathrm{j}\langle \mathbf{k}_l, \mathbf{r}\rangle}, \tag{3}$$

and spherical waves,

$$\phi_l(\mathbf{r}, \omega) = \frac{\mathrm{e}^{\mathrm{j}\mathbf{k}_l \cdot (\mathbf{r}-\mathbf{r}_l)}}{4\pi\|\mathbf{r}-\mathbf{r}_l\|}. \tag{4}$$

Here, $\langle \cdot, \cdot \rangle$ is the inner product operator, $\mathbf{k}_l$ represents the wave vector, defined by the direction of propagation $\hat{\mathbf{e}}_l$ and the wavenumber k, and $\mathbf{r}_l$ denotes the origin of the l^{th} spherical wave.

Given the wavefront divergence established by the aforementioned models, spherical waves are typically adopted when extrapolating the early part of the RIR, and plane waves are well suited for the late tale. More refined models exist in the sound field reconstruction literature—including but not limited to other analytical wave functions, data-driven methods, numerical approaches or statistical reconstruction.

3. EXTRAPOLATION METHODS

Four relatively simple extrapolation methods are considered, which are intended to examine specific modelling aspects commonly found across predictive methods.

3.0 Input data

In this paper, the RIRs are obtained using a rigid spherical microphone array. In this scenario, the sound field is effectively described using spherical harmonics (SH), which provide the mathematical foundation for Ambisonics encoding. Considering the scattering introduced by the body of the array, the SH coefficients p_{nm} relate to the measurements p via [20, 21]

$$p(\Omega, kr) = \sum_{n=0}^{N} \sum_{m=-n}^{n} p_{nm}(kr) b_n(kr) \mathrm{Y}_n^m(\Omega), \tag{5}$$

where $\Omega = \{\theta, \phi\}$ represents the spherical angular coordinates, N is the SH order of truncation, $b_n(kr)$ represents the analytical model of the array scattering, and Y_n^m is the spherical harmonic of degree n and order m. The process of computing p_{nm} involves the inversion of the radial functions b_n, which often requires regularisation to avoid excessive noise amplification at low frequencies. In this work, we follow the implementation by [22] with Tikhonov regularisation, allowing a maximum noise boost of 15 dB.

In addition to the SH representation of the RIRs, the position of the sound source is also considered to be known. For methods that consider running signals as the input data the reader is referred to [2, 7].

3.1 Plane Wave Decomposition (PWD)

The first method assumes that the sound field can be described as a linear combination of monochromatic plane waves, as in Eq. (3). A total of $L = 2000$ uniformly-distributed plane waves are considered. Once estimated, the wave parameters $\hat{w}_l$ are utilised to evaluate the sound field at the target position $\mathbf{r}_\bullet$ via Eq. (2). For a thorough description of the method, the reader is referred to [23].

3.2 Point Source and PW (PS-PWD)

This approach refines the *PWD* method by incorporating the time structure of the RIR and modelling the direct sound independently. Specifically, spherical wave propagation is adopted as in Eq. (4), assuming the loudspeaker, whose position is known, to behave as an ideal point source. The rest of the RIR is modelled using plane waves.

3.3 Point Source for the Direct Sound (PS-Dir)

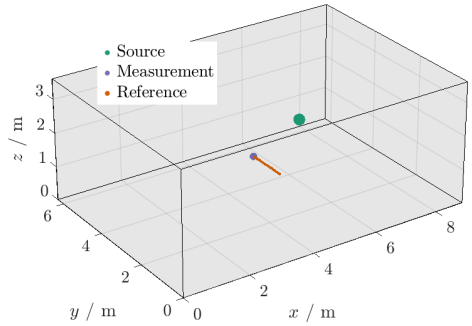
The third method is a pragmatic approach that prioritises perceptual goals. The direct sound is modelled as a spherical wave, and the remaining response is regarded as uniform across the room. This is, the synthesised late part is identical to that of the measurements, although shifted proportionally to the time-of-flight (TOF) between the source and the reconstruction points, thus preserving causality. A similar method was used in [24].

3.4 Point Sources for the Early Part (PS-Early)

This method expands on *PS-Dir* by also considering individual reflections. Here, point sources are used to characterise salient wavefronts during the early part of the RIR. Previous studies have employed this philosophy for single position reproduction [25], and for interpolation [4] and extrapolation [5] in 6DOF scenarios. The latter implementation, [5], is followed in this work.

The first 10 reflections of the early part are considered. The method: a) identifies the largest peaks in the energy response, b) estimates their corresponding directions-of-arrival (DOAs) by computing the pseudo-intensity vector, c) calculates their range by considering the TOF, and d) obtains the point source coefficients by assigning the observed pressure weighted by the $1/r$-law (cf. Eq. (4)). The reconstruction of the direct sound and early reflections is estimated via Eq. (2). Lastly, the residual is treated as uniform and causal, identically to *PS-Dir*.

(a) Panoramic view of the room.



(b) Sketch of the experimental setup.

Figure 1. Experimental setup in the variable-acoustics room "Arni" (Aalto University) in the *dry* configuration. (a) A panoramic view of the setup showing the receiver on the left and three sources on the right. Only the first source from the left-hand size is used in this paper. (b) Sketch of the setup.

4. DATA

All four methods described in the previous section were evaluated using experimental data. The dataset was measured in the variable-acoustics room "Arni" at the Aalto Acoustics Lab. The room is rectangular, with dimensions $8.9\,\text{m} \times 6.3\,\text{m} \times 3.6\,\text{m}$. All four walls and ceiling are covered with painted metallic boxes that contain absorptive acoustic material and can be opened and closed automatically. The room was in its *dry* configuration, i.e. all acoustic panels were open, resulting in an averaged reverberation time of $T_{60} \approx 0.41$ s. Figure 1(a) depicts a panoramic view of the room with the panels exposed. During the measurements, the panels were covered with curtains to minimise reflected high frequency components.

The source was a two-way loudspeaker (Genelec 8030B), driven with a 3 s-long exponential sweep ranging from 20 Hz to 20 kHz, at a sampling rate of $F_s = 48$ kHz. The receiver was a spherical microphone array (Eigen-

mike em32), with 32 microphones distributed over a rigid surface. As shown in Fig. 1(b), the measurement set is comprised of a single array position. The reference set, utilised here to evaluate the reconstruction, encompasses 25 array positions uniformly spaced every 5 cm. A motion capture system (Optitrack) was employed to track the receiver positions. These measurements are part of a larger dataset publicly available [26].

5. EVALUATION METRICS

Both physically- and perceptually-inspired metrics are employed to assess model performance. The evaluation is carried out over the predicted SH coefficients $\hat{\mathbf{p}}_{nm}$ against the reference $\mathbf{p}_{nm}$ at every reconstruction position. The main attributes evaluated here are spectral colouration and source localisation.

5.1 Physics-based metrics

The **normalised mean squared error** (NMSE) is the first objective metric considered. It is defined as

$$\text{NMSE} = \frac{1}{(N+1)^2} \sum_{n,m} \frac{\|\mathbf{p}_{nm} - \hat{\mathbf{p}}_{nm}\|^2}{\|\mathbf{p}_{nm}\|^2}, \tag{6}$$

where $\sum_{n,m}$ denotes the sum over $n \in [0, N]$ and $m \in [-n, n]$. It can be obtained either in the time or the frequency domain, providing an error in the $\ell = 2$-norm sense averaged over all SH signals.

The energy distribution over the unit sphere is evaluated via the **spatial correlation**, $\rho \in [0, 1]$, defined as

$$\rho = \frac{|\langle \mathbf{p}_{nm}, \hat{\mathbf{p}}_{nm} \rangle|^2}{\|\mathbf{p}_{nm}\|^2 \|\hat{\mathbf{p}}_{nm}\|^2}. \tag{7}$$

This metric evaluates the similarity between the reference and the reconstructed SH component, which is informative of the directional properties of the energy distribution around the SH expansion centre [27].

Finally, the **DOA of the direct sound** is estimated at every reconstruction position and compared to the reference. Here, the DOA is estimated via PWD and selecting the direction corresponding to the PW coefficient exhibiting the largest magnitude.

5.2 Perception-based metrics

The first psychoacoustic metric is the **energy vector**, $\mathbf{r}_E$. For its computation, a PWD decomposition is performed, obtaining the wave coefficients $\mathbf{w} \in \mathbb{C}^L$. Then, the $\boldsymbol{r}_E$-vector is obtained as

$$\boldsymbol{r}_E = \frac{1}{\|\mathbf{w}\|^2} \sum_{l=1}^{L} |w_l|^2 \hat{\mathbf{e}}_l. \tag{8}$$

Here, we obtain the energy vector in short time-instances throughout the response excluding the direct sound, which is informative of the perceived directionality of the reverberation. For more refined versions of the $\boldsymbol{r}_E$-vector model that take temporal masking into account, the reader is referred to [18, 28].

The subsequent metrics are applied onto binaural signals, denoted by $p_{l,r}$, where the subscript refers to the ear. In this work, we convolve the Ambisonic coefficients with the head-related-impulse-response (HRIR) dataset collected at TH Köln [29]. The binaural responses were created from the SH responses using the state-of-the-art Magnitude Least Squares (MagLS) decoder approach [30].

The second metric is the **composite loudness level** (CLL) [31], which measured perceived colouration. The CLL computation relies on a binaural model that includes a gammatone filterbank, half-wave rectification, and binaural summation. Colouration is then obtained as $\Delta\text{CLL}(f_c) = \text{CLL}(f_c) - \hat{\text{CLL}}(f_c)$, where f_c is the central frequency of the corresponding band.

A simple spatial metric for reverberation is the **interaural energy ratio** (IER), defined as $\text{IER} = 10 \log_{10}(p_l^2/p_r^2)$, which measures the energetic imbalance between binaural signals in dB.

The fourth metric is the **interaural cross-correlation** (IACC). It is defined as the maximum of the cross-correlation function between $p_l(t)$ and $p_r(t)$ within the delay time interval $|\tau| \leq 1$ ms. Its 'early' version, restricted to the first 80 ms of the response, is denoted by IACC_e. This metric is related to the apparent source width [17, 32], defined as $\text{ASW} = 1 - \text{IACC}_e$.

6. RESULTS

Results from the NMSE and ρ as a function of distance from the measurement array, Δr, are presented in Fig. 2. As expected, all methods perform worse as the distance increases, with the best results at the measurement position. Notably, the two methods that involve PWD show lower NMSE errors even at far distances. This is due to the systematic error introduced in *PS-Dir* and *PS-Early*, which utilise the same measured late tail for all reconstruction positions. Conversely, in the spatial correlation,

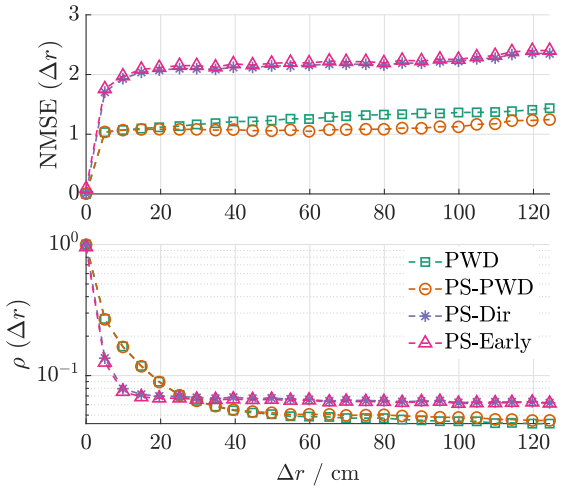


Figure 2. Time-averaged NMSE and Spatial correlation as a function of distance.

PS-Dir and *PS-Early* perform worse than *PWD* and *PS-PWD* for $\Delta r < 25\,\text{cm}$, but better for $\Delta r > 25\,\text{cm}$. This higher correlation is due to how these models preserve the spatial properties of the RIR far from the measurement array. Whilst both *PS-Dir* and *PS-Early* preserve the highly-energetic direct sound and the relative properties of the late tail, the extrapolation provided by PWD methods progressively loses its temporal structure as Δr increases. Finally, *PS-Early* does not offer significant improvements over *PS-Dir*, likely due to the scarce number of salient reflections in the room.

The DOA estimation for the direct sound is shown in Fig. 3. All models that include an individual treatment of the direct sound preserve the DOA even far from the array. Only *PWD* is not capable of extrapolating non-planar wavefronts due to model mismatch, leading to an erratic localisation at positions where $\Delta r > 35\,\text{cm}$. This is also supported by the spatial correlation values in this range seen in Fig 2.

Frequency-dependent NMSE is provided in Fig. 4. The grey area indicates the frequency range above the spatial aliasing frequency of the array, where the main source of error is not the modelling approach but the SH order truncation. The figure depicts selected positions from the reconstruction domain, with brighter colours corresponding to larger Δr. Here, *PWD* performs best. At the reference position (darkest green), the error progressively increases until the spatial aliasing frequency. Worse performance is seen for the other three methods, which include comb-filtering effects at high frequencies due to the wavefront windowing process.

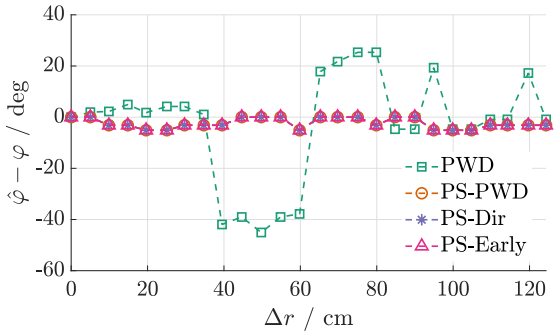


Figure 3. Frequency-averaged azimuth localisation error of the direct sound as a function of distance.

A different perspective is offered by Fig. 5, where the CLL is shown for different frequency bands at selected reconstruction positions. Again, brighter colours represent positions farther from the measurement array. In contrast to Fig. 4, *PWD* exhibits the largest colouration of all methods, particularly for large Δr and high frequencies. The spectral colouration is somewhat reduced by *PS-PWD* as this method characterises the direct sound with a more suitable sound field model. Finally, the perceptually-motivated methods exhibit the lowest colouration at all frequencies. Here, *PS-DIR* enables a reliable reconstruction even at the farthest position from the measurement array, displaying the best results as it preserves most of the binaural properties of the original measured data. *PS-Early* introduces more colouration as a result of the more complex processing of early reflections.

Fig. 6 shows the error based on the $\boldsymbol{r}_E$-vector, the IER and the IACC_e as a function of distance, comprising most of the perceptual metrics evaluated in this study. As seen before, PWD methods do not extrapolate well at large distances, resulting in a larger $\boldsymbol{r}_E$ error than methods that preserve most of the late tail of the measured response. It is worth noting that the $\boldsymbol{r}_E$-vector computation excludes the direct sound, largely covered by the DOA metric. Regarding the IER, similar conclusions can be drawn, resulting in the *PS-Dir* and *PS-Early* methods performing significantly better than PWD-based approaches. Furthermore, the low values for *PS-PWD* indicate that the IER difference is primarily influenced by the response after the direct sound processing. Finally, *PS-Dir* and *PS-*

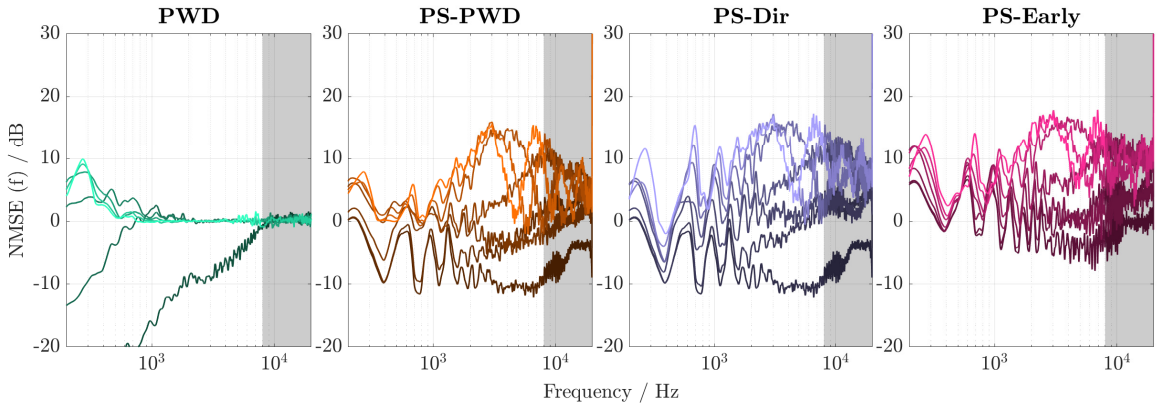


Figure 4. NMSE as a function of frequency for different distances. Brighter colours correspond to larger distances. The grey shaded area indicates frequencies above the spatial aliasing frequency of the array.

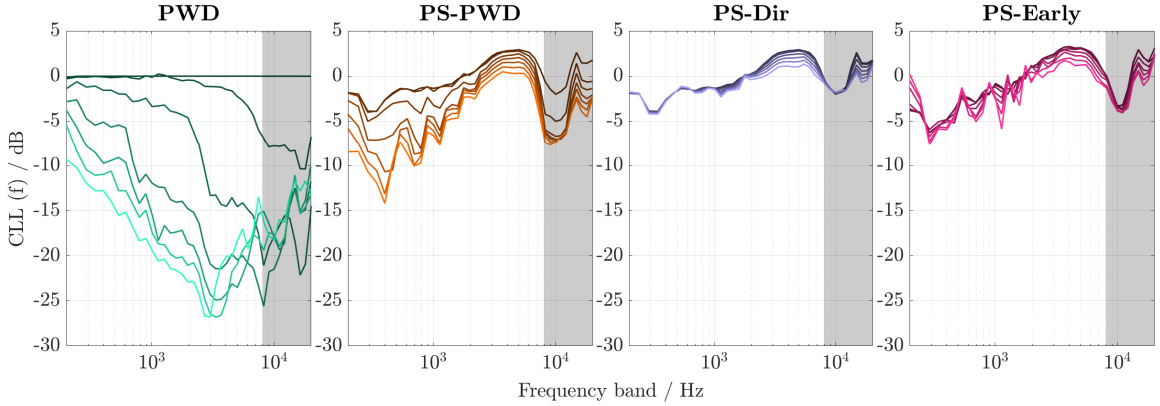


Figure 5. Composite loudness level (CLL) as a function of frequency for different distances. Brighter colours correspond to larger distances. The grey shaded area indicates frequencies above the spatial aliasing frequency of the array.

Early offer relatively stable and low IACC_e values even far from the measurement array. *PS-PWD*, and especially *PWD*, introduce errors when extrapolating farther than $\Delta r > 35\,\mathrm{cm}$, which strongly correlates to what is seen for the DOA analysis, potentially influencing the listener perception of the scene's width.

7. DISCUSSION

All in all, the results presented in the previous section establish a clear distinction in model performance depending on the target application. Whilst PWD-based methods achieve the best results using objective metrics near the measurement domain, maintaining the measured late tail at all positions preserves relevant psychoacoustic properties. Physical metrics like NMSE are not directly representative of spectral colouration or perceived reverberation, requiring additional considerations if the target application is audio reproduction. The methods implemented in this initial study are fairly simplistic, and modelling choices tailored to the specific acoustic scenarios need to be explored in future research. Finally, it is worth noting that, even though most of the psychoacoustic results are promising using simple perceptual methods, the study

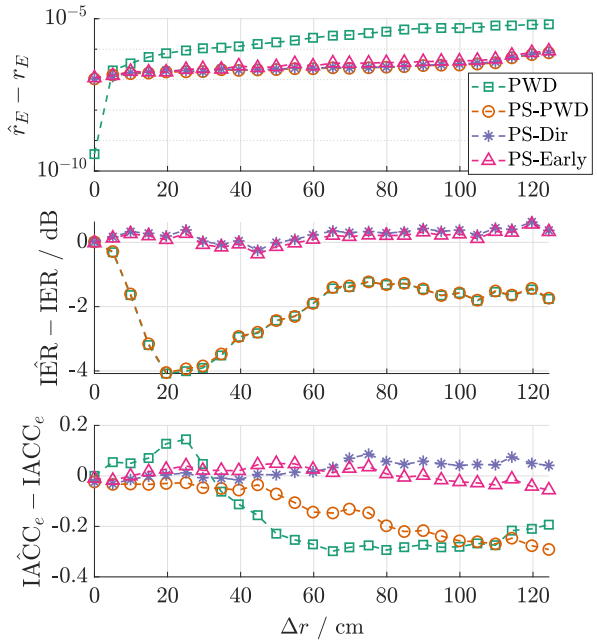


Figure 6. Averaged errors for the $\boldsymbol{r}_E$-vector, interaural energy ratio (IER), and early interaural cross-correlation (IACC$_e$) as a function of distance.

focused on perceived colouration and source localisation, and stronger conclusions about their suitability for audio reproduction can only be drawn after assessing other attributes and performing listening experiments.

8. CONCLUSIONS AND OUTLOOK

This paper presents an experimental comparison of four simple extrapolation methods from a single microphone array, assessed using physical and perceptual metrics. The methods are based on plane waves, spherical waves, and perceptually-inspired assumptions. The results show that the two extrapolation methods that rely on plane waves provide low physical errors near the measurement domain, as opposed to the methods that introduce minimal or no modification to the reverberant tail. Contrarily, these methods offer best performances under perceptual metrics, even at far distances, showing that physical metrics are not necessarily representative of spectral colouration or perceived reverberation. These preliminary results are to be complemented in the future with more complex acoustic scenarios, more representative methods, larger reconstruction domains, and listening experiments.

9. ACKNOWLEDGMENTS

This work has been supported by the VILLUM Foundation under the grant number 19179, 'Large scale Acoustic Holography', and by the Beatriz Galindo Senior Fellowship under the grant number BG22-00177.

10. REFERENCES

[1] J. G. Tylka and E. Y. Choueiri, "Domains of practical applicability for parametric interpolation methods for virtual sound field navigation," *J. Audio Eng. Soc.*, vol. 67, no. 11, pp. 882–893, 2019.

[2] E. Patricio, A. Rumiński, A. Kuklasiński, Januszkiewicz, and T. Żernicki, "Toward Six Degrees of Freedom Audio Recording and Playback Using Multiple Ambisonics Sound Fields," in *146th AES Convention*, 2019.

[3] F. Zotter, M. Frank, C. Schörkhuber, and R. Höldrich, "Signal-independent approach to variable-perspective (6DoF) audio rendering from simultaneous surround recordings taken at multiple perspectives," in *DAGA*, 2020.

[4] K. Müller and F. Zotter, "Auralization based on multi-perspective ambisonic room impulse responses," *Act Acust*, vol. 4, no. 6, 2020.

[5] K. Müller and F. Zotter, "The PerspectiveLiberator-An Upmixing 6DoF Rendering Plugin for Single-Perspective Ambisonic Room Impulse Responses," in *DAGA*, 2021.

[6] T. Mckenzie, N. Meyer-Kahlen, R. Daugintis, L. Mccormack, S. J. Schlecht, and V. Pulkki, "Perceptually informed interpolation and rendering of spatial room impulse responses for room transitions," in *ICA*, 2022.

[7] L. McCormack, A. Politis, T. McKenzie, C. Hold, and V. Pulkki, "Object-Based Six-Degrees-of-Freedom Rendering of Sound Scenes Captured with Multiple Ambisonic Receivers," *J. Audio Eng. Soc.*, vol. 70, no. 5, pp. 355–372, 2022.

[8] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[9] E. Fernandez-Grande and A. Xenaki, "Compressive sensing with a spherical microphone array," *J Acoust Soc Am*, vol. 139, no. 2, pp. EL45–EL49, 2016.

[10] F. M. Heuchel, E. Fernandez-Grande, F. T. Agerkvist, and E. Shabalina, "Active room compensation for sound reinforcement using sound field separation techniques," *J Acoust Soc Am*, vol. 143, no. 3, pp. 1346–1354, 2018.

[11] S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari, "Spatial Active Noise Control Based on Kernel Interpolation of Sound Field," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 3052–3063, 2021.

[12] R. Mignot, L. Daudet, and F. Ollivier, "Room reverberation reconstruction: Interpolation of the early part using compressed sensing," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 11, pp. 2301–2312, 2013.

[13] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J Acoust Soc Am*, vol. 143, no. 6, pp. 3770–3779, 2018.

[14] M. Nolan, S. A. Verburg, J. Brunskog, and E. Fernandez-Grande, "Experimental characterization of the sound field in a reverberation room," *J Acoust Soc Am*, vol. 145, no. 4, pp. 2237–2246, 2019.

[15] E. Fernandez-Grande, D. Caviedes-Nozal, M. Hahmann, X. Karakonstantis, and S. A. Verburg, "Reconstruction of room impulse responses over extended domains for navigable sound field reproduction," in *I3DA*, 2021.

[16] A. Figueroa-Duran and E. Fernandez-Grande, "Reconstruction of reverberant sound fields over large spatial domains," *J Acoust Soc Am*, vol. 157, no. 1, pp. 180–190, 2025.

[17] D. de Vries, E. M. Hulsebos, and J. Baan, "Spatial fluctuations in measures for spaciousness," *J Acoust Soc Am*, vol. 110, no. 2, pp. 947–954, 2001.

[18] J. G. Tylka and E. Y. Choueiri, "Models for evaluating navigational techniques for higher-order ambisonics," in *POMA*, vol. 30, 2017.

[19] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Springer, 2012.

[20] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J Acoust Soc Am*, vol. 116, no. 4, pp. 2149–2157, 2004.

[21] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

[22] A. Politis, *Microphone array processing for parametric spatial audio techniques*. PhD thesis, Aalto University, 2016.

[23] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head-movements," in *AES SFC*, 2013.

[24] S. A. Wirler, N. Meyer-Kahlen, and S. J. Schlecht, "Towards transfer-plausibility for evaluating mixed reality audio in complex scenes," in *AES AVAR*, 2020.

[25] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 61, 2013.

[26] A. Figueroa-Duran, N. Meyer-Kahlen, E. Fernandez-Grande, and T. Lokki, "Spatial Impulse Response Dataset: Aalto University - Arni," 2025.

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

[28] P. Stitt, S. Bertet, and M. van Walstijn, "Extended energy vector prediction of ambisonically reproduced image direction at off-center listening positions," *J. Audio Eng. Soc.*, vol. 64, no. 5, pp. 299–310, 2016.

[29] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *DAGA*, 2013.

[30] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *DAGA*, 2018.

[31] V. Pulkki, M. Karjalainen, and J. Huopaniemi, "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model," in *104th AES Convention*, 1998.

[32] T. Okano, . Leo, L. Beranek, T. Hidaka, and L. L. Beranek, "Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls," *J Acoust Soc Am*, vol. 104, no. 1, pp. 255–265, 1998.