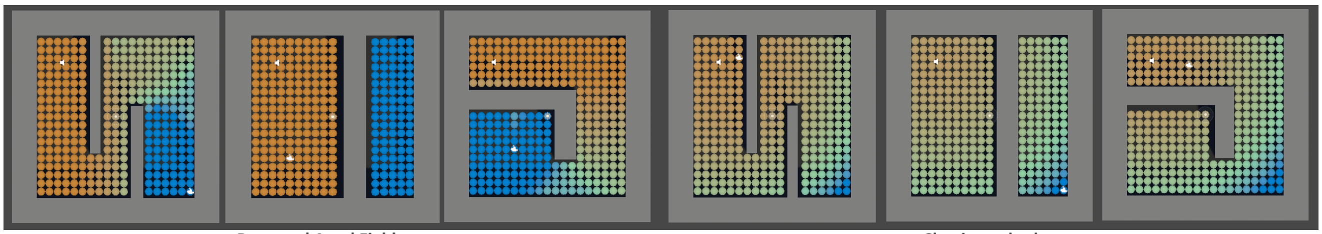
Aural Fields: A Real-Time Spatial Audio System Using Grid-Based Probes and Occlusion-Aware Attenuation

Raghav Suriyashekar
Kingston University, London, UK
K2314767@kingston.ac.uk

Vasileios Argyriou
Kingston University, London, UK
Vasileios.Argyriou@kingston.ac.uk



Proposed Aural fields (left 3-images) dynamically adapt to source positions and account for occlusions and environmental variability. Classic methods (right 3-images) use fixed zones and assume uniform, unobstructed environments

ABSTRACT

Audio plays a pivotal role in interactive applications, spanning games, simulations, virtual environments, architectural acoustics, and XR training. Even with advances in computing, real-time, physically based audio propagation is still not widely adopted, as traditional offline or precomputed methods continue to be used due to their lower computational requirements. We present *Aural Fields*, a novel runtime spatial audio system that explicitly accounts for occlusion without reliance on baked data. Inspired by global illumination in computer graphics, our method deploys volumetric acoustic probes that continuously perform ray-traced sampling of direct and transmitted sound energy. These probes define a spatially coherent acoustic field, which is interpolated at the listener to adapt seamlessly to dynamic sources and evolving geometry. Preliminary results indicate that our approach delivers perceptually coherent and stable spatial sound while balancing physical accuracy and computational efficiency for real-time and resource-constrained platforms.

KEYWORDS

Spatial audio, acoustic simulation, raytracing, real-time 3D audio, environmental acoustics

ACM Reference Format:
Raghav Suriyashekar and Vasileios Argyriou. 2025. Aural Fields: A Real-Time Spatial Audio System Using Grid-Based Probes and Occlusion-Aware Attenuation. In *Proceedings of SIGGRAPH Asia 2025, Posters (SIGGRAPH Asia 2025).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3563045.3563055

SIGGRAPH Asia 2025, December 15–18, 2025, Hong Kong

ACM ISBN /25/01
https://doi.org/10.1145/3563045.3563055

1 INTRODUCTION

Audio is essential for immersive interactive applications like games and simulations. However, despite the recent surge in neural and hardware-accelerated techniques for spatial acoustics, existing methods remain limited. Related real-time techniques, such as wave-based sound propagation using beam or wave tracing [5] and precomputed wave simulation for dynamic scenes [4], demonstrate the viability of such methods. The work in [1] uses learned implicit representations to predict impulse responses at arbitrary positions, but relies on offline training and lacks real-time adaptability to dynamic scenes. In [3] a hardware-based sound propagation method, achieves real-time performance via dedicated acceleration, yet it typically depends on pre-baked data or limited occlusion modeling, compromising flexibility. The demo in [2] integrates spatial audio in AR performances, but is tailored to fixed performance setups, not general, dynamic environments. The work in [6] enables gradient-based optimisation via differentiable rendering of GA path tracing—but is computationally heavy and unsuited for low-latency, runtime audio processing.

In contrast, Aural Fields (AF) runs fully at runtime, using a volumetric grid of probes that perform continuous, ray-traced sampling of direct and permeated sound propagation, capturing occlusion without baking. By interpolating acoustic data at listener positions, our approach supports dynamic scene changes in real time, offering a uniquely balanced solution between physical accuracy and performance.

2 METHODOLOGY FOR SPATIAL ENVIRONMENTALLY-AWARE SOUND

The proposed method deploys a volumetric grid of acoustic probes $P_{\mathbf{t}}$ positioned throughout the environment, where $\mathbf{t} \in \mathbb{R}^2$ (or $\mathbb{R}^3$ in full spatial configurations). The total number of probes is $T \geq 4$, arranged with either uniform spacing d_u or position-dependent sparse spacing $d_s(\mathbf{t})$. Each probe performs independent, continuous ray-traced sampling to estimate the acoustic visibility from its location to a set of audio sources A_m, $m \in \mathbb{Z}^+$, with $M \geq 1$.

The propagation model distinguishes between two components: direct rays (DR), representing unoccluded sound paths, and permeated rays (PR), representing attenuated transmission through or around obstacles (see Figure 1). Let $N > 0$ denote the number of rays cast per probe, and $B > 0$ the maximum number of bounces permitted per ray. In practical experiments, $B = 6$ provides a balance between computational cost and accuracy.

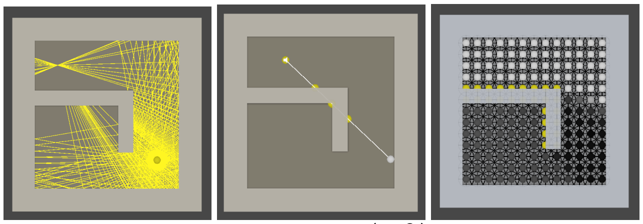


Figure 1: Raycasting attenuation. (Left) Direct Attenuation Raycasting, (Mid) Permeated Attenuation Raycasting, and (Right) Probe Values $E_{m,\mathbf{t}}$ (black-low, white-high)

2.1 Direct Ray Attenuation

From each probe $P_{\mathbf{t}}$, N rays are cast over incrementally varying azimuth and elevation angles. Each ray is propagated with specular reflection until the bounce limit B is reached. For the n-th reflected ray, represented by the unit vector $\boldsymbol{R}_n$, the geometric visibility to a source A_m is evaluated using the dot product

$$V_n = \boldsymbol{R}_n \cdot \boldsymbol{G}_n, \tag{1}$$

where $\boldsymbol{G}_n$ is the unit vector from the incident point $\boldsymbol{I}_n$ to the source. A ray is considered "visible" to A_m if $V_n \geq \cos(W)$, where W is the maximum allowed angular deviation. The direct attenuation factor for the probe–source pair $(m, \mathbf{t})$ is then

$$\mathrm{DR}_{m,\mathbf{t}} = \frac{\#\{n \mid V_n \geq \cos(W)\}}{N}. \tag{2}$$

2.2 Permeated Ray Attenuation

Permeated rays are evaluated along the line segment s connecting $P_{\mathbf{t}}$ to A_m. Let this segment be intersected by a sequence of obstacles O_i, each producing a sub-segment of length $D_{(s,i)}$ corresponding to its thickness. For an obstacle o_i, the attenuation per unit distance is given by $U_{o_i} > 0$. Starting from $E_{s,0} = 1$, the cumulative attenuation after passing through O_i is computed as

$$E_{s,i} = E_{s,i-1} \cdot \left[1 - U_{o_i} \cdot \left(1 - e^{-c \cdot D_{(s,i)}}\right)\right], \tag{3}$$

where c is a material-dependent constant. After all intersections are processed, the final permeated attenuation factor for $(m, \mathbf{t})$ is

$$\mathrm{PR}_{m,\mathbf{t}} = E_{s,\mathrm{final}}. \tag{4}$$

The direct and permeated components are combined to produce the total attenuation:

$$E_{m,\mathbf{t}} = \mathrm{PR}_{m,\mathbf{t}} + \left(1 - \mathrm{PR}_{m,\mathbf{t}}\right) \cdot \mathrm{DR}_{m,\mathbf{t}}. \tag{5}$$

2.3 Spatial Interpolation

Listener positions L_q with $q \geq 1$ do not perform raycasting directly; instead, their acoustic response is interpolated from surrounding probes. In uniform grids (d_u constant), bilinear interpolation in 2D (or 3D) is employed. In sparse configurations ($d_s(\mathbf{t})$ varying), non-uniform interpolation schemes such as inverse-distance weighting, natural neighbor interpolation, or radial basis function interpolation are applied. This produces a spatially continuous acoustic field that responds dynamically to changes in source positions and scene geometry without per-frame listener ray tracing.

Clip	Pos	Occ	SNR(dB)	RMS	ZCR	SC(Hz)	SF
Talking	Close	No	22.39	0.0295	0.1192	4840.11	0.0743
	Mid	No	12.13	0.0297	0.0463	3913.85	0.1299
	Far	No	8.99	0.0298	0.36	1662.19	0.1924
	Close	Yes	9.04	0.0298	0.0358	1645.22	0.1918
	Mid	Yes	8.82	0.0298	0.0295	2522.31	0.2099
	Far	Yes	8.94	0.0298	0.0361	1674.58	0.1929
Music	Close	No	25.95	0.1301	0.1968	6381.15	0.0591
	Mid	No	17.42	0.1299	0.1844	6276.43	0.0769
	Far	No	14.05	0.1297	0.1352	5217.13	0.0733
	Close	Yes	14.03	0.1297	0.1369	5347.51	0.0796
	Mid	Yes	14.06	0.1296	0.0679	5133.02	0.1589
	Far	Yes	14.02	0.1298	0.1353	5240.56	0.0744

Table 1: Audio metrics for Talking and Music clips.

3 RESULTS AND DISCUSSION

We evaluated the system using probe densities (low, mid, high) across listener placements from close to far, with and without occlusion. Metrics included SNR, RMS, Zero Crossing Rate, Spectral Centroid, and Spectral Flatness, capturing both signal energy and spectral behaviour (see Table 1). Without occlusion, SNR falls from over 22-25 dB at close positions to around 9-14 dB at far placements, alongside reduced spectral centroid and higher flatness, consistent with distance-related attenuation and low-pass filtering. With occlusion, SNR stabilises near 9-14 dB regardless of distance, while centroids remain around 1600/5100 Hz, producing perceptual muffling. Probe density has limited overall effect, though anomalies at lower probe counts suggest interpolation artifacts. Overall, the method reproduces plausible acoustic trends.

Method	Frame Time (ms)
Aural Fields (AF) without optimisation	7.76
Aural Fields (AF) with Frame Spreading	1.08
Aural Fields (AF) with Proximity Culling	1.02

Table 2: Performance and complexity of audio rendering.

Table 2 summarises the frame times for different Aural Fields rendering variations, measured both in the editor and the built application. Optimizations such as frame spreading and proximity culling significantly reduce computational cost, enabling smoother real-time performance.

REFERENCES

[1] Andrew Luo et. al. Learning neural acoustic fields. *NeurIPS*, 2022.
[2] Dávid Maruscsák et. al. Phases: Augmenting a live audio-visual performance with ar. In *SIGGRAPH 2024 Real-Time Live!* ACM, 2024.
[3] Eunjae Kim et. al. An architecture and implementation of real-time sound propagation hardware for mobile devices. In *SIGGRAPH Asia*, pages 1–9, 2023.
[4] Nikunj Raghuvanshi et. al. Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. In *SIGGRAPH*, 2010.
[5] Ravish Mehra et. al. Wave: Interactive wave-based sound propagation for virtual environments. In *IEEE VR/TVCG)*, pages 1–12, 2015.
[6] Ugo Finnendahl et. al. Differentiable geometric acoustic path tracing using time-resolved path replay backpropagation. *ACM Trans on Graphics*, 44(4), 2025.