Audio Engineering Society

Convention Paper

Presented at the AES 6th International Conference on Audio for Virtual and Augmented Reality and Immersive Games
June 30 - July 3, 2026, Sorbonne University/ d'Alembert & IRCAM/STMS, Paris, France

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AT_WaveSpace: a Wave Field Synthesis Engine for Research and Authoring, Applied to Near-Field Distance Perception of Focused Sources

Antoine Gonot[1], Olivier Macherey[2], and Sabine Meunier[3]

[1] *Aix Marseille Univ, CNRS, PRISM, Marseille, France*

[2] *Aix Marseille Univ, CNRS, Centrale Med, LMA, Marseille, France*

[3] *Aix Marseille Univ, CNRS, Centrale Med, LMA, Marseille, France*

Correspondence should be addressed to Antoine Gonot (antoine.gonot@univ-amu.fr)

ABSTRACT

Wave Field Synthesis is a well-established spatialization technique that solves the sweet-spot limitation of conventional sound reinforcement and uniquely allows the synthesis of focused sources — virtual sources positioned between the loudspeaker array and the listener. Despite its potential for extended reality (XR), WFS has remained confined to specialized environments such as live performance, installation or post-production workflows, with no accessible open-source tooling for research and creative authoring. We present AT_WaveSpace, an open-source WFS engine built on JUCE, distributed under the MIT licence, and integrated into Unity game engine, designed to democratize WFS for researchers, developers and creators. Building on the methodological framework of the SoundScape Renderer — which combined WFS engine development with a perceptual research platform — AT_WaveSpace serves simultaneously as a spatial audio delivery tool and as an experimental tool.

A perceptual evaluation of near-field distance perception of focused WFS sources was conducted using this framework — a dimension absent from prior literature. Using a Midpoint Comparison procedure, participants were unable to rank sources at 40–100 cm consistently, while they ranked sources at 120–150 cm in correct order. Spectral centroid analysis reveals a distance-dependent timbral variation in the proximal zone whose physical origin remains unclear. Low-frequency ILD remains the primary candidate cue for correct ranking at 120–150 cm. Perspectives for further studies are outlined.

1 Introduction to Wave Field Synthesis

Since the pioneering work of Berkhout [1], Wave Field Synthesis has been extensively studied (see [2], [3], [4], [5], [6] for reviews) and recently expressed in its simplest form, called *Planewave-Based Angle Panning (PBAB),* by Julius Smith [7], using individual loudspeaker delays. For a linear array of speaker and a virtual source S behind the array (Figure 1), the driving function $x_j(n)$ of the j[th] speaker is given by:

$$x_j(n) = s\left(n - \frac{r_{j,s}}{c}\right) \qquad (1)$$

where $r_{j,s}$ is distance between the virtual source S and the j^{th} speaker, c the speed of sound.

Focused sources require a simple time-reversal of the driving function of eq. (1), inverting the per-speaker delay [8]:

$$x_{j,focused}(n) = s\left(n - \left(\frac{r_{max}}{c} + \frac{r_{min}}{c} - \frac{r_j}{c}\right)\right) \quad (2)$$

where r_{max} and r_{min} are respectively the largest and the smallest source-speaker distances.

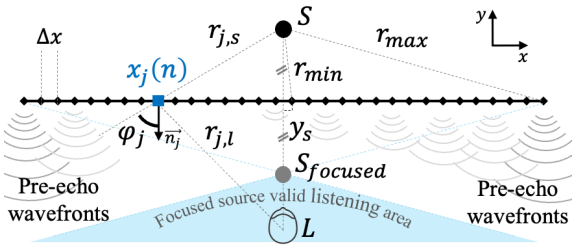


Figure 1. No-focused and focused source synthesis with a linear loudspeaker array. For the focused source, edge loudspeakers emit first, producing pre-echo wavefronts. The shaded area indicates the valid listening area (diverging region).

The 2.5D Kirchhoff-Helmholtz equation introduces three other components in the driving function of Eq (1). Expressed in the frequency domain (as a filter), the full driving function of the j[th] speaker is given by:

$$X_j(w) = \sqrt{\frac{r_{j,l}}{r_{j,s}+r_{j,l}}}\sqrt{\frac{ik}{2\pi}}\frac{\cos\varphi_j}{\sqrt{r_{j,s}}}e^{-ikr_{j,s}}S(w) \quad (3)$$

where:

- $e^{-ikr_{j,s}}$ is the delay, introduced in Eq. (1);
- $\frac{\cos\varphi_j}{\sqrt{r_{j,s}}}$ is a gain depending of the position of the virtual source, reducing the contributions of the speakers which are far away from the virtual source. This gain is sign-inverted when time-reversal is applied for focused sources;
- $\sqrt{\frac{ik}{2\pi}}$ is a half-order differentiator pre-filter (3 dB/octave high-pass and 45° phase lead), independent from the virtual source position. It can be approximated as a cascade of IIR filters [9] in the time domain;
- $\sqrt{\frac{r_{j,l}}{r_{j,s}+r_{j,l}}}$ is a gain depending on the position of the virtual source ($r_{j,s}$) and of the listening position ($r_{j,l}$). This term can be omitted when $r_{j,l} \gg r_{j,s}$. It is also common to consider a constant position for the listening point.

The 2.5D WFS driving function can be extended to a 2D loudspeaker layout. As illustrated in Figure 2, active loudspeakers are selected via a geometric criterion applied to each loudspeaker [6]:

- for non-focused sources, a loudspeaker is muted if $\overrightarrow{r_{j,s}}.\overrightarrow{n_j} < 0$, i.e. if it faces the virtual source;
- for focused sources, a loudspeaker is muted if $\overrightarrow{r_{l,s}}.\overrightarrow{r_{j,s}} < 0$, i.e. if the loudspeaker lies on the opposite side of the focused source from the listener.

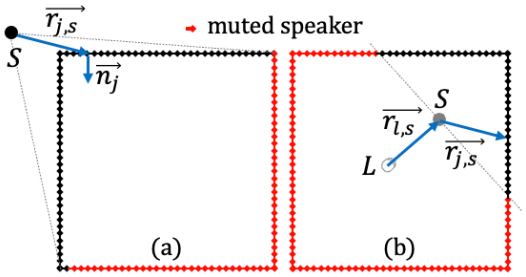


Figure 2. (a) Non-focused source: a speaker is muted if $\overrightarrow{r_{j,s}}.\overrightarrow{n_j} < 0$. (b) Focused source: a speaker is muted if $\overrightarrow{r_{l,s}}.\overrightarrow{r_{j,s}} < 0$.

2 Perception of focused sources

Focused sources produce a converging wave field between the loudspeaker array and the focus point, and a diverging wave field beyond it. Only the diverging region — where the synthesized field radiates outward from the focus as if from a point source — constitutes a valid listening area [8] (shaded area in Figure 1).

Spatial sampling and array truncation

A discrete loudspeaker array spatially samples the continuous driving function at intervals Δx, which is artifact-free only for frequencies below the aliasing frequency [8]:

$$f_{al} \leq \frac{c}{2\Delta x} \quad (4)$$

Below f_{al}, loudspeaker contributions cancel each other across the listening area except at the desired wavefront. Above f_{al}, this cancellation fails and additional wavefronts reach the listener before the desired signal — so-called pre-echos [8], [10]. Since they exist solely above f_{al}, pre-echos carry only high-frequency energy and are perceived either as a high-pass filtered distortion or, when their amplitude exceeds the precedence effect threshold, as a distinct additional source [8], [10]. Truncation of the loudspeaker array introduces a further trade-off: a shorter array reduces pre-echo artifacts but widens the focal point and introduces erroneous Interaural Level Difference (ILD) via diffraction at the array edges, thereby corrupting binaural localization cues [8], [10].

Aliasing-free zone around the focus
A remarkable property of focused sources is that aliasing does not affect the entire listening area uniformly. At the focus point itself, all loudspeaker contributions superimpose with equal phase regardless of frequency, producing constructive interference; moving away from the focus, this phase alignment degrades and aliasing artifacts progressively emerge. The size of this aliasing-free region shrinks with increasing frequency, and its radius is quantified by [10]:

$$r_{\mathrm{al}} = \frac{y_s \cdot c}{f \cdot \Delta x} \qquad (5)$$

where y_s is y-position of the focus source (Figue 1), c the speed of sound, f the frequency, and Δx the inter-loudspeaker spacing.

Perceptual studies and methodology
Formal perceptual investigations by Wierstorf et al. [10] identified two main concerns when looking at the rendering of WFS: the amount of audible artifacts and the perceived azimuthal localization of the focused source — both depending on array length and listener position. These studies were made possible by the SoundScape Renderer [11], a WFS rendering engine developed concurrently by the same research group both as a standalone spatial audio tool and as an experimental platform. By leveraging this infrastructure for dynamic binaural resynthesis with head tracking, Wierstorf et al. demonstrated that binaural virtualization — rendering the full loudspeaker array over headphones — provides a valid and reproducible methodology for investigating the perception of focused sources without requiring a physical array [10]. This dual development — advancing both the rendering engine and the perceptual methodology in parallel — proved instrumental in enabling rigorous experimental investigation, and underlines the value of conducting rendering tool design, physical measurements, and perceptual evaluations as a unified research effort.

3 The need for a new rendering tool

Existing tools — whether academic (IRCAM Spat [12], WFSCollider [13], SoundScape Renderer [11]) or commercial (SPAT Revolution, Holophonix) — are primarily designed for live performances, installations, or high-end post-production workflows. They operate either as standalone processes communicating via inter-process protocols (OSC/UDP), or as host-plug-in architectures (VST/AU/AAX) embedded within digital audio workstations, and typically require specialized hardware infrastructure — such as the Adamson Fletcher Machine or the Holophonix dedicated processor — or proprietary environments (Max/MSP). Neither paradigm provides native integration within a real-time 3D engine, leaving WFS difficult to access for the game development and interactive media communities.

Unlike these standalone systems, a game engine provides native scene-graph integration that is fundamentally better suited to interactive 3D audiovisual design. Virtual sources are represented as scene graph nodes sharing position, lifecycle, and event system with the visual and physics simulation. This architectural choice unlocks several capabilities unavailable to middleware-based solutions:

- **dynamic instantiation and destruction** of spatialized sources at runtime;
- **native peripheral integration**, as game engines expose a unified XR abstraction layer supporting HMDs, motion capture, and gesture sensors;
- **scripting extensibility**, with no architectural limit on custom interaction logic (C# for Unity, C++ for Unreal);
- **networked deployment**, as game engine networking frameworks (Mirror and Netcode for GameObjects in Unity, NetDriver in Unreal) expose replicated scene objects whose spatial audio properties propagate transparently across clients.

As noted by Fırat & Masullo [14], game engines are preferred platforms for VR designers precisely because of this hardware interfacing and component reusability.

Finally, from an academic perspective — whether for research or teaching — a useful tool should be (i) open-source, (ii) cross-platform, (iii) editable, (iv) ideally built on a well-established and efficient audio library, and (v) designed to facilitate signal processing experimentation. While IRCAM's Spat is a powerful and widely used tool in research and creative projects, it relies on closed-source Max/MSP externals, making its DSP internals opaque and non-editable. The SoundScape Renderer — which, as discussed in Section 2, played a central role in WFS perceptual research by enabling binaural virtualization — is designed as a standalone application rather than as an embeddable library. Its audio processing core is tightly coupled to the JACK audio layer and Qt-based GUI, and its legacy build system (GNU Autotools) with multiple external dependencies makes compilation non-trivial and cross-platform deployment — particularly on

Windows — unreliable. Reusing its WFS rendering engine in an independent context therefore requires decoupling components that were not initially designed for modular reuse.

The present work is directly inspired by this research tradition — developing a rendering engine not merely as a delivery tool but as an experimental platform tightly coupled to physical measurements and perceptual evaluation. AT_WaveSpace can be understood as a modernized successor to the SoundScape Renderer: an open-source WFS rendering engine built on the JUCE framework — a cross-platform C++ library widely adopted in the audio research community — distributed under the MIT licence, and designed from the ground up for modular DSP experimentation and straightforward compilation across Windows, macOS and Linux.

4 The At_WaveSpace Audio Engine

AT_WaveSpace is a C++ dynamic library exposing a plain C interface, built on the JUCE 8 framework. JUCE provides both fundamental signal processing primitives — parameter smoothing, fractional delay, filtering, convolution, multichannel audio file decoding — and higher-level features such as multithreading, which plays a key role in the engine's performance: handling large numbers of simultaneous audio players and parallelizing HRTF convolutions across WFS output channels when binaural virtualization is enabled. The engine's main features include:

- **real-time WFS rendering** for arbitrary loudspeaker configurations, with configurable driving function parameters (amplitude weighting and pre-filtering);
- **binaural HRTF virtualization** of the full loudspeaker array, enabling headphone-based monitoring without a physical array;
- **a Distance Variation Function (DVF)** for near-field binaural rendering;
- **a plain C interface** for straightforward integration into any host environment.

The full source code, Unity package, and documentation are available in the dedicated repository[1]. It provides two build targets: a dynamic library for Unity integration, and a standalone console application — the latter was used to drive the loudspeaker array and to generate the stimulus corpus for the perceptual evaluation described below. Readers interested in implementation details are encouraged to consult it directly.

The Unity integration is illustrated in Figure 3. Virtual loudspeakers are standard Unity GameObjects; the At_MasterOutput, At_Player, and At_Listener MonoBehaviours synchronize their transforms to the DSP layer each frame and persist scene state as a JSON file.

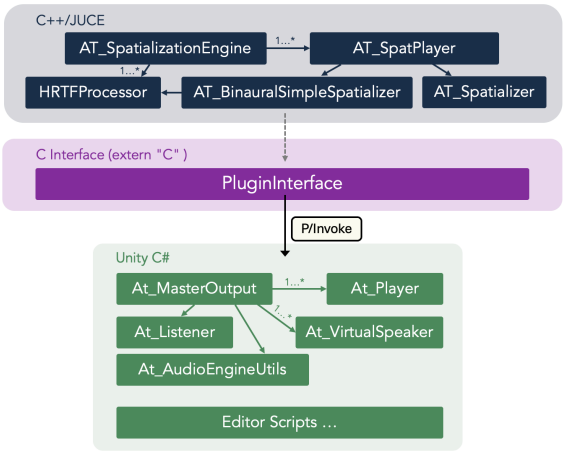


Figure 3. AT_WaveSpace: architecture of the C++/JUCE engine and its Unity C# integration.

The binaural virtualization feature deserves particular attention in the context of this study. Binaural monitoring of multichannel formats has become common practice in post-production (e.g. Dolby Atmos Renderer, Holophonix) and play a key role in perceptual experiments using the SoundScape Renderer, as mentioned earlier. AT_WaveSpace provides equivalent binaural monitoring by convolving each virtual loudspeaker's driving signal with the HRTF at the corresponding azimuth and further summing contributions from all loudspeakers. While Wierstorf et al. [10] showed that such a virtualization allows a listener to infer the azimuth of a focused source over headphones, it suffered from the same limitations as classical binaural synthesis concerning distance perception in the proximal region. This is due to the fact that virtualized focused sources do not reproduce the increase in ILD with source proximity, a cue whose importance in near-field auditory distance perception has already been established by Brungart [15]. To address this limitation, a Distance Variation Function (DVF) was implemented following Kan et al. [16]. It consists of a first-order IIR filter derived from the low-frequency approximation of the rigid-sphere head model [17],

[1] https://github.com/agonotamu/at-wavespace-unity-sdk

first introduced as a structural binaural filter by Brown & Duda [18]. This filter allows scaling far-field HRTFs to arbitrary near-field distances (Eq. 6)

$$H(s) = \frac{s + \frac{c}{r_{s,l}}}{s + \frac{c}{r_0}} \qquad (6)$$

where r_0 is the HRTF measurement distance and $r_{s,l}$ the target source distance from the listener in the scene.

The binaural rendering mode — with or without DVF — will not be evaluated in the present study. The perceptual evaluation presented below is a preliminary investigation serving two purposes: (1) validating the WFS rendering pipeline of AT_WaveSpace under controlled experimental conditions, and (2) identifying open research questions related to the perceptual cues involved in near-field distance perception of focused sources — in particular the respective contributions of low-frequency ILD and spectral centroid variations — that will be addressed in future work.

5 Evaluation of static distance perception of focused sources

Background

As discussed in Section 2, prior perceptual studies of focused WFS sources [8], [10] addressed azimuthal localization and perceptual artifacts, but none has investigated distance perception.

The only prior investigation related to distance perception was presented by Wittek et al. [19], who showed that ILD deviations between focused WFS sources and real sources become significant below 65 cm — a finding that motivates the present investigation of distance perception in the proximal region. The present study is grounded in the hypothesis that low-frequency ILD may not constitute a reliable cue for distance perception of focused sources, and that listeners may consequently be unable to infer distance in the absence of other cues, such as intensity or apparent azimuth variations. We therefore propose this preliminary study investigating distance perception of focused WFS sources rendered over a physical loudspeaker array, for source-listener distances below 150 cm.

Motivations of the protocol and apparatus

Before addressing this question, we first sought to verify and extend the observations of Wittek et al. by recording binaural signals of both focused and real sources using a KEMAR mannequin at source-listener distances ranging from 20 to 150 cm (from the center of the head). The stimuli were sequences of five rectangularly gated 150-ms pulses of white Gaussian noise, separated by 30-ms intervals of silence. This stimulus type is appropriate both for ILD measurement and for distance perception tasks, as established by Brungart [15]. These measurements informed the choice of the two experimental conditions — C1 and C2 — evaluated in the perceptual study described below.

Note that the amplitude weighting $\sqrt{\frac{r_{j,l}}{r_{j,s}+r_{j,l}}}$ of the WFS driving function in Eq. (3) was omitted from the focused source stimuli, so that overall level remains constant across distances and distance judgements can only be based on binaural and spectral cues.

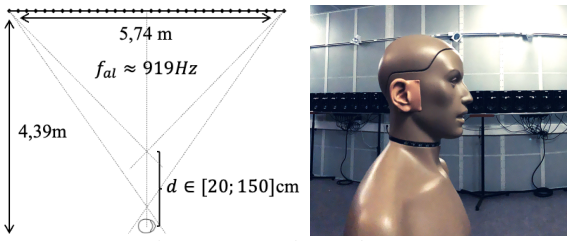


Figure 4. Experimental setup.

A 32-loudspeaker JBL 305P MkII linear array (18.5 cm spacing, 5.74 m aperture, $f_{al} \approx 927$ Hz) was deployed in a normalized listening room (T60 < 150 ms). The KEMAR mannequin was positioned at 4.39 m from the array (azimuth −90°), left ear facing the array center (Figure 4).

Stimuli of the experiment

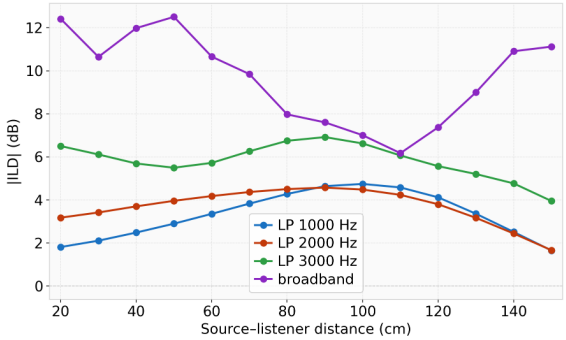


Figure 5. Absolute values of ILD (dB) across distance for a focused source at distance $d \in [20; 150]$ cm in four frequency bands.

The ILD was analyzed across source-listener distances from 20 to 150 cm in four frequency bands: [20–1000 Hz], [20–2000 Hz], [20–3000 Hz], and [20–20000 Hz] (broadband).

$$ILD = 10 \times log_{10}\left(\frac{\langle s_{bp,R}^2 \rangle}{\langle s_{bp,L}^2 \rangle}\right) \qquad (7)$$

where $s_{bp,L}$ and $s_{bp,R}$ are the filtered left and right channels of the binaural recordings. An 8th order band-pass Butterworth filter is applied in zero-phase mode. The mean value is computed over the entire signal.

As illustrated in Figure 5, the ILD of focused sources as a function of distance exhibits a non-monotonic pattern at low frequencies, in sharp contrast with real sources (Figures 6).

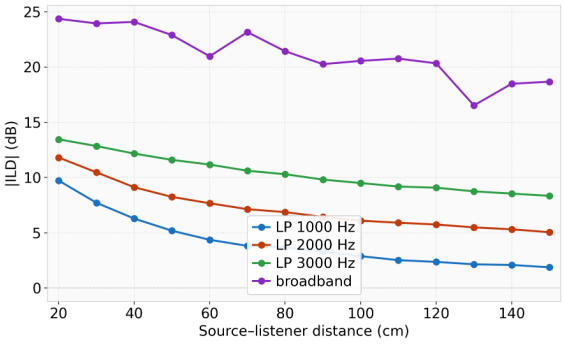


Figure 6. Absolute values of ILD (dB) across distance for a real source at distance $d \in [20; 150]$cm in four frequency bands.

For real sources (JBL 305P MkII loudspeaker at various distances), low-frequency ILD increases monotonically with decreasing distance, reaching 9.7 dB at 20 cm in the [20–1000 Hz] band. For focused sources, the same band peaks at 4.7 dB around 90–100 cm, then decreases toward shorter distances, reaching only 1.8 dB at 20 cm. This reversal, observed consistently below 2 kHz, defines two distinct distance regimes:

- above 90–100 cm: ILD magnitude increases as distance decreases — consistent with real source behavior;
- below 90–100 cm: ILD magnitude decreases as distance decreases — opposite to real source behavior.

Broadband ILD, by contrast, exhibits irregular non-monotonic behavior for both source types.

We recorded the focused source with and without the amplitude weighting component of the 2.5D WFS driving function ($\frac{\cos \varphi_j}{\sqrt{r_{j,s}}}$) of Eq. (3). Comparing low-frequency ILD ([0–1000 Hz]) in both cases (Table 1), the no-weighting condition yields a larger ILD dynamic range across the selected distances (4.17 dB vs. 3.09 dB peak-to-minimum)

The distances were divided into two conditions — C1 (up to and including 100 cm, “reversal” zone) and C2 (above 100 cm, “correct" zone) — with 20-cm spacing for C1 and 10-cm spacing for C2 to adequately sample the ILD decay towards the far field. ★ denotes the reversal peak at 100 cm. Distances of 20 and 30 cm were excluded to maintain an equal number of stimuli per condition (four each), and because the ILD step between consecutive distances is smaller in this range, reducing inter-stimulus discriminability within the condition. Although the metric spacing is larger in C1 (20 cm) than in C2 (10 cm), Table 1 shows that the resulting ILD increments between consecutive distances are comparable across both conditions (0.88–1.24 dB for C1, 1.01–1.20 dB for C2), ensuring equivalent perceptual discriminability.

Distance (cm)	ILD (dB) no-weighting	ILD (dB) weighting	Δ	Condition
20	1.92	1.81	0.11	excluded
30	2.27	2.11	0.16	excluded
40	2.70	2.48	0.22	C1
60	3.76	3.35	0.41	C1
80	5.00	4.28	0.72	C1
100 ★	5.88	4.74	1.14	C1
120	5.12	4.12	1.00	C2
130	4.11	3.36	0.75	C2
140	2.91	2.51	0.40	C2
150	1.71	1.65	0.06	C2

Table 1. ILD (dB) comparison for the focused sources with and without amplitude weighting.

For such source and listener positions, the effect of spatial aliasing should be reduced, because the sources and the listener are always centred relative to the speaker array and close from each other. Table 2 give an estimation of the aliasing frequency according to Eq. 5, given the frequency above which the distance from the focus source to the left ear is smaller than r_{al}. It shows that condition C2 could be much more affected than condition C1.

d	40	60	80	100	120	130	140	150
f_{al}	>20	13.7	9.3	6.8	5.3	4.7	4.2	3.7

Table 2. Aliasing frequency (kHz) across distance, corresponding to the frequency above which the distance from the focus source to the left ear is smaller than r_{al} (Eq. 5).

Evaluation and results

The 8 binaural recordings of the focused sources were separated to form two experimental conditions, with four stimuli that participants had to rank in terms of perceived distance. The order of the condition was

randomized. Distance ranking was assessed using the optimally efficient Midpoint Comparison (MPC) procedure, as introduced by Long et al. [20]. Participants made a series of two-interval forced-choice comparisons ("Which sound comes from farther away?") without feedback. We used the best-of-three modification as implemented in de Groote et al. [21]. The procedure was run up to three times per condition, each time with the stimuli entered in a different random order. For each condition, the mean distance rank and standard deviation were calculated across runs, and compared to the physical distance order using Spearman's ρ.

A training experiment was run with the stimuli recorded from the real source at several distances (20, 50, 90, 130 cm). To remove loudness differences between stimuli, the corpus was loudness-normalized via iterative scalar gain on both channels (Zwicker ISO 532-1 [22], using MOSQITO Toolbox[2]), thereby preserving spectrum and binaural cues. The application used for the test (including corpus) is available on a dedicated repository on GitHub[3].

Twenty-three participants took part in the evaluation (N = 23; 10 acoustic experts, 13 non-experts). Three participants were excluded from the analysis after reporting known hearing difficulties or a recent event likely to have affected their hearing. The remaining 20 participants formed the analyzed sample. As shown in Table 3, participants ranked sources significantly better in C2 than in C1 (ρ = +0.67 vs −0.22, Wilcoxon paired test: Δρ = +0.89, p < 0.01).

Series	Group	Mean ρ	SD	p	Sign.
C1	All	-0.22	0.64	< 0.05	No
C2	All	+0.67	0.38	< 0.001	High
C2>C1	Wil. paired	$\Delta\rho$=+0.89	-	< 0.01	Yes

Table 3. Mean Spearman ρ correlation with physical distance order, by condition. Results are consistent across expert and non-expert groups.

The strong positive correlation in C2 (p < 0.001) confirms that participants consistently ranked sources in the correct distance order, in agreement with the monotone low-frequency ILD profile observed for this condition. Table 4 shows the individual ρ values for C1: 7 participants showed a tendency toward inverse order, 3 toward correct order, and 10 remained ambiguous — with no dominant strategy emerging across the group.

	Mean ρ	SD	Class
S01	-0.9333	0.0943	inverse order
S02	-0.0667	0.5249	ambiguous
S04	0.6	0.1633	correct order
S05	1.0	0.0	correct order
S06	-1.0	0.0	inverse order
S07	-0.3333	0.5735	ambiguous
S08	0.4	0.0	ambiguous
S09	-0.8667	0.0943	inverse order
S10	0.6667	0.3399	correct order
S11	-0.4667	0.4714	ambiguous
S12	-0.7333	0.2494	inverse order
S13	0.2667	0.4989	ambiguous
S15	-0.2667	0.0943	ambiguous
S16	-0.9333	0.0943	inverse order
S18	-0.8	0.0	inverse order
S19	-0.4	0.0	ambiguous
S20	-0.0667	0.66	ambiguous
S21	-0.4	0.5657	ambiguous
S22	0.3333	0.411	ambiguous
S23	-0.8	0.0	inverse order

Table 4. Individual Spearman ρ and classification for condition C1 (40–100 cm). Participants are classified as correct order (ρ ≥ 0.6), inverse order (ρ ≤ −0.6), or ambiguous otherwise.

6 Discussion and perspectives

In this static listening condition, distance inference may rely on two distinct cues: low-frequency ILD, which increases with decreasing source distance (Brungart and Rabinowitz [23]), and spectral centroid, which increases as real sources approach the listener (Coleman [24]). These results raise two specific questions:

- is low-frequency ILD the cue responsible for correct distance ranking in condition C2?
- what are the causes of the confusion observed in condition C1?

More broadly, does the confusion in C1 reflect a fundamental inability to infer distance without head movement?

The spectral centroid of the stimuli is shown in Figure 7. In condition C1, the mean spectral centroid increases with source distance (5719 Hz at 40 cm to 6318 Hz at 100 cm), driven primarily by the spectral centroid of right ear. This trend is opposite to the heuristic expected for real sources, where a closer source produces a brighter timbre. Surprisingly, this spectral variation occurs despite the fact that, as shown in Table 2, the aliasing frequency remains above 9 kHz for all C1 distances, meaning the left ear

[2] https://github.com/Eomys/MoSQITo

[3] https://github.com/agonotamu/at-wavespace-eval-app

stays well within the aliasing-free zone. The physical origin of this spectral centroid variation in C1 therefore remains unclear and deserves further investigation. In condition C2, the mean spectral centroid shows no consistent trend (6218–6436 Hz, a variation of only 218 Hz), providing no reliable spectral cue for distance judgment. This leaves low-frequency ILD as the primary candidate to explain the correct rankings observed in C2.

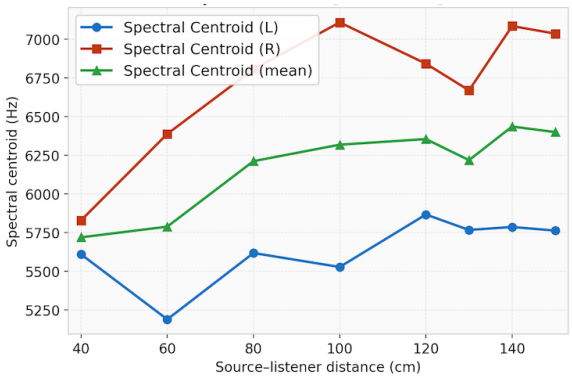


Figure 7. Left, Right and mean spectral centroid of the 8 stimuli of the perceptual evaluation

These observations are preliminary and call for further investigation. Whether spectral centroid actively contributes to distance judgments — or merely introduces conflicting cues — cannot be determined from the present data alone.

Several directions emerge from this study. First, repeating the evaluation with stimuli lowpass-filtered below the aliasing frequency ($f_{al} \approx 927$ Hz) would isolate the low-frequency ILD cue and test whether correct ranking in C2 is maintained when spectral aliasing artifacts are removed. Second, the same protocol should be conducted with physical loudspeaker rendering, allowing free head movement. Measuring head rotations during the task would provide a direct indicator of listeners' reliance on dynamic cues — in particular auditory parallax, manifested as apparent azimuth variation with head movement — and enable a comparison between static and dynamic distance perception strategies for focused sources. In the longer term, the knowledge gained from these studies on near-field distance perception of focused sources could inform the improvement of binaural WFS virtualization for proximal source rendering — a capability already implemented in AT_WaveSpace through the Distance Variation Function (DVF).

7 Conclusions

This paper presented AT_WaveSpace, an open-source WFS rendering engine built on the JUCE framework, distributed under the MIT licence, and integrated into Unity as native scene-graph component. Designed to serve simultaneously as spatial audio delivery platform and as an experimental research tool, it provides modular access to all driving function parameters, straightforward cross-platform compilation, and a standalone console application mode for corpus generation.

The perceptual evaluation of near-field distance perception of focused sources was conducted entirely within this framework, illustrating the value of the tight coupling between rendering tool and experimental design. In particular, the decision to suppress the amplitude weighting component of the WFS driving function — motivated by the need to isolate binaural and spectral cues from level variations — was made possible by the direct accessibility of the engine's DSP parameters, and resulted in a measurable improvement in inter-stimulus ILD discriminability. This back-and-forth between engine design and experimental constraints is precisely the workflow that AT_WaveSpace was conceived to support.

The results of this preliminary evaluation raise open questions regarding the perceptual cues underlying near-field distance judgment of focused sources. Each of the perspectives outlined above will benefit from the same iterative approach — advancing the rendering tool and the perceptual methodology in parallel — as demonstrated here.

References

[1] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis", *Journal of the Acoustical Society of America*, vol. 93, n° 5, p. 2764-2778, 1993, doi: 10.1121/1.405852.

[2] R. Nicol, « Restitution sonore spatialisée sur une zone étendue: Application à la téléprésence », Thèse de doctorat, Université du Maine, 1999.

[3] E. Corteel, « Caractérisation et extensions de la wave field synthesis en conditions réelles », Thèse de doctorat, Paris 6, 2004.

[4] T. Caulkins, « Caractérisation et contrôle du rayonnement d'un système de Wave Field Synthesis pour la situation de concert », Thèse de doctorat, Paris 6, 2007.

[5] E. N. G. Verheijen, "Sound Reproduction by Wave Field Synthesis", PhD Thesis, Delft University of Technology, Delft, The Netherlands, 1997.

[6] R. Oldfield, "Objective and Subjective Analysis of Wave Field Synthesis with Focused Sources", PhD Thesis, University of Salford, 2013.

[7] J. O. Smith, "A Spatial Sampling Approach to Wave Field Synthesis: PBAP and Huygens Arrays", 2018. doi: 10.48550.

[8] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and Perceptual Properties of Focused Virtual Sources in Wave Field Synthesis", in *Proceedings of the 127th AES Convention*, New York, USA, 2009.

[9] C. D. Salvador, "Wave Field Synthesis Using Fractional Order Systems and Fractional Delays", in *Proceedings of the 128th AES Convention*, London, UK, 2010.

[10] H. Wierstorf, A. Raake, M. Geier, and S. Spors, "Perception of Focused Sources in Wave Field Synthesis", *Journal of the Audio Engineering Society*, vol. 61, nº 1/2, p. 5-16, 2013.

[11] M. Geier, J. Ahrens, and S. Spors, "The SoundScape Renderer: A Unified Spatial Audio Rendering Framework for Arbitrary Rendering Methods", in *Proceedings of the 124th AES Convention*, Amsterdam, The Netherlands, 2008.

[12] T. Carpentier, M. Noisternig, and O. Warusfel, "Twenty years of Ircam Spat: looking back, looking forward", in *Proceedings of the International Computer Music Conference (ICMC)*, 2015, p. 270-277.

[13] Snoei, W., Negrao, M., Ganchrow, R., and Truetzler, J., *GameOfLife/WFSCollider-Class-Library*. (2026). SuperCollider. GameOfLife. https://github.com/GameOfLife/WFSCollider-Class-Library

[14] H. B. Fırat and M. Masullo, "3D sound spatialization with game engines: the virtual acoustics performance of a game engine and a middleware for interactive audio design", *Virtual Reality*, 2021, doi: 10.1007/s10055-021-00589-0.

[15] D. S. Brungart, "Auditory localization of nearby sources. II. Localization of a broadband source", *Journal of the Acoustical Society of America*, vol. 106, nº 4, p. 1956-1968, 1999, doi: 10.1121/1.428062.

[16] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function", *Journal of the Acoustical Society of America*, vol. 125, nº 4, p. 2233-2242, 2009, doi: 10.1121/1.3082108.

[17] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model", *Journal of the Acoustical Society of America*, vol. 104, nº 5, p. 3048-3058, 1998, doi: 10.1121/1.423886.

[18] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis", *IEEE Transactions on Speech and Audio Processing*, vol. 6, nº 5, p. 476-488, 1998, doi: 10.1109/89.709673.

[19] H. Wittek, S. Kerber, F. Rumsey, and G. Theile, "Spatial Perception in Wave Field Synthesis Rendered Sound Fields: Distance of Real and Virtual Nearby Sources", in *Proceedings of the 116th AES Convention*, Berlin, Germany, 2004.

[20] C. J. Long *et al.*, "Optimizing the clinical fit of auditory brain stem implants", *Ear Hear*, vol. 26, nº 3, p. 251-262, 2005, doi: 10.1097/00003446-200506000-00002.

[21] E. de Groote, O. Macherey, J. M. Deeks, S. Roman, and R. P. Carlyon, "Temporal Pitch Perception of Multi-Channel Stimuli by Cochlear-Implant Users", *JARO*, vol. 26, nº 3, p. 301-315, 2025, doi: 10.1007/s10162-025-00983-4.

[22] International Organization for Standardization, "ISO 532-1:2017 Acoustics – Methods for calculating loudness – Part 1: Zwicker method". Geneva, Switzerland, 2017.

[23] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions", *Journal of the Acoustical Society of America*, vol. 106, nº 3, p. 1465-1479, 1999, doi: 10.1121/1.427180.

[24] P. D. Coleman, "Dual role of frequency spectrum in determination of auditory distance", *Journal of the Acoustical Society of America*, vol. 44, p. 631-632, 1968, doi: 10.1121/1.1911130.