A PERCEPTUALLY INSPIRED SINGLE-PARAMETER AUDITORY DISTANCE RENDERER FOR MUSIC PRODUCTION

Stefanos Biliousis and Cumhur Erkut

Department of Architecture, Design and Media Technology
Aalborg University
Copenhagen, Denmark
sbilio25@student.aau.dk, cer@create.aau.dk

ABSTRACT

Conveying auditory distance in a digital audio workstation requires balancing several uncoupled tools (reverb, gain, equalization, pre-delay) by hand, a workflow that is cognitively demanding and easily produces spatially incoherent results. We demonstrate a real-time VST3 plugin, that derives five correlated distance cues from a single normalized control and keeps them mutually coherent by construction, grounded in the psychoacoustics of auditory distance perception. A headphone listening test with 18 participants showed that the coupled renderer roughly halves distance placement error relative to an uncoupled manual mix and was unanimously preferred on composite spatial quality. Users sweep one knob and hear sources move convincingly from near to far on multitrack material, compare the result against a manual uncoupled mix, and toggle an optional binaural externalization stage. Source code is available on github[1].

1. INTRODUCTION

Auditory distance is a holistic percept: listeners fuse several correlated acoustic cues, with the direct-to-reverberant energy ratio (DRR) dominant in enclosed spaces, supported by sound intensity, high-frequency air absorption, reverberation onset time, and a level-dependent spectral tilt [1, 2, 3]. These cues reinforce one another, so combining DRR and intensity yields more reliable distance judgments than either cue alone [4]. Conversely, when the cues disagree, listeners perceive the result as spatially unnatural even when they cannot identify the specific cause of the inconsistency [5].

In contemporary music production, these cues are nonetheless manipulated through independent plugin instances with no intrinsic coupling. Placing a source at a perceived distance means adjusting reverb send, high-frequency roll-off, gain, and pre-delay separately and by ear. While directional localization is extensively served by binaural technologies [6], distance rendering has received comparatively little attention in production tools, and no existing commercial tool couples all five cues through a single physically grounded parameter.

This demonstration presents a renderer that removes that burden by exposing distance as a single normalized parameter d and deriving all five cues from it through physically and psychoacoustically grounded mapping functions. Cue coherence thereby becomes an invariant of the system rather than a property the engineer must maintain by judgment. The renderer is a cross-platform VST3 plugin implemented in JUCE [7, 8], simulating a $1\,\mathrm{m}$ to $20\,\mathrm{m}$ range with an optional binaural stage for headphone externalization.

[1] https://github.com/stefbil/distance-space

2. SINGLE-PARAMETER DISTANCE RENDERING

2.1. Cue mapping

Let $d \in [0, 1]$ denote the normalized distance parameter, mapped to physical distance

$$r(d) = 1 + d\,(R_{\max} - R_{\text{ref}}) = 1 + 19d \quad (\mathrm{m}), \qquad (1)$$

with $R_{\text{ref}} = 1\,\mathrm{m}$ and $R_{\max} = 20\,\mathrm{m}$, a span chosen to match the staging depths typical of musical scenes. All five cues derive from $r(d)$. Source level follows the free-field inverse-square law [9],

$$G(d) = -20 \log_{10}\bigl(r(d)/R_{\text{ref}}\bigr) \quad (\mathrm{dB}), \qquad (2)$$

subtending $26\,\mathrm{dB}$ across the control range. The reverb wet level follows a square-root law approximating the power-law relation between DRR and perceived distance established by Zahorik [1],

$$W(d) = W_{\max}\sqrt{d}, \qquad W_{\max} = 0.45, \qquad (3)$$

and is mixed against the dry path with a constant-power crossfade ($G_{\text{dry}} = \sqrt{1 - W}$, $G_{\text{wet}} = \sqrt{W}$) so that energy is preserved during distance sweeps [10].

High-frequency air absorption is rendered by a Butterworth high shelf whose gain is linear in d, reaching $-14\,\mathrm{dB}$ at $d = 1$, and whose corner frequency sweeps logarithmically from $8\,\mathrm{kHz}$ to $2.5\,\mathrm{kHz}$, matching the logarithmic frequency resolution of hearing [11]. The attenuation is deliberately hyper-real, exceeding the physical ISO 9613-1 prediction over these distances [12], which listeners tolerate and indeed expect for musical material [13]. Reverberation onset is controlled by a pre-delay $\tau(d) = 30\,\mathrm{ms} \cdot d$, used as a production-oriented depth cue with its ceiling at the echo-perception threshold [14]. Finally, a $2.5\,\mathrm{dB}$ parametric peak at $300\,\mathrm{Hz}$ ($Q = 1$) compensates the spectral thinning of distant, quieter sources predicted by the equal-loudness contours of Fletcher and Munson [15] as revised in ISO 226 [16]. Table 1 lists the mapped values at representative d.

The five mappings are structurally heterogeneous: shelf gain and pre-delay are linear in d, whereas intensity, wet level, and shelf frequency compress the distance axis in physically or perceptually motivated ways (Figure 1). As a result, the dominant character of a

Table 1: Mapped DSP parameter values at representative values of the distance parameter d, computed from the mappings of Sec. 2.1.

d	Gain (dB)	HF Shelf (dB)	Shelf (Hz)	Wet (%)	Pre-delay (ms)
0.00	0.0	0.0	8000	0	0.0
0.10	−9.2	−1.4	7122	14	3.0
0.25	−15.2	−3.5	5981	22	7.5
0.50	−20.4	−7.0	4472	32	15.0
0.75	−23.7	−10.5	3344	39	22.5
1.00	−26.0	−14.0	2500	45	30.0

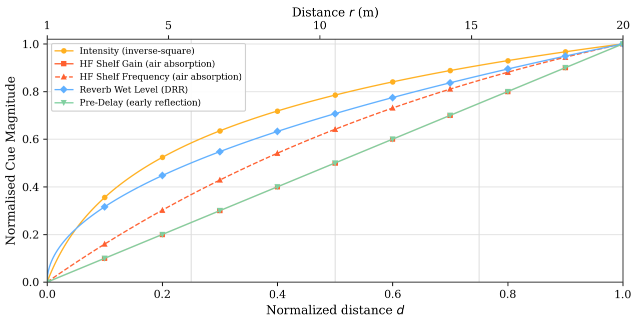

Figure 1: Normalized cue magnitudes as a function of the distance parameter d (bottom axis) and corresponding physical distance r in meters (top axis). Each curve is normalized to unity at $d = 1$.

distance step shifts from level-driven near the listener to reverberant and spectral at far range, consistent with the finding that DRR and spectral cues gain relative weight with distance [4, 2].

2.2. Signal topology

The processing chain (Figure 2) follows a parallel direct/reverberant topology after gain and spectral shaping. The reverberant branch is a stereo pre-delay ring buffer followed by a wet-only Freeverb stage [17], whose room size scales with d and with a discrete Small/Medium/Large selector; the constant-power dry/wet law of Sec. 2.1 is applied at the final mixer. All mapped targets are smoothed before each stage, so automating d produces artifact-free sweeps with no zipper noise.

An optional binaural stage, intended purely as a headphone externalization layer rather than a direction-resolved renderer, convolves the direct path only with ear-specific KEMAR impulse responses [18] (user impulse responses may be substituted), keeping the diffuse field outside the HRTF. Its contribution is blended in with a sigmoid weight centered at $d = 0.5$, so externalization emerges gradually in the perceptually intermediate range.

3. INTERFACE AND DEMONSTRATION

The interface (Figure 3) is organized around a single large distance dial, labeled in meters, alongside the discrete Room Size selector and the HRTF toggle. The right-hand panel gives live visual feedback that is decoupled from the audio thread: the composite EQ curve, a signal scope, and cue-activity indicators that show how each cue responds as the single control moves. This makes the otherwise hidden cue coupling directly observable while it is being heard.

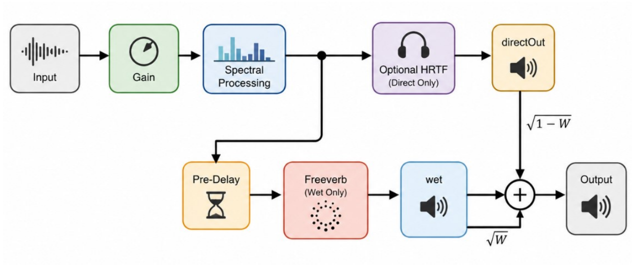

Figure 2: Real-time signal topology. The reverberant branch is stereo and non-directional; optional binaural filtering applies to the direct path only, and the two branches recombine under a constant-power law.

Figure 3: Plugin interface. A single distance dial drives all five cues; the cue-activity display makes the coupling visible as the control moves.

Users can interact with the plugin hands-on through three short activities:

Starting from prepared multi-track sessions, visitors move the single distance control and hear individual sources travel coherently from near to far, including during live automation.

A toggle compares the coupled renderer against an uncoupled, manually staged version of the same scene, the same contrast used in the listening test of Sec. 4, making the difference in spatial coherence immediately audible.

Over headphones, users can switch the binaural stage on and off and, if desired, load their own HRTF impulse responses or stems to place in the scene.

Throughout, the cue-activity display and live EQ curve let users watch the five cues update together, reinforcing the central idea: one gesture, one coherent distance percept.

4. EVALUATION

The plugin has been evaluated both technically and perceptually; we summarize the headline results here, with further data and analysis in the repository.[2]

The spectral stage reproduces its analytic target curves with a maximum deviation below 0.005 dB across the spectrum, far below the ≈0.5 dB just-noticeable difference for broadband level

[2] `https://github.com/stefbil/distance-space`

Table 2: Distance accuracy on the shared three-source music scene. MAE is in slider points (0–100); ρ is mean Spearman correlation between rated and intended distances; Order is the proportion of trials preserving the near-to-far source order.

Condition	MAE	Mean ρ	Order correct
Plugin (coupled)	7.3	0.86	89%
Baseline (manual)	15.3	0.56	50%
Anchor	40.3	−0.22	17%

changes [19]. Running as a VST3 instance at 48 kHz, CPU load stays at 1.5 % to 2.5 % without the binaural stage and peaks near 10.5 % with it enabled (worst-case block size), well below a 20 % real-time ceiling. Latency comprises the intentional, distance-dependent pre-delay plus a fixed 10.7 ms partition latency when the binaural stage is active.

A remote headphone listening test ($N = 18$; nine audio engineers, four musicians, five non-experts) compared the coupled renderer (*plugin*) against an uncoupled manual-style mix (*baseline*), a degraded mono *anchor*, and, for quality ratings, a hidden commercial-style *reference* mix. In the distance task on the shared three-source music scene (Table 2), the renderer roughly halved mean absolute placement error relative to the baseline and raised both rank correlation and near-to-far ordering accuracy. On composite spatial quality (mean of immersion, externalization, and spatial realism), the plugin was preferred over the baseline by all 18 participants.

5. CONCLUSIONS

We demonstrate a VST3 plugin that turns auditory distance into a single, automatable control by enforcing cue coherence as a system invariant, and whose benefit over conventional uncoupled staging is supported by a listening test. The demonstration lets users experience that coherence directly, hands-on and over headphones. Ongoing work extends the diffuse field toward a de-correlated, room-anchored binaural model and calibrates the cue mappings against physics-based room simulations.

6. REFERENCES

[1] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2110–2117, Oct. 2002.

[2] P. Zahorik, D. Brungart, and A. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, pp. 409–420, 05 2005.

[3] A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, vol. 78, no. 2, pp. 373–395, Feb. 2016.

[4] A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, no. 6719, pp. 517–520, Feb. 1999.

[5] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Localizing nearby sound sources in a classroom: Binaural room impulse responses," *The Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3100–3115, Apr. 2005.

[6] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, Oct. 1996.

[7] Raw Material Software Limited. (2026) Juce: Home. Accessed 2026-03-29. [Online]. Available: https://juce.com/

[8] Steinberg Media Technologies GmbH. (2025, Oct.) Steinberg releases vst 3.8 under mit license. Press release announcing the VST 3.8 SDK availability under the MIT Open Source License. [Online]. Available: https://www.steinberg.net/press/2025/vst-3-8/

[9] H. Kuttruff, *Room Acoustics*, 7th ed. London and New York: Routledge, 2024, first published 1973; this edition covers new measurement and simulation techniques, spatial/directional analysis, and recent psychophysical approaches to auditory perception in concert halls. [Online]. Available: https://www.routledge.com/Room-Acoustics/Kuttruff-Vorlander/p/book/9781032478258

[10] M. Fink, M. Holters, and U. Zölzer, "Signal-matched power-complementary cross-fading and dry-wet mixing," in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, Brno, Czech Republic, 2016, pp. 109–112. [Online]. Available: https://www.dafx.de/paper-archive/2016/dafxpapers/16-DAFx-16_paper_07-PN.pdf

[11] D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," *The Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592–2605, Jun. 1990.

[12] International Organization for Standardization, "Iso 9613-1:1993 acoustics — attenuation of sound during propagation outdoors — part 1: Calculation of the absorption of sound by the atmosphere," 1993, international Standard. [Online]. Available: https://www.iso.org/standard/17426.html

[13] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratioa)," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, Jul. 2008.

[14] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," *Journal of the Audio Engineering Society*, vol. 20, no. 2, pp. 146–159, mar 1972.

[15] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *The Journal of the Acoustical Society of America*, vol. 5, no. 2, pp. 82–108, Oct. 1933.

[16] International Organization for Standardization, "Iso 226:2023 acoustics - equal-loudness-level contours," 2023, international Standard. [Online]. Available: https://www.iso.org/standard/83117.html

[17] J. O. S. III, *Physical Audio Signal Processing: for Virtual Musical Instruments and Digital Audio Effects*. Stanford, CA: W3K Publishing, 2010, online book. [Online]. Available: https://ccrma.stanford.edu/~jos/pasp/

[18] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," https://sound.media.mit.edu/resources/KEMAR.html, 1995.

[19] H. Fastl and E. Zwicker, "Loudness," in *Psychoacoustics: Facts and Models*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 175–238.