Directional Sources and Listeners in Interactive Sound Propagation using Reciprocal Wave Field Coding

CHAKRAVARTY R. ALLA CHAITANYA*, Microsoft Research and McGill University
NIKUNJ RAGHUVANSHI*, Microsoft Research
KEITH W. GODIN, Microsoft Mixed Reality
ZECHEN ZHANG, Microsoft Research and Cornell University
DEREK NOWROUZEZAHRAI, McGill University
JOHN M. SNYDER, Microsoft Research

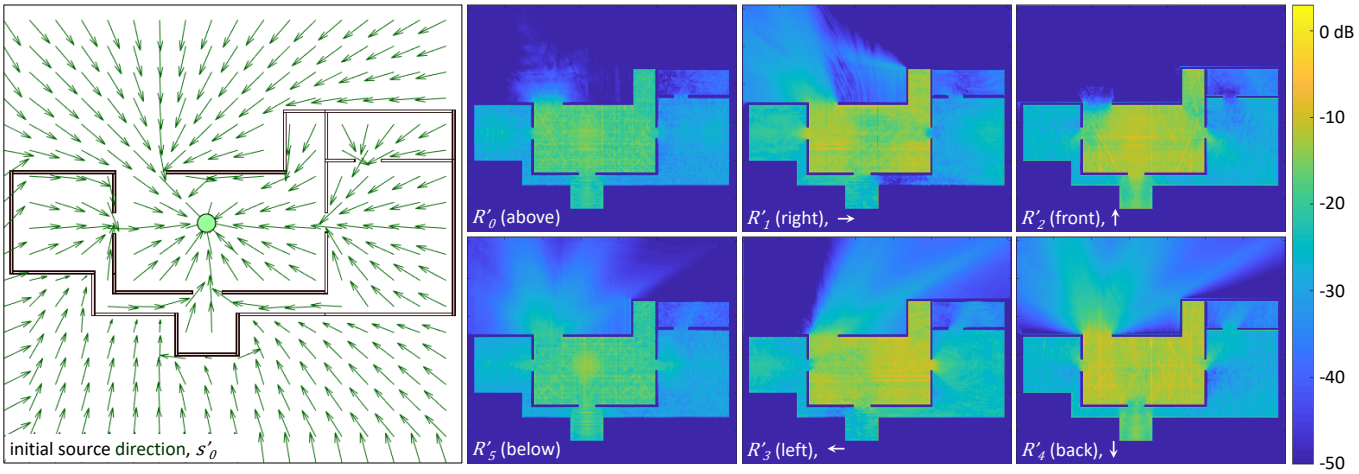

Fig. 1. Summarized parameter fields for HouseScene. For each listener position (green circle), we precompute and store directional parameter fields that vary over 3D source position. We visualize slices at listener height through these fields. Initial source direction (left) encodes the direction of radiation at each source location for the shortest (earliest arriving or "direct") path arriving at the listener. Similarly an arrival direction at listener is also encoded for each source location, not shown. Indirect energy transfer between source and listener is encoded in a 6×6 "reflections transfer matrix" (RTM) that aggregates about six signed axes in world space around both source and listener. RTM images (right) are summarized here by summing over listener directions via (16), representing source transfer anisotropy for an omnidirectional microphone. Our encoding and runtime use the full matrix, shown in Figure 11. Overall our fully reciprocal encoding captures directionality at both source and listener.

Common acoustic sources, like voices or musical instruments, exhibit strong frequency and directional dependence. When transported through complex environments, their anisotropic radiated field undergoes scattering, diffraction, and occlusion before reaching a directionally-sensitive listener. We present the first wave-based interactive auralization system that encodes and renders a complete reciprocal description of acoustic wave fields in general scenes. Our method renders directional effects at freely moving and rotating sources and listeners and supports any tabulated *source directivity function* and head-related transfer function. We represent a static scene's global acoustic transfer as an 11-dimensional *bidirectional impulse response* (BIR) field, which we extract from a set of wave simulations. We parametrically encode the BIR as a pair of radiating and arriving directions for the perceptually-salient initial (*direct*) response, and a compact 6×6 *reflections transfer matrix* capturing indirect energy transfer with scene-dependent anisotropy. We render our encoded data with an efficient and scalable algorithm – integrated in the Unreal Engine™ – whose CPU performance is agnostic to scene complexity and angular source/listener resolutions. We demonstrate convincing effects that depend on detailed scene geometry, for a variety of environments and source types.

CCS Concepts: • **Applied computing** → **Sound and music computing**; • **Computing methodologies** → **Virtual reality**.

Additional Key Words and Phrases: bidirectional impulse response, equalization, head-related transfer function (HRTF), spatial audio, sound propagation, source directivity, virtual acoustics, wave simulation

*Equal contribution.

Authors' addresses: Chakravarty R. Alla Chaitanya, chakravarty.alla@gmail.com, Microsoft Research and McGill University; Nikunj Raghuvanshi, nikunjr@microsoft.com, Microsoft Research; Keith W. Godin, kegodin@microsoft.com, Microsoft Mixed Reality; Zechen Zhang, zz335@cornell.edu, Microsoft Research and Cornell University; Derek Nowrouzezahrai, derek@cim.mcgill.ca, McGill University; John M. Snyder, johnsny@microsoft.com, Microsoft Research.

0730-0301/2020/7-ART44 $15.00
https://doi.org/10.1145/3386569.3392459

ACM Reference Format:
Chakravarty R. Alla Chaitanya, Nikunj Raghuvanshi, Keith W. Godin, Zechen Zhang, Derek Nowrouzezahrai, and John M. Snyder. 2020. Directional Sources and Listeners in Interactive Sound Propagation using Reciprocal Wave Field Coding. *ACM Trans. Graph.* 39, 4, Article 44 (July 2020), 14 pages. https://doi.org/10.1145/3386569.3392459

1 INTRODUCTION

Radiation patterns of real-world sound sources are complex, exhibiting strong direction and frequency dependence which our binaural hearing system detects and analyzes. Without looking, we can tell when someone talking turns toward us, drawing our attention. The surrounding environment also plays a salient role. When shouting between rooms, we know we must face the open door to be heard. Speech from someone walking ahead of us scatters off nearby surfaces, at once improving audibility and reinforcing the presence of surfaces. Adjoining chambers function as sound reservoirs whose reflected energy communicates to us only through open portals. In both cases, the directional profile of reflected energy is scene-dependent and often highly anisotropic. Such effects are critical for audio-visual immersion in games and virtual reality.

The challenge for interactive applications is to minimize CPU use while robustly capturing audible long-wavelength diffraction and scattering effects. Precomputed wave-based methods offer a promising approach recently adopted in games [Raghuvanshi et al. 2017], and are able to model listener directivity [Raghuvanshi and Snyder 2018]. But these remain limited to omnidirectional sources. We introduce the first wave-based system to interactively render translating/rotating directional sources and listeners in general scenes.

Extending [Raghuvanshi and Snyder 2018], we formalize the problem as encoding the scene's 11D wave function, called the *bidirectional impulse response* (BIR) field, that varies with 3D source and listener position, time, arrival direction at the listener, and radiation direction at the source (Figure 2). This captures a complete, reciprocal description of directional wave propagation between any two points in a static 3D scene. We show how to efficiently extract, encode, and render the BIR field. We perform extraction by extending the efficient flux density formulation in [2018] with minimal additional precomputation.

We perceptually encode the BIR by augmenting the parameter set in [Raghuvanshi and Snyder 2018] with the radiated direction of initial energy flowing from source to listener (first 10ms), and the aggregate directional reflected energy exchange between source and listener (next 80ms). Refer to Figure 1. We represent the latter as a 6×6 transfer matrix in terms of smooth lobes centered around the six Cartesian directions at the source and listener. The user provides a tabulated *source directivity function* (SDF) for each source, capturing its far-field radiation pattern. This representation is standard, per-octave, neglects phase, and amenable to recent fast equalization methods that significantly reduce per-source rendering cost.

We integrate our technique into the Unreal Engine™ and demonstrate a variety of complex scenes with noticeable scene-dependent source and listener directivity effects. By efficiently precomputing and perceptually encoding a fully reciprocal description of bidirectional energy propagation within the scene, our system for the first time provides lightweight, directional wave-acoustic effects for arbitrary source and listener motion in large game scenes. Run-time rendering is scalable. The CPU cost does not depend on either the amount of detail in scene geometry, or the angular resolution of source radiation pattern and head-related transfer function.

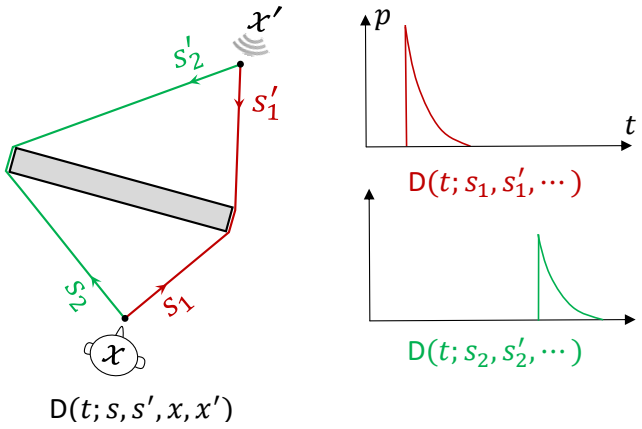

Fig. 2. Bidirectional impulse response. Sound energy from a pulse at source position x' radiates, propagates through a scene, and then arrives at listener position x. In this simple scene consisting of a single occluder slab, there are two relevant paths, emanating from directions s'_1, and s'_2 and arriving from directions s_1 and s_2, respectively. Corresponding impulse responses for these two paths are plotted on the right, parameterized in terms of time and direction at both source and listener. We encode the BIR's perceptual properties for all pairs of source (x') and listener (x) positions. At runtime, given an emitted sound and a location and orientation for each source, and the listener's head location and orientation, we efficiently render the BIR as heard by the listener. The BIR encapsulates all wave propagation interactions with the scene including multiple diffraction of salient paths around obstacles as shown here. Following conventions in spatial audio literature, source direction points in the direction of propagation for radiant energy while the listener direction points from the listener position towards arriving energy (negative of propagation direction at listener).

2 BACKGROUND AND FORMULATION

Throughout the paper, for any quantity $(*)$ at the listener, we use prime $(*')$ to denote a corresponding quantity at the source. E.g., x is listener location and x' source location (rather than time derivative).

Modeling. Interactive sound propagation aims to efficiently model the linear wave equation [Pierce 1989]:

$$\left[(1/c^2)\,\partial_t^2 - \nabla_x^2\right] p\left(t, x, x'\right) = \delta\left(t\right)\delta\left(x - x'\right), \tag{1}$$

where $c = 340$m/s is the speed of sound, ∇_x^2 the 3D Laplacian operator and δ the Dirac delta of an omnidirectional impulsive source. With boundary conditions provided by the scene shapes and materials, the solution $p(t, x, x')$ is the Green's function that fully describes the scene's global acoustic transport, including diffraction and scattering. For such an omnidirectional source, the principle of acoustic reciprocity ensures that

$$p(t, x, x') = p(t, x', x). \tag{2}$$

In general scenes, a numerical solver must be employed to discretely sample the Green's function; we employ the ARD time-domain wave solver [Raghuvanshi et al. 2009]. For stability and

accuracy, the resolution of spatio-temporal discretization depends on the frequency cutoff (1kHz in our tests), resulting in an update rate of 5882Hz and a cell size of 12.25cm. Computational complexity is linear in both the number of time steps and simulated air volume. Computational cost is insensitive to polygon count and detailed topology of the scene's free space, for a fixed air volume. Additional cost scales as surface area from numerical boundary modeling but in our tests it takes a small portion (< 5%) of the compute time.

In principle, directionality can be modeled via spatio-temporal convolution of $p(t, x, x')$ with volumetric source and listener distributions. Such an approach is conceptually valuable but too expensive for real-time evaluation on large scenes, requiring temporal convolution and spatial quadrature over sub-wavelength grids requiring re-evaluation on source or listener motion. A more efficient approach, pursued here, is to extract static directional information within the Green's function to allow a modular approach where source and listener directionality can be accounted for separately from global transport.

Monoaural rendering. Given an arbitrary pressure signal $q'(t)$ radiating omnidirectionally from a sound source located at x', the resulting signal at a monoaural listener located at x can be computed using a temporal convolution, denoted by $*$:

$$q(t; x, x') = q'(t) \; * \; p(t; x, x'). \tag{3}$$

Here $p(t; x, x')$ is the impulse response, obtained by fixing the parameters after semi-colon in $p(t; x, x')$, namely the listener and source locations (x, x'). This modularizes the problem by separating source signal from environmental modification but ignores directional aspects of propagation.

Directional listener. The notion of (9D) listener directional impulse response $d(t, s; x, x')$ [Embrechts 2016] generalizes the impulse response $p(t; x, x')$ to include direction of arrival, s. A tabulated head-related transfer function (HRTF) consisting of two spherical functions $H^{l/r}(s, t)$ specifies the angle-dependent acoustic transfer in the free field to the left and right ears. Per [Raghuvanshi and Snyder 2018], this allows directional rendering at the listener via

$$q^{l/r}(t; x, x') = q'(t) * \int_{\mathcal{S}^2} d\left(t, s; x, x'\right) \; * \; H^{l/r}\left(\mathcal{R}^{-1}(s), t\right) \, ds \, , \tag{4}$$

where $\mathcal{R}$ is a rotation matrix mapping from head to world coordinate system, and $s \in \mathcal{S}^2$ represents the space of incident spherical directions forming the integration domain.

Directional source and listener. The above still leaves out directionality at the source. This paper poses the natural extension via the bidirectional impulse response (BIR), an 11D function of the wave field, $D(t, s, s'; x, x')$. Analogous to the HRTF, the source's radiation pattern is tabulated in a source directivity function (SDF), $S(s, t)$. With this information, we obtain the virtual acoustic rendering equation for point-like sources:

$$q^{l/r}(t; x, x') = q'(t) * \iint D\left(t, s, s'; x, x'\right) * H^{l/r}\left(\mathcal{R}^{-1}(s), t\right) * S\left(\mathcal{R}'^{-1}(s'), t\right) \, ds \, ds', \tag{5}$$

where $\mathcal{R}'$ is a rotation matrix mapping from the source to world coordinate system, and the integration becomes a double one over the space of both incident and emitted directions $s, s' \in \mathcal{S}^2$.

The BIR is convolved with the source and listener's free-field directional responses S and $H^{l/r}$ respectively, while accounting for their rotation since (s, s') are in world coordinates, to capture modification due to directional radiation and reception. The integral repeats this for all combinations of (s, s'), yielding the net binaural response. This is finally convolved with the emitted signal $q'(t)$ to obtain a binaural output that should be delivered to the entrances of the listener's ear canals.

Our goal is to efficiently precompute the BIR field $D(t, s, s'; x, x')$ on complex scenes, compactly encode this 11D data using perception, and approximate (5) for efficient rendering.

Specializations. The bidirectional impulse response generalizes the listener directional impulse response (LDIR) used in (4) via

$$d(t, s; x, x') \equiv \int_{\mathcal{S}^2} D(t, s, s'; x, x') \, ds'. \tag{6}$$

In other words, integrating over all radiating directions s' yields directional effects at the listener for an omnidirectional source. A source directional impulse response (SDIR) can be reciprocally defined as

$$d'(t, s'; x, x') \equiv \int_{\mathcal{S}^2} D(t, s, s'; x, x') \, ds. \tag{7}$$

representing directional source and propagation effects to an omnidirectional microphone at x via the rendering equation

$$q(t; x, x') = q'(t) * \int_{\mathcal{S}^2} d'\left(t, s'; x, x'\right) \; * \; S\left(\mathcal{R}'^{-1}(s'), t\right) \, ds'. \tag{8}$$

Properties of the bidirectional decomposition. Our formalization admits direct geometric interpretation. With source and listener located at (x', x) respectively, consider any pair of radiated and arrival directions (s', s). In general, multiple paths connect these pairs, $(x', s') \rightsquigarrow (x, s)$, with corresponding delays and amplitudes, all of which are captured by $D(t, s, s'; x, x')$. Figure 2 illustrates a simple case. The BIR is thus a fully reciprocal description of sound propagation within an arbitrary scene. Interchanging source and listener, all propagation paths reverse:

$$D(t, s, s'; x, x') = D(t, s', s; x', x). \tag{9}$$

This reciprocal symmetry mirrors that for the underlying wave field, $p(t; x, x') = p(t; x', x)$, a non-trivial, fundamental property not shared by the listener directional impulse response d in (6). Section 4 will show how the complete reciprocal description allows us to extract source directionality with minimal added precomputation cost.

Note how our formulation separates out source signal, listener directivity, and source directivity, arranging the BIR field in D to characterize scene geometry and materials alone. This decomposition allows for various efficient approximations subsuming existing real-time virtual acoustic systems; we are unaware of a prior unifying formalization. This modular abstraction comes however at the cost of assuming that sound propagates in a feed-forward manner from source, through scene, to listener. This breaks down when higher-order interactions between source/listener and scene

predominate. An extreme example is wind instruments where the mouthpiece is a source strongly coupled to propagation in the instrument body, requiring two-way coupled simulation [Allen and Raghuvanshi 2015]. Closer to our application, when the bell of a trumpet gets close to a wall, the resulting feedback can alter the resonant modes. Such effects must be modeled as boundary conditions within a fixed animation [Wang et al. 2018] which is currently prohibitive for interactive systems on large 3D scenes that we consider.

3 RELATED WORK

Methods in geometric acoustics (GA) [Savioja and Svensson 2015] employ a short-wavelength approximation via the Eikonal equation to construct explicit paths connecting x' to x. They typically ignore phase and find diffraction difficult to model. Diffraction effects at audible wavelengths are perceptually significant, for example when the line of sight between source and listener is blocked and the sound thus heard attenuated and through portals or around obstacles - a ubiquitous occurrence in games and virtual reality. GA is conceptually simple, able to handle source and listener directivity trivially, and, potentially, to admit dynamic scenes. Yet shooting enough rays/paths to ensure convergence especially in complex scenes is computationally demanding [Cao et al. 2016], requiring more study to realize this potential for interactive applications.

In a related application, GA via bidirectional path tracing has been applied to synthesize new directional audio for a recorded panoramic video [Li et al. 2018]. Tens of seconds of computation are required to produce audio given the listener's canned animation path through the scene. In other words, the computation is fast but still offline and does not support interactive movement of source and listener.

GA inherits its approach from light (radiance) rendering, where decompositions similar to our own have been applied. Dobashi et al. [1995] used spherical harmonics (SH) as a basis for the direction-dependent radiance emitted by a local light source whose position is fixed in a scene; the radiation pattern can then be changed interactively as with our SDF. Also using a low-order SH basis, precomputed radiance transfer (PRT) [Lehtinen and Kautz 2003; Sloan et al. 2002] expresses the linear relation of incident into exiting radiance after reflectance and self-shadowing effects on the surface of glossy objects using a precomputed matrix, similar to our reflections transfer. PRT represents lighting at infinity and thus need to only tabulate over receiver locations on scene geometry; we tabulate over the 6D space of source and listener positions (x, x'). Our basis for directional transfer from acoustic reflections is also lower dimensional to match acoustic rather than visual perception, exploiting just six cosine squared basis functions around the coordinate axes to yield lightweight 6×6 transfer matrices. Finally, we extract directionality from a wave-based PDE solution rather than from geometrically traced rays or paths.

Wave-based methods discretize over time and space to solve (1) directly, elegantly including diffraction effects while remaining computationally insensitive to the scene's geometric complexity. But they are costly. Realtime simulations have been demonstrated in small rooms [Savioja 2010] or for 2D wind instruments using GPUs [Allen and Raghuvanshi 2015]. Wideband offline simulation of near-field 3D acoustic synthesis and propagation effects [Wang et al. 2018] have also been demonstrated. For the large virtual game scenes we consider, adverse computational scaling with frequency currently limits offline computation to about 1kHz.

Bilbao et al. [2019] introduced spherical harmonic sources directly into an FDTD simulation, intended for offline applications where source location and orientation are fixed functions of time. In contrast, our BIR formulation separates properties of the environment from the source, enabling practical precomputation that supports interactive movement and rotation of sources at runtime. Our explicit encoding of perceptually-salient direction of the initial response also spatializes more sharply, compared to order-truncated spherical harmonics.

Mehra et al. [2014] captured wave-based source directional effects using a low-order spherical harmonics (SH) representation in the frequency domain. The technique is specialized to outdoor scenes comprising a few separable objects, has high runtime cost that increases with scene complexity and scales adversely with angular resolution of source directivity and head-related transfer function.

Directional sound propagation for extended ambient sources such as ocean surf has also been studied [Zhang et al. 2018, 2019]. The entire scene can effectively be in the near field of such sources, and low-order directional effects at the listener captured for a source of fixed size and shape in relation to the static scene. Here we consider localized but dynamic directional sources.

The BIR acoustic field is closely related to the plenoptic function [Chai et al. 2000] for light, accounting for the directional energy distribution at various receiver (camera) locations, while extending it to add information about time-of-flight delays and response to directional radiation patterns at the source.

4 PRECOMPUTATION

We first describe how to precompute and encode the bidirectional impulse response field $D(t, s, s'; x, x')$ from a set of wave simulations.

Extracting directivity with flux. Our precomputation samples the 7D Green's function $p(t, x, x')$ and extracts directional information using the flux formulation first proposed by Pulkki [2007] and later demonstrated to be effective for listener directivity in simulated wave fields [Raghuvanshi and Snyder 2018]. Flux density, or "flux" for short, measures the directed energy propagation density in a differential region of the fluid. For each impulsive wavefront passing over a point, flux instantaneously points in its propagating direction. It is computed for any volumetric transient field $p(t, \alpha; \beta)$ with listener at α and source at β as

$$\begin{aligned} f_{\alpha \leftarrow \beta}(t, \alpha; \beta) &\equiv -\, p(t, \alpha; \beta)\, v(t, \alpha; \beta), \\ v(t, \alpha; \beta) &\equiv -\frac{1}{\rho_0} \int_{-\infty}^{t} \nabla_\alpha p(\tau, \alpha; \beta)\, d\tau, \end{aligned} \tag{10}$$

where v is the particle velocity, and ρ_0 is the mean air density (1.225kg/m^3). Note the negative sign in the first equation that converts propagating to arrival direction at α. Flux can then be normalized to recover the time-varying unit direction,

$$\hat{f}_{\alpha \leftarrow \beta}(t) \equiv f_{\alpha \leftarrow \beta}(t) / ||f_{\alpha \leftarrow \beta}(t)||. \tag{11}$$

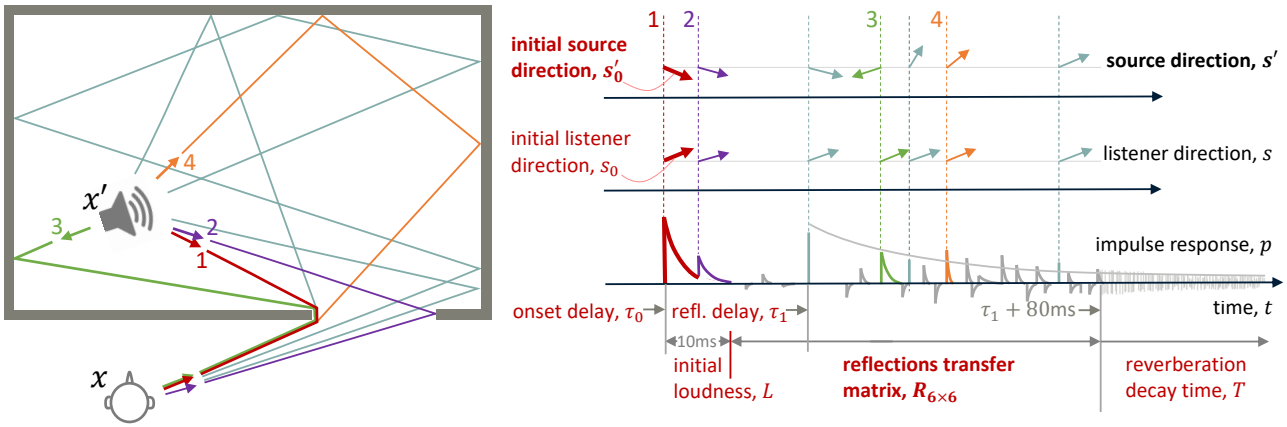

Fig. 3. Encoding bidirectional propagation. Each peak in the impulse response at delay t corresponds to a path connecting source and listener of length ct. We augment the impulse response with the radiated direction at the source (top right) and arrival direction at the listener (middle right), as shown for a few enumerated paths. Four of these are numbered to identify their corresponding directions in the time-varying directional signals shown on the right. Note that three arrows for listener direction (paths 1, 3, and 4) overlay each other on the bottom left, because these paths all emerge around the left part of the opening to get to the listener. Encoded parameters are designated by red text; those originally proposed in [Raghuvanshi and Snyder 2018] are unbolded while our extended parameters are shown in bold. The initial sound path from source to listener (labeled 1) arrives at time τ_0 and is assigned its radiated direction from the source, s'_0, in addition to its arrival direction at the listener, s_0. The reflection transfer matrix R summarizes the directional energy transfer between source and listener aggregating across reflections occurring in an 80ms window after the time when we detect that reflections start arriving, τ_1. Our encoding captures salient information allowing arbitrary translation, rotation, and directivity functions for both source and listener at runtime.

We extract the bidirectional impulse response as

$$D(t, s, s'; x, x') \approx \delta\left(s' - \hat{f}_{x' \leftarrow x}(t; x', x)\right)\ \delta\left(s - \hat{f}_{x \leftarrow x'}(t; x, x')\right) \\ p(t; x, x')\ . \tag{12}$$

At each instant in time t, the linear amplitude p is associated with the instantaneous direction of arrival at the listener $\hat{f}_{x \leftarrow x'}$ and direction of radiation from the source $\hat{f}_{x' \leftarrow x}$ as shown in Figure 2. The flux approximation has been used before for directional sound field encoding [Pulkki 2007] and has been shown to work well for impulsive fields against reference [Raghuvanshi and Snyder 2018].

Flux approximates the directionality of energy propagation which can be analyzed exactly with the much more costly reference of plane wave decomposition. Primary is the assumption of a single direction per time instant; two paths of equal length connecting source to listener cannot be disambiguated by flux and will be ascribed an averaged radiation and arrival direction. Fortunately, impulsive sound fields (those representing the response of a pulse) mostly consist of distinct moving wavefronts, especially in the initial, non-diffuse part of the response where directionality matters most. The resulting error is small in common cases, empirically analyzed in [Raghuvanshi and Snyder 2018].

Reciprocal discretization. As in parametric wave field coding [Raghuvanshi and Snyder 2014], we utilize reciprocity to make the precomputation more efficient and save memory by exploiting the fact that the runtime listener is typically more restricted in its motion than are sources. That is, we expect the listener to remain at roughly human height above floors or the ground in a scene.

We use the term "probe" for x representing listener location at runtime and source location during precomputation, and "receiver" for x', and thus assume x varies more restrictively than x' to save one dimension from the set of probes. We adaptively generate the set of probe locations $\{x\}$ [Alla Chaitanya et al. 2019], ensuring adequate sampling of walkable regions of the scene with spacing varying between 0.5m and 3.5m. Each probe is processed independently in parallel over many (~100) cluster nodes.

For each probe, the scene's volumetric Green's function $p(t, x'; x)$ is sampled on a uniform spatio-temporal grid with resolution $\Delta x =$ 12.5cm and $\Delta t = 170\mu$s, yielding a maximum usable frequency of $\nu_{\max} =$ 1kHz. Typical domain size is $90 \times 90 \times 30$m. The spatio-temporal impulse $\tilde{\delta}(t)\,\delta(x' - x)$ is introduced in the 3D scene and (1) is solved using the ARD pseudo-spectral solver [Raghuvanshi et al. 2009]. The perceptually equalized pulse $\tilde{\delta}(t)$ and directivity at the listener in (12) is computed exactly as in [Raghuvanshi and Snyder 2018], using additional discrete dipole source simulations to evaluate the gradient $\nabla_x p(t, x, x')$ required for computing $f_{x \leftarrow x'}$.

Source directivity. Exploiting reciprocity per (9), directivity at runtime source location x' can be obtained by evaluating flux $f_{x' \leftarrow x}$ via (10). Because the volumetric field for each probe simulation $p(t, x'; x)$ already varies over x', no additional simulations are required. To compute the particle velocity, we commute time integral and gradient yielding $v(t, x'; x) = -1/\rho_0\ \nabla_{x'} \int_{-\infty}^{t} p(\tau, x'; x)\, d\tau$. We maintain an additional discrete field $\int_{-\infty}^{t} p(\tau, x'; x)\, d\tau$ implemented as a running sum. Commutation saves memory by requiring additional storage for a scalar rather than a vector (gradient) field. It also relies on the zero-DC property of the introduced pulse, $\int d\tau \tilde{\delta}(\tau) = 0$, without which the integral can become large and required information lost to numerical underflow in our single-precision computation. The gradient is evaluated at each step using centered differences. Overall, this provides a lightweight streaming implementation to compute $f_{x' \leftarrow x}$ in (12).

Perceptual encoding. Extracting and encoding a directional response $D(t, s, s'; x, x')$ proceeds independently for each (x, x') which we hereafter drop from the notation. At each solver time step t, the encoder receives the instantaneous radiation direction $f_{x' \leftarrow x}(t)$, the listener arrival direction $f_{x \leftarrow x'}(t)$, and the amplitude $p(t)$. Figure 3 schematically illustrates BIR encoding in an example scene. Our encoded parameters augment the set in [Raghuvanshi and Snyder 2018]; new ones are shown in bold.

Encoding assumes a few temporal integration intervals from the literature as discussed in [Raghuvanshi and Snyder 2018]. Precedence effect studies [Litovsky et al. 1999] show that 1ms is the threshold for summing localization, and 10ms is the echo fusion threshold for impulsive sounds. We use these for encoding the initial sound direction and loudness respectively. For early reflections duration, 80ms is a commonly used value in room acoustics [Gade 2007] assuming running fusion of multiple arrivals. Finally, note that the reverberation time parameter is estimated from the full response beyond 80ms, as in [Raghuvanshi and Snyder 2018].

Initial source direction, s'_0. We compute the initial source direction

$$s'_0 \equiv \int_0^{\tau_0 + 1ms} f_{x' \leftarrow x}(t)\, dt, \tag{13}$$

where the delay of first arriving sound, τ_0, is computed as in [Raghuvanshi and Snyder 2018]. We retain the (unit) direction as the final parameter after integrating directions over a short (1ms) window after τ_0 to reproduce the precedence effect [Litovsky et al. 1999].

Reflections transfer matrix, R. We aggregate the directional loudness of reflections for 80ms after the time when reflections first start arriving, denoted τ_1. Directional energy is collected using coarse cosine-squared basis functions fixed in world space and centered around the six Cartesian directions $X_* = \{\pm X, \pm Y, \pm Z\}$,

$$w(s, X_*) \equiv (\max(s \cdot X_*, 0))^2, \tag{14}$$

yielding the reflections transfer matrix

$$R_{ij} \equiv 10 \log_{10} \int_{\tau_0 + 10\text{ms}}^{\tau_1 + 80\text{ms}} w(\hat{f}_{x \leftarrow x'}(t), X_i)\, w(\hat{f}_{x' \leftarrow x}(t), X_j)\, p^2(t)\, dt. \tag{15}$$

Matrix component R_{ij} encodes the loudness of sound emitted from the source around direction X_j that undergoes global transport to arrive at the listener around direction X_i. RTM components from the above formula frequency-average over the 1KHz simulation bandwidth but are applied to the full audible bandwidth during rendering, implicitly performing frequency extrapolation. Each of the 36 fields $R_{ij}(x'; x)$ is spatially smooth and compressible. It is quantized at 3dB, smoothed and down-sampled with spacing 1-1.5m, passed through running differences along each X scanline, and finally losslessly compressed with LZW.

The pressure $p(t)$ in (15) is measured at x from an omnidirectional impulse emitted at x'. This signal is the same as the response measured at x' from an impulse emitted at x, due to reciprocity (2). So directional subscripts for $p(t)$ are elided, but we cannot do so for flux which doesn't obey reciprocity so simply, but instead as in (9).

We also define the total reflection energy arriving at an omnidirectional listener for each directional basis function at the source

$$R'_j \equiv 10 \log_{10} \sum_{i=0}^{5} 10^{R_{ij}/10}, \tag{16}$$

used in the visualization in Figure 1. Figure 11 shows the complete matrix in an example scene.

Reciprocal coding. Our formulation preserves the reciprocal symmetry of the BIR per (9). The parameters we encode satisfy

$$s_0(x, x') = s'_0(x', x) \quad \text{and} \quad R_{ij}(x, x') = R_{ji}(x', x)\,. \tag{17}$$

These relations hold only up to numerical errors in simulation and encoding errors from spatial sub-sampling and quantization. The first relation can be interpreted geometrically as x and x' being the two end points of the unique geodesic path connecting them, with the two corresponding ending directions s_0 and s'_0. For the second relation, note that the RTM is not itself symmetric, $R_{ij}(x, x') \neq R_{ji}(x, x')$. To capture the symmetry, one must also interchange source and listener. This relation could in principle be used to reduce encoded data size but we currently compress and store $R_{ij}(x'; x)$ separately over probe locations $\{x\}$ to allow incremental decompression at runtime.

5 SOURCE DIRECTIVITY FUNCTION

Consider a directional sound source at the origin and let 3D position around it be expressed in spherical coordinates via $x = r\, s$. Outside the local source geometry, and neglecting for the moment interaction with the outside environment, its emitted field can be represented as $q'(t) * p(s, r, t)$ where $p(s, r, t)$ includes effects of self-scattering and self-shadowing, and $q'(t)$ is the sound signal emitted and modulated by such effects.

Past work has represented the detailed field radiated from modal vibrations of quasi-rigid objects, using the equivalent source method [James et al. 2006], and a faster approximation [Chadwick et al. 2009]. We wish to handle more general sources such as speech. As Chadwick et al. [2009] noted, the radiated field at sufficient distance from any source can be expressed via the spherical multipole expansion [Gumerov and Duraiswami 2005]

$$p(s, r, t) \approx \frac{\delta(t - r/c)}{r} * \sum_{m=0}^{M-1} \frac{\hat{p}_m(s, t)}{r^m}. \tag{18}$$

The above representation would still require M temporal convolutions at runtime to apply the source directivity in a given direction s, which is expensive for our application.

We simplify further to the commonly-employed far-field (large r) approximation by dropping all terms $m > 0$ yielding,

$$p(s, r, t) \approx \delta(t - r/c) \; * \; (1/r)\, \hat{p}_0(s, t). \tag{19}$$

The first two factors represent propagation delay and monopole distance attenuation, already contained in the simulated BIR, leaving us with $S(s, t) \equiv \hat{p}_0(s, t)$ which represents the angular radiation pattern at infinity compensated for free-field propagation. We further approximate S by removing phase information and averaging over perceptual frequency bands. For room-scale simulations where scene geometry is assumed not to be in the near-field of the source, phase is perceived primarily as propagation delay already captured

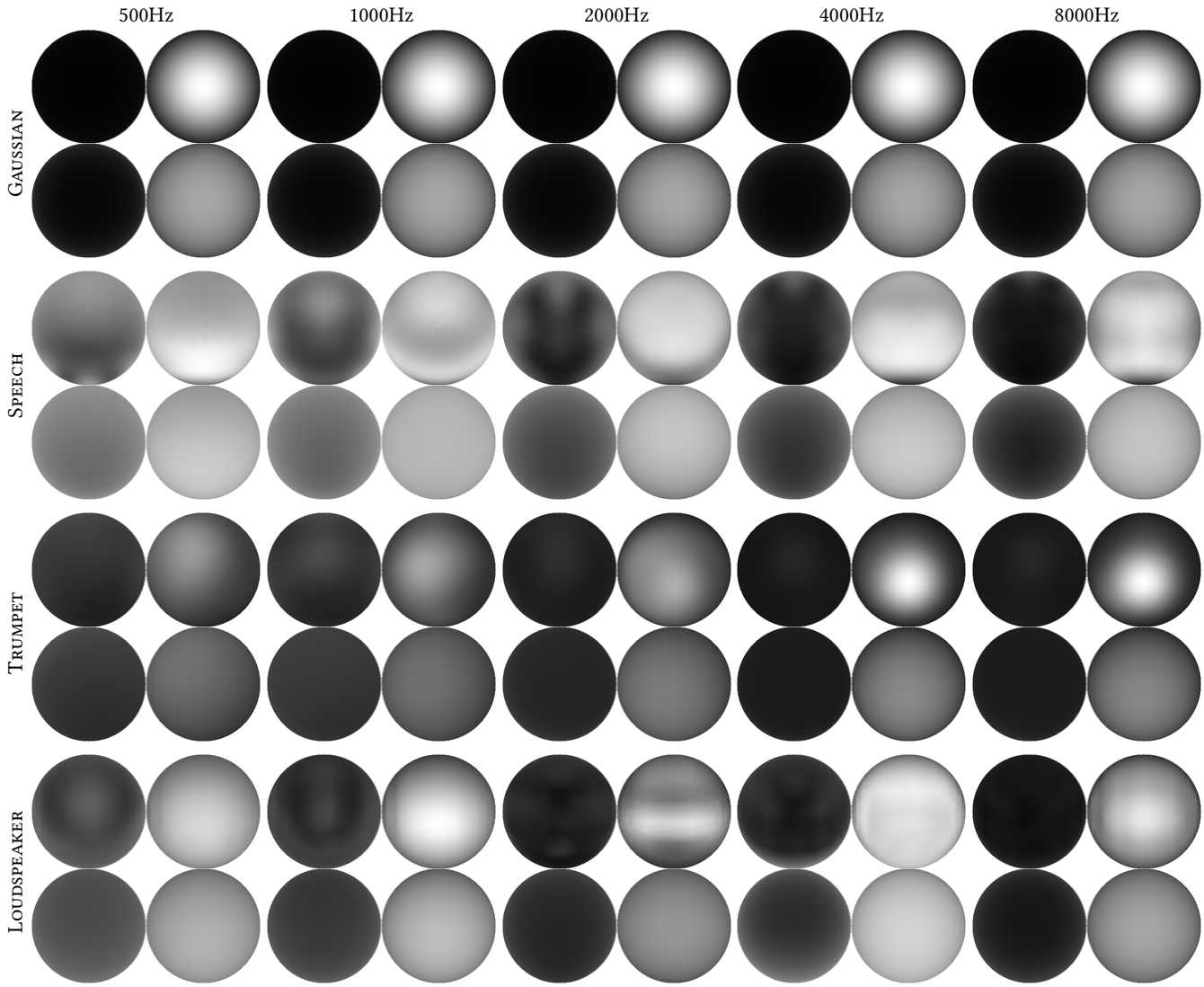

Fig. 4. Analytic and measured SDFs, mapping amplitude to brightness. Each pair of rows represents the original (top) and smoothed version (bottom) for four of our demo examples: Gaussian, Speech, Trumpet, and Loudspeaker. We use the unsmoothed SDF for direct transfer and the smoothed for indirect/reflected transfer. We visualize octave bands for five (of ten total) center frequencies, labeled on the top row. We plot each spherical function as an orthographic pair of the back (left) and front (right) hemispheres. For each hemisphere, the x axis represents azimuthal (i.e., left/right) directions and y elevation (up/down). More of the sound power is directed toward the front (right image of pair) in all examples, but the detailed directional profile varies significantly. Frequency-dependence is also easy to note: sources (except the Gaussian) generally get more sharply directional as frequency increases.

in (19). While detailed phase is critical for correctly modeling self-shadowing and scattering effects, such effects are already baked into the spherical magnitude variation. It is thus standard for measured directivity data to be presented in this form. For our application, such data is also compact and fast to render using modern graphic equalizer algorithms as we discuss later.

We average over ten octave bands spanning the full audible range with center angular frequencies: $\omega_k = 2\pi \times \{31.25, 62.5, 125, 250, 500, 1000, 2000, 4000, 8000, 16000\}$ rad/s. Denoting temporal Fourier transform with $\mathcal{F}$, we compute:

$$E^k(s) \equiv \frac{\int_{\omega_k/\sqrt{2}}^{\omega_k\sqrt{2}} |\mathcal{F}\{S\}(s,\omega)|^2 \, d\omega}{\omega_k(\sqrt{2} - 1/\sqrt{2})} \text{ and } S^k(s) \equiv 10 \log_{10} E^k(s). \quad (20)$$

The $\{S^k(s)\}$ thus form a set of real-valued spherical functions that are the input to our system, capturing salient source directivity information, such as the muffling of the human voice when heard from behind.

As a precomputation we also compute a smoothed version of the SDF to use for indirect (reflected) sound, by convolving it with the

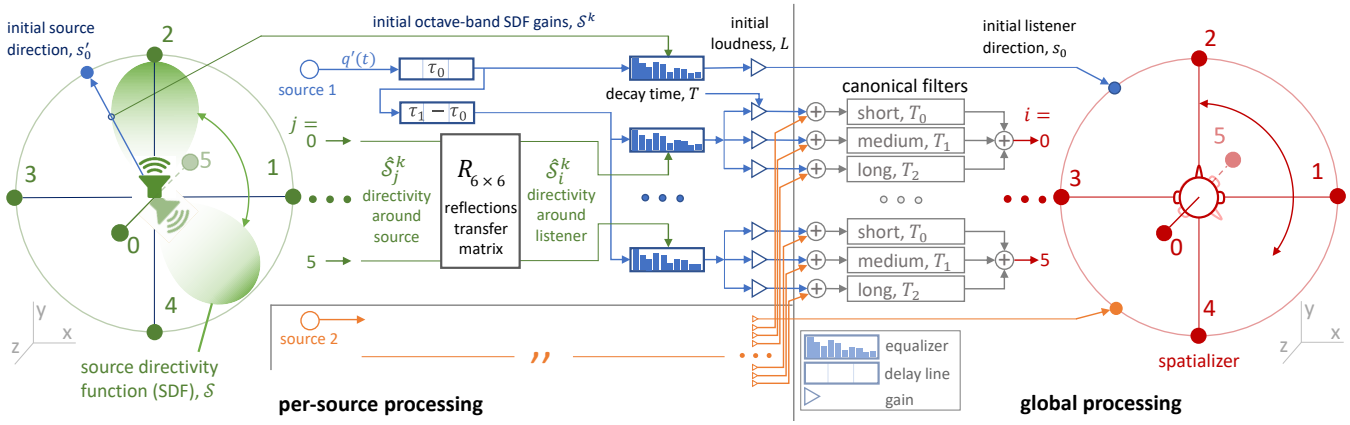

Fig. 5. Runtime signal processing.

cosine-squared lobe w in (14) to obtain

$$\hat{S}^k(s) \equiv 10\log_{10}\left(\int_{S^2} E^k(u)\, w(s,u)\, du \Big/ \int_{S^2} w(s,u)\, du\right). \quad (21)$$

Since w is partition-of-unity, the denominator above is just $4\pi/6$.

The SDF in each octave band, S^k, is quasi-uniformly sampled over a set of 2048 directions [Fliege and Maier 1999] obtained by numerically solving the Thomson problem of minimal electrostatic potential energy for unit charges restricted to the 2-sphere. This sampling requires 180KB per SDF, including the smoothed and unsmoothed data and a mapping of sample indices to directional locations. Multiple SDFs are kept loaded in memory and may be freely assigned to any source at runtime.

Figure 4 shows SDFs obtained from measured data and used in our experiments: speech directivity [Bellows et al. 2019], a trumpet [Shabtai et al. 2017; Weinzierl et al. 2017], and a loudspeaker. In the first two cases, data was provided in third-octave bands which we averaged to convert to octaves. Some datasets omit data for the lowest frequencies which we copy from the lowest available. We found in informal testing that high angular resolution SDF data is important for a convincing auralization. Low-resolution SDFs we tested (based on windowed 3rd-order spherical harmonics) wash out directional variation in direct sound and subtle anisotropy in reflected sound, especially for the speech and trumpet SDFs. Our datasets are based on 10th-order SH. More systematic study on human perception of source directivity remains for future work. Game audio engines typically provide simple, analytic directivity functions with designer controlled sharpness; we thus add a spherical Gaussian model given by

$$S^k_G(s;\mu) \equiv 10\log_{10} 10^{\lambda(\mu\cdot s-1)}, \quad (22)$$

where μ is the central direction and λ is the directional sharpness. We use $\lambda = 2$ in our tests which yields attenuations of -20dB and -40dB to the side and behind the source respectively. Such sharp falloffs are common in practice to provide a clear mix in complex soundscapes with multiple active sound sources.

6 RUNTIME

Figure 5 diagrams our runtime. It efficiently approximates (5) by augmenting the signal processing in [Raghuvanshi and Snyder 2018] to include source directivity. Our changes are restricted to "per-source processing" on the left of the figure. We focus on the modifications needed, summarizing prior work for completeness. In the following, index $i \in \{0, 1, \cdots, 5\}$ is used for reflection directions around the listener, index $j \in \{0, 1, \cdots, 5\}$ for reflection directions around the source, index $k \in \{0, 1, \cdots, 9\}$ for SDF octaves, and index $l \in \{0, 1, 2\}$ for reverberation decay times. Each sound source emits its own (monaural) signal which we denote $q'(t)$, and may be freely synthesized, replaced, or changed at runtime.

Initial (direct) sound. The radiated direction for the initial sound at the source, s'_0, is first transformed into the source's local reference frame. An SDF nearest-neighbor lookup is performed to yield the octave-band loudness values in dB,

$$L_k \equiv S^k(\mathcal{R}'^{-1}(s'_0)), \quad (23)$$

due to the source's radiation pattern. These add to the overall direct loudness encoded as a separate parameter, denoted L, along with $1/r$ distance attenuation, $-20\log_{10}(|x - x'|)$. As the source rotates, $\mathcal{R}'$ changes, L_k change accordingly, and are dynamically applied using a graphic equalizer described in the next section. The resulting filtered signal $q(t)$ is the initial sound that should then be spatialized as coming from the arrival direction s_0, accounting for the HRTF and current listener orientation $\mathcal{R}$ as detailed in [Raghuvanshi and Snyder 2018].

Equalization. Graphic equalization is a common signal processing task; see [Välimäki and Reiss 2016] for a review. Direct FFT-based techniques can be costly, especially as we need 7 instances of equalization per source: 1 for the initial sound plus 6 for reflections. They also introduce extra latency to avoid wrap-around artifacts. Recursive (IIR) digital filters improve both efficiency and latency. The primary building block is a digital biquadratic filter which we implement with the "direct-form 1" recursion [Smith 2007]:

$$y[n] \leftarrow b_0\cdot x[n] + b_1\cdot x[n-1] + b_2\cdot x[n-2] - a_1\cdot y[n-1] - a_2\cdot y[n-2]) \quad (24)$$

for input x, output y, and time sample index n.

The traditional approach employs filter-banks separating the signal into octave bands which are appropriately scaled per $\{L_k\}$ and summed. A test design we investigated cost 60 biquadratic

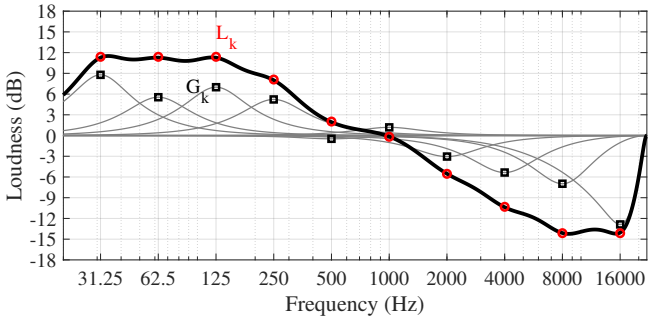

Fig. 6. Equalizer. Given a direction, the SDF provides octave gain targets $\{L_k\}$, shown as red circles. The example here is for Speech as heard behind the speaker's head; note the attenuation of high frequencies. The algorithm in [Oliver and Jot 2015] finds peak gains $\{G_k\}$ (black squares) specifying a set of biquadratic filters (with individual responses shown as the thin gray curves) that yield the total response targeted (heavy black curve). Reference implementation taken from [Välimäki and Liski 2017].

Butterworth filters to ensure about 1dB error. Such high cost is typical because the design seeks minimal overlap across octave bands, requiring steep falloff in each band's response at its limits.

We instead adopt a recent technique [Oliver and Jot 2015] that requires only one biquadratic filter per octave. As Figure 6 illustrates, rather than avoiding inter-band overlap the algorithm finds individual filter peak gains, $\{G_k\}$ so that their overlapping individual responses (gray traces) combine to approximate the desired overall equalization $\{L_k\}$ (thick black trace).

The filters have a proportional property (see [2015] for details) that allows a linear mapping in the dB domain: $\mathcal{L} = \mathcal{M}\mathcal{G}$, where $\mathcal{M}$ is a 10×10 diagonal-dominant matrix that captures filter interactions. At each visual frame, the vector $\mathcal{L} = \{L_k\}$ is obtained from the SDF, and the linear system solved to find $\mathcal{G} = \{G_k\}$. Filter coefficients $\{b_*^k, a_*^k\}$ for each biquadratic filter h_k can then be computed using analytic formulae given in [2015]. The algorithm works well for measured SDF data, with maximum fitting error of 1.5dB across all directions and datasets we tested. Though a more recent paper [Välimäki and Liski 2017] generally improves fitting errors, we still employ [2015] as its matrix $\mathcal{M}$ is constant and independent of the L_k. We exploit this property to precompute $\mathcal{M}^{-1}$, reducing computation to per-frame matrix-vector product $\mathcal{G} = \mathcal{M}^{-1}\mathcal{L}$.

Equalization is then applied via

$$q(t) = \circledast_{k=0}^{9} h_k\left(t; \{b_*^k, a_*^k\}\right) * q'(t), \tag{25}$$

where $\circledast$ denotes a nested series of convolutions. Each convolution is implemented using the recursion in (24), requiring each filter to keep two past inputs and outputs as history. We use a SIMD-optimized implementation performing in-place processing on the audio sample buffer to further reduce memory access latency. We find that the algorithm functions well in an interactive system with fast filter variations as sources rotate and move, requiring dynamic filter coefficient updates.

Reflections (indirect) transfer. Reflected energy transfer R_{ij} represents smoothed information over directions using the cosine-squared lobe w in (14). For each radiating direction X_j, we look up the smoothed SDF to compute the octave-band dB gains and corresponding energies via

$$\hat{S}_j^k \equiv \hat{S}^k(\mathcal{R}'^{-1}(X_j)) \quad \text{and} \quad \hat{E}_j^k \equiv 10^{\hat{S}_j^k/10}. \tag{26}$$

$\hat{S}_j^k$ represents frequency-dependent radiated loudness for octave-band index k radiated in cardinal direction X_j, accounting for the source's current rotation. $\hat{E}_j^k$ then converts this to energy. We pass these through the reflections transfer matrix to model global transport through the scene to compute energy distribution around listener world direction X_i using

$$\hat{F}_i^k \equiv \sum_{j=0}^{5} 10^{R_{ij}/10}\,\hat{E}_j^k \quad \text{and} \quad \hat{L}_i^k \equiv 10\log_{10}\hat{F}_i^k. \tag{27}$$

We sum energies rather than linear amplitudes. The distinction is important: summing linear amplitudes causes quieter directions to undergo perfect (but physically unrealistic) constructive interference which washes out audible anisotropy with rendering errors as large as $10\log_{10} 6 = 7.8$dB.

The source signal $q'(t)$ is first delayed by the reflections delay, τ_1, and then equalization performed for each listener direction X_i using the corresponding octave-band loudnesses $\hat{L}_i^k$ per (27) to compute output signals $q_i(t)$, representing indirect sound from the source arriving around the listener.

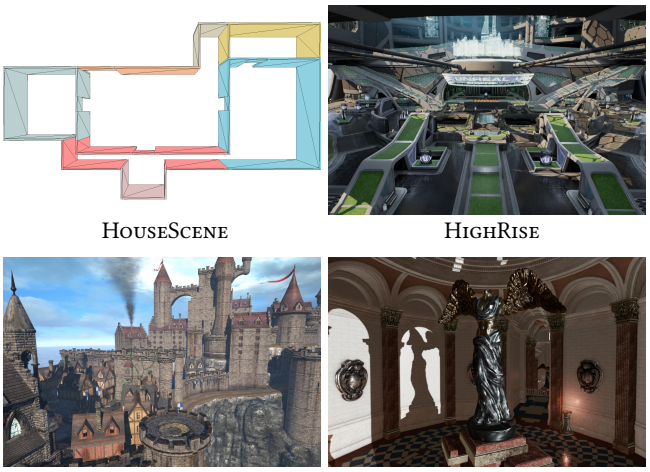

Fig. 7. Test scenes.

Reverberation. Our reverberation method follows [Raghuvanshi and Snyder 2018]; we summarize here. For each source, we obtain 6 time-varying signals $q_i(t)$ as above. Each of those 6 shares the same decoded reverb time T which interpolates between three canonical reverb filters indexed by l, representing canonical decay times that are short, medium, and long: $\{T_l\} = \{0.5, 1.5, 3.0\}$s.

We set up accumulation buffers representing sound signals for each of the six axial directions in world space around the listener (indexed by i) times three decay times (indexed by l) yielding 18 sound accumulation buffers, A_{il} (denoted by a circle-plus in the figure). For each source and each direction i, we add into A_{il} incoming $q_i(t)$ scaled by the coefficient interpolating it into the two adjacent decay times bracketing its actual decay time T. Thus each source performs 12 multiply-and-accumulate operations into appropriate buffers. These 18 buffers are accumulated over all sources. We then

perform partitioned frequency-domain convolution with the canonical reverb filter for each (i, l) (gray boxes in figure), summing results over all decay times (l) for each direction i, yielding 6 sound signals, $Q_i(t)$, shown as red arrows indexed 0 to 5 in the figure. Each $Q_i(t)$ is spatialized from world direction X_i.

Listener spatialization. Convolution with the HRTF $H^{l/r}$ in (5) is then evaluated as in [Raghuvanshi and Snyder 2018] to produce binaural output. For direct sound, s_0 is transformed to the local coordinate frame of the head, $s_0^H \equiv \mathcal{R}^{-1}(s_0)$, and $q(t)$ spatialized in this direction. Similar processing is done over all sources and the results summed to yield the initial binaural output signal $q(t)^{l/r}$. For indirect sound (reflections), we similarly transform each world coordinate axis to the listener's frame, $X_i^H \equiv \mathcal{R}^{-1}(X_i)$, and spatialize each $Q_i(t)$ in the direction X_i^H yielding $Q_i(t)^{l/r}$. Finally, we mix and play the binaural signal over headphones: $q(t)^{l/r} + \sum_i Q_i(t)^{l/r}$.

7 RESULTS

Test scene geometry is shown in Figure 7. We precompute and encode the BIR field in each scene to render the direct and indirect energy transfer between freely moving sources and listeners. Total data size is HouseScene: 32MB, SunTemple: 44MB, Citadel: 192MB, and HighRise: 650MB, with respective listener probe counts of {136, 114, 524, 1378}. As in [Raghuvanshi and Snyder 2018], each probe is processed in parallel on a compute cluster consisting of high-end 8-core CPUs. We add an overhead of 25% to [2018], with per-probe compute times of 6-8 hours at the 1kHz band-limit frequency used in our demos, and 25 minutes at 500Hz. The latter still produces plausible renderings in our informal listening. With optimizations, overhead could be reduced below 10%, based on FLOP estimates. Please listen to the supplementary video with headphones. Anechoic speech samples were obtained from [Wilkins et al. 2018].

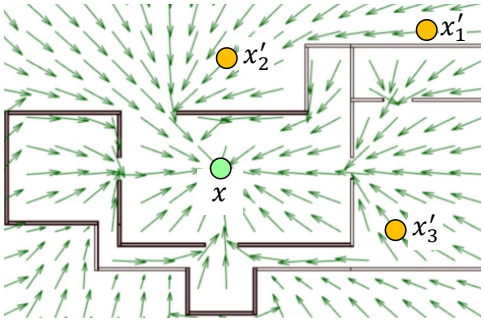

Fig. 8. Annotated snippet from Figure 1 for initial source direction.

To explain our parameter field visualizations, we zoom-in the initial source direction from Figure 1 showing results for HouseScene in Figure 8. Field visualizations show the parameter value at a fixed listener location, x (green circle at center), while varying the source location, x' over space. The visualizations restrict x' to a 2D slice passing through the listener. The arrows represent the radiated direct sound world direction, s_0', in which the wavefront must leave the source in order to arrive at the listener on the geodesic (shortest) path through the scene. For example, with source at x_1' the wavefront must first diffract at the nearby corner, then again around the central door's edge to arrive at x. Thus, the arrow at x_1', $s_0'(x, x_1')$ points to that corner. Not shown here is the direction at the listener, $s_0(x, x_1')$, which reciprocally points towards the door along the path constituting the last straight hop to the listener. Similarly, x_2' and x_3' point at the nearest portal through which the listener can be reached. In general $s_0'(x'; x)$ constitutes a flow field such that a particle advected along it always reaches x. Our system extracts this geometric information automatically from complex wave fields generated from "triangle-soup" input scene geometry.

Next we illustrate indirect energy variations due to scene geometry as captured in the RTM. To distill the complex information contained in the 36 fields for each RTM matrix element, assume a monoaural listener and sum across listener directions via (16). Intuitively, this yields the energy arriving at the listener regardless of direction, when the source radiates with a cosine-squared angular energy distribution (14) pointing in world direction X_j. These distilled fields are shown in the right six panels of Figure 1 from which Figure 9 excerpts the R_1' and R_3' fields. The complete reflection transfer matrix (RTM) for HouseScene is shown in Figure 11.

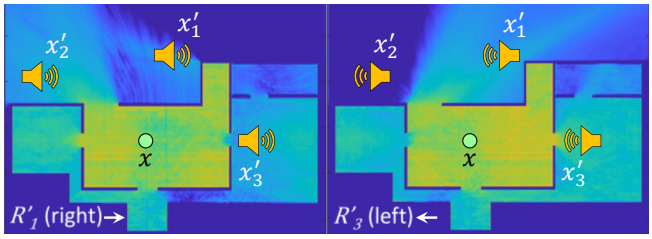

Fig. 9. Annotated snippet from Figure 1 showing listener-integrated RTM.

Consider source location x_1' in Figure 9. When the source radiates to the right (left image, R_1'), only weak scattering off the corner protruding to the source's right can enter the central room through the doorway to reach the listener at x. When the source instead radiates to the left (right image, R_3'), its energy can enter the door directly yielding a much larger value. The contrast is even bigger at x_2', because in R_3' there is no corresponding geometry present to the left of the source to redirect energy back towards the listener. At x_3', we see that when the source faces the portal leading to the listener in R_3', much more reflected energy arrives. Whereas in R_1' the source radiates away from the portal and its energy must first bounce in the adjoining room and diffract back again through the portal, becoming greatly attenuated before reaching the listener. Similar visualizations summing across listener directions are also shown for EpicCitadel (Figure 12) and SunTemple (Figure 13).

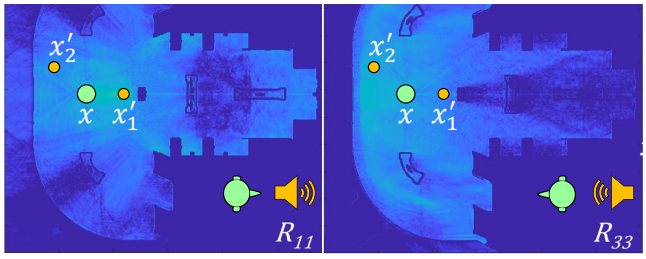

Fig. 10. Annotated snippet from Figure 14 showing RTM components.

We show the full set of parameters we encode in Figure 14 for HighRise, from which we excerpt in Figure 10 to illustrate scene-dependent information encoded in the RTM. Specifically fields for

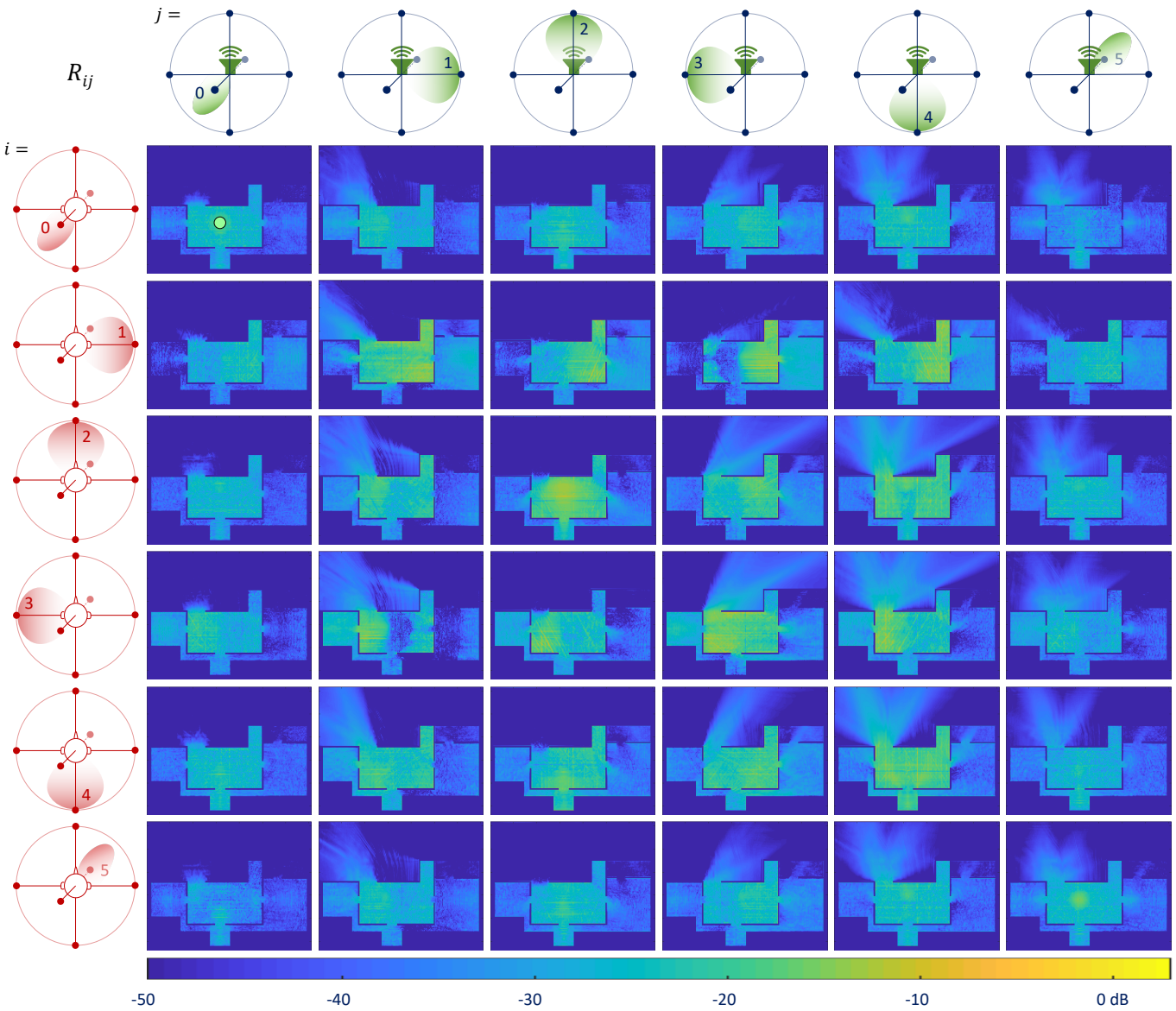

Fig. 11. Reflections transfer matrix for HouseScene. Colormap representing loudness in dB is shown at bottom.

components R_{11} (where both source and listener basis functions point right in world space) and R_{33} (where they both point left). At location x_1', sound energy scatters off the pillar to the right of x_1' to arrive at the listener at x from the right. The R_{11} field is thus much brighter in the region around x_1' compared to the same location in R_{33}, which represents energy radiated to the left of x_1' and arriving from the left of x. The situation reverses at source location x_2' which has a nearby railing to its left. The resulting scattered energy arriving at x from the left can be seen as a bright curved outline in R_{33}, a feature missing when sound instead radiates and is received to the right in R_{11}. This example illustrates bidirectional outdoor scattering; similar anisotropic effects also result from reverberation inside coupled chambers.

Audible anisotropy resulting from rendering these parameter fields can be heard in the accompanying video. The video visualizes s_0' with a yellow arrow (the source's pose is shown by a green arrow), and also plots rendered octave band dB gains for both initial and reflected sound. The latter further distills to a single equalization vector by summing (27) across listener directions X_i via $\hat{L}^k = 10 \log_{10} \sum_i \hat{F}_i^k$. It therefore accounts for the source's current pose, its SDF, and the scene's RTM.

Variation in the direct sound is easily audible in all examples as the source rotates. We start with HouseScene which demonstrates a Gaussian source. A listener inside the house and visually occluded from the source hears the initial sound loudest when the source faces the central door; i.e., when the green arrow indicating source

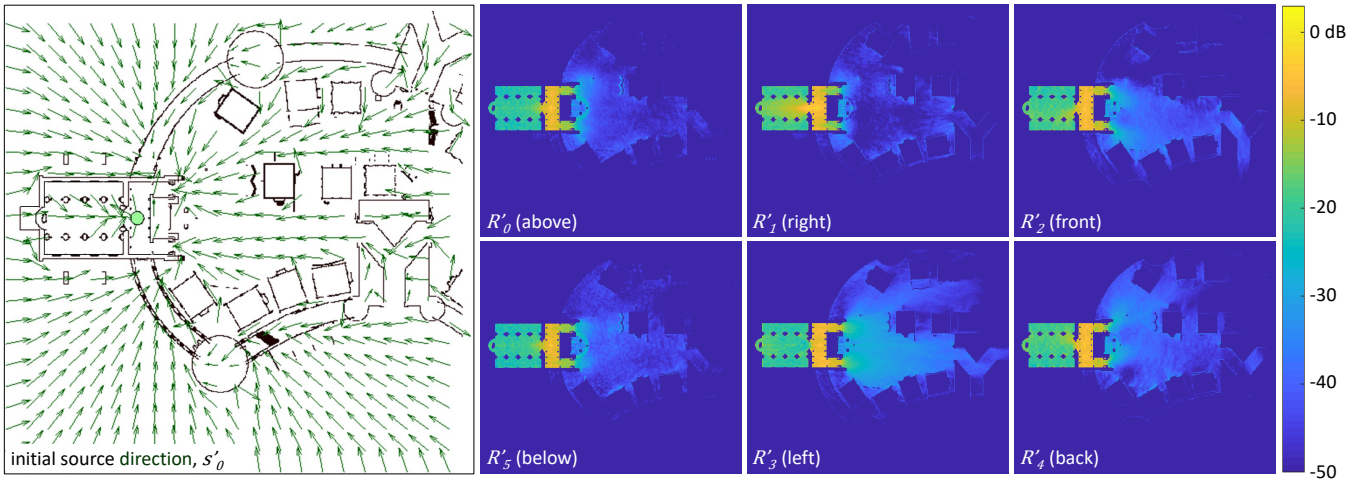

Fig. 12. Parameter field summary for EpicCitadel.

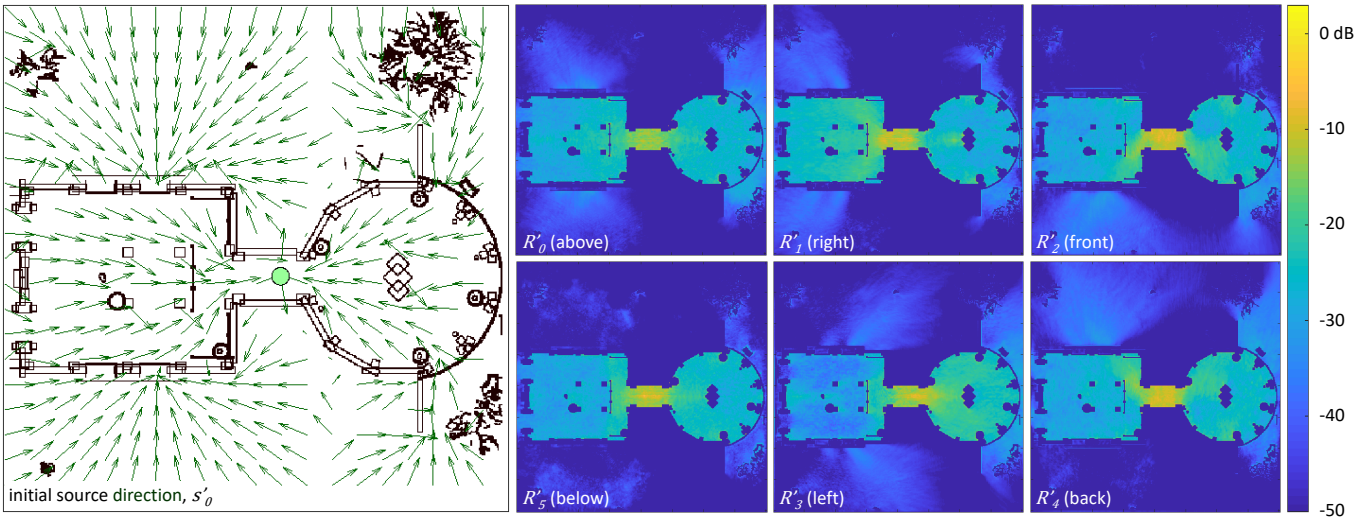

Fig. 13. Parameter field summary for SunTemple.

direction overlaps the yellow arrow for s'_0. Also note that our system remains robust in such non-line-of-sight cases which occur frequently in interactive architectural walkthroughs, games, and virtual reality.

Notice the strong variation in reflected loudness and its perceived arrival direction when the listener is outdoors and the source rotates in HouseScene. Similarly in HighRise, when the character faces towards vs. away from an adjoining chamber there is a marked difference in the loudness of reverberation as well as its spatialized direction. Capturing such effects requires a full bidirectional rendering. Also note that when source directivity is turned off in the HighRise scene with both source and listener in a reverberant chamber, the resulting sound is much louder and brighter, with strong high-frequency content in both the initial sound and reverberation. These cues result in the source to be perceived as much closer than its true distance. Source directivity is plainly necessary to capture the correct direct-to-reverberant ratio, a primary auditory cue for source distance perception [Blauert 1997].

Along with natural loudness and spatialization variations, we also render audible spectral cues contained in the SDF. Near the beginning of the SunTemple clip, note the trumpet's brighter, high frequency sound when it faces the doorway. This is also true for the reflections (auralized separately with initial sound off for illustration, but still audible in combination) as can also be seen in the octave gain visualization.

8 CONCLUSION

Our reciprocal decomposition of acoustic rendering in terms of the bidirectional impulse response in (5) and its approximation with flux in (12) yields a practical system that, for the first time, renders directional wave effects as both source and listener interactively move and rotate in complex, general scenes. The CPU cost of the technique is insensitive to scene complexity and angular resolution of source directivity functions or HRTFs.

Near-field radiation effects can be incorporated without much extra memory or computational cost. The encoded initial and reflected

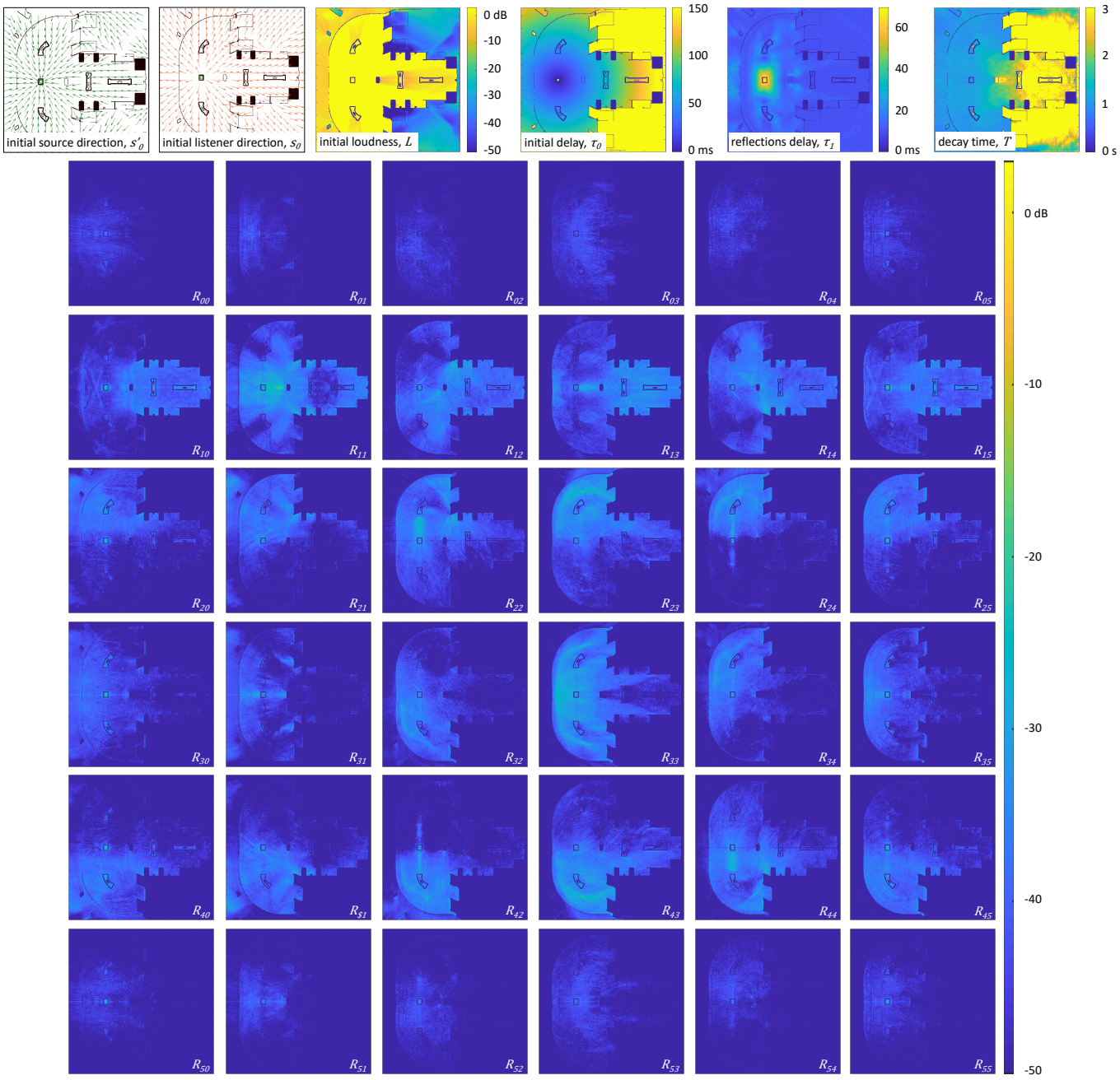

Fig. 14. Complete set of parameter fields for HighRise. Colormaps for parameter scale are shown at right of each parameter.

sound delay parameters can be used to derive propagation distances (by multiplication with the speed of sound) as additional lookup parameters into a distance-dependent source directivity function (SDF). From there rendering proceeds as described.

A natural extension is to encode frequency-dependent propagation effects by making initial loudness (L), the reflections transfer matrix (R_{ij}) and decay time (T) depend on frequency. We expect this should improve realism though obviously at the expense of data size. Exploring matrix compression for the RTM while maintaining spatial smoothness is a related direction for future work.

Outdoor reveberation remains a significant limitation of our system because we only capture echoes statistically (in the formulation of reverb filters) and fail to explicitly characterize discrete echoes present in simulation. Defining, extracting and encoding perceptually salient early reflections while maintaining spatial smoothness is a challenging problem.

A promising future direction is to combine our precomputed system for the static scene with flexible approximations for dynamic geometry like destructible walls or moving occluders, paralleling developments in lighting for games. Our reciprocal encoding contains complex topological (such as all-pairs shortest path) information useful for path planning, extracted from the systematic global exploration of the scene that wave propagation induces.

Our reciprocal decomposition might be applicable for visual rendering, to allow interactive change to both light source (including its location, direction, and intensity profile) and camera. The transfer operator must not be too high-resolution to be practical, yielding a small matrix that doesn't vary too fast in (x, x'), and probably more useful for indirect rather than direct transfer, as we use the RTM.

REFERENCES

Chakravarty R. Alla Chaitanya, John M. Snyder, Keith Godin, Derek Nowrouzezahrai, and Nikunj Raghuvanshi. 2019. Adaptive Sampling for Sound Propagation. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1846–1854. https://doi.org/10.1109/TVCG.2019.2898765

Andrew Allen and Nikunj Raghuvanshi. 2015. Aerophones in Flatland: Interactive Wave Simulation of Wind Instruments. *ACM Trans. Graph.* 34, 4 (July 2015), 11. https://doi.org/10.1145/2767001

Samuel D. Bellows, Claire M. Pincock, Jennifer K. Whiting, and Timothy W. Leishman. 2019. Averaged Speech Directivity. https://scholarsarchive.byu.edu/directivity/1 Online dataset. Retrieved Jan 5, 2020.

Stefan Bilbao, Jens Ahrens, and Brian Hamilton. 2019. Incorporating Source Directivity in Wave-based Virtual Acoustics: Time-domain Models and Fitting to Measured Data. *The Journal of the Acoustical Society of America* 146, 4 (2019), 2692–2703. https://doi.org/10.1121/1.5130194

J. Blauert. 1997. An Introduction to Binaural Technology. In *Binaural and Spatial Hearing in Real and Virtual Environments*, R. Gilkey and T. R. Anderson (Eds.). Lawrence Erlbaum, USA.

Chunxiao Cao, Zhong Ren, Carl Schissler, Dinesh Manocha, and Kun Zhou. 2016. Interactive Sound Propagation with Bidirectional Path Tracing. *ACM Trans. Graph.* 35, 6, Article 180 (Nov. 2016), 11 pages. https://doi.org/10.1145/2980179.2982431

Jeffrey N. Chadwick, Steven S. An, and Doug L. James. 2009. Harmonic Shells: A Practical Nonlinear Sound Model for Near-rigid Thin Shells. In *SIGGRAPH Asia '09: ACM SIGGRAPH Asia 2009 papers* (Yokohama, Japan). ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/1661412.1618465

Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum. 2000. Plenoptic Sampling. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH âĂŹ00)*. ACM Press/Addison-Wesley Publishing Co., USA, 307âĂŞ318. https://doi.org/10.1145/344779.344932

Yoshinori Dobashi, Kazufumi Kaneda, Hideki Nakatani, and Hideo Yamashita. 1995. A Quick Rendering Method Using Basis Functions for Interactive Lighting Design. *Comput. Graph. Forum* 14 (1995), 229–240.

Jean-Jacques Embrechts. 2016. Review on the Applications of Directional Impulse Responses in Room Acoustics. In *Proceedings of CFA 2016* (Le Mans, France). Société française d'acoustique (SFA). http://orbi.ulg.ac.be/handle/2268/193820

Jorg Fliege and Ulrike Maier. 1999. The Distribution of Points on the Sphere and Corresponding Cubature Formulae. *IMA J. Numer. Anal.* 19, 2 (April 1999), 317–334. https://doi.org/10.1093/imanum/19.2.317

Anders Gade. 2007. Acoustics in Halls for Speech and Music. In *Springer Handbook of Acoustics* (2007 ed.), Thomas Rossing (Ed.). Springer, Chapter 9. http://www.worldcat.org/isbn/0387304460

Nail A. Gumerov and Ramani Duraiswami. 2005. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions (Elsevier Series in Electromagnetism)* (1 ed.). Elsevier Science, Amsterdam. http://www.worldcat.org/isbn/0080443710

Doug L. James, Jernej Barbic, and Dinesh K. Pai. 2006. Precomputed Acoustic Transfer: Output-sensitive, Accurate Sound Generation for Geometrically Complex Vibration Sources. *ACM Transactions on Graphics* 25, 3 (July 2006), 987–995. https://doi.org/10.1145/1141911.1141983

Jaakko Lehtinen and Jan Kautz. 2003. Matrix Radiance Transfer. In *Proceedings of the 2003 Symposium on Interactive 3D Graphics* (Monterey, California) *(I3D âĂŹ03)*. Association for Computing Machinery, New York, NY, USA, 59âĂŞ64. https://doi.org/10.1145/641480.641495

Dingzeyu Li, Timothy R. Langlois, and Changxi Zheng. 2018. Scene-Aware Audio for 360° Videos. *ACM Trans. Graph.* 37, 4, Article Article 111 (July 2018), 12 pages. https://doi.org/10.1145/3197517.3201391

Ruth Y. Litovsky, Steven H. Colburn, William A. Yost, and Sandra J. Guzman. 1999. The Precedence Effect. *The Journal of the Acoustical Society of America* 106, 4 (1999), 1633–1654. https://doi.org/10.1121/1.427914

Ravish Mehra, Lakulish Antani, Sujeong Kim, and Dinesh Manocha. 2014. Source and Listener Directivity for Interactive Wave-Based Sound Propagation. *IEEE Transactions on Visualization and Computer Graphics* 20, 4 (April 2014), 495–503. https://doi.org/10.1109/tvcg.2014.38

Richard J. Oliver and Jean-Marc Jot. 2015. Efficient Multi-Band Digital Audio Graphic Equalizer with Accurate Frequency Response Control. In *Audio Engineering Society Convention 139*. http://www.aes.org/e-lib/browse.cfm?elib=17963

Allan D. Pierce. 1989. *Acoustics: An Introduction to Its Physical Principles and Applications*. Acoustical Society of America. http://www.worldcat.org/isbn/0883186128

Ville Pulkki. 2007. Spatial Sound Reproduction with Directional Audio Coding. *J. Audio Eng. Soc* 55, 6 (2007), 503–516. http://www.aes.org/e-lib/browse.cfm?elib=14170

Nikunj Raghuvanshi, Rahul Narain, and Ming C. Lin. 2009. Efficient and Accurate Sound Propagation Using Adaptive Rectangular Decomposition. *IEEE Transactions on Visualization and Computer Graphics* 15, 5 (2009), 789–801. https://doi.org/10.1109/tvcg.2009.28

Nikunj Raghuvanshi and John Snyder. 2014. Parametric Wave Field Coding for Precomputed Sound Propagation. *ACM Trans. Graph.* 33, 4, Article 38 (July 2014), 11 pages. https://doi.org/10.1145/2601097.2601184

Nikunj Raghuvanshi and John Snyder. 2018. Parametric Directional Coding for Precomputed Sound Propagation. *ACM Trans. Graph.* 37, 4, Article 108 (July 2018), 14 pages. https://doi.org/10.1145/3197517.3201339

Nikunj Raghuvanshi, John Tennant, and John Snyder. 2017. Triton: Practical precomputed sound propagation for games and virtual reality. *The Journal of the Acoustical Society of America* 141, 5 (2017), 3455–3455. https://doi.org/10.1121/1.4987164 arXiv:https://doi.org/10.1121/1.4987164

Lauri Savioja. 2010. Real-Time 3D Finite-Difference Time-Domain Simulation of Mid-Frequency Room Acoustics. In *13th International Conference on Digital Audio Effects (DAFx-10)* (Graz, Austria).

Lauri Savioja and U. Peter Svensson. 2015. Overview of Geometrical Room Acoustic Modeling Techniques. *The Journal of the Acoustical Society of America* 138, 2 (01 Aug. 2015), 708–730. https://doi.org/10.1121/1.4926438

Noam R. Shabtai, Gottfried Behler, Michael Vorländer, and Stefan Weinzierl. 2017. Generation and Analysis of an Acoustic Radiation Pattern Database for Forty-one Musical Instruments. *The Journal of the Acoustical Society of America* 141, 2 (2017), 1246–1256. https://doi.org/10.1121/1.4976071 arXiv:https://doi.org/10.1121/1.4976071

Peter-Pike Sloan, Jan Kautz, and John Snyder. 2002. Precomputed Radiance Transfer for Real-time Rendering in Dynamic, Low-frequency Lighting Environments. *ACM Trans. Graph.* 21, 3 (July 2002), 527–536. https://doi.org/10.1145/566654.566612

Julius O. Smith. 2007. *Introduction to Digital Filters with Audio Applications*. W3K Publishing, http://www.w3k.org/books/.

Vesa Välimäki and Juho Liski. 2017. Accurate Cascade Graphic Equalizer. *IEEE Signal Processing Letters* 24, 2 (Feb 2017), 176–180. https://doi.org/10.1109/LSP.2016.2645280

Vesa Välimäki and Joshua D. Reiss. 2016. All About Audio Equalization: Solutions and Frontiers. *Applied Sciences* 6, 5 (2016). https://doi.org/10.3390/app6050129

Jui-Hsien Wang, Ante Qu, Timothy R. Langlois, and Doug L. James. 2018. Toward Wave-based Sound Synthesis for Computer Animation. *ACM Trans. Graph.* 37, 4, Article 109 (July 2018), 16 pages. https://doi.org/10.1145/3197517.3201318

Stefan Weinzierl, Michael Vorländer, Gottfried Behler, Fabian Brinkmann, Henrik von Coler, Erik Detzner, Johannes KrÃđmer, Alexander Lindau, Martin Pollow, Frank Schulz, and Noam R. Shabtai. 2017. A Database of Anechoic Microphone Array Measurements of Musical Instruments. https://depositonce.tu-berlin.de/handle/11303/6305.2. Online dataset. Retrieved Jan 5, 2020.

Julia Wilkins, Prem Seetharaman, Alison Wahl, and Bryan Pardo. 2018. VocalSet: A Singing Voice Dataset. https://doi.org/10.5281/zenodo.1442513 Online dataset. Retrieved Jan 5, 2020.

Zechen Zhang, Nikunj Raghuvanshi, John Snyder, and Steve Marschner. 2018. Ambient Sound Propagation. *ACM Trans. Graph.* 37, 6, Article 184 (Dec. 2018), 10 pages. https://doi.org/10.1145/3272127.3275100

Zechen Zhang, Nikunj Raghuvanshi, John Snyder, and Steve Marschner. 2019. Acoustic Texture Rendering for Extended Sources in Complex Scenes. *ACM Trans. Graph.* 38, 6, Article 222 (Nov. 2019), 9 pages. https://doi.org/10.1145/3355089.3356566