Parametric Directional Coding for Precomputed Sound Propagation

NIKUNJ RAGHUVANSHI, Microsoft Research
JOHN SNYDER, Microsoft Research

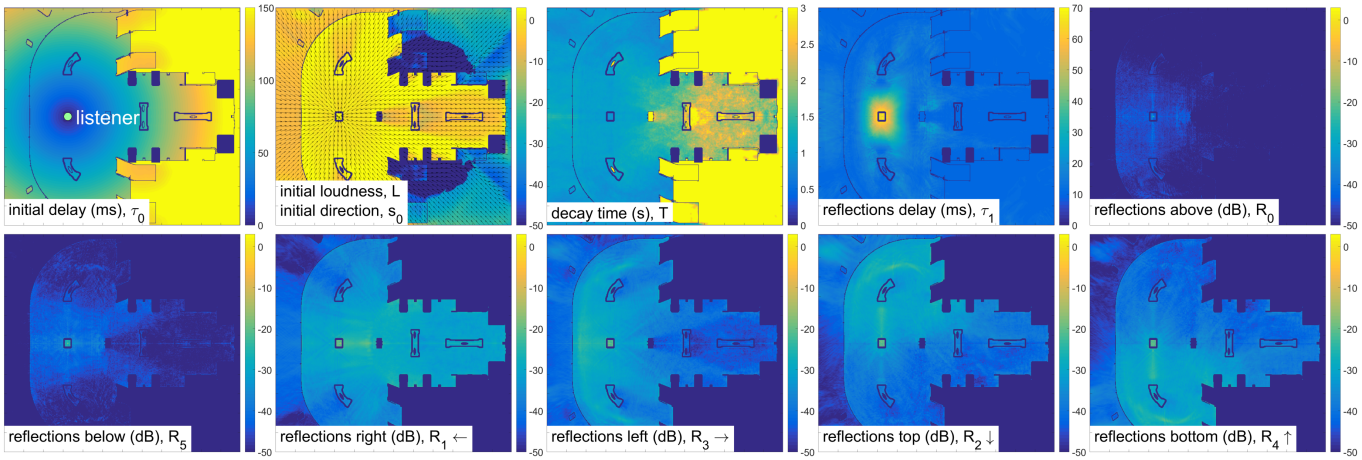


Fig. 1. Parameter fields for HighRise. Source position varies in 3D; one horizontal slice is shown. The listener is held fixed at the green circle.

Convincing audio for games and virtual reality requires modeling directional propagation effects. The initial sound's arrival direction is particularly salient and derives from multiply-diffracted paths in complex scenes. When source and listener straddle occluders, the initial sound and multiply-scattered reverberation stream through gaps and portals, helping the listener navigate. Geometry near the source and/or listener reveals its presence through anisotropic reflections. We propose the first precomputed wave technique to capture such directional effects in general scenes comprising millions of polygons. These effects are formally represented with the 9D directional response function of 3D source and listener location, time, and direction at the listener, making memory use the major concern. We propose a novel parametric encoder that compresses this function within a budget of ~100MB for large scenes, while capturing many salient acoustic effects indoors and outdoors. The encoder is complemented with a lightweight signal processing algorithm whose filtering cost is largely insensitive to the number of sound sources, resulting in an immediately practical system.

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

Additional Key Words and Phrases: virtual acoustics, sound propagation, spatial audio, directional impulse response, vector intensity, flux density, HRTF, plane wave decomposition, wave equation.

Authors' addresses: Nikunj Raghuvanshi, nikunjr@microsoft.com, Microsoft Research; John Snyder, johnsny@microsoft.com, Microsoft Research.

0730-0301/2018/8-ART108 $15.00
https://doi.org/10.1145/3197517.3201339

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

1 INTRODUCTION

Hearing is directional, complementing vision by detecting where sound events occur in our environment. Standing outside a hall, we're able to locate the open door through which streams a crowd's chatter even when the door is obscured or lies behind us. As we walk inside, the auditory scene wraps around us. Meanwhile the initial sound's direction smoothly resolves from the door to each of the speakers, helping us face and navigate to a chosen individual. While reflections envelope the listener indoors, partly open spaces yield anisotropic reflections, reinforcing the visual location of nearby scene geometry. When source and listener are close (e.g. with footsteps), the delay between the initial sound and first reflections can become audible, strengthening the perception of distance to walls.

Our goal is practical modeling and rendering of such directional acoustic effects for games and VR applications in scenes containing millions of polygons. Manual mesh simplification or scene decomposition must be avoided. Many physical effects are perceivable and must be simulated accurately and efficiently as the sound diffracts around obstacles and through portals and scatters many times. Transient effects are critical. Initial sound arriving in the first 1ms, possibly through multiple portals, determines its perceived direction. The directional distribution of later arriving reflections convey additional information about the listener's surroundings.

Ours differs from typical problems in room acoustics [Kuttruff 2000]. We handle real-time movement of both sources and listener, capturing variation in complex scenes without breaking the fraction-of-a-core CPU budget typical for games. Effects should be robust even in occluded cases where conventional path tracing demands

enormous sampling to produce smooth results, greatly exceeding what can be computed within budget.

We take a precomputed, wave-based approach that simulates frequencies up to 1000Hz in our experiments. Paths are not explicitly traced thus obviating aliasing problems. On-line CPU demands are modest. The challenge becomes compact encoding of the petabyte-scale wave fields generated. Comparing to previous work, [Mehra et al. 2013] capture directional acoustics but handle only outdoor scenes composed of 10-20 explicitly separated objects like building facades or boulders. The technique is CPU- and memory-intensive and doesn't support general scenes like the ones we demonstrate. [Raghuvanshi and Snyder 2014] limit memory and computation in general scenes but neglect directionality.

Directional audio codecs (Surround, Ambisonics, DiRaC) aim to compactly encode the *sound field* at a listener due to superposed propagated signals from many sound sources. They produce an audio stream for efficient playback, usually to accompany visual frames. More recent encodings (Dolby Atmos, MPEG-H) add limited interactivity during decoding with steerable point emitters ("audio objects") but the scene's acoustics are not modeled.

We instead encode the entire 9D spatially-varying *directional impulse response* (DIR) field for a static scene as a function of both source and receiver position, direction, and time. This lets us model acoustic effects for arbitrarily moving listener and sources that can emit any signal at runtime. The parameters we encode are shown in Figure 1: delay, direction, and loudness of the initial sound (where direction is coded directly as a 3D vector), delay and direction of the early reflections (where direction is coded in terms of 6 coarse directional basis functions labeled "above", "below", "right", "left", "front", and "back" in the figure), and 60dB decay time of response energy after onset of reflections.

No published work has yet exploited the spatial coherence we show inheres in perceptual coding of DIR fields. Unlike a sound field which combines multiple baked-in sources and so entangles source signal content with propagation effects, the DIR separates these. We show with virtual experiments that arrival directions in the DIR are independent of frequency to a good approximation, even in the presence of edge diffraction and scattering. Said another way, for a given source and listener position, most of the DIR's energy in any given transient phase of the response comes from a consistent set of directions across frequency. This lets us avoid encoding and rendering frequency-dependent directions.

Our work is close in motivation to DiRaC [Laitinen et al. 2012] which uses flux density (also called vector intensity) for directional analysis and harnesses perception to reduce memory. But DiRaC aims at encoding general sound fields and requires orders of magnitude too much memory if applied directly to our problem. We reduce memory demands by specializing perceptual coding to DIR fields rather than sound fields, extracting only a few salient parameters, and exploiting their spatial coherence. We also show for the first time the remarkable agreement flux density achieves with ground truth (linear) plane wave decomposition, despite its nonlinear formulation and faster computation.

We complement the encoder with a lightweight rendering technique that applies DIR filters for each sound source with cost largely insensitive to the number of sources. This results in a system that can handle large 3D scenes within practical RAM and CPU budget for games and VR.

2 PRELIMINARIES

We provide brief background on directional sound propagation needed to understand ours and related work.

2.1 Green's Function and the DIR Field

Sound propagation can be represented in terms of Green's function [Pierce 1989], p, representing pressure deviation satisfying the wave equation:

$$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right] p\left(t, x, x'\right) = \delta\left(t\right)\delta\left(x - x'\right), \tag{1}$$

where $c = 340$m/s is the speed of sound and δ the Dirac delta function representing the PDE's forcing impulse. Holding (x, x') fixed, $p\left(t;\, x, x'\right)$ yields the *impulse response* at a 3D receiver point x due to a spatio-temporal impulse introduced at point x'. Thus p forms a 6D field of impulse responses capturing global propagation effects like scattering and diffraction determined by the boundary conditions which comprise the geometry and materials of the scene. In nontrivial scenes, analytical solutions are unavailable and p must be sampled via computer simulation or real-world measurements. The principle of *acoustic reciprocity* holds that under fairly general conditions Green's function is invariant to interchange of source and receiver: $p(t, x, x') = p(t, x', x)$.

We confine our attention to omni-directional point sources. The response at x due to a source at x' emitting a pressure signal $\tilde{q}(t)$ can be recovered from Green's function via a temporal convolution, denoted by $*$, as

$$q\left(t;\, x, x'\right) = \tilde{q}(t) * p\left(t;\, x, x'\right). \tag{2}$$

As we discuss in Section 4, $p\left(t, x;\, x'\right)$ in any finite, source-free region centered at x can be uniquely expressed as a sum of plane waves, which form a complete basis for free-space propagation. The result is a decomposition into signals propagating along plane wavefronts arriving from various directions, called the *directional impulse response (DIR)* [Embrechts 2016]. Refer to Figure 2. Applying this decomposition at each (x, x') yields the directional impulse response field, denoted $d(s, t, x, x')$, where s parameterizes arrival direction. Our goal is to compute and compactly encode the DIR field so that it can be perceptually reproduced for any number of sound sources and associated signals, efficiently at runtime.

2.2 Binaural Rendering with the HRTF

The response of an incident plane wave field $\delta(t + s \cdot \Delta x/c)$ from direction s can be recorded at the left and right ears of a person. Δx denotes position with respect to the listener's head centered at x. Assembling this information over all directions yields the person's Head-Related Transfer Function (HRTF), denoted $h^{L/R}(s, t)$. Low-to-mid frequencies (<1000Hz) correspond to wavelengths much larger than the head and diffract around it, creating a detectable time difference between the two ears. Higher frequencies are shadowed, causing a significant loudness difference. These phenomena, respectively called the *interaural time difference* (ITD) and the *interaural level difference* (ILD), let us localize sources [Blauert 1997]. Both

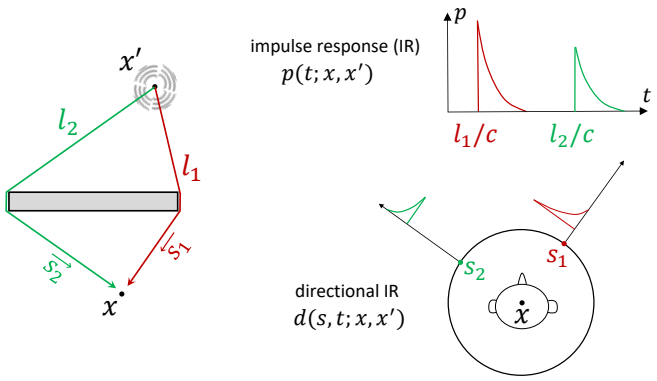


Fig. 2. Directional impulse response (DIR). This simplified diagram shows a pulse emanating from a source at x' and two diffracted wavefronts subsequently arriving at listener position x from directions s_1 and s_2. The resulting IR is plotted in the upper right and the corresponding DIR on the lower right, parameterizing the response in terms of both time and direction. Our system aims to encode the DIR's perceptual properties for all pairs of source (x') and listener (x) positions. At runtime, given the listener head location, orientation, source location and emitted sound, we efficiently render the DIR's audible effect on the sound as heard by listener.

are functions of direction as well as frequency, and depend on the particular geometry of the person's pinna, head and shoulders.

Given the HRTF, rotation matrix $\mathcal{R}$ mapping from head to world coordinate system, and DIR field absent the listener's body, *binaural rendering* reconstructs the signals entering the two ears, $q^{L/R}$, via

$$q^{L/R}(t; x, x') = \tilde{q}(t) * p^{L/R}(t; x, x') \tag{3}$$

where $p^{L/R}$ is the *binaural impulse response*

$$p^{L/R}(t; x, x') \equiv \int_{S^2} d\left(s, t;\, x, x'\right) * h^{L/R}\left(\mathcal{R}^{-1}(s), t\right) ds. \tag{4}$$

S^2 indicates the spherical integration domain and ds the differential area of its parameterization, $s \in S^2$. Note that in audio literature the terms "spatial" and "spatialization" refer to directional dependence (on s) rather than source/listener dependence (on x and x').

We use a generic HRTF dataset combining measurements across many subjects using the setup described in [Bilinski et al. 2014]. It samples binaural responses for $N_H = 2048$ discrete directions $\{s_j\}, j \in [0, N_H - 1]$ uniformly spaced over the sphere. Any HRTF dataset may be used with our technique.

3 RELATED WORK

Room acoustics has intensively studied sound propagation, with special focus on single-chamber indoor spaces such as concert halls [Kuttruff 2000]. Such spaces can be auralized in real-time [Vorländer 2007] but can monopolize desktop scale computation, suitable for walk-through auralizations in architectural acoustics. Games and VR applications require techniques that are faster, with approximations motivated by more general scenes including outdoor and multi-chamber spaces. Occluded cases such as diffracted sound arriving around doorways are of central importance.

3.1 Geometric and wave solvers

Room acoustics usually takes a geometric/Lagrangian approach [Kuttruff 2000; Rindel and Christensen 2013], propagating rays in a high-frequency approximation to the wave equation. Handling sub-wavelength geometric features (in practice, smaller than a few meters) necessitates user-guided scene simplification. Modeling all physical paths containing edge diffractions and surface scattering of arbitrary order remains an unsolved problem; see [Savioja and Svensson 2015, Table 1] for a survey.

Time-domain wave solvers take an Eulerian approach [Hamilton et al. 2017; Murphy et al. 2007; Raghuvanshi et al. 2009a], producing a 4D slice of Green's function $p(t, x;\, x')$ for a given source location x'. Aliasing is eliminated by bandlimiting the source signal. All propagation paths of length less than the simulation duration are included without being generated explicitly. Directional information must be extracted with careful processing, as we will describe in Section 7. Complex scene geometry can be treated without user intervention by voxelizing it into the simulation grid.

Computation scales as ν_m^4 where ν_m is the upper limit on simulation frequency. While this cost motivates the traditional preference for geometric approximations in room acoustics [Siltanen et al. 2010a], wave methods have seen increasing interest as CPU/GPU computing power has grown and algorithms improved. Time-domain simulations on desktop machines can now handle up to middle frequencies (~1000Hz) on concert hall sized scenes [Hamilton et al. 2017; Mehra et al. 2012]. These solutions are usually extrapolated to higher frequencies as we do here or can be combined with geometric techniques [Yeh et al. 2013].

3.2 Precomputed simulation

Precomputed approaches analyze static scene geometry offline and store a compressed encoding of some portion of Green's function. Increased memory is thus traded for reduced runtime computation.

[Tsingos 2009] describes one of the first practical precomputed techniques for approximating reflections and directional reverberation in games. Diffraction is ignored so that correct initial arrival direction or attenuation due to occlusion cannot be modeled. Reliance on the image source method means input geometry must consist of a few large planar facets. Acoustic radiance transfer [Antani et al. 2012; Siltanen et al. 2010b] is analogous to radiosity for light transport. Under the geometric approximation, it efficiently models diffuse energy transport in complex scenes, compressing the global operator using singular value decomposition (SVD). The technique is designed for coarsely directional late reverberation but not the (highly directional) initial sound or specular early reflections.

The Equivalent Source Method (ESM) approximates $p(t, x, x')$ as a linear superposition of elementary multipole solutions individually satisfying (1) with free-field boundary conditions. [James et al. 2006] introduced ESM to computer graphics to model directional radiation including self-shadowing and self-scattering from an isolated vibrating object. Extensions focus on accelerating runtime cost of summing multipole evaluations [Chadwick et al. 2009] and reducing data size [Li et al. 2015].

ESM has also been applied to global sound propagation [Mehra et al. 2013, 2015]. The latter supports directional effects for a moving source and listener but is limited to sparse outdoor scenes consisting of a few well-separated objects. Scene acoustics are encoded as a per-frequency global transport operator using SVD-based lossy compression similar to [Antani et al. 2012]. Runtime memory usage

increases with the frequency limit and number of objects, taking 15GB for 5 objects (rock faces) at ν_m = 1000Hz [Mehra et al. 2015, Fig. 4 middle, Table 3]. Typical game scenes we target include indoor/mixed spaces that do not admit such decomposition or outdoor spaces containing thousands of such objects. At runtime, the global transport matrix must be applied per frequency. This is expensive, requiring 100ms of computation that saturates the GPU and all CPU cores for just one sound source.

Volumetric approaches [Raghuvanshi and Snyder 2014; Raghuvanshi et al. 2010] offload nearly all computation by precomputing Green's function for a discrete set of probes $\{x'\}$. Instead of representing the highly oscillatory Green's function p directly, it is perceptually coded in terms of its frequency-averaged energy and decay in each transient phase. The resulting perceptual parameter fields are smooth and highly compressible. Initial sound is rendered in the line-of-sight direction; reflections and reverberation are rendered as arriving isotropically at the listener. Our work extends this technique by extracting and perceptually encoding the Directional Impulse Response (DIR) field.

3.3 Online simulation

Online techniques compute the directional response at runtime and so handle dynamic geometry. For single-scattering, the Kirchoff approximation can be applied [Tsingos et al. 2007] but most methods use geometric acoustics.

RAVEN [Schröder 2011] is perhaps the most complete geometric acoustic system, allowing real-time architectural walkthroughs for appropriately simplified (few thousand polygon) scenes using multiple desktop machines. Closer to our application, path tracing has been used for sound propagation [Chandak et al. 2008; Schissler et al. 2014; Taylor et al. 2009], where the challenge is to obtain consistent, aliasing-free results while staying within CPU budget. Compared to light transport, convergence is needed for the sum within each time bin of the energy response, making path length an additional search dimension. Terminating path search too soon yields incoherent "pivots" in source direction that perceptually derives from the shortest path direction, or "pops" in loudness as the source or listener moves. Convergence issues are studied in [Cao et al. 2016].

3.4 Sound field coding

Ambisonics [Gerzon 1973] represents the sound field around a point using spherical harmonic coefficients and independently of the reproduction setup (speakers or headphones). Parametric surround approaches, such as MPEG-Surround [Breebaart et al. 2005], assume a known speaker configuration around the listener. The encoder's input is a multi-channel signal: raw waveforms meant to be played back on corresponding speakers. These channels are summed into one or two downmix streams to generate an aggregate sound coming from all directions, and compressed using standard waveform coding. In parallel, time-frequency processing is performed to extract a stream of perceptual parameters describing the directional properties of each input channel relative to the downmix. These parameters are based on binaural directional perception, such as level difference, phase difference and inter-aural coherence (diffuseness) in each time-frequency bin.

MPEG-H [Herre et al. 2015] extends these ideas to allow encoding agnostic to the reproduction setup and support higher-order Ambisonics and binaural rendering. A few point source locations can also be specified along with their mono-aural signals. Environmental acoustics is outside the standard's scope and must be supplied explicitly in the form of binaural responses [Herre et al. 2015, Fig.3]. Computing these is the focus of our paper.

3.5 Directional Impulse Response (DIR) coding

Existing techniques encode directional responses at one or few points (see [Embrechts 2016] for a survey); ours is the first work to compress the entire response field $d(s, t, x, x')$ and exploit its spatial coherence once transformed to directional parameters.

The Spatial Decomposition Method (SDM) [Tervo et al. 2013] fits an image source model to responses measured with a microphone array, approximating it at a point with multiple delayed spherical wavefronts. The intent is analysis rather than compact encoding. Closer to our work, [Laitinen et al. 2012] propose Directional Audio Coding (DiRaC). The input is the directional sound signal at a listener, which is a superposition of all sound source signals in a scene convolved with the corresponding DIRs. DiRaC computes direction and a diffuseness parameter for each of many time-frequency bins, using the same flux density formulation we apply. Since the input signal can contain many sound events at arbitrary times, time is discretized uniformly to adequately capture each event's onset. The time bin size must be determined carefully (usually ~10ms). Finer bins better resolve sound event onsets, but increase encoded size and limit frequency resolution via Fourier uncertainty.

The techniques underlying DiRaC generalize to DIR encoding because a DIR is a special directional sound signal generated by a single, impulsive point source. This application of DiRaC is Spatial Impulse Response Rendering (SIRR) [Merimaa and Pulkki 2005]. Our approach instead specializes to DIRs, resulting in two advantages. First, as we show in Section 4.4, while frequency-dependent directions are necessary to encode directional sound signals that mix multiple sources, directional information is largely frequency-independent in DIRs. This reduces memory use and rendering cost. Second, we focus temporal resolution where it matters by explicitly locating the first arrival time, and recording the initial direction with high (1ms) temporal resolution matching human auditory perception, and coarser resolution later.

4 DIRECTIONAL ANALYSIS

We describe directional analysis of sound fields, comparing a reference solution (plane wave decomposition or PWD), to the approximation our encoder actually employs (flux density).

4.1 Plane Wave Decomposition (PWD)

Let Δx denote relative position in a volume centered around the listener at x where the local pressure field is to be directionally analyzed. For any source position x' (hereafter dropped), we denote the local IR field by $p(\Delta x, t)$ and the Fourier transform of the time-dependent signal for each Δx by $P(\Delta x, \omega) \equiv \mathcal{F}[p(\Delta x, t)]$. In general, we denote the Fourier transform of $g(t)$ as $G(\omega) \equiv \mathcal{F}[g(t)] \equiv \int_{-\infty}^{\infty} g(t)\, e^{i\omega t}\, dt$, assuming time-harmonic dependence of the form

$e^{-i\omega t}$. We drop angular frequency ω from the notation in the following; it is understood that the directional analysis we describe must be performed for each value of ω.

Parameterizing in terms of spherical coordinates, $\Delta x = r\, s(\theta, \phi)$ where $s(\theta, \phi) \equiv (\sin\theta\, \cos\phi, \sin\theta\, \sin\phi, \cos\theta)$ represents a unit direction and $r \equiv \|\Delta x\|$. This coordinate system yields orthogonal solutions (modes) of the Helmholtz equation, allowing representation of the solution P in any source-free region via

$$P(\Delta x) = \sum_{l,m} P_{l,m}\; b_l(\kappa\, r)\; Y_{l,m}(s), \tag{5}$$

where the mode coefficients $P_{l,m}$ determine the field uniquely. The function b_l is the (real-valued) spherical Bessel function; $\kappa \equiv \omega/c \equiv 2\pi\nu/c$ is the wavenumber where ν is the frequency. The notation $\sum_{l,m} \equiv \sum_{l=0}^{n-1} \sum_{m=-l}^{l}$ indicates the sum over all integer modes where $l \in [0, n-1]$ is the order, $m \in [-l, l]$ the degree, and n the truncation order. Lastly, $Y_{l,m}$ are the n^2 complex spherical harmonic (SH) basis functions defined as

$$Y_{l,m}(s) \equiv \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\; \mathcal{P}_{l,m}(\cos\theta)\, e^{im\phi} \tag{6}$$

where $\mathcal{P}_{l,m}$ is the associated Legendre function.

Diffraction limit. Suppose the sound field is observed by an ideal microphone array within a spherical region $\|\Delta x\| \le r_o$ free of sources and boundary. The mode coefficients can be estimated by inverting the linear system represented by (5) to find the unknown (complex) coefficients $P_{l,m}$ in terms of the known (complex) coefficients of the sound field, $P(\Delta x)$. The angular resolution of any wave field sensor is *fundamentally* restricted by the size of the observation region, which is the diffraction limit. This manifests mathematically as an upper limit on the SH order n dependent on r_0 to keep the linear system well-conditioned.

Such analysis is standard in fast multipole methods for 3D wave propagation [Gumerov and Duraiswami 2005] and for processing output of spherical microphone arrays [Rafaely 2015]. One must compensate for the scattering that real microphone arrays introduce in the act of measuring the wave field. Our synthetic case avoids these difficulties since "virtual microphones" simply record pressure without scattering. Nevertheless, prior work is sparse on directional analysis of sound fields produced by wave simulation. Low-order decomposition is proposed in [Southern et al. 2012], while [Sheaffer et al. 2015] propose high-order decomposition that samples the synthetic field over the entire 3D volume $\|\Delta x\| \le r_o$ rather than just its spherical surface, estimating the modal coefficients $P_{l,m}$ via a least-squares fit to the over-determined system (5).

We follow a similar technique using a frequency-dependent SH truncation order of

$$n(\omega, r_o) \equiv \left\lfloor \frac{\kappa\, r_o\, e}{2} \right\rfloor, \tag{7}$$

where $e \equiv \exp(1)$; see Eq. 10 in [Zhang et al. 2010].

Solution. We found unnecessary the regularization in [Sheaffer et al. 2015]. We speculate this is because the solver we employ is different from FDTD. We simply solve the linear system in (5) using QR decomposition to obtain $P_{l,m}$. This recovers the (complex) directional amplitude distribution of plane waves that best matches the observed field around x, known as the *plane wave decomposition*,

$$D_{l,m} = \frac{i^l}{4\pi}\, P_{l,m}. \tag{8}$$

Assembling these coefficients over all ω and transforming from frequency to time domain reconstructs the directional impulse response (DIR), $d(s,t) = \mathcal{F}^{-1}[D(s,\omega)]$ where

$$D(s,\omega) \equiv \sum_{l,m} D_{l,m}(\omega)\, Y_{l,m}(s). \tag{9}$$

Binaural impulse responses for our PWD reference are generated by (4), performing convolution in frequency space. For each angular frequency ω, we compute the spherical integral multiplying the frequency-space PWD with each of the N_H (=2048, Sec. 2.2) spherical HRTF responses transformed to the frequency domain via

$$P^{L/R}(\omega) = \sum_{j=0}^{N_H-1} D\left(\mathcal{R}(s_j), \omega\right)\; H^{L/R}\left(s_j, \omega\right), \tag{10}$$

where $H^{L/R} \equiv \mathcal{F}[h^{L/R}]$ and $P^{L/R} \equiv \mathcal{F}[p^{L/R}]$, followed by a transform to the time domain to yield $p^{L/R}(t)$.

4.2 Acoustic Flux Density

Suppressing source location x', the impulse response is a function of receiver location and time representing (scalar) pressure variation, denoted $p(x,t)$. The flux density, $f(x,t)$, is defined as the instantaneous power transport in the fluid over a differential oriented area, analogous to irradiance in optics. It follows the relation

$$f(x,t) = p(x,t)\, v(x,t), \quad v(x,t) = -\frac{1}{\rho_0}\int_{-\infty}^{t} \nabla p(x,\tau)\, d\tau \tag{11}$$

where v is the particle velocity and ρ_0 is the mean air density (1.225kg/m^3). We use central differences on immediate neighbors in the simulation grid to compute spatial derivatives for ∇p, and midpoint rule over simulated steps for numerical time integration.

Flux density (or simply, flux) estimates the direction of a wavefront passing x at time t. When multiple wavefronts arrive simultaneously, PWD is able to tease apart their directionality (up to angular resolution determined by the diffraction limit) while flux is a differential measure, necessarily merging their directions. This merging is not as problematic as it might seem. Wavefronts arriving within 1ms are fused in human perception anyway. We do an explicit comparison in the next section.

To reconstruct the DIR from flux for a given time t (and suppressing x), we form the unit vector $\hat{f}(t) \equiv f(t)/\|f(t)\|$ and associate the corresponding pressure value $p(t)$ to that single direction, yielding

$$d(s,t) = p(t)\, \delta(s - \hat{f}(t)). \tag{12}$$

Note that this is a nonlinear function of the field, unlike (9). We compute binaural responses using the spherical integral in (4), by plugging in the DIR $d(s,t)$ from (12) and doing a temporal Fourier transform, which simplifies to

$$P^{L/R}(\omega) = \int_0^\infty p(t)\, e^{i\omega t}\, H^{L/R}\left(\mathcal{R}^{-1}\left(\hat{f}(t)\right), \omega\right)\, dt. \tag{13}$$

The time integral is carried out at the simulation time step, and HRTF evaluations employ nearest-neighbor lookup. The result is

then transformed back to binaural time-domain impulse responses used for comparing flux with PWD. Section 7 details how we use flux to extract DIR perceptual parameters actually used by our system at runtime.

4.3 Directional Analysis Comparison

The two alternatives we've presented for the directional IR in (9) and (12) are not equivalent physically, mathematically, or computationally. Plane wave decomposition is diffraction-limited, requiring a finite 3D ball of the wave field around the receiver point and solving a large compute-intensive linear system for each ω. Flux is nonlinear, working in the time domain on a differential neighborhood around the receiver point. It can be computed using finite difference, making it suitable for fast, streaming encoding. We present controlled virtual experiments to test how well the two match based on wave simulation with ν_m = 2000Hz.

The leftmost column of Figure 3 shows the six scenes in our experiment. They introduce complexity incrementally: SCENE1 begins with a single wall between source and listener; SCENE2 further completes that wall to form a window; SCENE3 adds a ground plane; SCENE4 opens a second window; SCENE5 adds a back wall; and finally SCENE6 adds side walls and a ceiling to form an enclosure of size 10m×15m×6m.

In all six scenes, the source (red sphere) is outside the front wall at a distance of 5m while the listener (green sphere) is inside it at the same distance of 5m. Windows are square openings of size 1m×1m. Wall thickness is 0.25m. Acoustic energy absorptivity of all geometry is 0.1. The first window (to the right as the listener faces it) provides a fairly direct but still occluded geodesic path from source to listener bending 15° at the window's left edge. The second window (to the listener's left) yields a longer, more diffracted and thus weaker path bending an angle of 45° at the window's center. For PWD, we use a microphone array of radius r_o = 1m (drawn to scale as the green sphere in the scene renderings). This is conservatively large to contain human head and shoulders.

The truncation order for PWD results (n at top of Figure 3) is determined by (7) applied at the center frequency of each of the three Bark bands. For flux, we match SH order to the highest Bark band used for PWD. Energy distribution sums $|D(s,\omega)|^2$ from (9) over all frequencies ω in the Bark band, followed by a square root to emphasize low-amplitude directions.

4.4 Directional Analysis Results

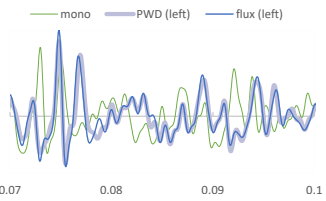
Results from our experiment are shown in Figure 3 (directional energy) and Figure 4 (binaural responses). The listener directly faces the source in all cases. Directional energy is visualized using an orthographic "twin hemispheres" spherical plot. The left hemisphere shows arrivals from behind the listener's head and right shows front. Arrivals to the left or right of the listener map as such to the images.

In Figure 3, energy distribution forms a single spot for SCENE1 and SCENE2, corresponding to the direct path diffracting around the edge of the wall or single window, and becoming elongated due to scattering off the right edge of the window in SCENE2. Adding a ground plane yields an additional, dim spot near the bottom of the right hemisphere in SCENE3. SCENE4's additional window introduces another spot to the left, while SCENE5 and SCENE6 increase the number of reflection spots including ones behind the listener.

Results confirm our two main hypotheses.

First, IR directionality is similar for different frequencies. This is best demonstrated when we integrate directionality over perceptual frequency intervals (Bark bands, see Section 7) just as human hearing does. Compare spherical plots for the three PWD Bark bands in Figure 3: lower frequency bands are just a blurrier version of the higher frequency ones, with peaks centered around the same directions. This blurriness is a consequence of the diffraction limit, rather than being a physical property of the field. Obtaining directional detail at lower frequencies with PWD requires an array much bigger than the human head.

Second, flux matches PWD surprisingly well. Spherical energy diagrams show a good match; binaural impulse responses match even better. Note that the "mono" curves in Figure 4 represent the original IR for reference; i.e., p rather than $p^{L/R}$. PWD results are plotted with a thicker line and grayer color than flux so that the two can be compared, as shown in the inset. As expected, the two methods start to diverge a little when wavefront arrival gets chaotic/simultaneous, as in the dimmer reflection spots in the spherical energy distribution of SCENE5 and SCENE6, or their corresponding binaural responses after about 40ms. The inset figure details the later, more chaotic transients from the left ear's binaural impulse response for SCENE6, showing these minor differences. An aural comparison is available in the supplemental video.



Note that the arriving *energy* in each perceptual frequency band varies over frequency due to scattering and shadowing. Our encoding in Section 7 thus loses audible detail by averaging energy over all simulated frequencies to save RAM as in [Raghuvanshi and Snyder 2014]). What these results support is that frequency-independent encoding of *directions* derived from flux loses little audible detail.

5 PRECOMPUTATION

Our framework is based on [Raghuvanshi and Snyder 2014] henceforth cited as [2014]. Ordinary restrictions on listener position (such as atop walkable surfaces) can be exploited by *reciprocal simulation* to significantly shrink precompute time, runtime memory and CPU. Such simulation exchanges source and listener position between precomputation and runtime so that runtime source and listener correspond respectively to (x, x') in (1). The first step is to generate a set of probe points $\{x'\}$ with typical spacing of 3-4m. For each probe point in $\{x'\}$, we perform 3D wave simulation using a wave solver [Raghuvanshi et al. 2009b] in a volume centered at the probe (90m × 90m × 30m in our tests), thus yielding a 3D slice $p(x, t; x')$ of the full 6D field of acoustic responses. The constrained runtime listener position reduces the size of $\{x'\}$ significantly. We extend this framework to extract and encode directional responses.

Reciprocal Dipole Simulation. We use flux (Section 4.2) to compute the directional response, requiring the spatial derivative of the pressure field for the runtime listener at x'. But the solver

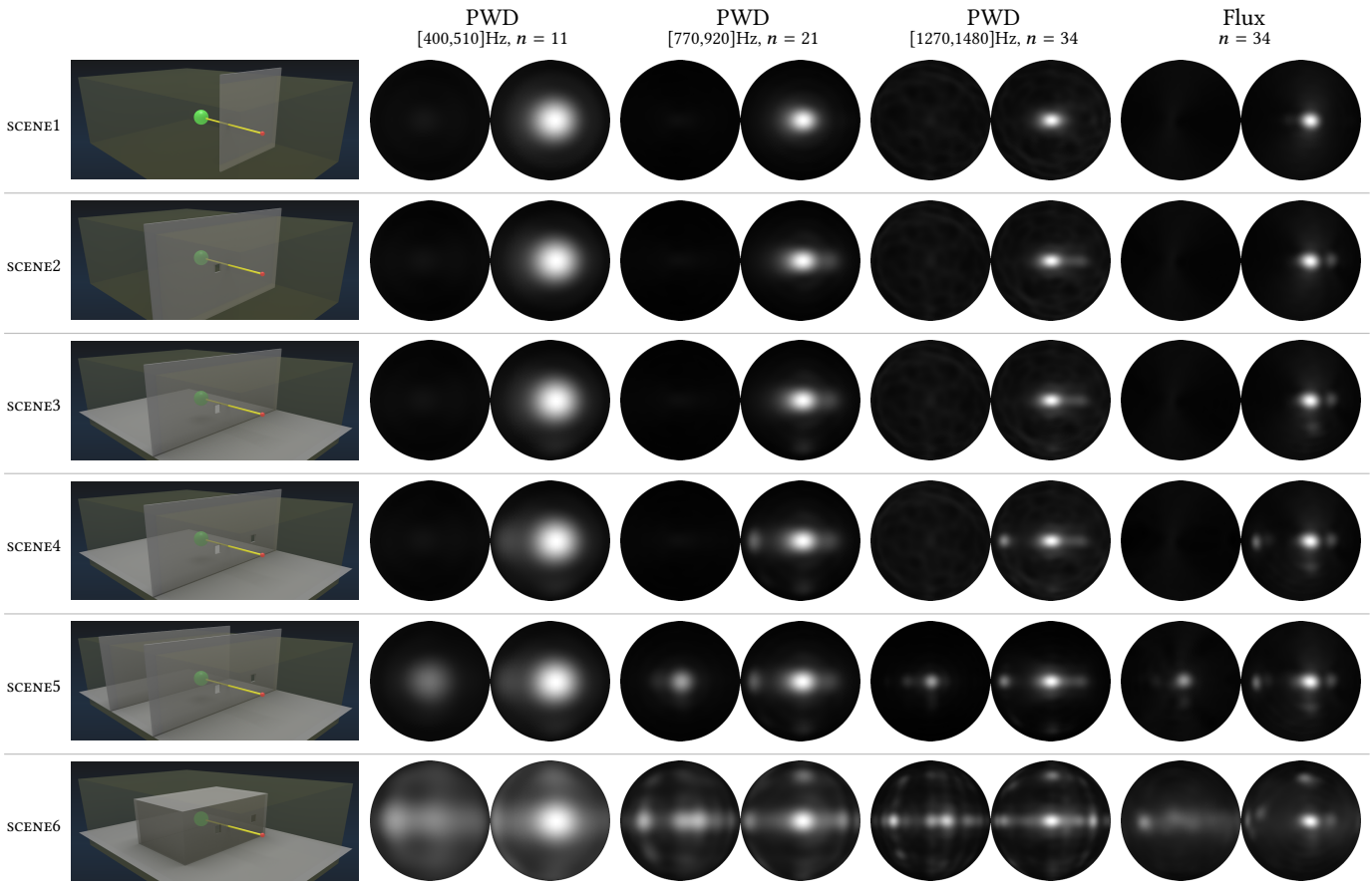


Fig. 3. Spherical energy distribution of the DIR, d, in six example scenes. Visual renderings on left show listener with green sphere, looking at the source in red. Simulation domain is shown with translucent box. Left hemisphere shows arrivals from behind listener and right hemisphere from the front.

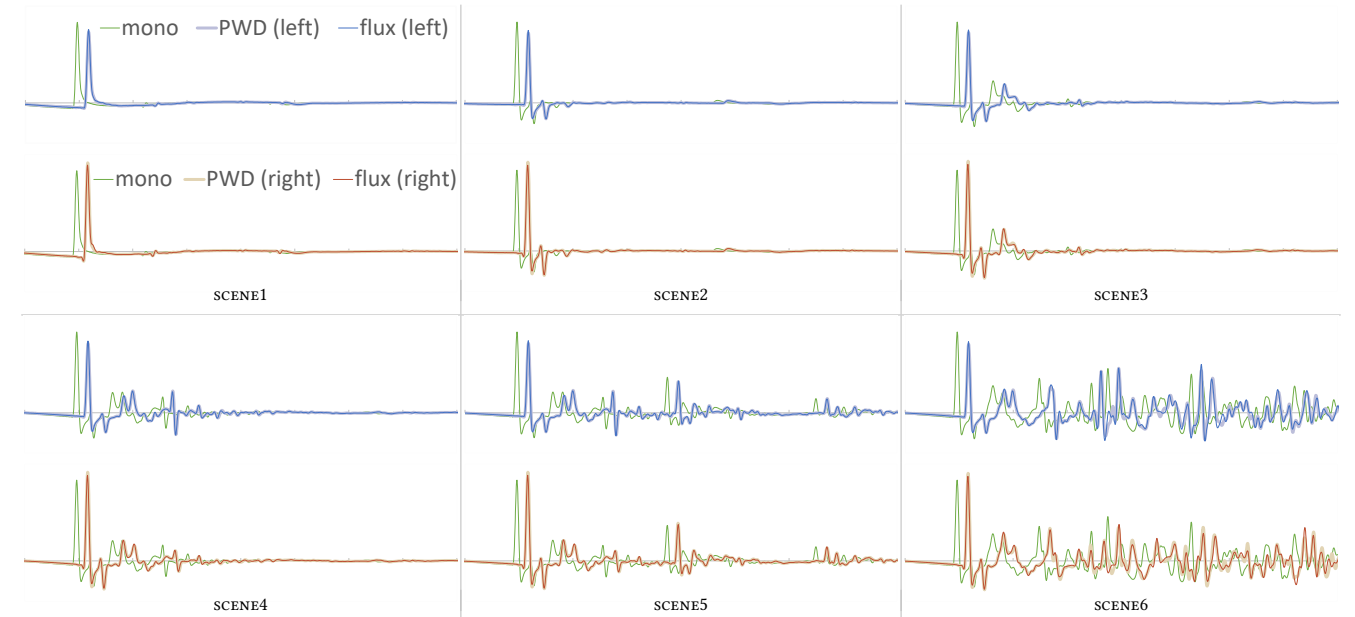


Fig. 4. Binaural impulse responses, $p^{L/R}$, for the six scenes above with listener looking at the source. Showing initial 80ms of response.

yields $p(x, t; x')$; i.e., the field varies over runtime source positions (x) instead. We present a solution that lets us compute flux at the runtime listener location while retaining the benefits of reciprocal simulation. For some grid spacing h, we wish to compute $\nabla_{x'} p(x, x') \approx [p(x; x'+h) - p(x; x'-h)]/2h$ via centered differencing. Due to the linearity of the wave equation, this can be obtained as response to the spatial impulse $[\delta(x - x' - h) - \delta(x - x' + h)]/2h$. In words, flux at a fixed runtime listener (x') due to a 3D set of runtime source locations (x) is obtained by simulating discrete dipole sources at x'. The three Cartesian components of the spatial gradient require three separate dipole simulations. The above argument extends to higher-order derivative approximations but we have found centered differences sufficient.

Time integration. To compute particle velocity via (11), we need the time integral of the gradient $\int_t \nabla p$ which commutes to $\nabla \int_t p$. Since the wave equation is linear, $\int_t p$ can be computed by replacing the temporal source factor in (1) with $\int_t \delta(t) = H(t)$, the Heaviside step function. The full source term is therefore $H(t)[\delta(x - x' + h) - \delta(x - x' - h)]/2\rho_0 h$, for which the solver's output directly yields particle velocity, $v(t, x; x')$. The three dipole simulations are complemented with a monopole simulation with source term $\delta(t)\,\delta(x - x')$, resulting in four simulations to compute the response fields $\{p(t, x; x'), f(t, x; x')\}$.

Bandlimiting. Discrete simulation must bandlimit the forcing impulse in space and time. We set the cutoff at $\nu_m = 1000$Hz in most of our experiments, requiring a grid spacing of $h = \frac{3}{8}c/\nu_m \equiv \frac{1}{2}c/\nu_M = 12.75$cm. This discards the highest 25% of the simulation's entire Nyquist bandwidth ν_M due to its large numerical error. DCT spatial basis functions in our solver (adaptive rectangular decomposition [Raghuvanshi et al. 2009b]) naturally convert delta functions into sincs bandlimited at wavenumber $\kappa = \pi/h$, simply by emitting the impulse at a single discrete cell. The source pulse must also be temporally bandlimited, denoted $\tilde{\delta}(t)$. Temporal source factors are modified to $\tilde{\delta}(t)$ and $H(t) * \tilde{\delta}(t)$ for the monopole and dipole simulations respectively; $\tilde{\delta}$ will be defined precisely in Section 7. Quadrature needed for the convolution $H(t) * \tilde{\delta}(t)$ can be precomputed to arbitrary accuracy and input to the solver.

Streaming. Past work on precomputed wave simulation uses a two stage approach in which the solver writes a massive spatio-temporal wave field to disk which the encoder then reads and processes. Disk I/O bottlenecks the processing of large game scenes, becoming impractical for mid-frequency ($\nu_m = 1000$Hz) simulations. It also complicates cloud computing and GPU acceleration.

Our new streaming encoder executes entirely in RAM. Processing for each runtime listener location x' proceeds independently across machines. For each x', four instances of the wave solver are run simultaneously to compute monopole and dipole simulations. The time-domain wave solver naturally proceeds as discrete updates to the global pressure field. At each time step t, 3D pressure and flux fields are sent in memory to the encoder coprocess which extracts the parameters. The encoder is SIMD across all grid cells. It cannot access field values beyond the current simulation time t, unlike prior work where the entire time response was available. Furthermore, the encoder must retain intermediate state from prior time steps (such as accumulators); this per-cell state must be minimized to keep RAM requirements practical. In short, the encoder must be causal with limited history. Section 7 shows how to design such an encoder for the parameters we propose.

Cost. Typical simulations we perform for $\nu_m = 1000$Hz have $|\{x\}|$=120 million cells. The total size of the discrete field across a typical simulation duration of 0.5s is 5.5TB which would take 30 hours just for disk I/O at 100MB/s. Our system executes in 5 hours taking 40GB RAM with no disk use. Compared to [2014, Table 2], our precompute on Citadel at $\nu_m = 500$Hz is 3 times faster despite our three additional dipole simulation and directional encoding.

6 DIRECTIONAL ACOUSTIC PERCEPTION

We briefly describe human auditory perception relevant to encoding directional impulse responses [Gade 2007; Litovsky et al. 1999]. The directional impulse response can usefully be divided into three successive phases in time: initial arrivals, followed by early reflections, which smoothly transition into late reverberation.

Precedence. In the presence of multiple wavefront arrivals carrying similar temporal signals, our perception non-linearly favors the first to determine the primary direction of the sound event. This is called the *precedence effect* [Litovsky et al. 1999]. Referring to Fig. 2, if the mutual delay $(l_2 - l_1)/c$ is less than 1ms we perceive a direction intermediate between the two arrivals, termed *summing localization* and representing the temporal resolution of directional hearing. Directions from arrivals lagging beyond 1ms are strongly suppressed and must be as much as 10dB louder to move the perceived direction significantly, called the *Haas effect*.

Extracting the correct direction for the potentially weak and multiply-diffracted first arrival is thus critical for faithfully rendering perceived direction of the sound event. It forms the primary cue guiding the listener to visually occluded sound sources. Our encoder is designed to extract the onset time robustly and uses a short 1ms window after onset to integrate the first arrival direction.

Panning. Summing localization is exploited by traditional speaker amplitude panning, which plays the same signal from multiple (usually four to six) speakers surrounding the physical listener. By manipulating the amplitude of each signal copy, the perceived direction moves smoothly between the speakers. The downside is that the arrival acquires a perceptible angular extent, lacking the crispness of a single plane wavefront. We adapt this idea and exploit summing localization to efficiently encode and render directional reflections, which we assume are less crisp directionally than the initial arrival.

Echo threshold. When a sound follows the initial arrival after a delay called the *echo threshold*, it is perceived as a separate event; otherwise it is fused. The echo threshold varies between 10ms for impulsive sounds, through 50ms for speech, to 80ms for orchestral music [Litovsky et al. 1999, Tbl.1]. We conservatively fuse using a 10ms window to aggregate loudness for initial arrivals.

Initial time delay gap. Initial arrivals are followed by stronger reflections reflected off big features like walls, mixed with weaker arrivals scattered from smaller, more irregular geometry. If the first

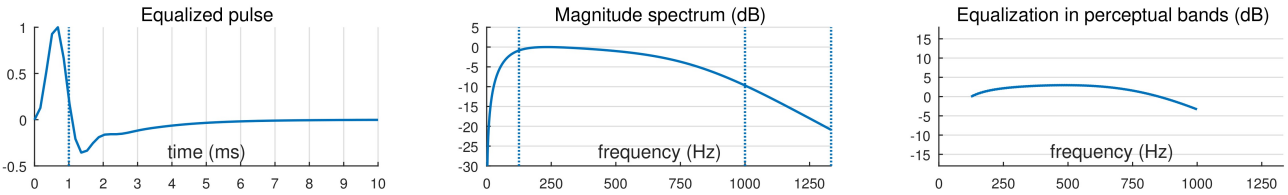


Fig. 5. Equalized pulse $\tilde{\delta}(t)$ for ν_l=125Hz, ν_m=1000Hz and ν_M=1333Hz. The pulse is designed to have a sharp main lobe (~1ms) to match auditory perception (left), while having limited energy outside $[\nu_l, \nu_m]$ (middle) with smooth falloff to minimize ringing in time domain. Within these constraints, it is designed to have matched energy (to within ±3dB) in equivalent rectangular bands centered at each frequency (right).

strong reflection arrives beyond the echo threshold, its delay becomes audible. Its delay is called the *initial time delay gap*, with perceptual just-noticeable-difference of about 10ms [Gade 2007]. Audible gaps arise easily, e.g. when the source and listener are close but far from surrounding geometry. Prior work has extracted this parameter for a few responses semi-manually [Fujii et al. 2004]. We require a fully automatic technique that produces smooth fields.

Reflections. Once reflections begin arriving, they typically bunch closer than the echo threshold due to environmental scattering and are perceptually fused. We use a value of 80ms following the initial time delay gap as the duration of early reflections. Their aggregate directional distribution conveys important detail about the environment around the listener and source. The ratio of energy arriving horizontally and perpendicular to the initial sound is called *lateralization* and conveys spaciousness and apparent source width. Anisotropy in reflected energy arising from surfaces close to the listener provides an important proximity cue [Paasonen et al. 2017]. When source and listener are separated by a portal, reflected energy arrives mostly through the portal and is strongly anisotropic, localizing the source to a different room than the listener's. We encode this anisotropy in the aggregate reflected energy.

Reverberation. As time progresses, scattered energy gets weaker but arrives more frequently so that the response's tail resembles decaying noise. This characterizes the (late) reverberation phase. Its decay rate conveys overall scene size, typically measured as RT60 or the time taken for energy to decay by 60dB. The aggregate directional properties of reverberation affect listener "envelopment". We simplify by assuming that the directional distribution of reverberation is the same as that for reflections.

7 ENCODING

At each time step t, the encoder receives $\{p(t, x; x'), f(t, x; x')\}$ representing the pressure and flux at runtime listener x' due to a 3D field of possible runtime source locations, x, for which it performs independent, streaming processing. We suppress positions below.

Notation. $t_k \equiv k\,\Delta t$ denotes the k^{th} time sample with time step Δt, where $\Delta t = 0.17$ms for $\nu_m = 1000$Hz. First-order Butterworth filtering with cutoff frequency ν in Hz is denoted $\mathcal{L}_\nu$. A signal $g(t)$ filtered through $\mathcal{L}$ is denoted $\mathcal{L} * g$. Its cumulative time integral is denoted $\int g \equiv \int_0^t g(\tau)\,d\tau$.

7.1 Equalized Pulse

Encoder inputs $\{p(t), f(t)\}$ are responses to an impulse $\tilde{\delta}(t)$ provided to the solver. Response properties are typically computed by deconvolving out this impulse, a costly operation requiring the entire response. But the streaming encoder has access only to the current and few past samples in time. We show it's possible to design an impulse function (Figure 5) to conveniently estimate the IR's energetic and directional properties without undue storage or costly convolution. The pulse must satisfy several properties:

(1) **equalized** to match energy in each perceptual frequency band. $\int p^2$ thus directly estimates perceptually weighted energy averaged over frequency.
(2) **abrupt** in onset, critical for robust detection of initial arrival. We need an accuracy of about 1ms or better when estimating the initial arrival time, matching auditory perception.
(3) **sharp** in main peak with a half-width of less than 1ms. Flux merges peaks in the time-domain response; this property ensures such mergers happen only when they would undergo summing localization in our perception anyway.
(4) **anti-aliased** to control numerical error, with energy falling off steeply in the frequency range $[\nu_m, \nu_M]$.
(5) **mean-free**. Sources with substantial DC energy yield residual particle velocity after curved wavefronts pass, making flux inaccurate. Reverberation in small rooms can also settle to a non-zero value, spoiling energy decay estimation.
(6) **quickly decaying** to minimize interference between flux from neighboring peaks. Abrupt cutoffs at ν_m for (4) or at DC for (5) cause non-compact ringing and should be avoided.

Human pitch perception can be roughly characterized as a bank of frequency-selective filters, with frequency-dependent bandwidth known as Equivalent Rectangular Bandwidth (ERB). The same notion underlies the Bark psychoacoustic scale consisting of 24 bands equidistant in pitch and utilized by our PWD visualizations in Section 4.4.

A simple model for ERB around a given center frequency ν in Hz is given by $B(\nu) \equiv 24.7\,(4.37\,\nu/1000 + 1)$ [Moore and Glasberg 1996]. Condition (1) above can then be met by specifying the pulse's energy spectral density (ESD) as $1/B(\nu)$ but this violates properties (4) and (5). We therefore substitute the modified ESD

$$E(\nu) = \frac{1}{B(\nu)} \frac{1}{\left|1 + 0.55\,(2i\nu/\nu_h) - (\nu/\nu_h)^2\right|^4} \frac{1}{|1 + i\nu/\nu_l|^2} \qquad (14)$$

where $\nu_l = 125$Hz is the low and $\nu_h = 0.95\,\nu_m$ the high frequency cutoff. The second factor is a second-order low-pass filter designed to attenuate energy beyond ν_m per (4) while limiting ringing in the time domain via the tuning coefficient 0.55 per (6). The last factor combined with a numerical derivative in time attenuates energy near DC, as explained more below.

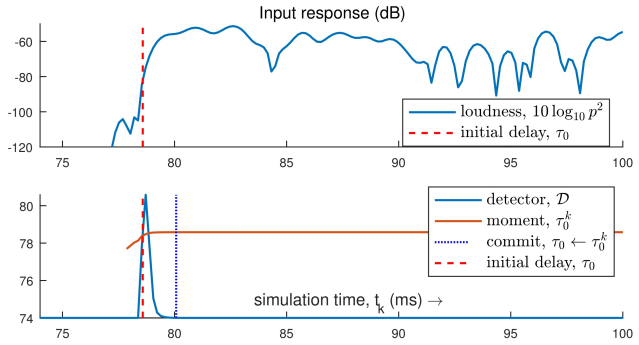


Fig. 6. Initial delay estimation. $\mathcal{D}$ is scaled to span the y axis.

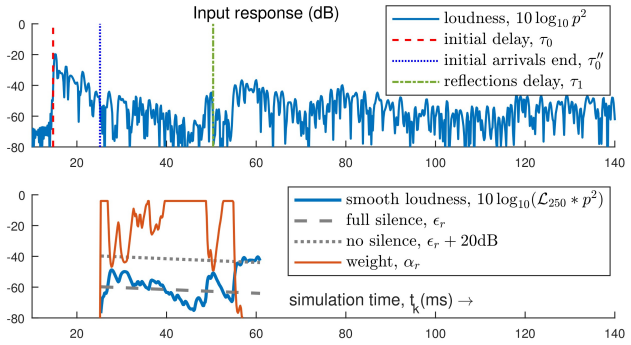


Fig. 7. Reflections delay estimation. $\alpha_r \in [0, 1]$ is scaled to span the y axis.

We then design a minimum-phase filter [Smith 2007] with $E(\nu)$ as input. Such filters manipulate phase to concentrate energy at the start of the signal, satisfying properties (2) and (3). To make DC energy 0 per (5), we compute a numerical derivative of the pulse output by minimum-phase construction. The ESD of the pulse after this derivative is $4\pi^2\,\nu^2\,E(\nu)$. Dropping the $4\pi^2$ and grouping the ν^2 with the last factor in (14) yields $\nu^2/|1 + i\nu/\nu_l|^2$, representing the ESD of a first-order high-pass filter with 0 energy at DC per (5) and smooth tapering in $[0, \nu_l]$ to control the negative side lobe's amplitude and width per (6). The output is passed through another low-pass $\mathcal{L}_{\nu_h}$ to further reduce aliasing, yielding the final pulse shown in Figure 5.

7.2 Initial Delay (Onset), τ_0

Figure 6 illustrates processing with an actual response from the HighRise scene. The solver fixes the emitted pulse's amplitude so the received signal at 1m distance in the free field has unit energy, $\int p^2 = 1$. Initial delay could be computed by comparing incoming energy p^2 to an absolute threshold, as in [2014]. But in occluded cases, a weak initial arrival can rise above threshold at one location and stay below at a neighbor, causing distracting jumps in rendered delay and direction at runtime.

We design a more robust detector $\mathcal{D}$. Initial delay is computed as its first moment, $\tau_0 \equiv \int t\mathcal{D}(t)/\int \mathcal{D}(t)$, where

$$\mathcal{D}(t) \equiv \left[\frac{d}{dt}\left(\frac{E(t)}{E(t-\Delta t)+\epsilon}\right)\right]^n, \tag{15}$$

$E(t) \equiv \mathcal{L}_{\nu_m/4} * \int p^2$, and $\epsilon = 10^{-11}$. E is a monotonically increasing, smoothed running integral of energy in the pressure signal. The ratio in (15) looks for jumps in energy above a noise floor ϵ. The time derivative then peaks at these jumps and descends to zero elsewhere, as shown in the figure. For the detector to peak, energy must abruptly overwhelm what has been accumulated so far. We use $n = 2$ to control the detector's peakedness.

This detector is streamable. $\int p^2$ is implemented as a discrete accumulator. $\mathcal{L}$ is a recursive filter, requiring internal history of one past input and output. One past value of E is needed for the ratio, and one past value of the ratio kept to compute the time derivative via forward differences. However, computing onset via first moment poses a problem as the entire signal must be processed to produce a converged estimate.

We observe that the detector is allowed some latency, namely 1ms for summing localization. We keep a running estimate of the moment, $\tau_0^k = \int_0^{t_k} t\,\mathcal{D}(t)/\int_0^{t_k} \mathcal{D}(t)$, and commit a detection $\tau_0 \leftarrow \tau_0^k$ when it stops changing; that is, it's latency satisfies $t_{k-1} - \tau_0^{k-1} <$ 1ms and $t_k - \tau_0^k >$ 1ms (dotted line in figure). This detector can trigger more than once, indicating the arrival of significant energy relative to the current accumulation in a small time interval and letting us treat the last as definitive. Each commit resets the subsequent processing state as necessary.

7.3 Initial Loudness and Direction, (L, s_0)

Initial loudness and its 3D direction are estimated via

$$L \equiv 10\,\log_{10}\int_0^{\tau_0''} p^2(t)\,dt, \quad s_0 \equiv \int_0^{\tau_0'} f(t)\,dt \tag{16}$$

where $\tau_0' = \tau_0 +$ 1ms and $\tau_0'' = \tau_0 +$ 10ms. We retain only the (unit) direction of s_0 as the final parameter. This assumes a simplified model of directional dominance where we suppress directions outside a 1ms window but let their energy contribute to loudness for 10ms.

7.4 Reflections Delay, τ_1

Reflections delay is the arrival time of the first significant reflection. Its detection is complicated by weak scattered energy almost always present after onset. A binary classifier based on a fixed amplitude threshold performs poorly. We instead aggregate the duration of silence in the response, where "silence" is given a smooth definition discussed shortly. Silent gaps are usually concentrated right after the initial arrivals but before reflections from surrounding geometry have become sufficiently dense in time from repeated scattering. The combined duration of this silence is a new parameter roughly paralleling the notion of initial time delay gap discussed in Section 6.

Figure 7 shows estimation which starts after initial arrivals end at τ_0''. The duration of silence is initialized as $\Delta\tilde{\tau}_1 =$ 10ms. The reflections delay estimate is defined as $\tilde{\tau}_1 \equiv \tau_0 + \Delta\tilde{\tau}_1$. We define a threshold for silence relative to the initial sound's peak energy as $\epsilon_r =$ -40dB + $10\log_{10}\left(\max\{p^2(t)\}, t \in [0, \tau_0'']\right)$. The incoming energy is smoothed and loudness computed as $10\,\log_{10}\left(\mathcal{L}_{250} * p^2\right)$ then passed through the linear mapping $[\epsilon_r, \epsilon_r +$ 20dB$] \rightarrow [1, 0]$. This produces a weight, α_r that is clamped to $[0, 1]$, with $\alpha_r = 1$ indicating complete silence. The silence duration estimate is then updated as $\Delta\tilde{\tau}_1 \leftarrow \Delta\tilde{\tau}_1 + \alpha_r\,\Delta t$. The estimate is considered converged

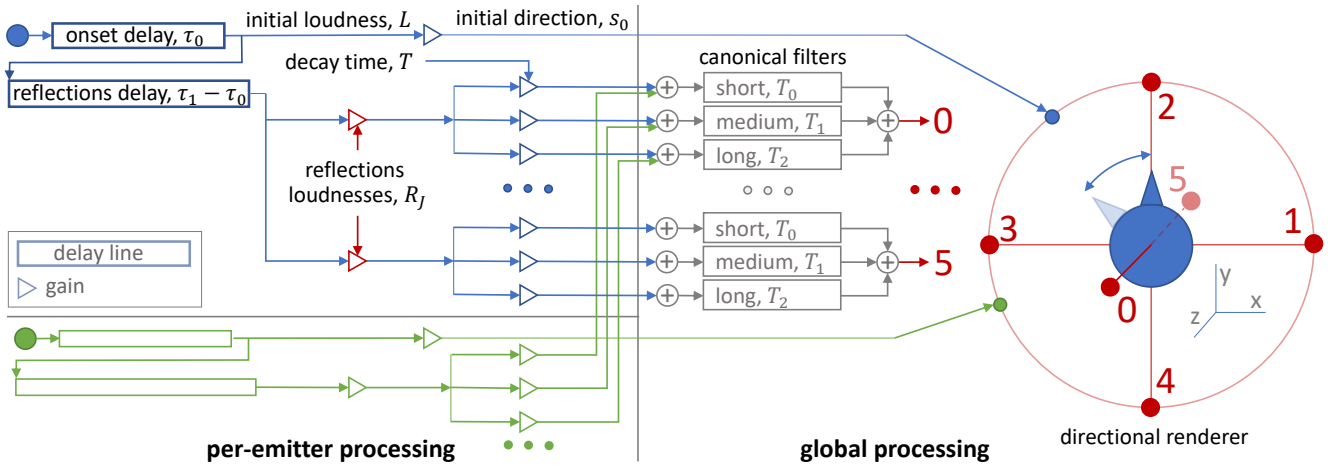


Fig. 8. Runtime signal processing. Per-emitter processing is limited to two (optional) variable delay lines, directional rendering for the initial-arriving sound, and scale-and-sum at the inputs to global canonical filters. These are mono filters whose output is rendered as incoming from fixed axial directions in world space while accounting for dynamic listener head pose.

when the latency $t - \tilde{\tau}_1$ increases above 10ms for the first time, at which point we set $\tau_1 \leftarrow \tilde{\tau}_1$.

In cases of fast energy decay, later arriving reverberant energy can get classified as silence. We bias estimation to the beginning of the response by modifying the silence threshold to fall off as $\epsilon_r \leftarrow \epsilon_r + d_r\,(t - \tau_0'')$ with a fast decay rate of d_r = -120dB/s.

7.5 Directional Reflection Loudnesses, R_J

We aggregate loudness and directionality of reflections for 80ms after the reflections delay (τ_1). Waiting for energy to start arriving after reflecting from proximate geometry gives a more consistent energy estimate than [2014], which collects energy for a fixed interval after direct sound arrival (τ_0). We collect directional energy using coarse cosine-squared basis functions fixed in world space and centered around the coordinate axes S_J, yielding six directional loudnesses indexed by J

$$R_J \equiv 10\,\log_{10} \int_{\tau_0+10\text{ms}}^{\tau_1+80\text{ms}} p^2(t) \max^2\left(\hat{f}(t)\cdot S_J, 0\right) dt. \qquad (17)$$

Since $|\hat{f}(t)| = 1$, this directional basis forms a partition of unity which preserves overall energy, and does not ring to the opposite hemisphere like low-order spherical harmonics.

Our approach allows flexible control of RAM and CPU rendering cost not afforded by spherical harmonics. For example, elevation information could be omitted by summing energy in $\pm z$ equally in the four horizontal directions. Alternatively, one could preferentially increase azimuthal resolution with suitable weights.

7.6 Decay Time, T

Impulse response decay time is usually computed as a backward time integral of p^2 but a streaming encoder lacks access to future values. With appropriate causal smoothing, robust decay estimation can be performed via online linear regression on the smoothed loudness $10\,\log_{10}\left(\mathcal{L}_{20} * p^2\right)$. We avoid estimation of separate early and late decays, instead computing an overall 60dB decay slope starting at the reflection delay, τ_1.

7.7 Spatial Compression

The preceding processing results in a set of 3D parameter fields varying over x for a fixed runtime listener location x'. As in [2014], each field is spatially smoothed and subsampled on a uniform grid with 1.5m resolution. Fields are then quantized and each z-slice sent through running differences followed by a standard byte-stream compressor (Zlib). The novel aspect is treating the vector field of primary arrival directions, $s_0(x; x')$.

Singularity. $s_0(x; x')$ is singular at $|x - x'| = 0$. Small numerical errors in computing the spatial derivative for flux yield large angular error when $|x - x'|$ is small. Denoting the line of sight direction as $s_0' \equiv (x' - x)/|x' - x|$, we replace the encoded direction with $s_0(x; x') \leftarrow s_0'$ when the distance is small and propagation is safely unoccluded; i.e., if $|x - x'| <$ 2m and $L(x; x') >$ -1dB. When interpolating, we use the singularity-free field $s_0 - s_0'$, add back s_0' to the interpolated result, and renormalize to a unit vector.

Compressing directions. Since s_0 is a unit vector, encoding its 3D Cartesian components wastes memory and yields anisotropic angular resolution. This problem also arises when compressing normal maps for visual rendering. We tailor a simple solution to our case which first transforms to an elevation/azimuth angular representation: $s_0 \rightarrow (\theta, \phi)$. Simply quantizing azimuth, ϕ, results in artificial incoherence when ϕ jumps between 0 and 2π. We observe that only running differences are needed for compression and use the update rule $\Delta\phi \leftarrow \text{argmin}_{x \in \{\Delta\phi, \Delta\phi+2\pi, \Delta\phi-2\pi\}} |x|$. This encodes the signed shortest arc connecting the two input angles, avoiding artificial jumps.

Quantization. Discretization quanta for $\{\tau_0, L, s_0, \tau_1, R_*, T\}$ are given by $\{2\text{ms}, 2\text{dB}, (6.0^\circ, 2.8^\circ), 2\text{ms}, 3\text{dB}, 3\}$. The primary arrival direction, s_0, lists quanta for (θ, ϕ) respectively. Decay time T is encoded as $\log_{1.05}(T)$ as in [2014].

8 RENDERING

Figure 8 diagrams our runtime signal processing. Note the middle partition separating per-emitter processing from global. Per-emitter

processing is determined by dynamically decoded values for the parameters based on runtime source and listener location. Decompression and spatial interpolation operate similarly as [2014]. Although the parameters are computed on bandlimited simulations, rendering applies them for the full audible range, thus implicitly performing frequency extrapolation.

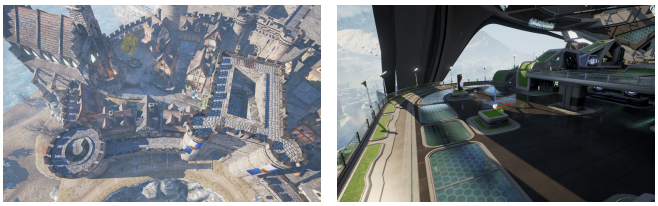
Initial sound. Starting at the top left of Fig. 8, the mono source signal is sent to a variable delay line [Smith 2007] to apply the initial arrival delay, τ_0. This also naturally captures environmental Doppler shift effects based on the shortest path through the environment. Next we apply a gain driven by the initial loudness, L (as $10^{L/20}$) and send the resulting signal for rendering at the primary arrival direction, s_0.

Directional canonical filters. To avoid the cost of per-source convolution, we extend the idea of canonical filters [2014] to incorporate directionality. For all combinations of the world axial directions S_J and possible RT60 decay times $\{T_I\} = \{0.5s, 1.5s, 3s\}$, we build a *mono* canonical filter as a collection of delta peaks whose amplitude decays exponentially, mixed with Gaussian white noise that increases quadratically with time. We match the peak delays across all $\{S_J\}$ to allow coloration-free interpolation and, as discussed shortly, ensure summing localization. The same pseudo-random signal is used across $\{T_I\}$ with S_J held fixed. However, we use *independent* noise signals across directions $\{S_J\}$ to achieve inter-aural decorrelation that aids in natural, enveloping reverberation.

For each direction S_J, the output across filters for various decay times $\{T_I\}$ is summed and then rendered as arriving from world direction S_J. This is different from multi-channel surround encodings where the canonical directions are fixed in the listener's frame of reference rather than in the world. Because all canonical filters share time delays for peaks, interpolating between them across $\{S_J\}$ results in summing localization, creating the perception of reverberation arriving from an intermediate direction. This exploits summing localization in the same way as speaker panning, discussed in Section 6.

Reflections and reverberation. The output of the onset delay line is fed into a reflection delay line that renders the variable delay $\tau_1 - \tau_0$, thus realizing the net reflection delay of τ_1 on the input signal. The output is then scaled by the gains $\{10^{R_J/20}\}$ to render the directional amplitude distribution. To incorporate the decay time T, we compute three weights corresponding to canonical decay times $\{T_I\}$ as in [2014], which further multiply the directional gains. The results are summed into the inputs of the 18 canonical filters (6 directions × 3 decay times). To reduce the cost of scaling and summing into 18 filter inputs, we observe that only 12 of these are nonzero, corresponding to the two decay times in $\{T_I\}$ that bracket the actual decay time T decoded.

Spatialization. Directional rendering (Figure 8, right) is device dependent and our technique is agnostic to its details. It renders the impression that an input mono signal arrives from its associated input world direction, producing multiple signals for playback on the user's output hardware. Recall that directions arise either from the per-emitter primary arrival direction, s_0, or the fixed canonical directions, S_J. These incoming world directions, denoted s_w, are first transformed into the listener's reference frame, $s_l = \mathcal{R}^{-1}(s_w)$.



Citadel HighRise

Fig. 9. Demo scenes.

Our results binaurally render using generic HRTFs for headphones. We perform nearest-neighbor look up in the HRTF dataset to the direction s_l, and then convolve (using partitioned, frequency-domain convolution) the input signal with the per-ear HRTFs to produce a binaural output buffer at each audio tick. To avoid popping artifacts, the input signal's audio buffer is cross-faded with complementary sigmoid windows and fed to HRTFs corresponding to s_l at the previous and current audio tick. Other spatialization approaches can easily be substituted. Instead of HRTFs, one could compute panning weights given s_l to produce multi-channel signals for speaker playback in a stereo, 5.1 or 7.1 surround, or with-elevation setups.

9 RESULTS

Our approach produces smooth auralizations in complex game scenes shown in Figure 9 that can be heard in the supplemental video. Both scenes were precomputed at $\nu_m = 1000$Hz. Citadel had 780 listener (probe) samples, with a total data size of 80MB. HighRise had 1350 listener samples with data size of 160MB. Each listener probe centers a simulation around x' of dimension 90m×90m×30m. Probes can be computed in parallel with each one taking 5-6 hours at 1kHz and 20 minutes at 500Hz on a single 8-core machine.

We visualize the encoder's raw output estimating parameters *independently* at each cell in Figures 1 and 10. These raw parameter fields are quite smooth, affording spatial coherence and good compression. In Figure 10, the listener is located inside the cathedral, shown with a green dot, with two doors leading outside. Observing the initial delay field (τ_0), arrivals take a circuitous route to get behind the cathedral and are extremely attenuated, yet our delay estimation stays robust and produces smooth values.

To interpret the initial direction field (s_0), recall that we perform reciprocal dipole simulations. The figure visualizes arrival direction of the sound emanating from a field of possible runtime source locations x, with runtime listener held fixed at x' located symmetrically with respect to the two doors. The field demonstrates our modeling of the precedence effect. Directions show a piecewise constant, symmetric distribution in the upper and lower halves of the parameter image. Sources in the top half of the figure pass through the portal at the top, and sources at the bottom through the bottom portal. In the middle, within the 1ms summing localization window, the directions smoothly merge to an intermediate. A piecewise constant distribution of directions occurs frequently when the runtime listener is indoors: the sound for any source outside must choose between directions to one of the portals to get to the listener first. The resulting compression is an additional advantage of using reciprocity.

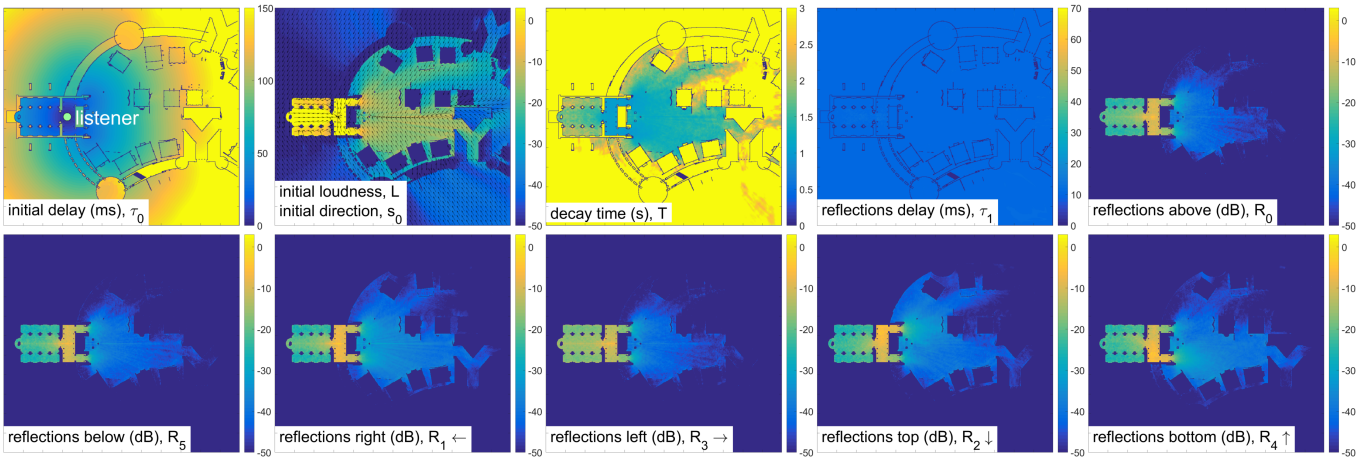


Fig. 10. Parameter fields for CITADEL.

The reflection delay field shows little variation in CITADEL because geometry close to the listener causes dense scattering. In a more open environment like HIGHRISE (Figure 1) we can see large reflection delays when the source and listener are close together but far from scene geometry. We observed in our experiments that reflection delay is most easily heard when sounds emanate from near the listener, like footsteps.

In Figure 10, the main chamber of the cathedral is bright only for the "reflections left" (R_3) field: reflected energy for any source in the main chamber must arrive through the door to the listener's left in the figure. When source and listener occupy different spaces, intervening portals cause marked anisotropy as reflections must funnel through them. Rendering isotropic reverberation [2014] in such cases creates the incorrect perception of source and listener being enclosed in the same room, especially when the listener is outdoors. In the last two plots ("reflections top", R_2, and "reflections bottom", R_4), brightness is mirrored in the top and bottom halves of the plots; i.e., when the sound source is in the entrance towards the top, reflected energy arrives from the same direction, and similarly for the bottom. In open spaces, anisotropy from geometry close to the source can be pronounced as seen in the last two plots in Figure 1 (R_2 vs. R_4). Anisotropy also arises near archways and corners which act as local reservoirs of energy with directional output.

Experimental auralization. The supplemental video compares our reference PWD and flux binaural results for $\nu_m = 2000$Hz (Section 4) with our system's fully encoded/decoded runtime result based on a lower frequency simulation of $\nu_m = 1000$Hz. PWD and flux sound virtually identical on all scenes. The simplest experimental scenes (SCENE1-SCENE4) lack geometry to hold sound and so don't reverberate. We auralize SCENE4 in the video as representative. Note the strong precedence effect: even though sound comes through both windows (and is rendered as such in the reference auralization), the perceived direction is that of the closer window. Our rendering produces too much reverberation in this extremely simple scene: some of the later arriving sound from the farther window contributes to a small amount of reflected energy which is then rendered through our shortest canonical filter which is still 0.5s. In SCENE5, flutter echoes are distinctly audible as sounds bounce back and forth between the scene's parallel walls. While flux generally agrees with PWD, in this case, flux merges directions for overlapping reflections from the two walls, creating the incorrect perception of reverberation progressing to between the walls. Such detailed temporal structure is lost entirely by our system, but it does reproduce the directional cues for the initial sound and reflection from the back wall. Our system's result in SCENE6 matches well.

10 CONCLUSION

We present the first system to capture 9D directional acoustics in large, complex but static scenes in real time for moving sources and listener. Our perceptual encoding makes memory use manageable by extracting a few parameters from each directional impulse response in a precomputed wave simulation. A novel streaming encoder allows feasible precomputation up to 1kHz on large game scenes. Our results demonstrate many directional effects, including correct initial sound direction in occluded cases and anisotropic reflections that provide immersive auditory cues about surrounding geometry. We show for the first time how well the flux approximation matches ground truth plane wave decomposition, and how well our system matches these references, in a controlled virtual experiment.

In future work, we're interested in improving realism especially outdoors and investigating other parameters including echo density, more directional detail in early reflections, and independently directional late reverberation. Our system's use of dipole sources to introduce a listener's head into an impulsive sound field after the fact could be similarly exploited to handle directional sources which can rotate at runtime. We note that our underlying wave solver can be improved in many ways, e.g. to handle frequency-dependent absorption and sound transmission through geometry.

ACKNOWLEDGMENTS

We wish to acknowledge Hannes Gamper and Keith Godin for the HRTF dataset and related code. Thanks to John Morgan for providing some of the audio clips. Thanks also to the SIGGRAPH reviewers for their constructive comments.

REFERENCES

Lakulish Antani, Anish Chandak, Micah Taylor, and Dinesh Manocha. 2012. Direct-to-Indirect Acoustic Radiance Transfer. *IEEE Transactions on Visualization and Computer Graphics* 18, 2 (Feb. 2012), 261–269. https://doi.org/10.1109/tvcg.2011.76

P. Bilinski, Ahrens J., Thomas M. R. P., Tashev I. J., and Platt J. C. 2014. HRTF magnitude synthesis via sparse representation of anthropometric features. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence*. 4468–4472. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6854447&isnumber=6853544

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.

Jeroen Breebaart, Sascha Disch, Christof Faller, Jürgen Herre, Gerard Hotho, Kristofer Kjörling, Francois Myburg, Matthias Neusinger, Werner Oomen, Heiko Purnhagen, and Jonas Rödén. 2005. MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status. In *Audio Engineering Society Convention 119*. http://www.aes.org/e-lib/browse.cfm?elib=13333

Chunxiao Cao, Zhong Ren, Carl Schissler, Dinesh Manocha, and Kun Zhou. 2016. Interactive Sound Propagation with Bidirectional Path Tracing. to appear. *ACM Transactions on Graphics (SIGGRAPH Asia 2016)* (2016).

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*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/1661412.1618465

Anish Chandak, Christian Lauterbach, Micah Taylor, Zhimin Ren, and Dinesh Manocha. 2008. AD-Frustum: Adaptive Frustum Tracing for Interactive Sound Propagation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1707–1722. https://doi.org/10.1109/tvcg.2008.111

Jean-Jacques Embrechts. 2016. Review on the applications of directional impulse responses in room acoustics. In *Proceedings of CFA 2016*. Société française d'acoustique (SFA). http://orbi.ulg.ac.be/handle/2268/193820

Kenji Fujii, Takuya Hotehama, Kosuke Kato, Ryota Shimokura, Yosuke Okamoto, Yukio Suzumura, and Yoichi Ando. 2004. Spatial Distribution of Acoustical Parameters in Concert Halls: Comparison of Different Scattered Reflections. 4 (01 2004).

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

Michael A. Gerzon. 1973. Periphony: With-Height Sound Reproduction. *J. Audio Eng. Soc* 21, 1 (1973), 2–10. http://www.aes.org/e-lib/browse.cfm?elib=2012

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. http://www.worldcat.org/isbn/0080443710

Brian Hamilton, Stefan Bilbao, Brian Hamilton, and Stefan Bilbao. 2017. FDTD Methods for 3-D Room Acoustics Simulation With High-Order Accuracy in Space and Time. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 25, 11 (Nov. 2017), 2112–2124. https://doi.org/10.1109/TASLP.2017.2744799

Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties. 2015. MPEG-H Audio - The New Standard for Universal Spatial/3D Audio Coding. *J. Audio Eng. Soc* 62, 12 (2015), 821–830. http://www.aes.org/e-lib/browse.cfm?elib=17556

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

Heinrich Kuttruff. 2000. *Room Acoustics* (4 ed.). Taylor & Francis. http://www.worldcat.org/isbn/0419245804

Mikko V. Laitinen, Tapani Pihlajamäki, Cumhur Erkut, and Ville Pulkki. 2012. Parametric Time-frequency Representation of Spatial Sound in Virtual Worlds. *ACM Trans. Appl. Percept.* 9, 2 (June 2012). https://doi.org/10.1145/2207216.2207219

Dingzeyu Li, Yun Fei, and Changxi Zheng. 2015. Interactive Acoustic Transfer Approximation for Modal Sound. *ACM Trans. Graph.* 35, 1 (Dec. 2015). https://doi.org/10.1145/2820612

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, Nikunj Raghuvanshi, Lakulish Antani, Anish Chandak, Sean Curtis, and Dinesh Manocha. 2013. Wave-based Sound Propagation in Large Open Scenes Using an Equivalent Source Formulation. *ACM Trans. Graph.* 32, 2 (April 2013). https://doi.org/10.1145/2451236.2451245

Ravish Mehra, Nikunj Raghuvanshi, Lauri Savioja, Ming C. Lin, and Dinesh Manocha. 2012. An efficient GPU-based time domain solver for the acoustic wave equation. *Applied Acoustics* 73, 2 (Feb. 2012), 83–94. https://doi.org/10.1016/j.apacoust.2011.05.012

Ravish Mehra, Atul Rungta, Abhinav Golas, Ming Lin, and Dinesh Manocha. 2015. WAVE: Interactive Wave-based Sound Propagation for Virtual Environments. *IEEE transactions on visualization and computer graphics* 21, 4 (April 2015), 434–442. http://view.ncbi.nlm.nih.gov/pubmed/26357093

Juha Merimaa and Ville Pulkki. 2005. Spatial Impulse Response Rendering I: Analysis and Synthesis. *J. Audio Eng. Soc* 53, 12 (2005), 1115–1127. http://www.aes.org/e-lib/browse.cfm?elib=13401

Brian Moore and Brian Glasberg. 1996. A Revision of Zwicker's Loudness Model. 82 (03 1996), 335–345.

D. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley. 2007. Acoustic Modeling Using the Digital Waveguide Mesh. *IEEE Signal Processing Magazine* 24, 2 (March 2007), 55–66. https://doi.org/10.1109/msp.2007.323264

Juhani Paasonen, Aleksandr Karapetyan, Jan Plogsties, and Ville Pulkki. 2017. Proximity of Surfaces - Acoustic and Perceptual Effects. *J. Audio Eng. Soc* 65, 12 (2017), 997–1004. http://www.aes.org/e-lib/browse.cfm?elib=19365

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

Boaz Rafaely. 2015. *Fundamentals of Spherical Array Processing (Springer Topics in Signal Processing)* (2015 ed.). Springer. http://www.worldcat.org/isbn/9783662456644

Nikunj Raghuvanshi, Rahul Narain, and Ming C. Lin. 2009a. 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, Rahul Narain, and Ming C. Lin. 2009b. 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 (July 2014). https://doi.org/10.1145/2601097.2601184

Nikunj Raghuvanshi, John Snyder, Ravish Mehra, Ming C. Lin, and Naga K. Govindaraju. 2010. Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes. *ACM Transactions on Graphics* 29, 3 (July 2010).

Jens H. Rindel and Claus L. Christensen. 2013. The use of colors, animations and auralizations in room acoustics. In *Internoise 2013*.

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

Carl Schissler, Ravish Mehra, and Dinesh Manocha. 2014. High-order Diffraction and Diffuse Reflections for Interactive Sound Propagation in Large Environments. *ACM Trans. Graph.* 33, 4 (July 2014). https://doi.org/10.1145/2601097.2601216

Dirk Schröder. 2011. *Physically Based Real-Time Auralization of Interactive Virtual Environments*. Logos Verlag. http://www.worldcat.org/isbn/3832530312

Jonathan Sheaffer, Maarten Van Walstijn, Boaz Rafaely, and Konrad Kowalczyk. 2015. Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 23, 12 (Dec. 2015), 2125–2135. https://doi.org/10.1109/taslp.2015.2468066

S. Siltanen, T. Lokki, and L. Savioja. 2010a. Rays or Waves? Understanding the Strengths and Weaknesses of Computational Room Acoustics Modeling Techniques. In *Proc. Int. Symposium on Room Acoustics*. Melbourne, Australia.

Samuel Siltanen, Tapio Lokki, and Lauri Savioja. 2010b. Room acoustics modeling with acoustic radiance transfer. *Proc. ISRA Melbourne* (2010).

Julius O. III Smith. 2007. Introduction to Digital Filters with Audio Applications. (2007). https://ccrma.stanford.edu/~jos/filters/

Alex Southern, Damian T. Murphy, and Lauri Savioja. 2012. Spatial Encoding of Finite Difference Time Domain Acoustic Models for Auralization. *Trans. Audio, Speech and Lang. Proc.* 20, 9 (Nov. 2012), 2420–2432. https://doi.org/10.1109/tasl.2012.2203806

Micah T. Taylor, Anish Chandak, Lakulish Antani, and Dinesh Manocha. 2009. RESound: interactive sound rendering for dynamic virtual environments. In *Proceedings of ACM conference on Multimedia*. ACM, New York, NY, USA, 271–280. https://doi.org/10.1145/1631272.1631311

Sakari Tervo, Jukka Pätynen, Antti Kuusinen, and Tapio Lokki. 2013. Spatial Decomposition Method for Room Impulse Responses. *J. Audio Eng. Soc* 61, 1/2 (2013), 17–28.

Nicolas Tsingos. 2009. Pre-computing geometry-based reverberation effects for games. In *35th AES Conference on Audio for Games*.

Nicolas Tsingos, Carsten Dachsbacher, Sylvain Lefebvre, and Matteo Dellepiane. 2007. Instant Sound Scattering. In *Rendering Techniques (Proceedings of the Eurographics Symposium on Rendering)*. http://www-sop.inria.fr/reves/Basilic/2007/TDLD07

Michael Vorländer. 2007. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality (RWTHedition)* (1 ed.). Springer. http://www.worldcat.org/isbn/3540488294

Hengchin Yeh, Ravish Mehra, Zhimin Ren, Lakulish Antani, Dinesh Manocha, and Ming Lin. 2013. Wave-ray Coupling for Interactive Sound Propagation in Large Complex Scenes. *ACM Trans. Graph.* 32, 6 (Nov. 2013). https://doi.org/10.1145/2508363.2508420

Wen Zhang, Thushara D. Abhayapala, Rodney A. Kennedy, and Ramani Duraiswami. 2010. Insights into head-related transfer function: Spatial dimensionality and continuous representation. *The Journal of the Acoustical Society of America* 127, 4 (01 April 2010), 2347–2357. https://doi.org/10.1121/1.3336399