Scene-Aware Audio for 360° Videos

Dingzeyu Li, Timothy R. Langlois, Changxi Zheng

Denis THY

Review:Product Importance Sampling for Light Transport Guiding

- Evaluation of MonteCarlo Ray Tracing
- Previous work: Guiding technique
- Gaussian Mixed Model (GMM)
- EM Algorithm
- Process
  - Preprocessing BRDF
  - Training
  - Rendering

Contents

- Motivation & goal
- Previous works
- Main idea
- Method overview
- Acoustic model & ambisonic audio
- Strength & weaknesses
- Results & application

Motivation & goal

- Technological advances
- 360° video
- Lack of immersive audio

- Adding scene-aware spatial audio to 360° videos

Previous work: Spatial Audio in VR [Schissler et al.] 2016

- Novel analytical formulation for large area and volumetric sound sources in outdoor environments.
- Spatial Room IR

Simulation of sound propagation

- Wave-based: expensive computation [Raghuvanshi et al]
- Geometric acoustics: cheaper
- Bidirectional path tracing [Cao et al.]
- Precomputation of impulse response [Raghuvanshi et al]

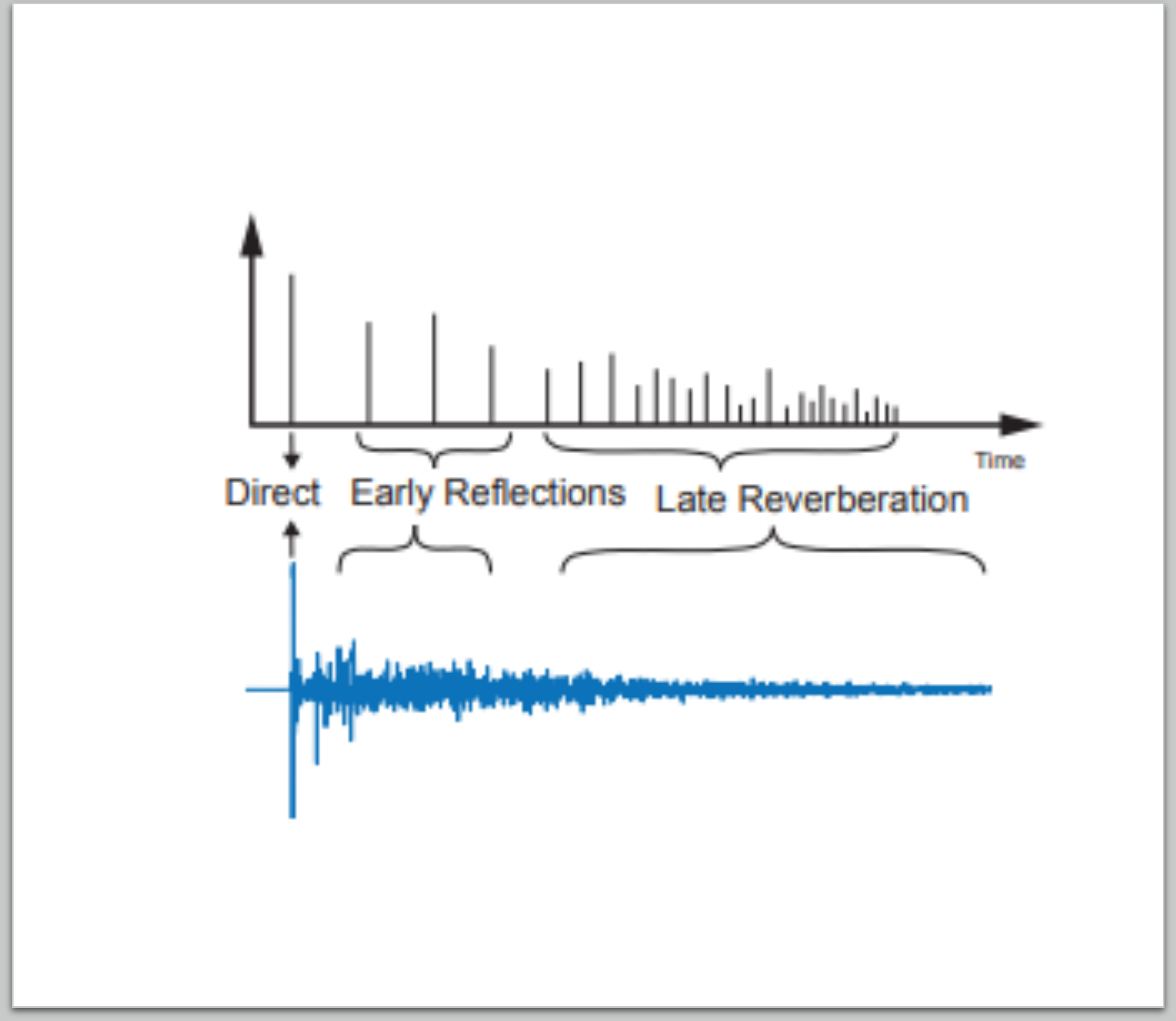
Current solution

Ambisonic microphone

Main idea

Constructing a directional Impulse Response (IR)

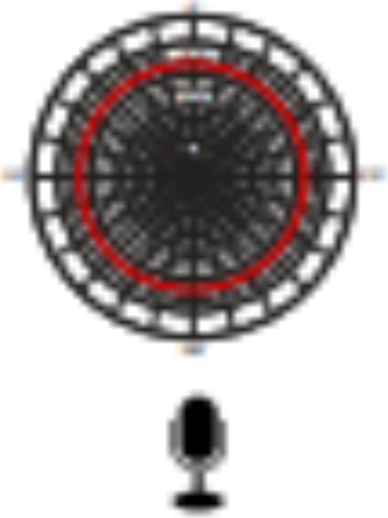
- Direct sound
- Early reflection impulse response (ERIR)
- Late reflections impulse response (LRIR)



Direction

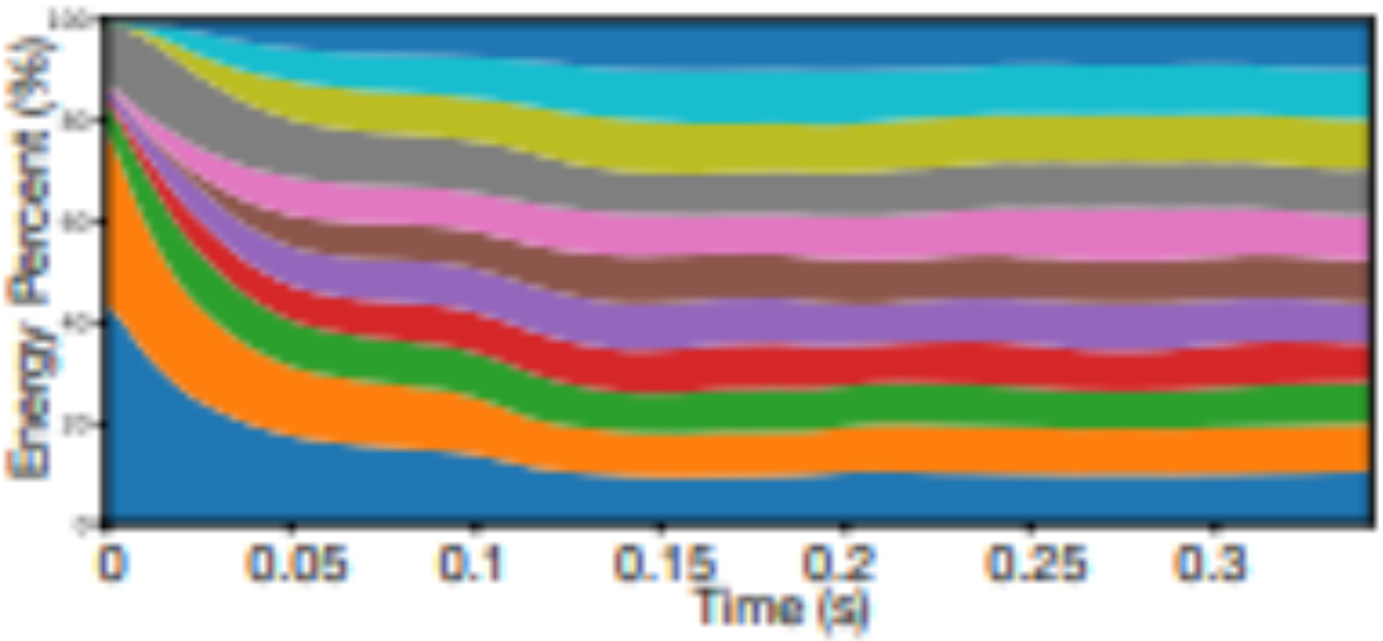
- The direction of the sound source is given by the direct sound and the early reflection
- Precedence effect

Late reverberation's isotropy



Non-Directional

Highly Directional



Method overview

360° video analysis

- 3D Scene reconstruction
- Structure-from-motion [Huang et al. 2017]
- Point cloud scene

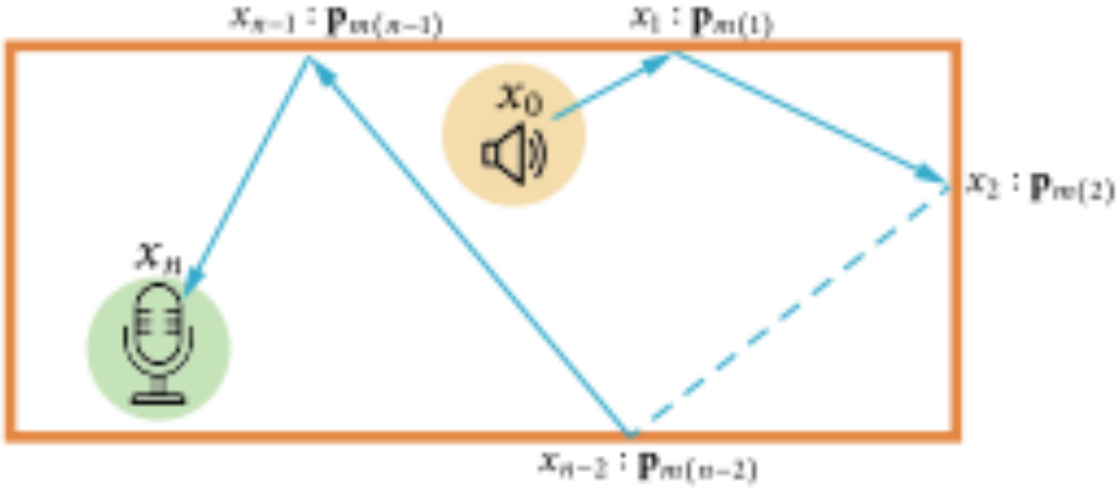
Room IR analysis

- IR measurement – sine sweep technique [Farina 2000]
- $s_e(t) = \sin\left[\frac{\omega_1 T}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\left(e^{\frac{t}{T}\ln\frac{\omega_1}{\omega_2}} - 1\right)\right]$
- Late Reverberation

Measurement Setup

Sense the acoustic material properties

$p = [p_{m(1)}, p_{m(2)}, \dots, p_{m(k)}]$ is the material absorption coefficient



$$e_j(p) = \beta_j \prod_{i=1}^{N_j} p_{m(i)}$$

$$J(p) = \sum_{j=1}^{M} [\log_{10}\left(\frac{e_j(p)}{e_0}\right) - \log_{10}(\frac{\tilde{h}(t_j)}{\tilde{h}(t_0)})]^2$$

Room IR analysis

- Late reverberation part
- Sense the acoustic material properties
- Transition point between ERIR & LRIR

Spatial audio generation

- Compute ERIR from source and positions of the reconstructed camera path
- Combine ERIR and LRIR

Room acoustic analysis for 360° scene

Room acoustic

- Geometric acoustic model
- Bidirectional path tracing to simulate sound propagation
- Ignoring wave behaviors

Room resonance

- Frequency modulation method
- Fourier transform:

$$\widetilde{H}(\omega) = F[\widetilde{H}(\mathrm{t})] \text{ and } H(\omega) = F[H(t)]$$
$$for\ t_0 < t < t_0 + \Delta t$$

$$M(\omega) = \left|\frac{H(\omega)}{\widetilde{H}(\omega)}\right|$$

Ambisonic Audio

Direction-Aware IR construction

- Trajectory analysis : structure-from-motion [Huang et al. 2017]
- Simulating ER: collect a set of incoming rays
- Constructing IR
  - ERIR: Linkwitz-Riley 4th order crossover filter + Modulation → $H_{r,\theta}$
  - LRIR: $H_L(t) = K \ * H(t) \ for \ t > t_{ER}$

Generating ambisonic audio

Background

Helmholtz equation:

$$(\Delta + k^2)p = -f_k(\psi)\frac{\delta(r - r_L)}{{r_L}^2}$$

Sound pressure:

$$p_k(r,\psi) = -ik\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\phi_{k,nm}Y_n^m(\psi)h_n(kr_L)j_n(kr)$$

Generating ambisonic channels

Early: $\phi_{nm} = Y_n^m(\Theta)\left(s_i(t) * H_{r,\theta}(t)\right)$

Late: $s_L(t) = s_i(t) * H_L(t)$

Strength & limitations

- Cheap
- Fast

- Outdoors scene
- Smaller obstacles

Results & Applications

Settings

- Ricoh Theta V 360 Camera
- TA-1 3D Audio Microphone
- Zoom H2n Recorder
- Presonus Eris E3.5 Reference Speaker
- 4-core Intel i7 CPU

Summary

- The key idea is combining simulated early reflections with recorded late reverberation
- Directional information from ERIR
- Recording an IR provides many information such as material acoustics properties
- This method converts a mono-channel input into a spatial audio

Thank you for listening
Any question ?