Interactive Sound Rendering

Ming C. Lin
Department of Computer Science
lin@cs.umd.edu
http://www.cs.umd.edu/~lin
http://gamma.cs.umd.edu/

Themes

- Exploiting analytical solutions using Modal Analysis to accelerate numerical simulation and reducing runtime computation
- Capture only perceptually important auditory cues to perform real-time sound synthesis and acoustic propagation on complex 3D scenes

2

Overview

- Interactive Sound Synthesis
- Efficient Numerical Acoustic Simulation
- Interactive Sound Propagation
- Conclusion and Future Work

3

Overview

- Interactive Sound Synthesis
  - Modal Analysis
  - Perceptually-based acceleration techniques
  - Hundreds of sounding objects
  - Sound from image textures
- Efficient Numerical Acoustic Simulation
- Interactive Sound Propagation
- Conclusion and Future Work

4

Physically-based Sound Synthesis

- Aim: Take object geometry and material as input and produce sound
- Current trend: Recorded sounds
- Problems with recorded sounds:
  - Difficult, expensive or dangerous to record (eg. Explosions)
  - Repetitiveness
  - Complex interactions (impact/rolling)

A typical foley studio*

* Image taken from: http://www.marblehead.net/foley/index.html

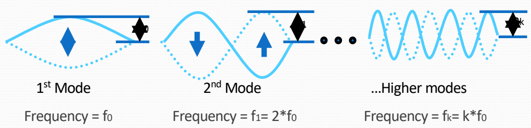
5

Physical Simulation

- Elastic deformable model
- Typical simulation time-steps must be ~10^{-5} s
- Direct simulation infeasible
- Efficient method: Modal Analysis

6

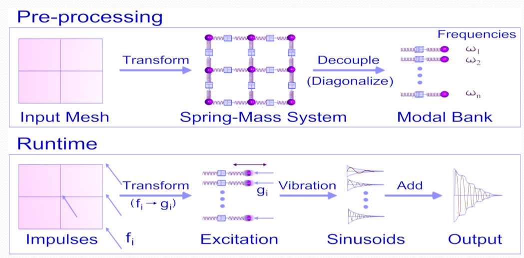
Modal Analysis



- Each mode represents a resonant mode of vibration
- Frequency of a mode is fixed
- Applying impulse excites modes of vibration
- Position of impact determines relative amplitude of modes

7

Overview of Technique



9

Approach

- Simpler model: Spring-mass system
- Fast: Supports hundreds of sounding objects
- Runs in real-time, low CPU utilization (~10%), graceful degradation in quality with limited computation
- Exploit human auditory perception
  - Mode Compression
  - Mode Truncation
  - Quality Scaling

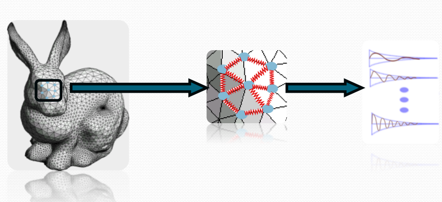
10

Modal Analysis

- Deformation modeling
  - Vibration of surface generates sound
  - Sound sampling rate: 44100 Hz
  - Impossible to calculate the displacement of the surface at sampling rate
  - Represent the vibration pattern by a bank of damped oscillators (modes)
- Standard technique for real-time sound synthesis

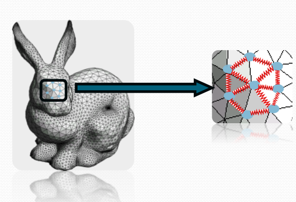
Modal Analysis

- Discretization
  - An input triangle mesh → a spring-mass system
  - A spring-mass system → a set of decoupled modes



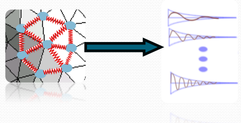
Modal Analysis

- The spring-mass system set-up
  - Each vertex is considered as a mass particle
  - Each edge is considered as a damped spring



Modal Analysis

- Coupled spring-mass system to a set of decoupled modes

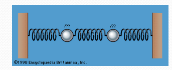


Modal Analysis

- A discretized physics system
  - We use spring-mass system

$$K(d) + C(d, \dot{d}) + M(\ddot{d}) = f$$

Stiffness Damping Mass



- Small displacement, so consider it linear

$$Kd + C\dot{d} + M\ddot{d} = f$$

Stiffness Damping Mass

Modal Analysis

- Solve the Ordinary Differential Equation (ODE)

$$Kd + C\dot{d} + M\ddot{d} = f$$

- Rayleigh damping

$$Kd + (\gamma M + \eta K)\dot{d} + M\ddot{d} = f$$

And diagonalizing

$$K = GDG^{-1}$$

- Now, solve this ODE instead

$$DG^{-1}d + (\gamma G^{-1}M + \eta DG^{-1})\dot{d} + G^{-1}M\ddot{d} = G^{-1}f$$

Modal Analysis

- Solve the ODE

$$DG^{-1}d + (\gamma G^{-1}M + \eta DG^{-1})\dot{d} + G^{-1}M\ddot{d} = G^{-1}f$$

- Substitute $z = G^{-1}d$ the modes)

Now, solve this ODE instead

$$Dz + (\gamma M + \eta D)\dot{z} + M\ddot{z} = G^{-1}f$$

Modal Analysis

- General solution

$$z_i = c_i e^{\omega_i^+ t} + \bar{c}_i e^{\omega_i^- t}$$

$$\omega_i^{\pm} = \frac{-(\gamma\lambda_i + \eta) \pm \sqrt{(\gamma\lambda_i + \eta)^2 - 4\lambda_i}}{2}$$

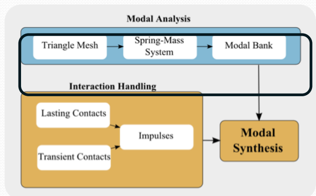
- External excitation defines the initial conditions

Modal Analysis

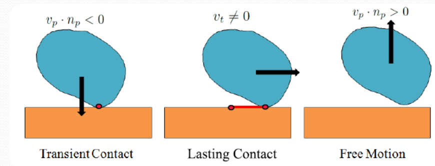
- Assumptions
  - In most graphics applications, only surface representations of geometries are given
  - A surface representation is used in modal Analysis
  - Synthesized sound appears to be “hallow”

Modal Analysis Summary

- An input triangle mesh →
  A spring-mass system →
  A set of decoupled modes



State Detection



State Detection

- Distinguishing between lasting and transient contacts
  - In contacts?

$$\begin{cases} v_p \cdot n_p < 0 & \text{in contact} \\ v_p \cdot n_p > 0 & \text{not in contact} \end{cases}$$

  - In lasting contacts?

$$\begin{cases} v_t \neq 0 & \text{lasting contact} \\ v_t = 0 & \text{not in lasting contact} \end{cases}$$

Interaction Handling

- Lasting contacts → a sequence of impulses
- Transient contacts → a single impulse

Impulse Response

- Dirac Delta function as impulse excitation
  - General solution

$$z_i = c_i e^{\omega_i^+ t} + \bar{c}_i e^{\omega_i^- t}$$

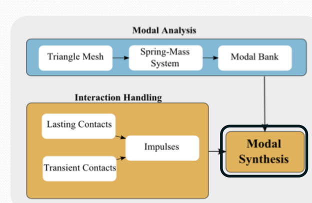
with initial condition given by the impulse, we have:

$$c_i' = c_i e^{\omega_i^+ t_0} + \frac{g_i}{m_i(\omega_i^+ - \omega_i^-)}$$

$$\bar{c}_i' = \bar{c}_i e^{\omega_i^- t_0} - \frac{g_i}{m_i(\omega_i^+ - \omega_i^-)}$$

Impulse Response

$$z_i = c_i e^{\omega_i^+ t} + \bar{c}_i e^{\omega_i^- t}$$



Handling Lasting Contacts

- The interaction simulation has to be stepped at the audio sampling rate: 44100 Hz
- The update rate of a typical real-time physics simulator: on the order of 100's Hz
- Not enough simulation is provided by the physics engine
- An customized interaction model for sound synthesis

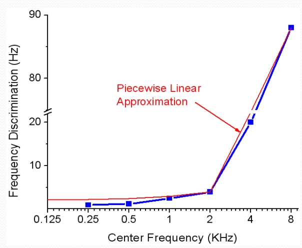
Mode Compression: Principle

- Humans can't distinguish two frequencies arbitrarily close to each other [Sek et. al., 1995*]
- Accuracy in discriminating frequencies depends on the frequency in question
- Different frequencies were played in succession to find if the subject could distinguish between them

*Sek, A., and Moore, B. C. 1995. Frequency discrimination as a function of frequency, measured in several ways. J. Acoust. Soc. Am. 97, 4 (April), 2479–2486.

27

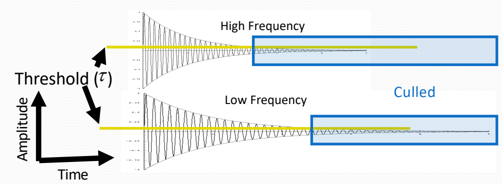
Mode Compression: Auditory Perception



Frequency Discrimination at 2 KHz is about 4 Hz -- We can't tell apart frequencies within the range 1998 - 2002 Hz

28

Mode Truncation



- Impact Sounds: Attack + Decay
- Key Point: Critical to capture attack properly
- Stop mixing mode when its contribution falls below a prescribed threshold, τ (typically -60 to -80 dB of initial level)

29

Quality Scaling

- A typical audio scene consists of foreground and background sounds
- Higher intensity sounds are considered to be foreground
- Idea: Give more importance to foreground sounds
- Provides a graceful way to adapt to variable time constraints

30

Implementation Details

- System: 3.4 GHz Pentium 4 Laptop, 1 GB RAM
- Graphics: GeForce 6800 Go, 256 MB
- Sound: Creative Sound Blaster Audigy 2 ZS
- Physics: Pulsk (written by former student: Nico Galoppo), Rendering: G3D
- Also integrated with NVIDIA PhysX Engine recently

31

System Demonstration

VIDEO

32

Sound from Image Textures

VIDEO

33

Limitations

- Implementation: Model only the surface (not an inherent limitation of the approach)
- More approximate than an FEM-based formulation
  - Some tuning is required
- All sound synthesis techniques relying on Modal Analysis: Can only use linear damping models

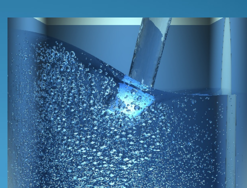
34

Sounding Liquids [Moss et al. 2009]

- Work in physics and engineering literature since 1917
  - Sound generated by resonating bubbles
- *Physically-based Models for Liquid Sounds* (van den Doel, 2005)
  - Spherical bubble model
  - No fluid simulator coupling
    - Hand tune bubble profile

Background (Fluid)

- Grid-based methods
  - Accurate to grid resolution
    - Bubbles can be smaller
  - Slow
  - Can be two-phase



Background (Fluid)

- Shallow Water Equations
  - Simulate water surface
    - No breaking waves
  - Real time
  - One phase
    - Explicit bubbles

Overview

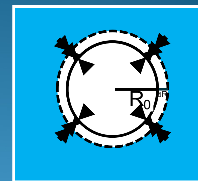
- Generate sound from existing fluid simulation
  - Model sound generated by bubbles
- Apply model to two types of fluid simulators
  - **Particle-Grid-based**
    - Extract bubbles
    - Process spherical and non-spherical bubbles
    - Generate sound
  - **Shallow Water Equations**
    - Processes surface
      - Curvature and velocity
    - Select bubble from distribution
    - Generate sound

Mathematical Formulations

- **Spherical Bubbles**

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{3\gamma p_0}{\rho R_0^2}}$$

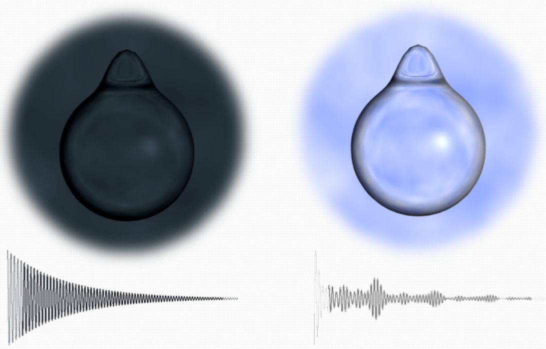
$$\tau(t) = Asin(2\pi f(t)t)e^{-dt}$$



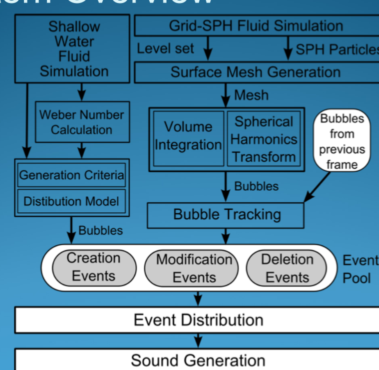
- **Non-spherical bubbles**
  - Decompose into a spherical harmonics

$$f_n^2 \approx \frac{1}{4\pi^2}(n-1)(n+1)(n+2)\frac{\sigma}{\rho R_0^3}$$

Without Spherical Harmonics | **With** Spherical Harmonics

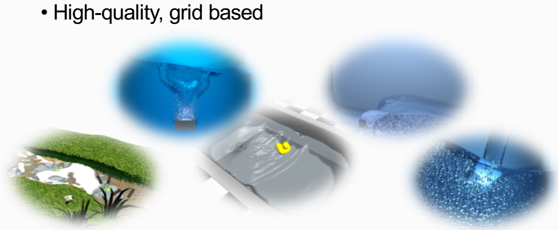


System Overview



Summary

- Simple, automatic sound synthesis
- Applied to two fluid simulators
  - Interactive, shallow water
  - High-quality, grid based



Video Demonstration

VIDEO

Overview

- Interactive Sound Synthesis
- Efficient Numerical Acoustic Simulation
- Interactive Sound Propagation
- Conclusion and Future Work

Overview

- Interactive Sound Synthesis
- Efficient Numerical Acoustic Simulation
  - Novel technique based on 3D Adaptive Rectangular Decomposition
  - Hundred times faster than Finite Difference Time Domain
- Interactive Sound Propagation
- Conclusion and Future Work

Acoustics: Governing Equation

- Solve the Linear Wave Equation:

$$\frac{\partial^2 p}{\partial t^2} - c^2 \nabla^2 p = F(\mathbf{x}, t)$$

∇^2 is the Laplacian operator in 3D
c is the speed of sound in air
p is the pressure field to solve

- The RHS is the forcing term, corresponding to volume sound sources in the scene

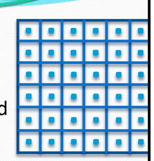
State of the art: Room Acoustics

- Geometric Techniques (e.g. Ray/Beam/Frustum Tracing) combined with explicit diffraction modeling
- Auralization software (e.g. ODEON, CATT): Hybrid image-source and ray-tracing along with (upcoming) explicit diffraction modeling
- Numerical acoustic simulation for complex 3D scenes has been explored only recently (2006)

Acoustics for Interactive Applications

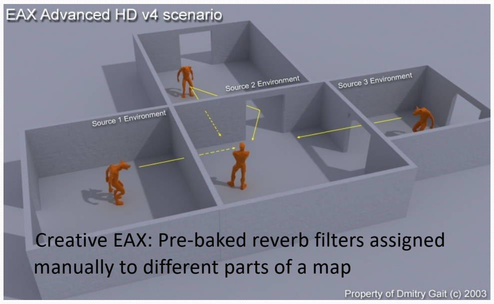
- Geometric Approaches
  - Beam Tracing (Funkhouser et. al.)
  - Phonon Tracing (Bertram and Deines et. al.)
  - Frustum Tracing (Chandak et. al.)
- Advantages: Efficient, easy to understand
- Difficulties: Diffraction / Scattering, high-order reflections

Numerical Acoustics



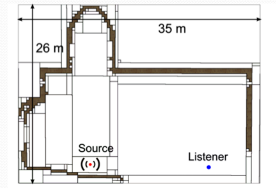
- Discretize and solve Wave Equation on a grid
- Explored for complex 3D scenes (eg. auditoria) only recently (2004 – 2006) by Sakamoto et. al.
- Disadvantage: Slow and memory-intensive
  - Simulations are band-limited
- Advantages: Diffraction / Scattering, high-order reflections

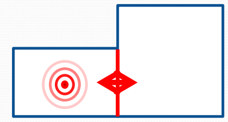
Acoustics in Games



50

Adaptive Rectangular Decomposition





Interface Handling

- Numerical Simulation of the Wave Equation
- Rectangular Decomposition of a 3D scene
- Exploit analytical solutions on rectangular spaces
- 6th order Finite Difference for interface transmission

51

Solution on a Rectangular Domain

- Rectangular space in 3D with size (l_x, l_y, l_z), and perfectly reflective boundary
- Modal Analysis can be done ***analytically*** –

$$p(x,y,z,t) = \sum_{i=(i_x,i_y,i_z)} m_i(t)\,\Phi_i(x,y,z)$$

$$\Phi_i(x,y,z) = \cos\left(\frac{\pi i_x}{l_x}x\right)\cos\left(\frac{\pi i_y}{l_y}y\right)\cos\left(\frac{\pi i_z}{l_z}z\right)$$

52

Leveraging GPU for Acoustics

- Solution of Wave Equation within each rectangle can be done using a 3D Discrete Cosine Transform (DCT)
- DCTs can be computed using FFT
- Use efficient FFT implementation on GPU
  - Govindaraju, N. K., Lloyd, B., Dotsenko, Y., Smith, B., and Manferdelli, J. **High-Performance Discrete Fourier Transforms on Graphics Processors**. In the *Proc. of 2008 ACM/IEEE Supercomputing*

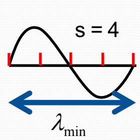
53

Computational Efficiency

- For a scene of size **L** in 3D and simulation duration **T** –

Memory : $\left(\frac{L s \nu_{max}}{c}\right)^3$

Time : $T\left(\frac{L}{c}\right)^3 (s\nu_{max})^4$



- Nyquist Limit: **s ≥ 2**
- FDTD: **s = 10**. My approach: **s = 2.6**
- Speedup with my technique : $(10/2.6)^4$ > 100

54

Demo

Video

55

Performance Comparison

Scene Name	Volume (m³)	Time: FDTD (CPU)	Time: My Technique (GPU)	**Speedup**
Corridor	375	365 min	4 min	**~ 90x**
House*	1,275	3.5 days	24 min	**~ 200x**
Cathedral	13,650	1 week (estimated)	29 min	**~ 300 x**

- Quad-core 2.8GHz Intel Xeon CPU with 8GB RAM, NVIDIA GeForce GTX 280

* This simulation was band-limited to 2 kHz, instead of 1 kHz

56

Summary

- Adaptive Rectangular Decomposition yields 100x improvement in performance over FDTD and consumes 10x less memory
- Source of Speedup: Modal Analysis of rectangular spaces as well as GPU-DCT
- Can feasibly simulate acoustics for large, complex scenes, such as a Cathedral

57

Overview

- Interactive Sound Synthesis
- Efficient Numerical Acoustic Simulation
- Interactive Sound Propagation
- Conclusion and Future Work

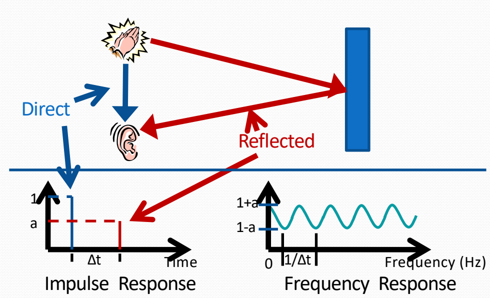
59

Overview

- Interactive Sound Synthesis
- Efficient Numerical Acoustic Simulation
- Interactive Sound Propagation
  - Perceptual aspects of acoustics
  - Novel perceptually-motivated techniques
  - Interactive auralization system: moving sources and listener
- Conclusion and Future Work

60

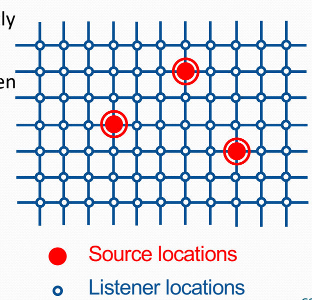
Impulse Response (IR)



61

Challenges

- Direct approach is costly
- Days of simulation (even with fast simulator)
- Terabytes of storage



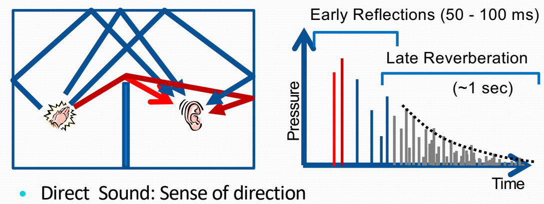
62

Contributions

- Approach –
  - Sample data at lower resolution in space (~1 m)
  - Novel perceptually-based scheme: Store each Impulse Response compactly
  - Spatially-interpolate Impulse Responses
- Audio engine that uses fast frequency-domain convolutions

63

Auditory perception of acoustic spaces

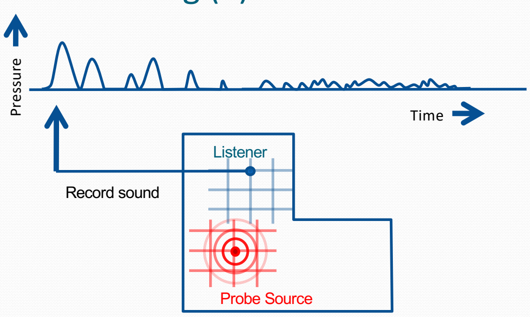


- Direct Sound: Sense of direction
- ER: Loudness, Timbre, "Envelopment". Perceivable spatial variation.
- LR: Only statistical properties perceivable – Decay Time (RT60), Periodicities (Flutter echoes)

Reference: "Room Acoustics" by Heinrich Kuttruff

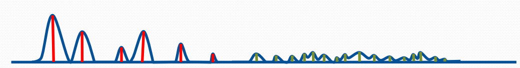
64

IR Factoring (1)



65

IR Factoring (2)

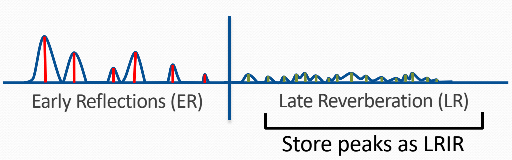


Peak Detection

- Finds peak delays and amplitudes

66

IR Factoring (3)

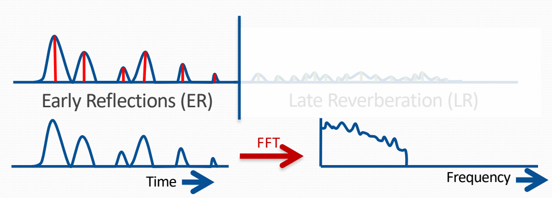


ER - LR Decomposition

- Compute and store only one Late Reverberation filter per room
- Reduces pre-computation time and runtime memory usage by 10 times

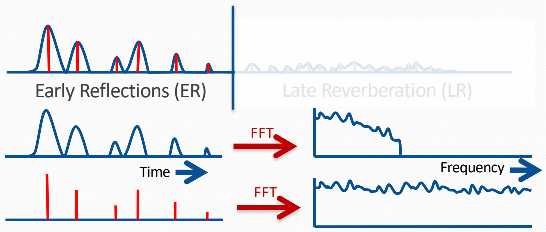
67

IR Factoring (4)



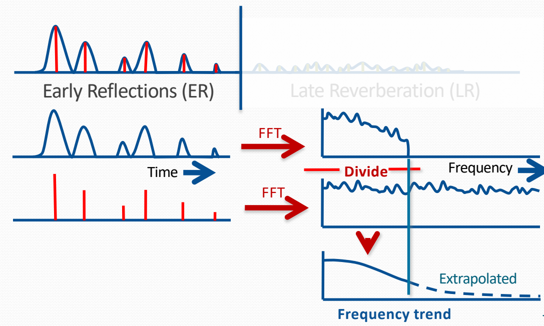
68

IR Factoring (5)



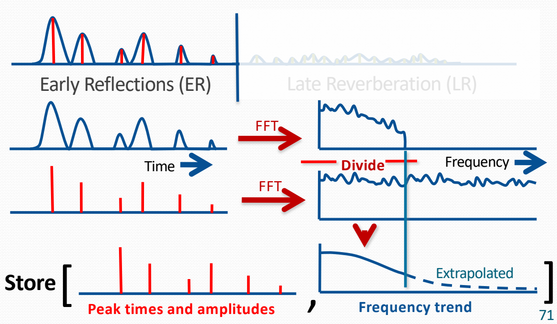
69

IR Factoring (6)



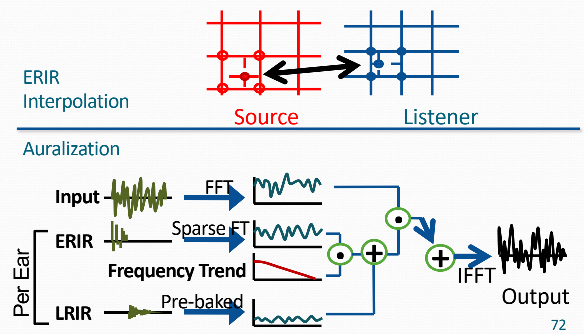
70

IR Factoring (7)



71

Runtime Processing



72

Performance

- Pre-processing times typically a few hours
- Can handle about 10 sound sources in real-time on a Quad-core Xeon system with 4 GB RAM
- Bottleneck: 1D FFT
  - Auralization system maps well to parallel processors

73

Walkthrough: Game Scene

- “Citadel” Scene from the game Half-Life 2
- Large Size: 3,500 m^3
- Complex geometry (fin-like structures)

74

System Demonstration

Video

http://www.youtube.com/watch?v=MQt1jtDBNK4

Summary

- The first interactive sound propagation system that leverages numerical simulation
- Can render important acoustic effects like Late Reverberation and Diffraction low-pass filtering in real-time
- Can handle multiple moving sound sources and listener
- Works for large, complex 3D scenes

Overview

- Sound Synthesis
- Efficient Numerical Acoustic Simulation
- Interactive Sound Propagation
- Conclusion and Future Work

Summary

- Interactive Sound Synthesis
  - Perceptually-motivated optimizations enabling hundreds of sounding objects in real-time
- Efficient Numerical Acoustics
  - A simulator 100 times faster than Finite Difference Time Domain for constant wave speed simulations
- Interactive Sound Propagation
  - Leverage fast numerical acoustics
  - Exploit auditory perception
  - Render wave-based acoustics for multiple moving sources and listener in real-time

Conclusion

- Physically-based Sound: Complex underlying physical processes require a lot of computational power
- Combination of efficient algorithms, perceptually-motivated optimizations and fast hardware

Future Work: Sound Synthesis

- Ultimate goal: Virtual Worlds with physically-based sounds for collisions, rolling, sliding, creaking, cloth, gunshots, water, automobiles, and so on
- Infer audio materials from video
- Virtual Musical Instruments using next generation user interfaces

Future Work: Sound Propagation

- Acoustics for Games and Virtual Worlds
- Accurate numerical predictions in auditorium design
- Efficient numerical solvers for high-performance computing applications
- Combine Sound Synthesis and Acoustics for a completely physically-based auralization system

82

References

- Nikunj Raghuvanshi and Ming C. Lin. **Interactive Sound Synthesis for Large Scale Environments**. In SI3D '06: Proceedings of the 2006 symposium on Interactive 3D graphics and games, 2006.
- Nikunj Raghuvanshi, Nico Galoppo, and Ming C. Lin. **Accelerated Wave-based Acoustics Simulation**. In SPM '08: Proceedings of the 2008 ACM symposium on Solid and physical modeling, 2008.
- Nikunj Raghuvanshi, Rahul Narain, and Ming C. Lin. **Efficient and Accurate Sound Propagation using Adaptive Rectangular Decomposition**. IEEE Transactions on Visualization and Computer Graphics, December 2009.
- Nikunj Raghuvanshi, Brandon Lloyd, and Ming C. Lin. **Efficient Numerical Acoustic Simulation on Graphics Processors Using Adaptive Rectangular Decomposition**. Proc. EAA Symposium on Auralization, 2009.

83

84