Reverberation and Room Simulation

Paris Smaragdis
paris@illinois.edu
paris.cs.illinois.edu
Overview

• Digital reverberation
  • Simple filter models
  • Convolution with real impulses

• Measuring reverb from rooms

• Room simulators
  • Image-source model
Reverberation

• Effects that rooms impose on sounds
  • Desirable effect for music, movies, video games
  • Undesirable effect for speech recordings!

• Couple of effects
  • Echoes (repeating the input after a while)
  • Coloring (changing the spectrum of the input)
Reverberation

- What you hear in a room is a sound and all of its echoes (and their echoes) from reflections
An exception

- Anechoic chambers
  - No noticeable reverberation due to special sound absorbing surfaces

- Very useful for acoustical testing and evaluation
Rooms as filters

- Each sound reflection will cause an echo
- We can create echoes with convolution

The set of these echoes is known as the *Room Impulse Response (RIR)*
Anatomy of reverb

- Three main parts
  - Direct sound; early reflections; tail
- Main characteristic: The $RT_{60}$
Anatomy of reverb

- Spectral view
  - Early reflections are generally broadband (why?)
  - Later echoes tend to muffle sound (why?)
Early reflections

- First ~80-100ms in reverb response
- Generally irregular
- Key part to get right when emulating rooms
The tail (fused reflections)

- Almost looks like white noise
- Echoes of echoes of echoes ... (1,000+ echoes/sec)
- Smoothly decaying energy
Some descriptors

- The Energy Decay Curve (EDC)
  - Measure how fast the reverb decays
  - The integral of the squared response over time
    \[ EDC(t) = \int_{t}^{\infty} |h(\tau)|^2 d\tau \]
    \[ RT_{60} = t : EDC(t) = EDC(0) - 60dB \]

- The Energy Delay Relief (ERD) curve
  - Energy at time \( t \) for frequency \( \omega \)
    \[ EDR(\omega,t) = \int_{t}^{\infty} |H(\omega,\tau)|^2 d\tau \]
Some more

- Clarity index
  - Amount of early reflections vs. reverb tail energies
    \[ C = 10 \log_{10} \frac{\int_{0}^{\infty} |h(t)|^2 \, dt}{\int_{T}^{\infty} |h(t)|^2 \, dt} \]

- Resonant modes
  - Frequency ranges where signal is more prominent
Simulating reverb

• Analog reverb simulators
  • The spring reverb
    • Use a vibrating spring to add reverb
  • The plate reverb
    • Use a resonating plate instead

• Adjustable rooms
  • Very expensive!!
Simulating reverb digitally

• We can emulate a room by matching the early reflections and the reverb tail

• Two main approaches
  • 1) Fake it with filters
    • Fast and easy
  • 2) Compute impulse response from room geometry
    • Computationally heavy, needs some more thinking
Modeling a very simple room

- How will this room’s impulse response look like?
Modeling a sole echo

- Impulse will be:
  \[ y[t] = x[t] + gx[t - N] \]
  \[ Y[k] = 1 + ge^{-jkN} \]
One step up

• How will this room impulse look like?
Comb filters

- We can model an infinite echo with:

\[ y[t] = x[t] + gy[t - N] \]
\[ Y[k] = \frac{1}{1 - ge^{-j k N}} \]

![Diagram showing input, delay, and attenuation]

- Impulse function
- Frequency response
Allpass filters

- Even better, has no spectral coloring

\[ y[t] = gx[t] + x[t - N] + gy[t - N] \]
How about a room with many walls?

- We can have multiple walls in a simple room.
- Many different orders of reflections.
The Schroeder model

- Combining multiple comb/allpass filters
  - Parallel combs model ERs and add “coloring”
  - Allpass filters in series perform diffusion
- All filter delays should be unrelated!
The other Schroeder model

- Combining multiple allpass filters in series
  - No ER model from comb
  - Smoother sounding, very diffuse
    - Allpass filters in series are also called diffusers

Input → Allpass → Allpass → Allpass → Allpass → Allpass → Output
Elaborations of the Schroeder model

- We can also account for scattering
  - Add a lowpass filter in the comb structure
    - We can also do the with the all pass structure
Wetness of reverb

- “Wet” means a lot of reverb
- “Dry” means no reverb
Shortcomings and issues

• Can sound too synthetic
  • Picking unrelated delays for the filters is key!
    • Schroeder’s suggestion:

\[
N_i = \frac{F_s}{10 \cdot 3^i}
\]

• Not really a room model
  • Doesn’t account for room geometry or acoustical effects
Reverberation via convolution

- Since the room is a filter, we can directly convolve with the room impulse response
  - Advantage: Exact model
  - Disadvantage: What’s the room’s impulse response?
Approach 1: Cheating

- We can make up a room response

- Things to look for:
  - First 100ms model the early reflections
    - Need a few prominent echoes, no coloring, not too busy
  - Rest of response needs a very dense echo count
    - Rule of thumb: 1,000+ echoes per sec
Making up responses

• Early reflections
  • Can use sparse noise
    • E.g., Cauchy noise (divide two Gaussian noise sequences)

• Reverb tail
  • Exponentially decaying noise
    • Make sure it’s white to avoid coloring
      • Or not, if you want coloring!
Using real room impulse responses

• Instead of making them up, use real responses

• Two ways to do this:
  • Go download one (e.g. here)
  • Estimate the response of a room
    • Either analytically, or by measuring recordings in it
Measuring room impulse responses

- The direct approach
  - Make an impulsive sound in the room (clap, balloon pop, gun)
  - Measure the resulting sound
How does that work?

• Assume that the sound we make is \([1,0,0,0,0,0,...]\)
• Output: \(y[t] = h[t] \ast [1,0,0,0,...] = h[t]\)
• Simulate with: \(z[t] = y[t] \ast x[t]\)
• We can use the recording as an impulse

\[
y[t] = h[t] \ast [1,0,0,0,0,...] = h[t]
\]
Some problems

- Extra noises will bias the measurement
  - What if someone drops a coin while recording?
    - You will get an extra metal-sounding echo!

- Hard to get an exact impulse sound
  - Excitation will most likely be “colored”
  - Will result in a biased RIR estimate
A more robust approach

- Deconvolution
  - Excite room with an appropriate signal
  - Knowing the excitation, estimate the RIR

- Repeat the experiment multiple times
  - Average out the results for robustness

- Allows for more accurate RIR estimates
Deconvolution process

- Excitation is:

\[ y[t] = h[t] \ast x[t] \]
\[ Y[\omega] = H[\omega]X[\omega] \]

- Solve for RIR:

\[ H[\omega] = \frac{Y[\omega]}{X[\omega]} \]
A better way

- **Maximum-Length Sequences** (MLS)
  - Sequences that autocorrelate well with themselves
    \[ X[\omega]X[\omega]^* \approx 1 \]
  - Binary sequences with length \(2^{N-1}\)
  - Generated via a simple iterative process
Using MLS to measure RIRs

- Rewrite deconvolution as:
  \[
  H[\omega] = \frac{Y[\omega]}{X[\omega]} = \frac{X^*[\omega]Y[\omega]}{X^*[\omega]X[\omega]} = X^*[\omega]Y[\omega]
  \]

- Removes potentially problematic division
  - Fast and easy to compute
  - Also works with other MLS-like signals
Alternative excitations

- The swept sine excitation
  - Pass through all frequencies individually

- The exponential swept sine
  - Has some better properties than swept wave

- White noise
  - A poor man’s MLS! :)


Example case - my office

![Input chirp](image1)

![Recorded chirp](image2)

![Estimated RIR](image3)
Tradeoffs and problems

• Longer excitations are better
  • More accurate spectra (more frequencies)

• Noise robustness via repetition
  • Average result out of many trials

• Be careful of loudspeakers and mic response!
  • You measure these as well
Physical models of rooms

• We can also make an analytical model of a room
  • i.e. estimate the room response analytically

• Multiple approaches
  • Finite Element (FE) models
  • Image-source methods
  • Path/Beam tracking
Tracing all paths

- Explicitly compute the delays from all echoes
  - Use them to construct the RIR

- How to we find the two delays here?
Image-source model

- Project source to virtual location through wall
- Calculate that distance instead
Do that for all walls
And repeat

- Keep going for as many reflections as you need
- Note that this becomes exponentially more complex!
And you also have to do it in 3D ...
Blending models

• Image-source model is great for early reflections
  • But gets complicated at the tail
    • Remember 1,000s of echoes per sec!

• Blend image-source model with noise convolution
  • Convolving with exponentially decaying noise can simulate the tail much faster than otherwise
More elaborate models

• Image-source isn’t perfect
  • No model for diffusion, problematic with complex room shapes, invalid virtual sources problems

• Instead we can use path or beam tracing
  • Randomly follow paths from source to microphone
    • Or beams instead of paths
  • Find hits and use for room model
Can handle complex environments
But there’s not just one RIR!

- Each spot inside a room has its own RIR
  - There isn’t a single true RIR!

- To really model a room we can:
  - Compute all sound beams explicitly
    - Costly!
  - Blend various RIRs as a source moves in room
    - Easier, but less precise
Compensating for physical rooms

• What happens when you play a simulated room recording inside a room?
  • Not what you wanted!

• We often measure room responses so that we can undo them and present the desired signal
  • Have to both equalize and remove echoes
Recap

• Reverberation via filtering
  • Simple filters
  • Room Impulse Responses
    • Measuring room responses

• Geometric models of rooms
  • The image-source model
Reference material

- Artificial reverberation
  - https://ccrma.stanford.edu/~jos/pasp/Artificial_Reverberation.html

- Beam tracing

- Image-source model
Next lab

- Designing simple room simulators
- Measuring reverberation
Adminstrivia

- Grading for lab 1 should be completed
  - Let me know if you haven’t heard back with a score

- Second lab due this Thursday!