Reverberation algorithms

Augusto Sarti

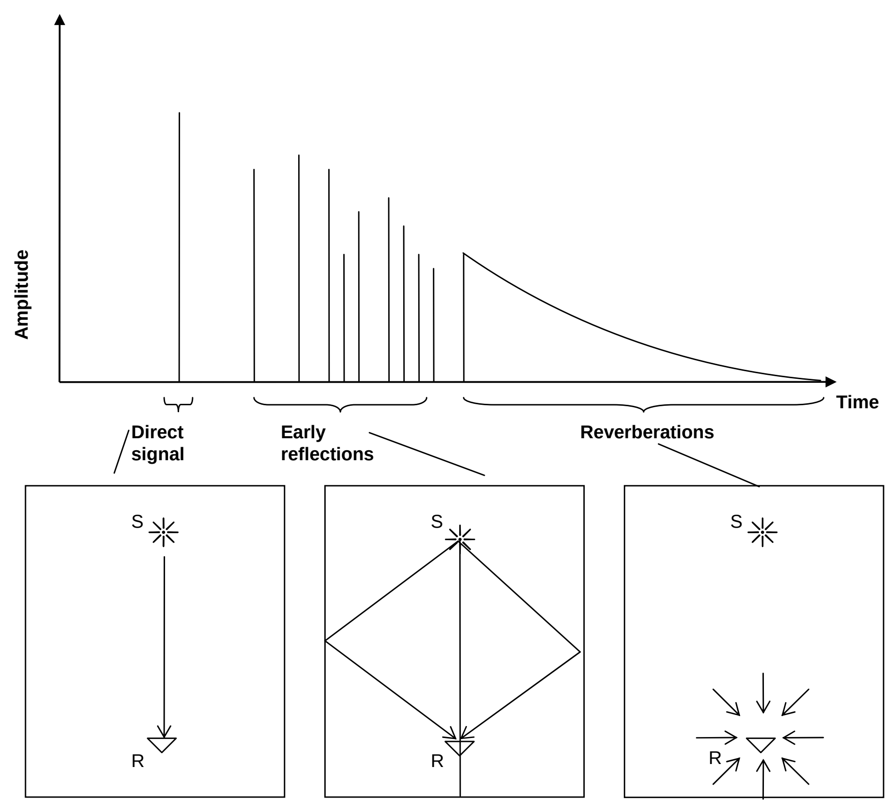
Summary

- The Reverb Problem
- Reverb Perception
- Acoustic impulse response:
  - Formation mechanisms
  - Parameters
- Early Reflections
- Late Reverb
- Numerical reverberation algorithms
  - Schroeder Reverbs
  - Feedback Delay Network (FDN) Reverberators
  - Waveguide Reverberators
- Geometrical reverberation algorithms

Impulse response

- The sounds we perceive heavily depend on the surrounding environment
- Environment-related sound changes are of convolutive origin (filtering)
  - Well-modeled by a space-varying impulse response

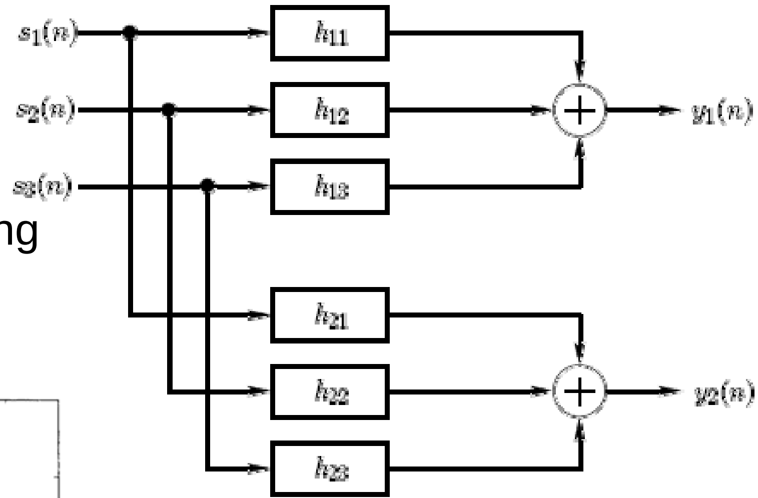
Impulse response

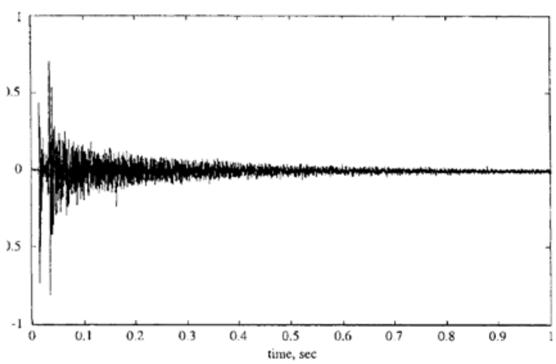


Reverberation tf Function

- Three sources, one listener (two ears)
- Filters should include pinnae filtering
- Filters change if anything in the room changes

(exact model)





Global descriptors

- Energy decay curve (EDC)

$$\text{EDC}(t) \triangleq \int_t^{\infty} h^2(\tau)d\tau$$

  - Introduced by Schroeder to define reverberation time
  - It measures the total signal energy remaining in the reverberator's impulse response at time *t*
  - It decays more smoothly than the impulse response, therefore it works better than the amplitude's envelope for defining the reverberation time
  - In reverberant environments a large amount of the total energy is contained in the last portion of the impulse response
- Reverberation time

$$T60 = \{t : EDC(t) = EDC(0) - 60dB\}$$

Global descriptors

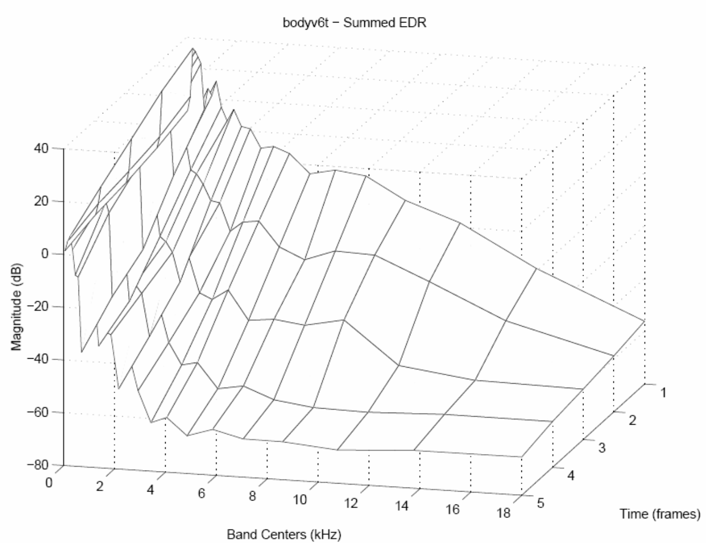
The *energy decay relief (EDR)* generalizes the EDC to multiple frequency bands:

$$\text{EDR}(t_n, f_k) \triangleq \sum_{m=n}^{M} |H(m,k)|^2$$

where $H(m,k)$ denotes bin k of the short-time Fourier transform (STFT) at time-frame m, and M is the number of frames.

- FFT window length $\approx 30 - 40$ ms
- $\text{EDR}(t_n, f_k)$ = total signal energy remaining at time t_n sec in frequency band centered at f_k

EDR of a violin body



- Energy summed over frequency within each "critical band of hearing" (Bark band)
- Violin body = "small box reverberator"

Global descriptors

- In the room's transfer function we can single out resonant modes
- The spacing between two resonant modes is given by

$$\Delta f_{\max} \approx \frac{4}{T_r} \text{ Hz.}$$

- which is valid above the threshold frequency

$$f_g \approx 2000\sqrt{\frac{T_r}{V}} \text{ Hz.}$$

Global descriptors

- Number of echoes in the impulse response before time *t*

$$N_t = \frac{4\pi(ct)^3}{3V}$$

- Derivative of N_t:

$$\frac{dN_t}{dt} = \frac{4\pi c^3}{V} t^2$$

- Clarity index: ratio btw early reflections energy and late reverberation energy

$$C = 10\log_{10}\left\{\frac{\int_0^{80\text{ ms}} p^2(t)dt}{\int_{80\text{ ms}}^{\infty} p^2(t)dt}\right\} \text{ dB}$$

Implementation

Let $h_{ij}(n)$ = impulse response from source j to ear i.
Then the output is given by *six convolutions*:

$$y_1(n) = (s_1 * h_{11})(n) + (s_2 * h_{12})(n) + (s_3 * h_{13})(n)$$
$$y_2(n) = (s_1 * h_{21})(n) + (s_2 * h_{22})(n) + (s_3 * h_{23})(n)$$

- For small n, filters $h_{ij}(n)$ are *sparse*
- Tapped Delay Line (TDL) a natural choice

Transfer-function matrix:

$$\begin{bmatrix} Y_1(z) \\ Y_2(z) \end{bmatrix} = \begin{bmatrix} H_{11}(z) & H_{12}(z) & H_{13}(z) \\ H_{21}(z) & H_{22}(z) & H_{23}(z) \end{bmatrix} \begin{bmatrix} S_1(z) \\ S_2(z) \\ S_3(z) \end{bmatrix}$$

Complexity of Exact Reverberation

Example:

- Let $t_{60} = 1$ second
- $f_s = 50$ kHz
- Each filter h_{ij} requires 50,000 multiplies and additions per sample, or 2.5 *billion* multiply-adds per second.
- Three sources and two listening points (ears) $\Rightarrow$ 30 billion operations per second

Conclusion: Exact implementation of point-to-point transfer functions is too expensive for real-time computation.

Possibility of a Physical Reverb Model

In a complete *physical model* of a room,

- sources and listeners can be moved without affecting the room simulation itself,
- *spatialized* (in 3D) stereo output signals can be extracted using a "virtual dummy head"

How expensive is a room physical model?

- Audio bandwidth = 20 kHz $\approx 1/2$ inch wavelength
- Spatial samples every $1/4$ inch or less
- A 12'x12'x8' room requires $>$ 100 million grid points
- A lossless 3D finite difference model requires one multiply and 6 additions per grid point $\Rightarrow$ 30 billion additions per second at $f_s = 50$ kHz
- A 100'x50'x20' concert hall requires more than *3 quadrillion operations per second*

Conclusion: Fine-grained physical models are too expensive for real-time computation, especially for large halls.

Perceptual Aspects of Reverberation

Artificial reverberation is an unusually interesting signal processing problem:

- "Obvious" methods based on physical modeling or input-output modeling are too expensive
- We do not perceive the full complexity of reverberation
- What is important perceptually?
- How can we simulate only what is audible?

Perception of Echo Density and Mode Density

- For typical rooms
  - Echo density increases as t^2
  - Mode density increases as f^2
- Beyond some time, the echo density is so great that a *stochastic process* results
- Above some frequency, the mode density is so great that a *random frequency response* results
- There is no need to simulate many echoes per sample
- There is no need to implement more resonances than the ear can hear

Proof that Echo Density Grows as Time Squared

Consider a single spherical wave produced from a point source in a rectangular room.

- Tesselate 3D space with copies of the original room
- Count rooms intersected by spherical wavefront

Proof that Mode Density Grows as Freq. Squared

The resonant modes of a rectangular room are given by

$$k^2(l,m,n) = k_x^2(l) + k_y^2(m) + k_z^2(n)$$

- $k_x(l) = l\pi/L_x = l$th harmonic of the fundamental standing wave in the x
- L_x = length of the room along x
- Similarly for y and z
- Mode frequencies map to a uniform 3D Cartesian grid indexed by (l,m,n)
- Grid spacings are π/L_x, π/L_y, and π/L_z in x,y, and z, respectively.
- Spatial frequency k of mode (l,m,n) = *distance* from the (0,0,0) to (l,m,n)
- Therefore, the number of room modes having a given spatial frequency grows as k^2

Early Reflections and Late Reverb

Based on limits of perception, the impulse response of a reverberant room can be divided into two segments

- *Early reflections* = relatively sparse first echoes
- *Late reverberation*—so densely populated with echoes that it is best to characterize the response *statistically*.

Early Reflections and Late Reverb

Similarly, the *frequency response* of a reverberant room can be divided into two segments.

- Low-frequency sparse distribution of resonant modes
- Modes packed so densely that they merge to form a *random frequency response* with regular statistical properties

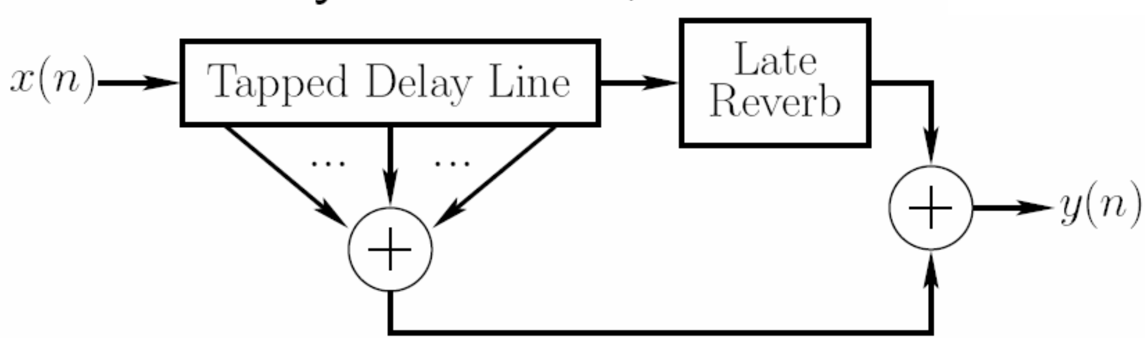
Perceptual Metrics for Ideal Reverberation

Some desirable controls for an artificial reverberator include

- $t_{60}(f)$ = desired reverberation time at each frequency
- $G^2(f)$ = signal power gain at each frequency
- $C(f)$ = "clarity" = ratio of impulse-response energy in early reflections to that in the late reverb
- $\rho(f)$ = *inter-aural correlation coefficient* at left and right ears

Perceptual studies indicate that reverberation time $t_{60}(f)$ should be independently adjustable in at least *three* frequency bands.

Reverb = Early Reflections + Late Reverb



- TDL taps may include lowpass filters (air absorption, lossy reflections)
- Several taps may be fed to late reverb unit, especially if it takes a while to reach full density
- Some or all early reflections can usually be worked into the delay lines of the late-reverberation simulation (transposed tapped delay line)

Early Reflections

The "early reflections" portion of the impulse response is defined as everything up to the point at which a statistical description of the late reverb takes hold.

- Often taken to be the first 100ms
- Better to test for *Gaussianness*
  - *Histogram* test for sample amplitudes in 10ms windows
  - *Exponential fit* to EDC (Prony's method, matrix pencil method)
  - *Crest factor* test (peak/rms)

Early Reflections

- Typically implemented using *tapped delay lines* (TDL) (suggested by Schroeder in 1970 and implemented by Moorer in 1979)
- Early reflections should be *spatialized* (Kendall)
- Early reflections influence *spatial impression*

Late Reverberation

Desired Qualities:

1. a smooth (but not too smooth) decay, and
2. a smooth (but not too regular) frequency response.

- Exponential decay no problem
- Hard part is making it *smooth*
  - Must not have "flutter," "beating," or unnatural irregularities
  - Smooth decay generally results when the echo density is sufficiently high
  - Some short-term energy fluctuation is required for naturalness

Late Reverberation

Desired Qualities:

- A smooth *frequency response* has no large "gaps" or "hills"
  - Generally provided when the mode density is sufficiently large
  - Modes should be spread out uniformly
  - Modes may not be too regularly spaced, since audible periodicity in the time-domain can result

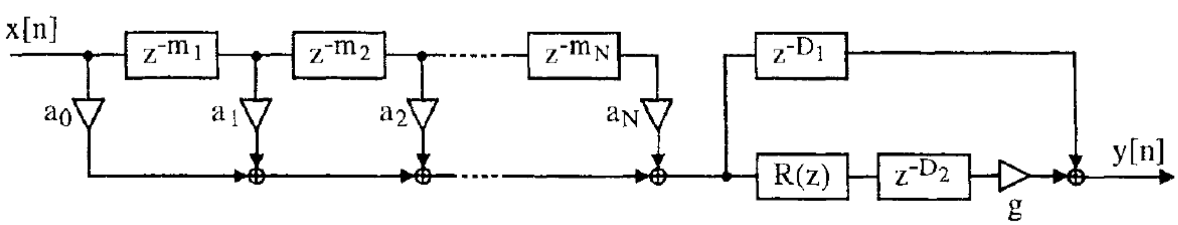
Late Reverberation

Desired Qualities:

- Moorer's ideal late reverb: *exponentially decaying white noise*
  - Good smoothness in both time and frequency domains
  - High frequencies need to decay faster than low frequencies
- Schroeder's rule of thumb for echo density in the late reverb is 1000 echoes per second or more
- For impulsive sounds, 10,000 echoes per second or more may be necessary for a smooth response

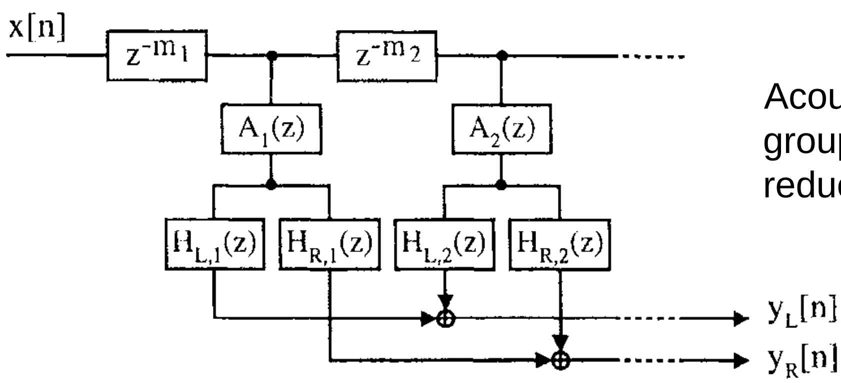
Moorer reverberator

- accounts for late reverberations by placing an IIR filter after the FIR filter (tapped delay line)



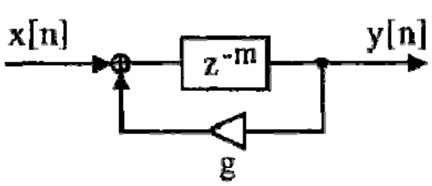
Binaural impulse response

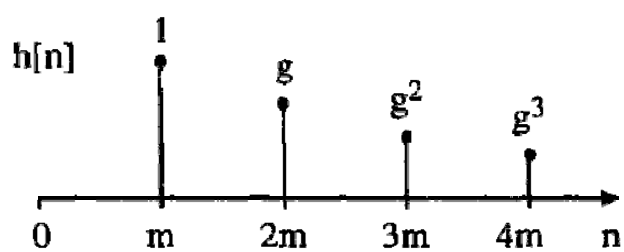
- Our sound perception is affected by our own body
  - Head Related Transfer Function (HRTF)

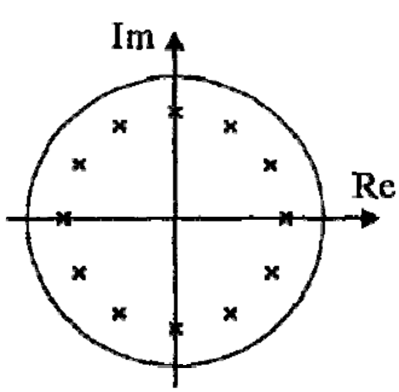


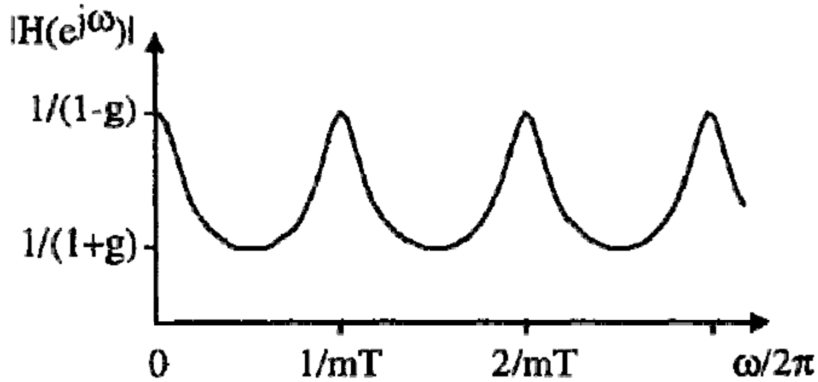
Acoustic paths can be grouped together to reduce cost

Comb filter

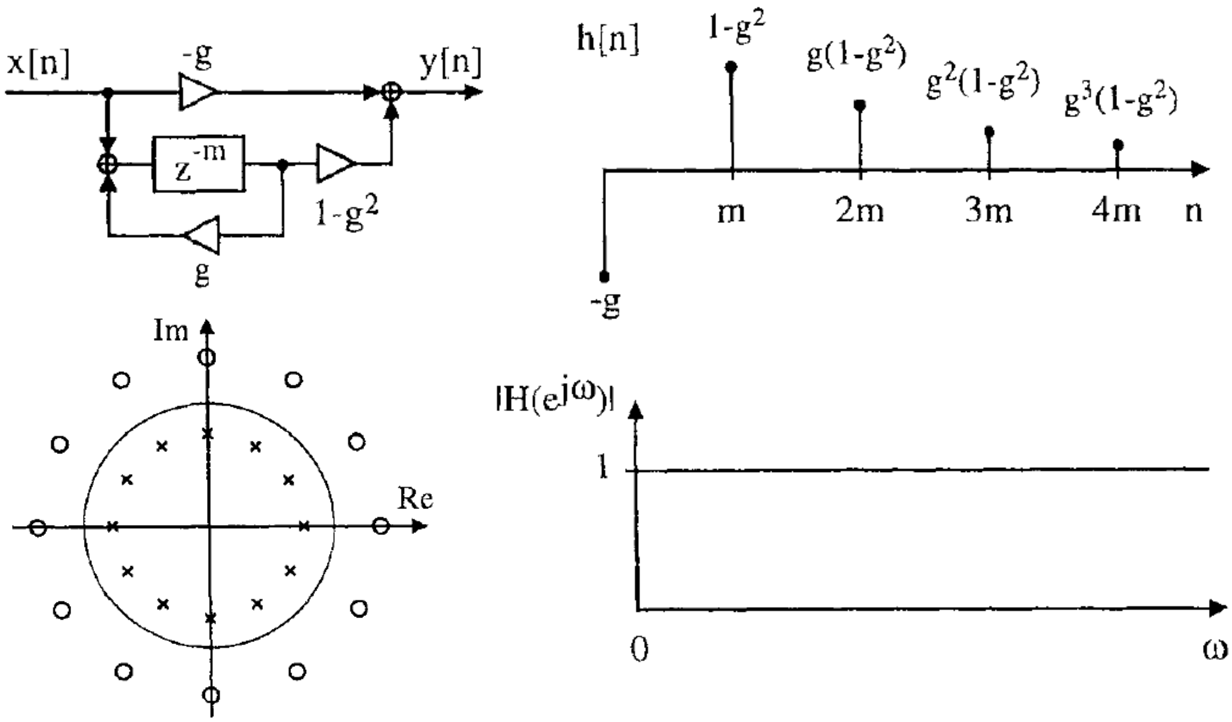








Allpass filter



Why Allpass?

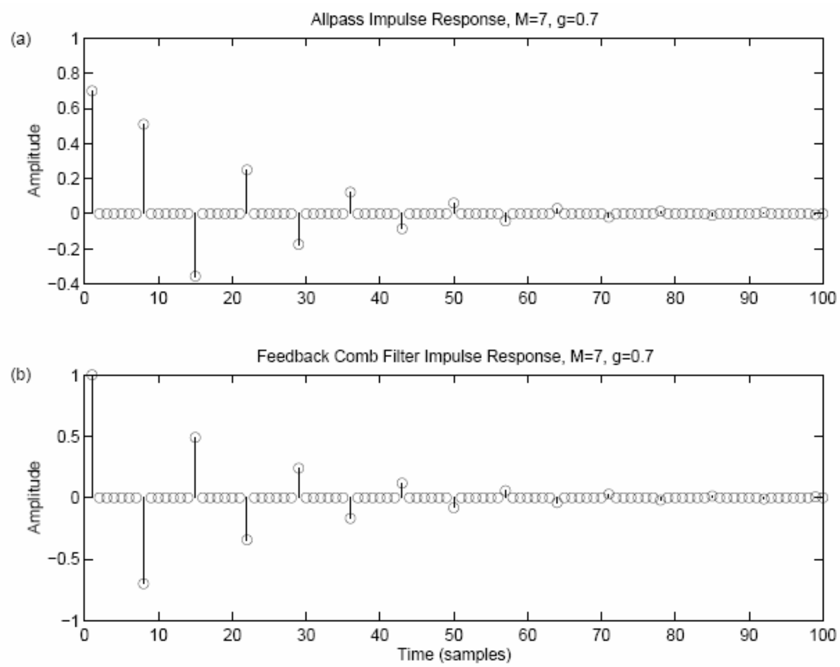
- Allpass filters do not occur in natural reverberation!
- "Colorless reverberation" is an idealization only possible in the "virtual world"
- **Perceptual factorization:** Coloration now orthogonal to decay time and echo density

Are Allpasses Really Colorless?

- Allpass impulse response only "colorless" when extremely short (less than 10 ms or so).
- Long allpass impulse responses sound like feedback comb-filters
- The difference between an allpass and feedback-comb-filter impulse response is *one echo*!

Are Allpasses Really Colorless?

Steady-state tones (sinusoids) really do see the same gain at every frequency in an allpass, while a comb filter has widely varying gains



(a) $H(z) = \frac{0.7+z^{-7}}{1+0.7z^{-7}}$ (b) $H(z) = \frac{1}{1+0.7z^{-7}}$

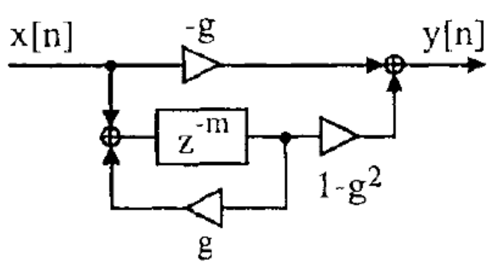
Comb filters and reverberation time

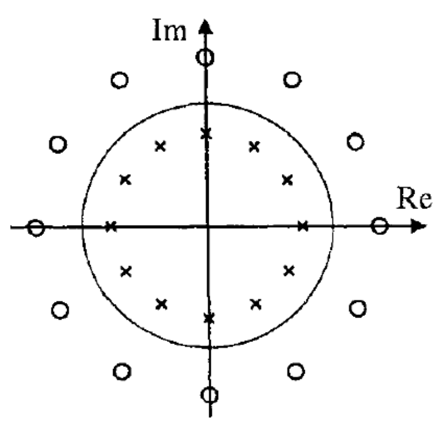
- The decay between successive samples in comb and allpass filters is described by the gain coefficient g_i
- In order for the comb filter's decay to correspond to a given reverberation time, we must have

$$\frac{20\log_{10}(g_i)}{m_i T} = \frac{-60}{T_r}$$

Combination of comb filters

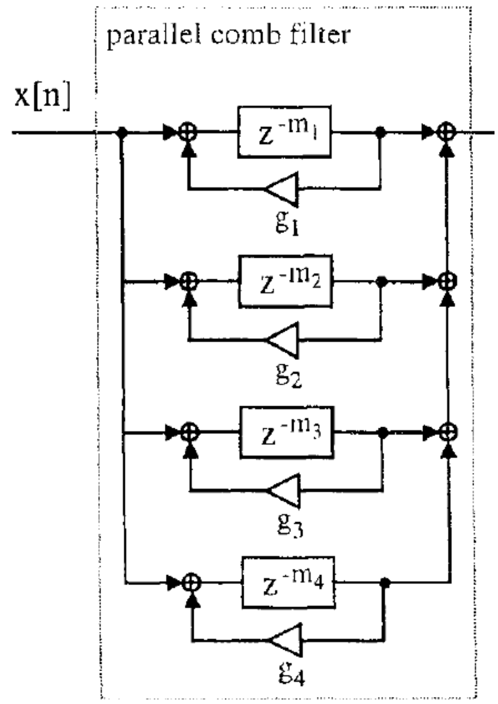
- Single comb filters do not provide sufficient echo density
- In order to improve the echo density, we need to combine multiple comb filters
  - Cascading comb filters corresponds to multiplying their transfer functions
  - Frequency peaks not shared by all comb filters are cancelled by multiplication





Combination of comb filters

- Better to place comb filters in parallel
  - Example



Parallel comb filters

- The poles of comb filters are given by

$$\prod_{i=1}^{N}(g_i - z^{m_i}) = 0$$

- The poles have the same magnitudes

$$\gamma_i = \sqrt[m_i]{g_i} = 10^{-3T/T_r}$$

- The modal density (No. of modes per Hz) is

$$D_m = \sum_{i=0}^{N-1} \tau_i = N \cdot \tau$$

Parallel comb filters

- Modal density turns out to be the same at all frequencies, unlike real rooms
- Above a threshold frequency, the modal density is constant
- The modal density of the comb filters is then set to the modal density above the threshold frequency

$$\sum_i \tau_i = D_m > D_f \approx \frac{T_{\max}}{4}$$

Parallel comb filters

- The echo density of the comb filters is approximatively given by

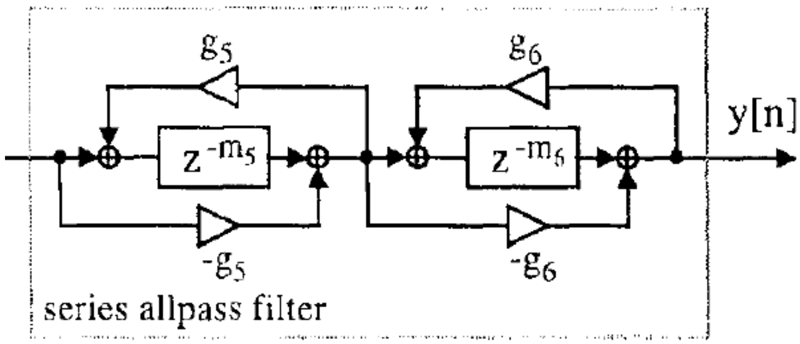
$$D_e = \sum_{i=0}^{N-1} \frac{1}{\tau_i} \approx \frac{N}{\tau}$$

- Relating echo density and modal density provides:

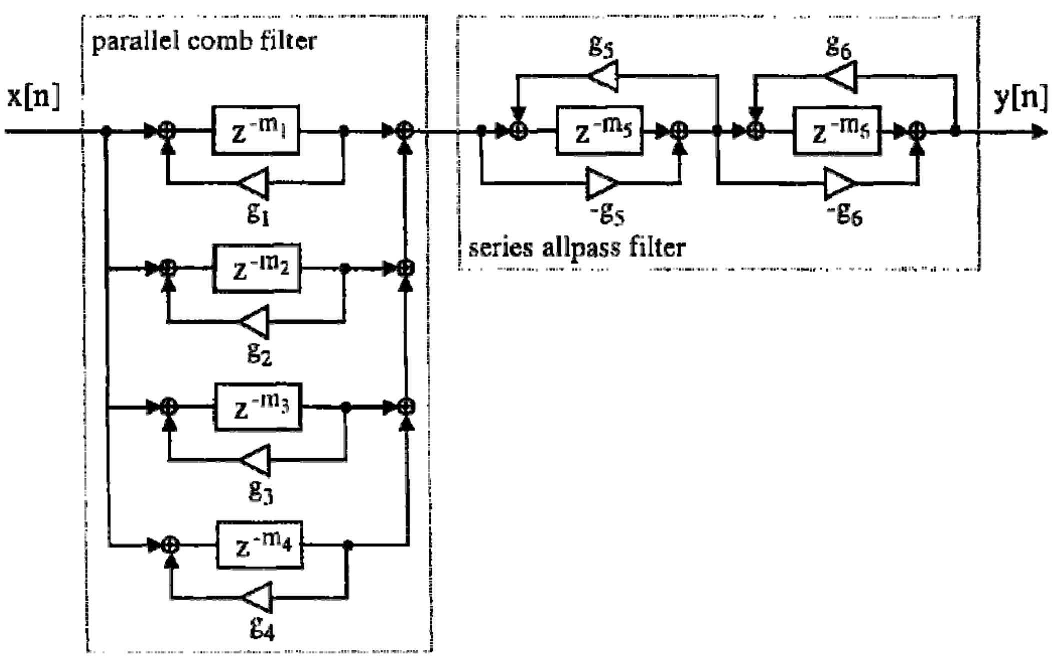
$$N \approx \sqrt{D_m D_e}$$

Combination of allpass filters

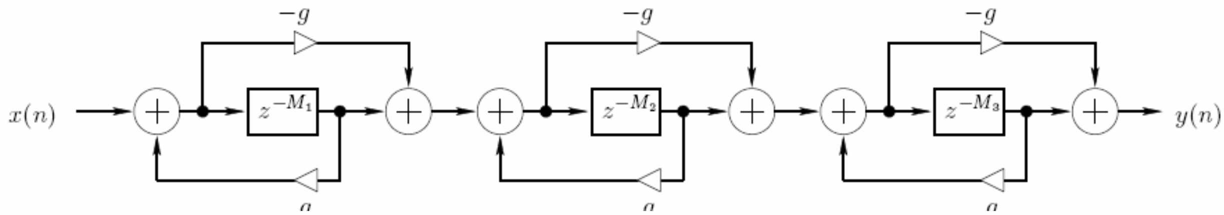
- Unlike comb filters, allpass filters must be cascaded
  - Multiplying freq. responses corresponds to adding phase responses



Schroeder's reverberator (1)



Schroeder Allpass Sections



- Typically, $g = 0.7$
- Delay-line lengths M_i mutually prime, and span successive orders of magnitude e.g., $1051, 337, 113$
- Allpass filters in series are allpass
- Each allpass *expands* each nonzero input sample from the previous stage into an entire infinite allpass impulse response
- Allpass sections may be called "*impulse expanders*", "impulse diffusers" or simply "diffusers"
- NOT a physical model of diffuse reflection, but single reflections are expanded into many reflections, which is qualitatively what is desired.

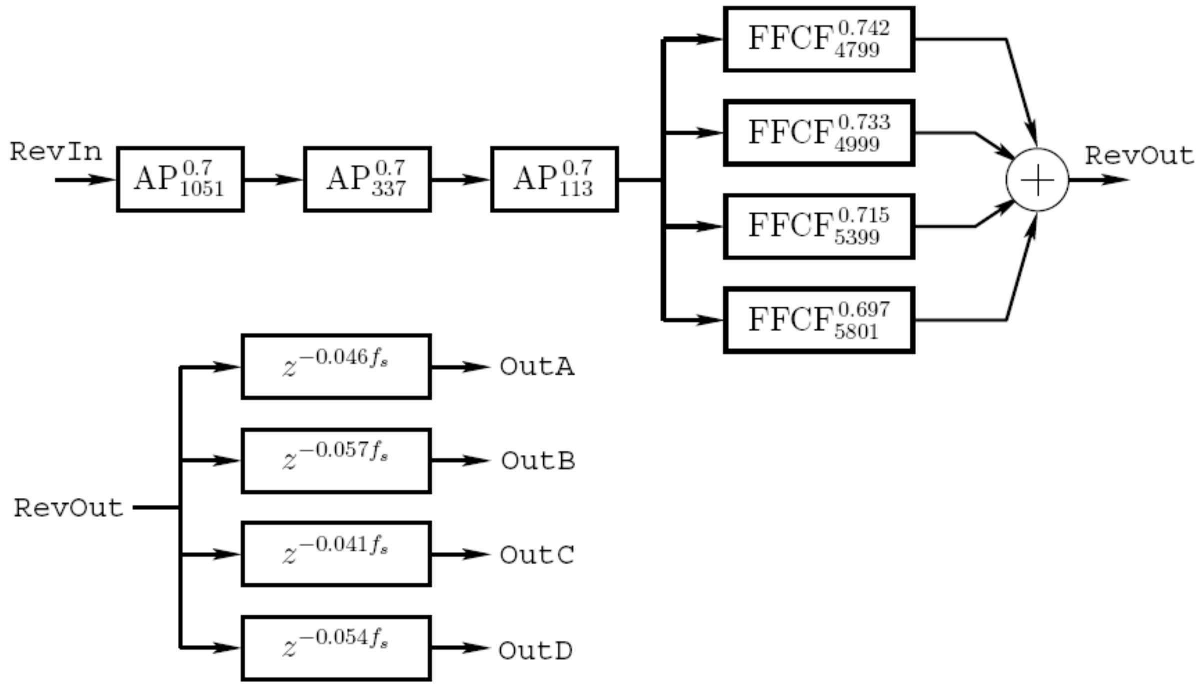
Schroeder's reverberator

- Delays of the comb and allpass filters are chosen so that the ratio of the largest and smallest delay is 1.5 (typically 30 and 45 ms)
- The gains g_i of the comb filters are chosen to provide a desired reverberation time T_r according to

$$g_i = 10^{-3m_i T/T_r}$$

- Allpass filters delays are set to 5 and 1.7 ms

A Schroeder Reverberator called JCRev



A Schroeder Reverberator called JCRev

- Three Schroeder allpass sections:

$$\mathsf{AP}^g_N \triangleq \frac{g + z^{-N}}{1 + g z^{-N}}$$

- Four feedforward comb-filters:

$$\mathsf{FFCF}^g_N \triangleq g + z^{-N}$$

- Schroeder suggests a progression of delays close to

$$M_i T \approx \frac{100 \text{ ms}}{3^i}, \quad i = 0, 1, 2, 3, 4.$$

- Comb filters impart distinctive coloration:
  - Early reflections
  - Room size
  - Could be one tapped delay line

A Schroeder Reverberator called JCRev

- Usage: Instrument adds scaled output to `RevIn`
- Reverberator output `RevOut` goes to four *delay lines*
  - Four channels *decorrelated*
  - *Imaging* of reverberation between speakers avoided
- For stereo listening, Schroeder suggests a *mixing matrix* at the reverberator output, replacing the decorrelating delay lines
- A mixing matrix should produce maximally rich yet uncorrelated output signals

Feedback Delay Networks…
…at a glance

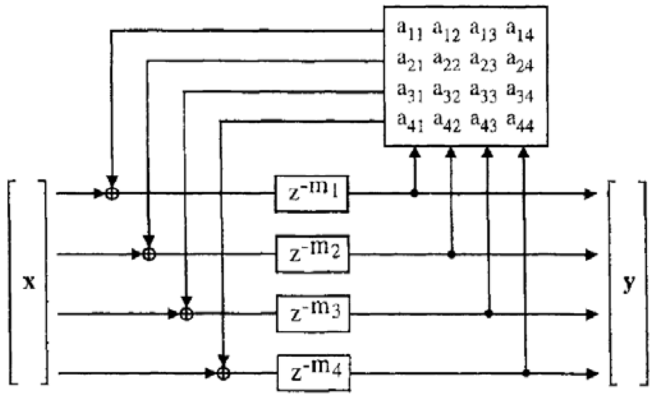
- Unitary matrix: definition
  - A matrix is unitary if :

$$\| \mathbf{M} \cdot \mathbf{u} \| = \| \mathbf{u} \|$$

  - We can also write that a matrix is unitary if

$$\| \mathbf{M}\mathbf{M}^T \| = \| \mathbf{M}^T\mathbf{M} \| = 1$$

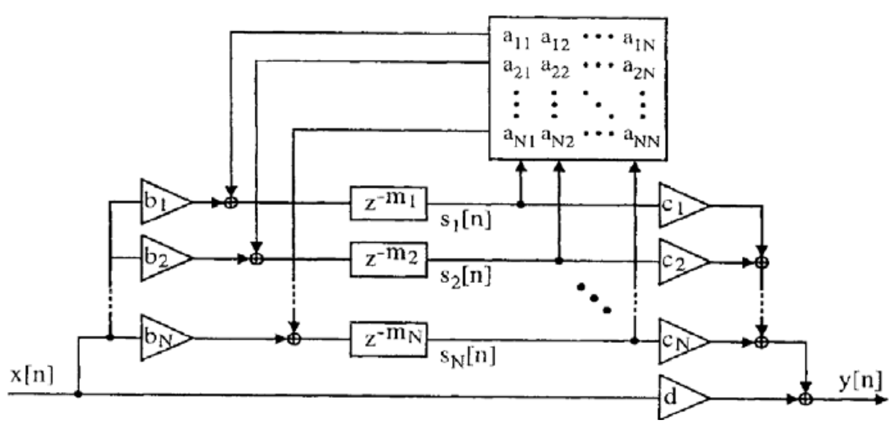
FDN



$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

- ❑Stability of the feedback loop is guaranteed if A = gM where M is an unitary matrix and |g|<1.
- ❑Outputs will be mutually incoherent: we can use the FDN to render the diffuse soundfield with a 4 loudspeaker system.
- ❑The early reverbeartions can be simulated by appropriately injecting the input signal into the delay lines.

Jot's reverberator



$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N1} & a_{42} & \cdots & a_{4N} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix}$$

Jot's reverberator

The input-output relation of Jot's reverberator is given by

$$y(z) = \mathbf{c}^T \mathbf{s}(z) + d x(z)$$

$$\mathbf{s}(z) = \mathbf{D}(z)[\mathbf{A}\mathbf{s}(z) + \mathbf{b}\, x(z)]$$

with $\mathbf{s}(z) = \begin{bmatrix} s_1(z) \\ \vdots \\ s_N(z) \end{bmatrix}$ and $\mathbf{D}(z) = \begin{bmatrix} z^{-m_1} & & 0 \\ & \ddots & \\ 0 & & z^{-m_N} \end{bmatrix}$

Jot's reverberator

- System transfer function:

$$H(z) = \frac{y(z)}{x(z)} = \mathbf{c}^T [\mathbf{D}(z^{-1}) - \mathbf{A}]^{-1} \mathbf{b} + d$$

- Zeros:

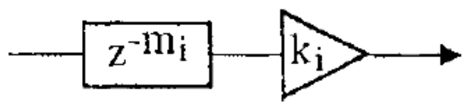
$$\det[\mathbf{A} - \frac{\mathbf{b}\mathbf{c}^T}{d} - \mathbf{D}(z^{-1})] = 0$$

- Poles:

$$\det[\mathbf{A} - \mathbf{D}(z^{-1})] = 0$$

Jot's reverberator

- Moorer noted that convolving exponentially decaying white noise with source signals produces a very natural sounding.
- As a consequence, by introducing absorptive losses into a lossless prototype, we should obtain a natural sounding reverberator.
- This is accomplished by associating a gain with each delay:



Jot's reverberator

- The logarithm of the gain is proportional to the length of the delay:

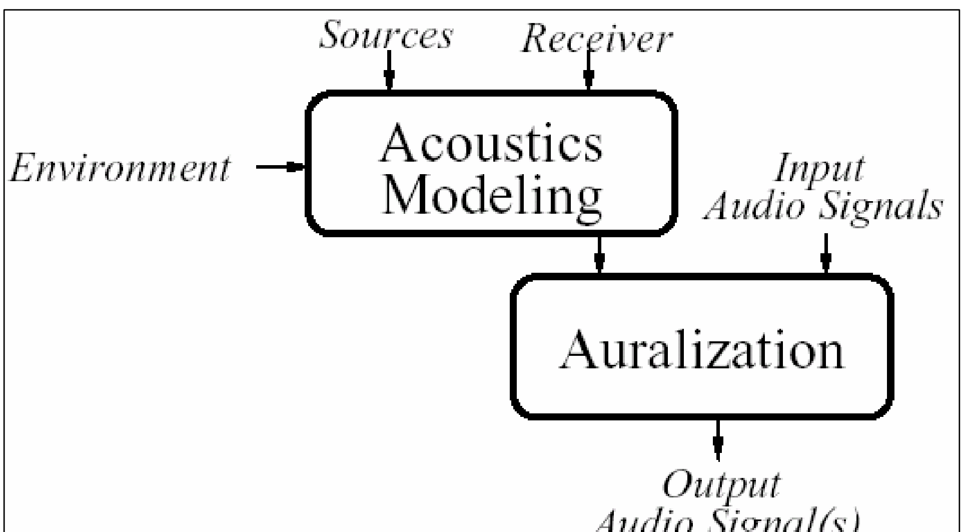
$$k_i = \gamma^{m_i}$$

- The above modification has the effect of replacing z with z/γ in the transfer function
- The lossless prototype response h[n] will be multiplied by an exponential envelope γ^n

Modeling the Environment

Modeling the environment

- Simulate reverberations due to environment



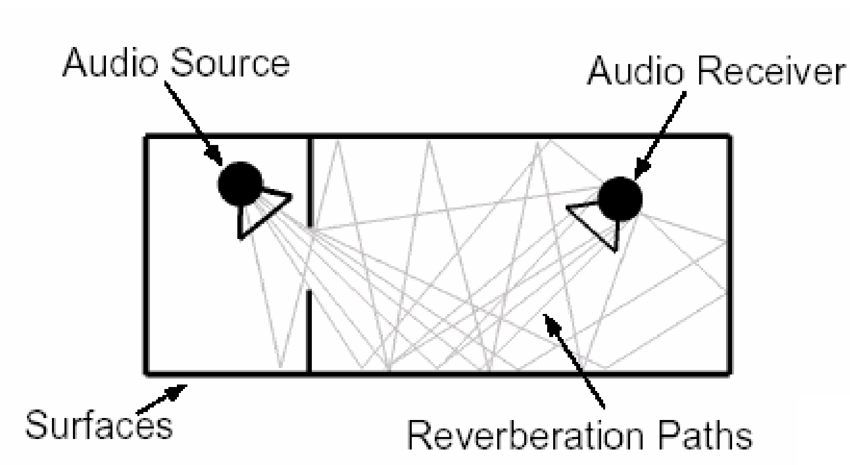
Motivations

Acoustical environment provides ...

- Sense of presence
- Comprehension of space
- Localization of auditory cues
- Selectivity of audio signals ("cocktail party effect")

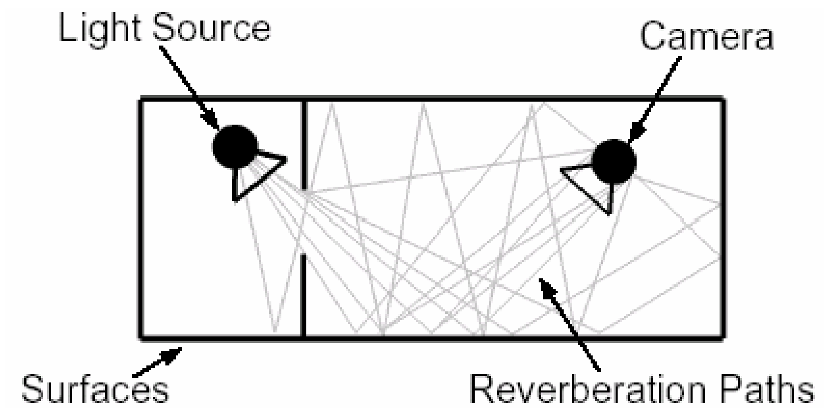
Geometric acoustic modeling

- Spatialize sound by computing reverberation paths from source to receiver



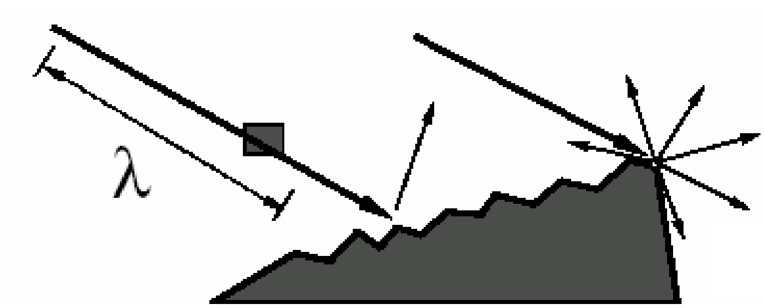
Similarities to Graphics

- Both model wave propagation



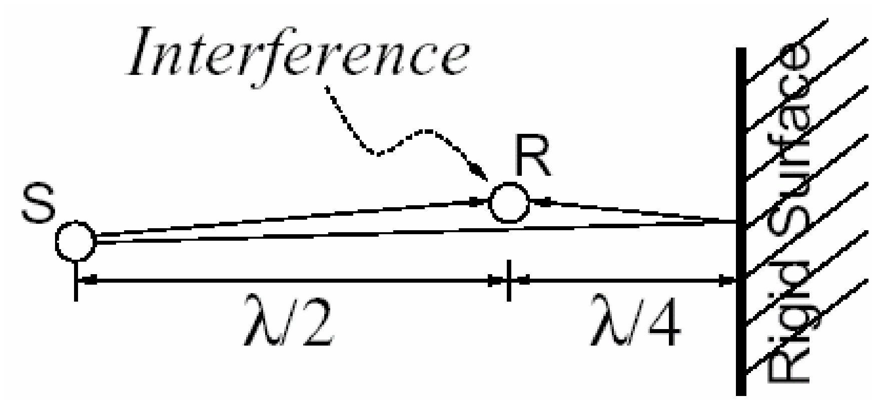
Differences from Graphics I

- Sound has longer wavelengths than light
  - Diffractions are significant
  - Specular reflections dominate diffuse reflections
  - Occlusions by small objects have little effect



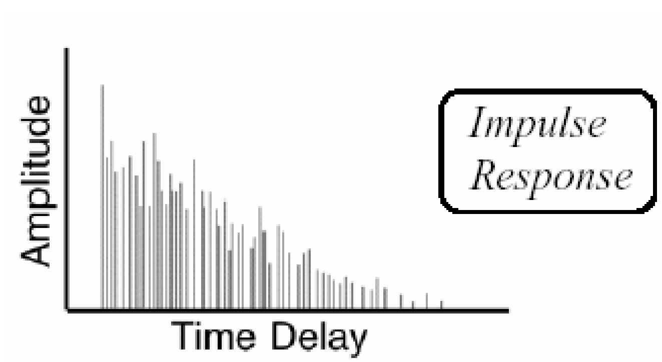
Differences from Graphics II

- Sound waves are coherent
  - Modeling phase is important



Differences from Graphics III

- **Sound travels more slowly than light**
  - Reverberations are perceived over time

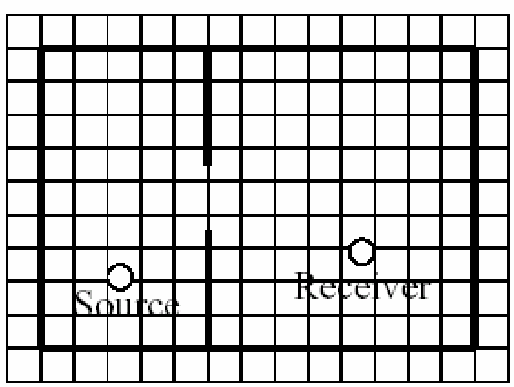


Overview of approaches

- **Finite element methods**
- **Boundary element methods**
- **Image source methods**
- **Ray tracing**
- **Beam tracing**

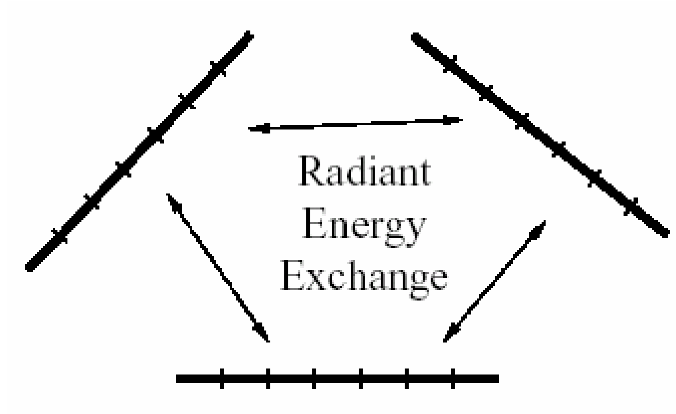
Finite element methods

- **Solve wave equation over grid-aligned mesh**



Boundary element methods

- **Solve wave equation over discretized surfaces**



Boundary Element Trade-offs

- **Advantages**
  - Works well for low frequencies
  - Simple formulation

$$B_i = E_i + \rho_i \sum B_j F_{ij}$$

Boundary Element Trade-offs

- **Disadvantages**
  - Complex function stored with each element
  - Form factors must model diffractions & specularities
  - Elements must be much smaller than wavelength

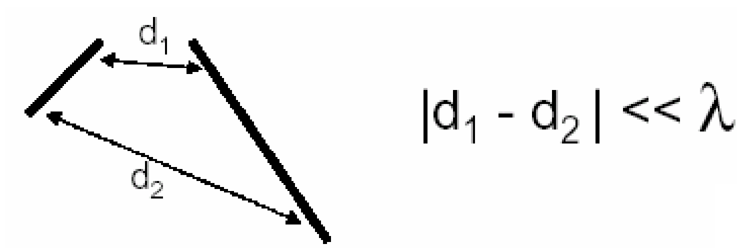


Image source methods

- **Consider direct paths from "virtual sources"**

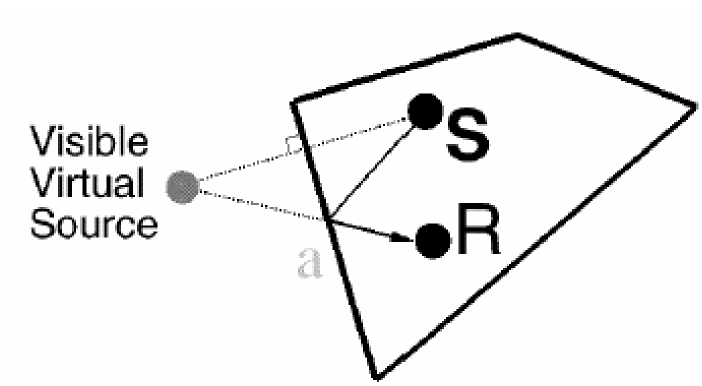


Image source trade-offs

- **Advantages**
  - Simple for rectangular rooms

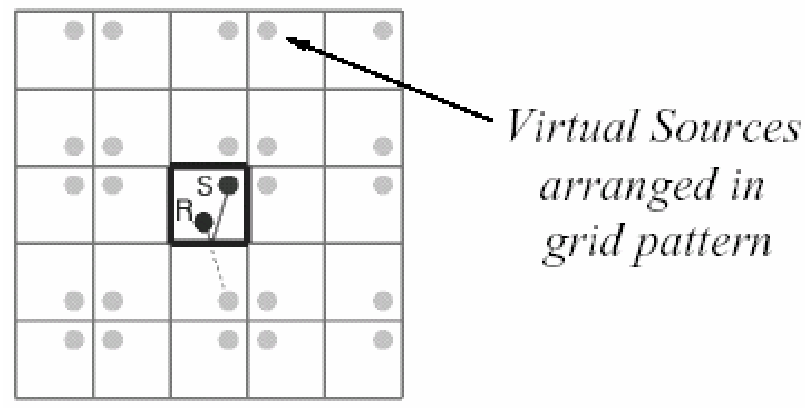
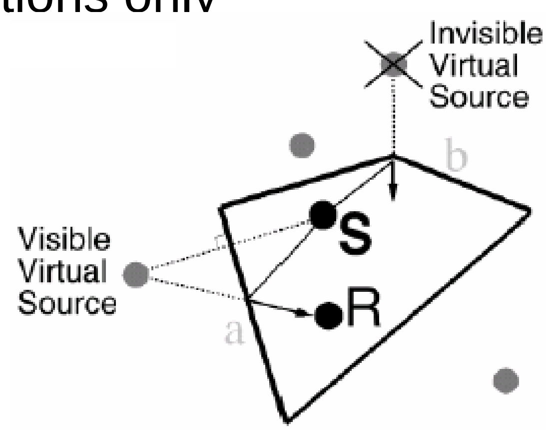


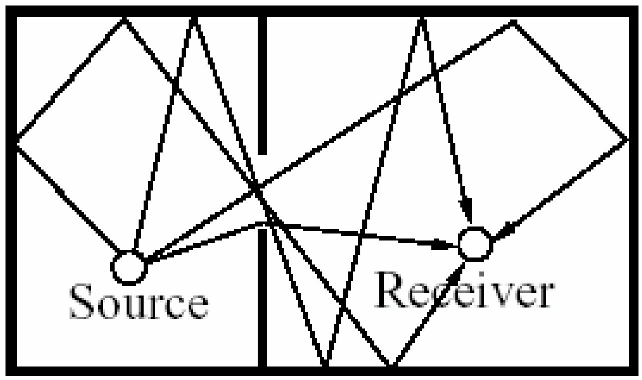
Image source trade-offs

- **Disadvantages**
  - O(n^r) visibility checks in arbitrary environments
  - Specular reflections only



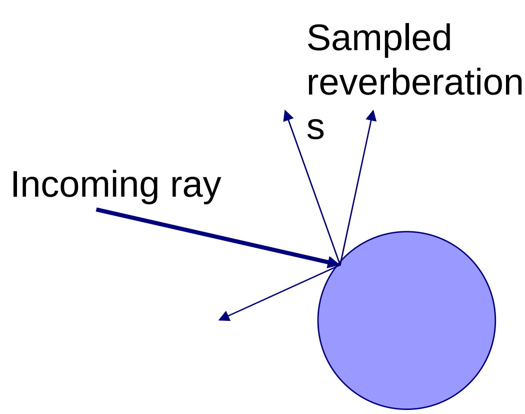
Path tracing

- **Trace paths between source and receiver**



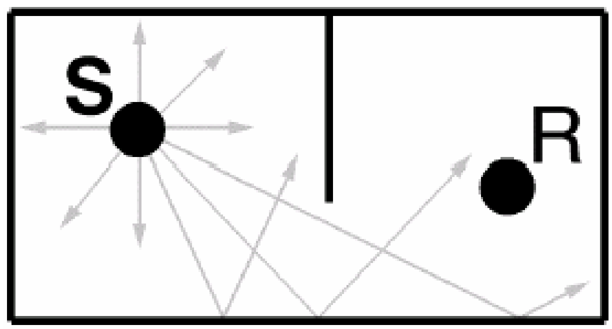
Path Tracing Trade-offs

- **Advantages**
  - Models all types of surfaces and scattering
  - Simple to implement



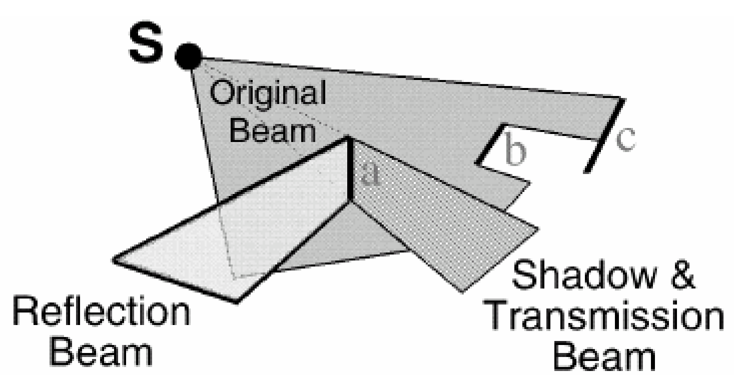
Path Tracing Disadvantages

- Disadvantages
  - Subject to sampling errors (aliasing)
  - Depends on receiver position



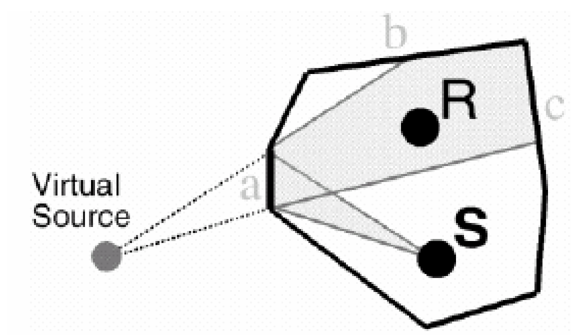
Beam Tracing

- Trace beams (bundles of rays) from source



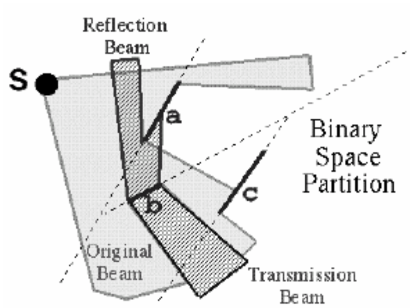
Beam Tracing Trade-offs

- Advantages
  - Takes advantage of spatial coherence
  - Predetermines visible virtual sources

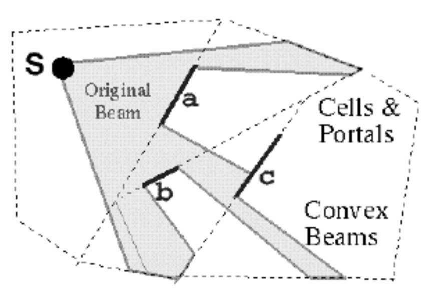


Beam Tracing Disadvantages

- Disadvantages
  - Difficult for curved surfaces or refractions
  - Requires efficient polygon sorting and intersection



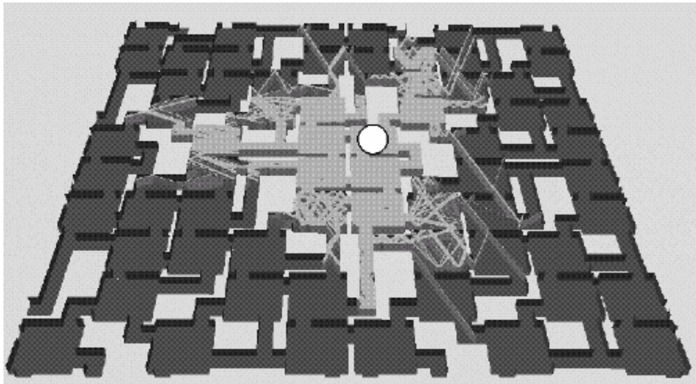
BSPs



Cell adjacency graphs

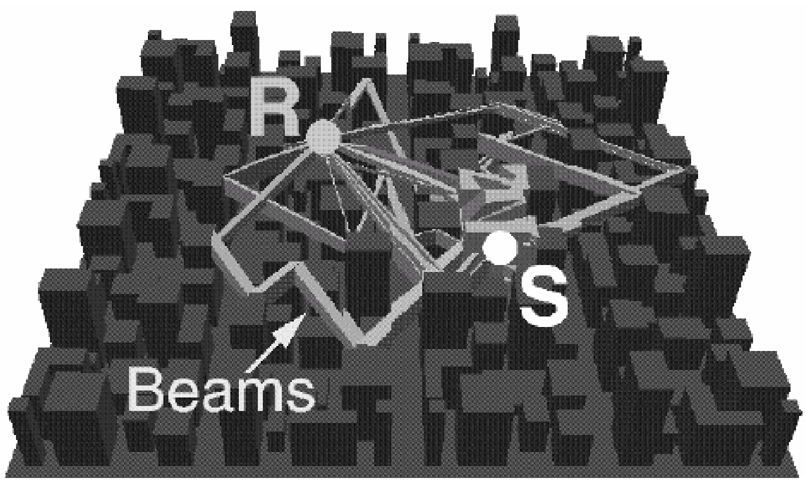
Complex 3D Environments

- Precompute beam tree for stationary source



Interactive Performance

- Lookup beams containing moving receiver



Summary

- **FEM/BEM**
  - best for low frequencies
- **Image source methods**
  - best for rectangular rooms (very common)
- **Path tracing**
  - best for high-order reflections (very common)
- **Beam tracing**
  - best for precomputation