MUMT 618: Week #3

1 Room Acoustics

The room in which we listen to sound has an important influence on what we hear. There are many methods and criteria for the evaluation and design of an acoustic space. Acoustic spaces can have many uses that may or may not share common desired acoustic characteristics.
In preparation for the subsequent discussion on artificial room reverberation, we consider common quantitative and qualitative measures used to describe sound propagation and reflection in rooms.
An excellent overview of geometric room acoustic modeling techniques is by

- Savioja, L. and Svensson, U. P. "Overview of geometrical room acoustic modeling techniques," in the Journal of the Acoustical Society of America, Vol. 138, No. 2, pp. 708–730, 2015.

1.1 Sound Propagation in a Room

- Sound waves travel at about 345 meters/second, so that the sound coming directly from a source within a large room will generally reach a listener after a time of anywhere from 0.01 to 0.2 seconds.
- Direct sound will decrease by 6 dB for each doubling of propagated distance (spherical spreading).
- Shortly after the arrival of the direct sound, a series of semi-distinct *early reflections* from various reflecting surfaces (walls and ceiling) will reach the listener.
- The early reflections should arrive within about 50 milliseconds (ms) for rapidly varying sounds, such as speech, or up to 80 ms for slowly varying music, in order that they not be heard as separate from the direct sound.
- The reflections that reach the listener after the early reflections are typically of lower amplitude and very closely spaced in time. These reflections merge into what is called the *reverberant sound* or *late reflections*.
- If the source emits a continuous sound, the reverberant sound builds up until it reaches an equilibrium level. When the sound stops, the sound level decreases at a more or less constant rate until it reaches inaudibility.
- For impulsive sounds (ex. a ballon pop), the reverberant sound begins to decay immediately.

1.2 Reverberation Time

- Reverberation time (RT60) is defined as the time it takes for the sound in a room to decrease by 60 dB from its original level, after the sound source is stopped. It generally varies with frequency.
- In a bare room, where all surfaces absorb the same fraction of the sound that reaches them, the theoretical reverberation time is proportional to the ratio of volume to surface area.

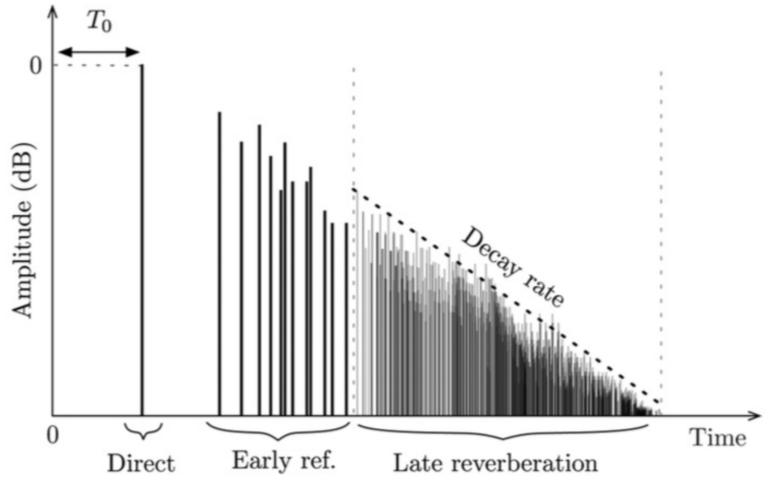

Figure 1: A theoretical room impulse response (from Välimäki et al. [2012]).

- In the late 1890s, Wallace Sabine proposed an empirical formula to calculate reverberation time from the dimensions (in meters) of a room as:

$$RT60 = 0.161V/A, \tag{1}$$

where V is the volume of the room in cubic meters and A is the effective "total absorption area."

- The "total absorption" area is calculated as the sum of all surface areas in the room, each multiplied by its respective absorption coefficient for a particular frequency.
- A low absorption coefficient indicates a more reflective material.
- Typical frequencies for which absorption coefficients are determined are 125, 250, 500, 1000, 2000, and 4000 Hz.
- Air contributes a substantial amount to the absorption of high frequency sound and can be accounted for (in a simplistic way) in the formula for reverberation time as RT60 = 0.161 $V/(A + mV)$, where m is a constant that varies with air temperature, humidity, and frequency.
- More accurate estimations for $RT60$ are nowadays found from modeled or measured room impulse responses.
- The optimum reverberation time of a room is dependent on the use for which it is designed (speaking, chamber music, opera, orchestral music, organ, etc.).
- Modern technologies, such as the Meyer Sound's Constellation system, can extend the apparent reverberation time of a room. If one wishes to lower the $RT60$ of a room, curtains or other acoustic absorbers can be introduced.

1.3 Assessing Room Characteristics

- Common requirements for good acoustics in halls include:
  - *adequate loudness:* the sound source(s) can be heard throughout the room;
  - *uniformity:* the sound level is similar thoughout the space;
  - *clarity:* the sound is easily perceived or understood;
  - *reverberance or liveness:* the listener should be bathed in sound from all directions, though still able to localize the sound source;

- *freedom from echoes:* no strong reflections or resonances;
- *minimal background noise:* the room should be free from undesirable sounds, such as from fans, electronics, mechanical systems and sounds coming from outside the room.

- Some quantifiable measures of room characteristics include:
  - *RT60:* quantifies aspects of *reverberance or liveness*;
  - *initial time delay gap (ITDG):* the time interval between the arrival of the direct sound and the first reflection at the listener;
  - *early decay time (EDT):* initial rate of sound decay in a room, based on the early reflections;
  - *clarity (C_{80}):* the difference (in dB) of the sound energy received at a listener in the first 80 milliseconds minus the (late) reverberant energy (all remaining sound energy);
  - *brilliance:* related to high-frequency damping, measured by EDT_{2000} / EDT_{mid} and EDT_{4000} / EDT_{mid}, where EDT_{mid} is the average of EDT values at 500 and 1000 Hz;
- Subjective aspects used when assessing rooms include:
  - *intimacy or presence:* the impression of the size of a space, evaluted using ITDG, more intimate halls (ITDG ¡ 20 ms) are generally preferred over wider, more remote halls;
  - *spaciousness: auditory source width (ASW):* results from the presence of lateral early reflections;
  - *spaciousness: listener envelopment (LEV):* primarily influenced by the late reverberant field of the room impulse response;
  - *warmth:* provided to a music hall via a slight increase in low frequency reverberation.
- A greater value of C_{80} gives music a sensation of *definition*, while decreased definition adds "fullness of tone" (or "muddiness" when excessive). In a study of 22 European concert halls, less definition was preferred.
- Optimum reverberation time is a compromise between clarity (requiring short reverberation time), sound intensity (requiring a high reverberant level), and liveness (requiring a long reverberation time).
- A two stage decay can satisfy two conflicting, but desirable, attributes of music. A short EDT provides "clarity" and a long RT provides liveness to music.

1.4 Room Acoustic Modeling

-
- Computational approaches to model the response of particular rooms (defined by complete geometrical specifications) are generally grouped into two categories:
  - Numerical Acoustic Techniques (wave-based, FEM, BEM, FDTD, DWG, ...)
  - Geometric Acoustic (GA) Techniques (ray tracing, image-source, beam tracing, surface-based, ...)
- The wave-based approaches generally require significantly more computational resources, though ray-tracing techniques can be quite demanding as well.
- In GA, wave properties are neglected and sound is assumed to propagate as rays, which is more valid when wavelengths are short compared to surface dimensions and overall dimensions of the space (at high frequencies) but less accurate at lower frequencies.
- GA approaches can be energy- or pressure-based, the former resulting in energy-time (or echogram) responses and the latter resulting in impulse responses.

- The *image-source (IS) method* is a ray-based model for room reflections that assumes specular reflections from large, smooth surfaces (as illustrated in Fig. 2). "Virtual sources" are determined at mirror image locations with respect to reflecting surfaces. Once the virtual sources are determined, propagation distances can be easily calculated from two- or three-dimensional Euclidean geometry. Note that path validity checks are necessary.

- The IS method can be used to efficiently compute early reflections but the technique becomes challenging to manage at higher orders.

- Various approaches to model reflections from curved or non-smooth surfaces, as well as diffraction at finite surface edges have been proposed in GA.

- In ray-tracing techniques, the main prinicple is to randomly cast rays from a sound source and to register valid paths to a listener. Source directivity patterns can easily be accommodated by weighting the ray distribution. Diffuse reflection properties can be modeled by casting numerous reflected rays at a boundary, as illustrated on the right side of Fig. 3.

- Another approach in GA is beam tracing, in which a volumetric region is tracked rather than a single ray. There are two different branches of beam tracing, one being similar to ray tracing (Fig. 4) and the other being more related to the image-source method (Fig. 5).

- Popular commercial room modeling software systems, such as CATT-Acoustic and Odeon, make use of GA modeling techniques.

2 Artificial Reverberation

There are a variety of approaches to synthesizing the effect of a reverberant space. Those based on direct measurement of a particular room response (convolution techniques) tend to be less extensible but gaining in popularity. The use of three-dimensional wave-based modeling techniques is limited by computational requirements. Most work in artifically simulating reverberation has been on "physically- and perceptually-informed" techniques that seek to create parametrically-controllable systems. These models can produce very good reverberant responses though they generally cannot be made to correlate with actual room measurements.
Two excellent overviews of artificial reverbation developments are given by

- Gardner, W. G. "Reverberation Algorithms," in Applications of Signal Processing to Audio and Acoustics, M. Kahrs and K. Brandenburg, Eds., Kluwer Academic, Norwell, MA, 1997.

- Välimäki, V., Parker, J., Savioja, L. Smith, J. O., Abel, J. "Fifty Years of Artificial Reverberation, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 5, July 2012.

Artificial reverberation methods can generally be grouped into the following categories:

1. convolutional: an input signal is convolved with a recorded or estimated impulse response of an acoustic space;

2. computational acoustics: an input signal is fed into a system that simulates acoustic propagation in a modeled geometry;

3. delay networks: an input signal is delayed, filtered and fed back along a number of simulated propagation paths to achieve a parameterized reverberation characteristic.

This section focuses primarily on the use of delay networks for articial reverberation.

2.1 Convolutional or Transfer-Function Approaches

- An acoustic space can be approximated as a linear, time-invariant system, in which case one can measure its impulse response and use that response as a filter to reproduce the effect of playing a given audio signal in that space.
- A key limitation of that approximation is that there are no changes in the number of listeners and sound sources, nor their positions.
- The simulation of room reverberation using a convolutional approach ideally involves two transfer functions per sound source per listener (one for each ear). The tranfer functions or filter representations will change if anything in the room changes.
- For the three source, one listener setup depicted in Fig. 7, the output signals would be computed via six convolutions:

$$y_i[n] = \sum_{j=1}^{3} s_j * h_{ij}[n] = \sum_{j=1}^{3} \sum_{m=0}^{M_{ij}} s_j[m] h_{ij}[n-m], \qquad i = 1, 2$$

  where $h_{ij}[n]$ is an FIR filter representation of the impulse response from source j to ear i and M_{ij} is the length of the filter.
- For impulse responses of one second (appropriate when the $t_{60} = 1$ second) and a sample rate $f_s = 50$ kHz, each filter would require 50,000 multiplies and additions per sample or 2.5 billion multiply-adds per second. For three sources and two listening points (ears), this corresponds to 30 billion operations per second.
- Efficiencies can be gained by performing the computations in the frequency domain and by making other simplifying assumptions.
- With the increasing availability and affordability of multi-processor computing platforms, convolutional approaches for artificial reverberation are becoming more popular and available. From a business perspective, companies making such systems can also profit by selling room responses (since recording impulse responses requires significant expertise and specialized equipment).
- Generally, it can be challenging to dynamically modify impulse responses to support flexible control schemes (such as shortening or lengthening reverb times) and still maintain good audio quality.

2.2 Perceptual Approach to Reverberation Simulation

- Based on perceptual limits, the impulse response of a reverberant space can be divided into two segments:
  - The beginning of the impulse response consists of distinct, relatively sparse, *early reflections*.
  - The remainder of the impulse response, called the *late reverberation*, consists of densely-packed echoes that become impossible to distinguish in time.
  - This region of high echo density (which increases as t^2) can be approximated by a random time distribution.
- The frequency response of a reverberant space can likewise be divided into two segments:
  - The low-frequency region consists of a relatively sparse distribution of resonant modes.
  - Higher-frequency modes are packed so densely that they are best characterized by a random frequency distribution with certain statistical properties.
- Parametric controls for an artificial reverberator would ideally include:
  - $t_{60}(f)$ in at least three frequency bands

- $G^2(f)$ = signal power gain
- $C(f)$ = "clarity" (ratio of impulse response energy in early reflections to that in the late reverb section)
- *inter-aural correlation coefficient* at the left and right ears

2.3 Early Reflections

- Early reflections, within the first 100 milliseconds or so, are typically implemented using *tapped delay lines* (suggested by Schroeder [1970] and implemented by Moorer [1979]).
- Early reflections should be calculated for a given geometry and spatialized.
- The delay-line tap outputs should be scaled in proportion to propagation distance.
- Most room surfaces are not perfectly flat, resulting in diffuse scattering. Thus, attempts to exactly reproduce the response of a given room via techniques such as *ray tracing* are generally unsuccessful.

2.4 Late Reverberation

- A good late reverberation should have a smooth decay and a smooth frequency response.
- Some fluctuation in the short-term energy is needed to achieve a natural sound [Dattorro, 1997, Blesser, 2001].
- Moorer's ideal late reverb: exponentially decaying white noise. But it would be better to say exponentially decaying "colored" noise, since the high-frequency energy should decay faster than the low-frequency energy.
- Schroeder [1962] suggested the use of parallel comb filters and cascaded allpass filters to synthesize reverberation.
- Allpass filters produce frequency-dependent time shifts, which help diffuse the sound. For this reason, Schroeder allpass sections are sometimes referred to as *impulse expanders* or *impulse diffusers*.
- The gain values are typically set around $g = 0.7$. The delay-line lengths M_i should be mutually prime and span successive orders of magnitude.
- The impulse response, calculated with the Matlab script allpass.m, of three cascaded Schroeder allpass sections is shown in Fig. 10.
- The feedback comb filters provide coloration and the delay-line lengths are set to mutually prime values.
- The STK classes `PRCRev`, `JCRev`, and `NRev` implement Schroeder reverberators of various complexities. In particular:
  - `PRCRev` implements two series allpass units and two parallel comb filters.
  - `JCRev` implements three series allpass units, four parallel comb filters, and two decorrelation delay lines in parallel at the output.
  - `NRev` implements six parallel comb filters, three series allpass units, a lowpass filter, another allpass filter in series, followed by two allpass filters in parallel at the output.

2.5 Feedback Delay Networks (FDN)

- Figure 12 illustrates an example FDN reverberator using three delay lines proposed by Jot [1992].
- An FDN can be seen as a *vector feedback comb filter*, with N feedback "channels" (N=3 in Fig. 12).
- The "mixing matrix" provides diffusion by "scattering" energy amongst the N channels. Assuming decay control is handled by the g_i coefficients, this matrix should be "lossless".
- To achieve frequency-dependent decay control, the g_i coefficients must be replaced by low-order digital filters.
- The "tonal correction" filter $E(z)$ is a low-order filter that serves to equalize modal energy across the spectrum.
- The delay-line lengths are generally chosen to be mutually prime. System "tuning" remains a manual, trial and error process.

2.6 Feedback ("Mixing") Matrix Stability

- A "3-channel" FDN feedback matrix can be represented as:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

- The inner loop calculations of the FDN shown in Fig. 12 can then expressed as:

$$\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{bmatrix} = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} x_1(n-M_1) \\ x_2(n-M_2) \\ x_3(n-M_3) \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} u(n)$$

  and the loop output given by

$$v(n) = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} x_1(n-M_1) \\ x_2(n-M_2) \\ x_3(n-M_3) \end{bmatrix}$$

- These expressions can also be written in frequency-domain vector notation as

$$\begin{aligned} \mathbf{X}(z) &= \mathbf{GMD}(z)\mathbf{X}(z) + \mathbf{BU}(z) \\ \mathbf{V}(z) &= \mathbf{CD}(z)\mathbf{X}(z) \end{aligned}$$

  where

$$\mathbf{D}(z) \triangleq \begin{bmatrix} z^{-M_1} & 0 & 0 \\ 0 & z^{-M_2} & 0 \\ 0 & 0 & z^{-M_3} \end{bmatrix}$$

- The matrix $\mathbf{A} = \mathbf{GM}$ is called the *state transition matrix*. $\mathbf{G}$ is typically a diagonal matrix of lowpass filters, each having gain no greater than 1.
- Stability of the FDN is assured when the norm of the state vector $\mathbf{x}[n]$ decreases over time when the input signal is zero:

$$\|\mathbf{x}(n+1)\| < \|\mathbf{x}(n)\|,$$

  for all $n \geq 0$, where

$$\mathbf{x}(n+1) = \mathbf{A} \begin{bmatrix} x_1(n-M_1) \\ x_2(n-M_2) \\ x_3(n-M_3) \end{bmatrix}.$$

- Stable feedback matrices can thus be parameterized in terms of $\mathbf{A} = \mathbf{G}\mathbf{M}$, where $\mathbf{M}$ is any *orthogonal matrix* and $\mathbf{G}$ is a diagonal matrix having entries less than 1 in magnitude.
- A feedback matrix $\mathbf{M}_N$ is lossless if and only if its eigenvalues have modulus 1 and its N eigenvectors are linearly independent.

2.7 Householder Feedback Matrices

- One choice of feedback matrix $\mathbf{M}_N$ for FDNs is a specific *Householder reflection* proposed by Jot [1992]:

$$\mathbf{M}_N = \mathbf{I}_N - \frac{2}{N}\underline{\mathrm{u}}_N\underline{\mathrm{u}}_N^T$$

  where $\underline{\mathrm{u}}_N^T = [1, 1, \ldots, 1]$ is the specific vector about which the input vector is reflected in N-dimensional space.
- In addition to being lossless and not requiring any multiplies when N is a power of 2 (for fixed-point implementations), the Householder matrix is attractive because the feedback matrix-times-channel-vector operation can be computed with only $2N - 1$ additions (by first forming $\underline{\mathrm{u}}_N^T$ times the input vector, applying the scale factor $2/N$, and subtracting the result from the input vector).

2.8 Delay Lengths

- FDN delay-line lengths are generally chosen to be mutually prime, which maximizes the psuedo-random behavior of the system.
- A rough guide to the average delay-line length is the "mean free path" of the desired reverberant environment, which is defined as the average distance a ray of sound travels before it encounters a reflecting obstacle.
- The mean free path can be approximated as $\bar{d} = 4V/S$, where V is the total volume and S is the total surface area enclosing the space.
- The desired *modal density* can guide the determination of the total sum of the delay line lengths M. Schroeder suggests a modal density of 0.15 modes per Hz for a 1 second t_{60}. This can be generalized to $M \geq 0.15 t_{60} f_s$.

2.9 Reverberation Time and Energy Equalization

- Reverberation time is controlled by lowpass filters implemented within each feedback channel (the $\mathbf{G}$ matrix discussed above).
- A lowpass filter in series with a length M_i delay line should approximate $H_i(z) = G^{M_i}(z)$, where $G(z)$ is the ideal per-sample decay filter. In terms of a desired t_{60}, this implies

$$\left|H_i(e^{j\omega T_s})\right|^{\frac{t_{60}(\omega)}{M_i T_s}} = 0.001.$$

- Jot proposes first-order filters of the form:

$$H_i(z) = g_i \frac{1 - a_i}{1 - a_i z^{-1}},$$

  where g_i is set to give a desired reverberation time at dc and a_i determines the reverberation time at high frequencies.

- From the expression above, we find
$$g_i = 10^{-3M_i T_s / t_{60}(0)}$$
and from [Jot and Chaigne, 1991]
$$a_i = \frac{\ln(10)}{4} \log_{10}(g_i) \left(1 - \frac{1}{\alpha^2}\right),$$
where
$$\alpha \triangleq \frac{t_{60}(\pi/T_s)}{t_{60}(0)}.$$

- Finally, overall energy through the FDN can be equalized (to account for the effects of the loop filters) by applying a tonal correction filter at the output. For the first-order loop filters described here, a good equalization filter is given by
$$E(z) = \frac{1 - bz^{-1}}{1 - b},$$
where
$$b = \frac{1 - \alpha}{1 + \alpha}.$$

References

B. Blesser. Interdisciplinary synthesis of reverberation viewpoint. *Journal of the Audio Engineering Society*, 49:867–903, 2001.

J. Dattorro. Effect design. *Journal of the Audio Engineering Society*, 45:660–684, 1997.

J.-M. Jot. *Etude et Réalization d'un Spatialisateur de Sons par Modèles Physiques et Perceptifs*. PhD thesis, French Telecom, Paris, 1992.

J.-M. Jot and A. Chaigne. Digital delay networks for designing artificial reverberators. *Audio Engineering Society Convention*, Feb. 1991.

J. A. Moorer. About this reverberation business. *Computer Music Journal*, 3(2):13–18, 1979.

L. Savioja and U. P. Svensson. Overview of geometrical room acoustic modeling techniques. *Journal of the Acoustical Society of America*, 138(2):708–730, 2015.

M. R. Schroeder. Natural-sounding artificial reverberation. *Journal of the Audio Engineering Society*, 10(3): 219–223, 1962.

M. R. Schroeder. Digital simulation of sound transmission in reverberant spaces (part 1). *Journal of the Acoustical Society of America*, 47(2):424–431, 1970.

V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel. Fifty years of artificial reverberation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20:1421–1448, 2012.

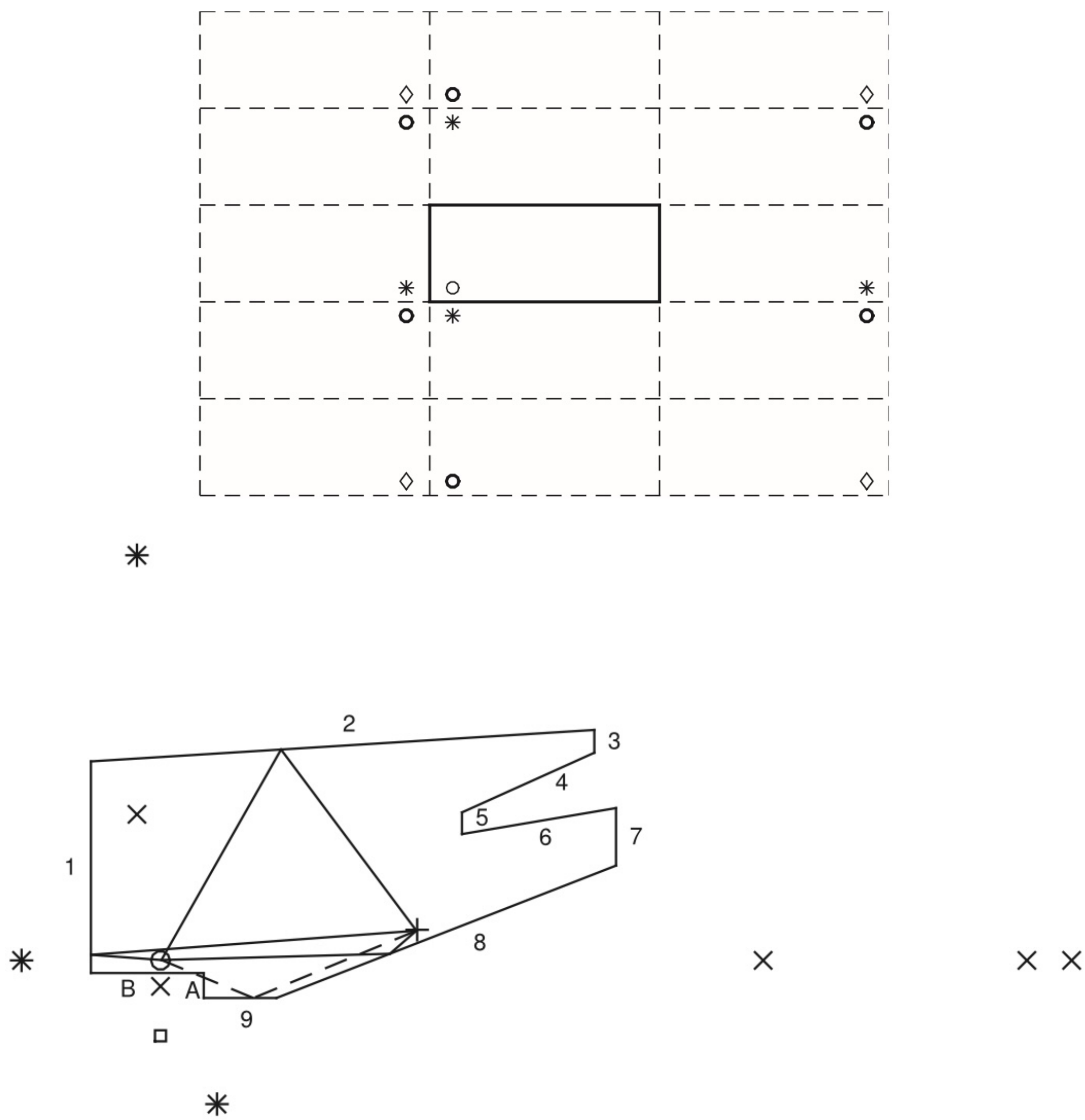

Figure 2: Image-source representation of a shoebox-shaped room (left) and a more complex geometry (right), from [Savioja and Svensson, 2015]. Original sources are designated by $\circ$, first-order ISs with $*$, second-order ISs with **o**, and third-order ISs with $\diamond$. For the right figure, the receiver is designated with $+$, ISs with reflection points outside the polygon by $\times$ and an IS with a valid reflection point but obstructed path with $\square$.

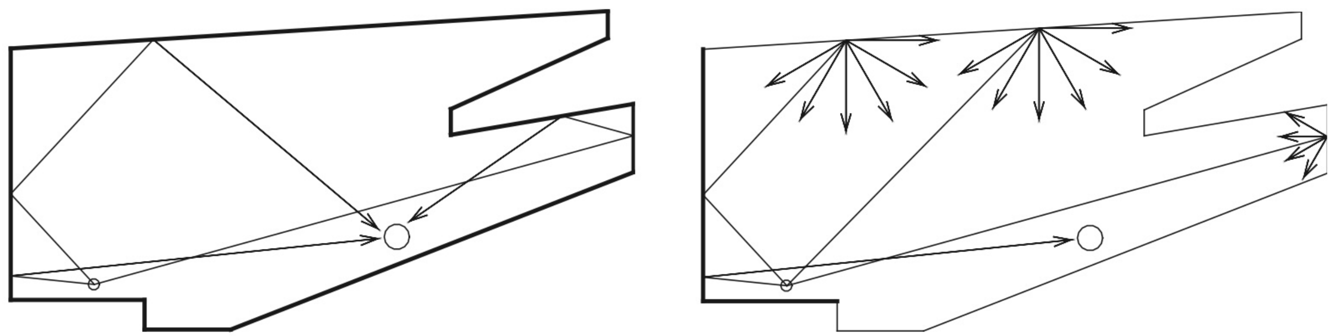

Figure 3: Specular (left) and diffuse reflection (right) ray tracing examples, from [Savioja and Svensson, 2015].

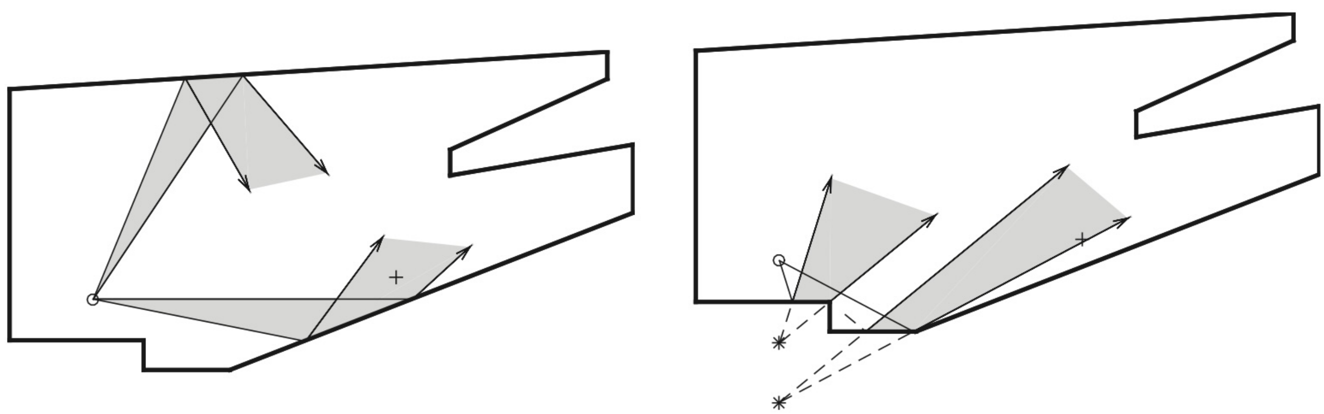

Figure 4: An example of the ray-tracing style of beam tracing, from [Savioja and Svensson, 2015].

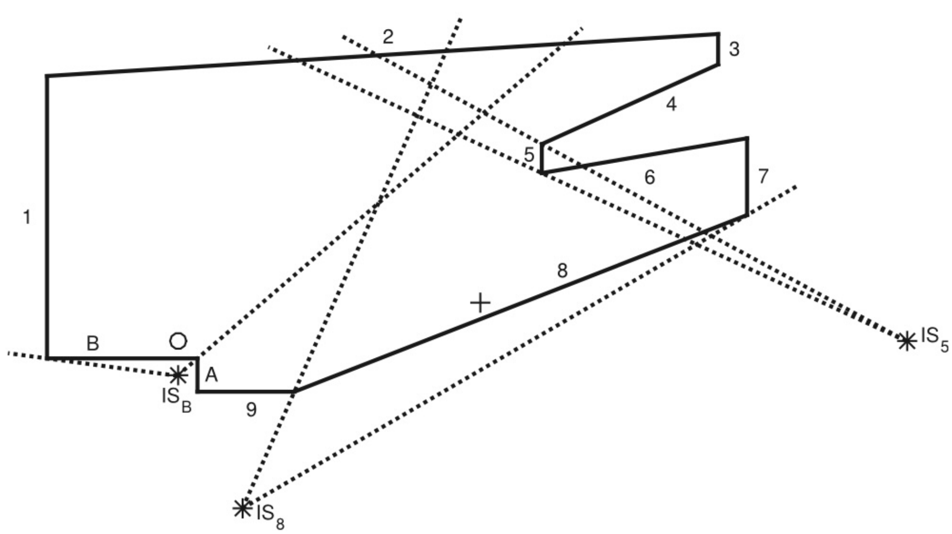

Figure 5: An example of the image-source style of beam tracing, from [Savioja and Svensson, 2015].

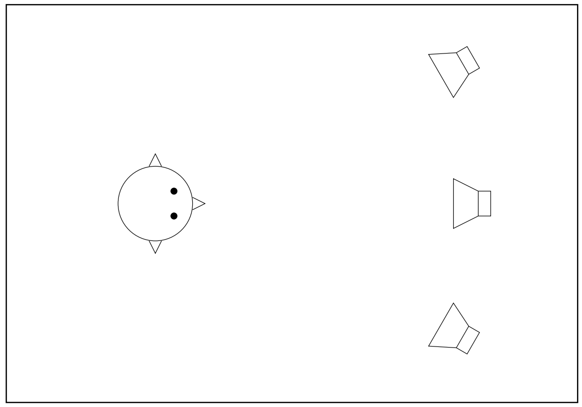

Figure 6: A listener-source setup in a room.

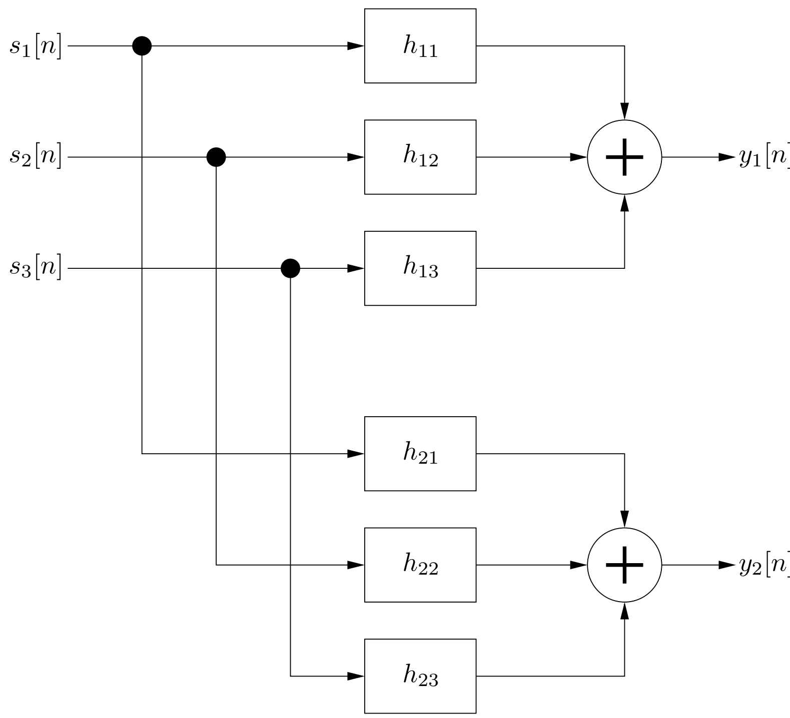

Figure 7: Transfer function approach to reverberation simulation for three sources and one listener.

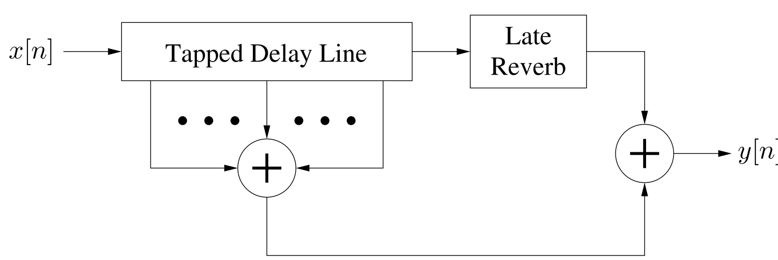

Figure 8: Early reverberation implemented with a tapped delay line, followed by a late reverberation processing block.

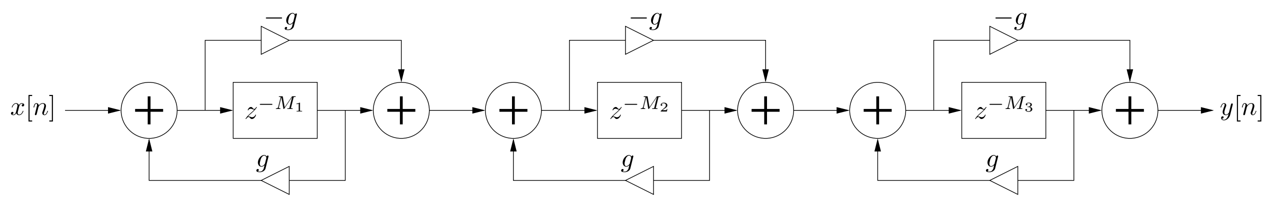

Figure 9: Cascaded Schroeder allpass sections.

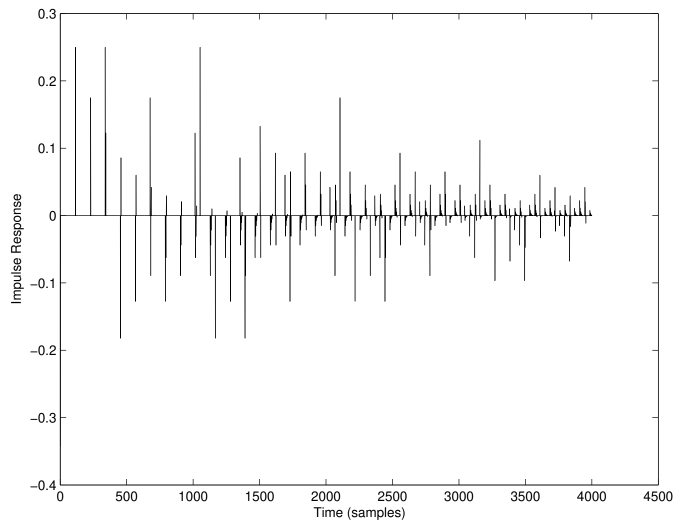

Figure 10: Impulse response of three cascaded Schroeder allpass sections ($g = 0.7$ and $M_i = [113, 337, 1051]$).

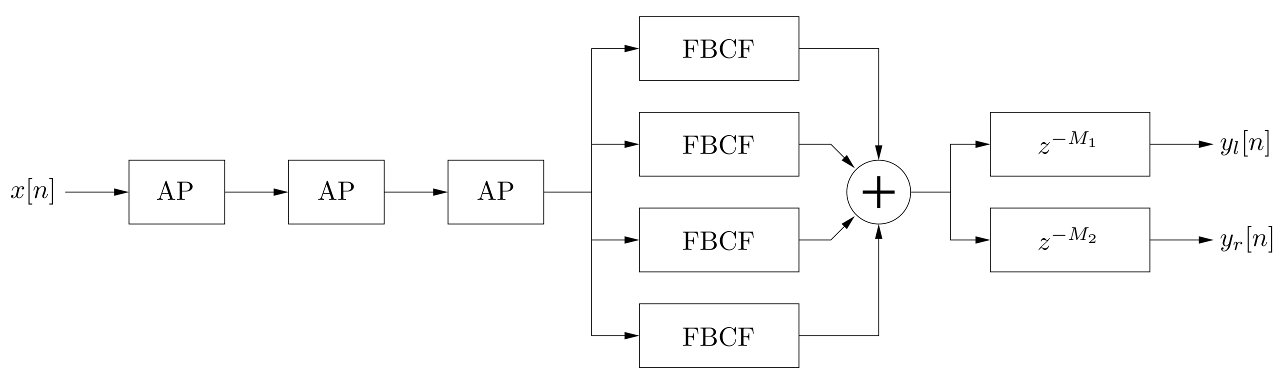

Figure 11: The `JCRev` reverberator from CCRMA (based on Schroeder/Moorer).

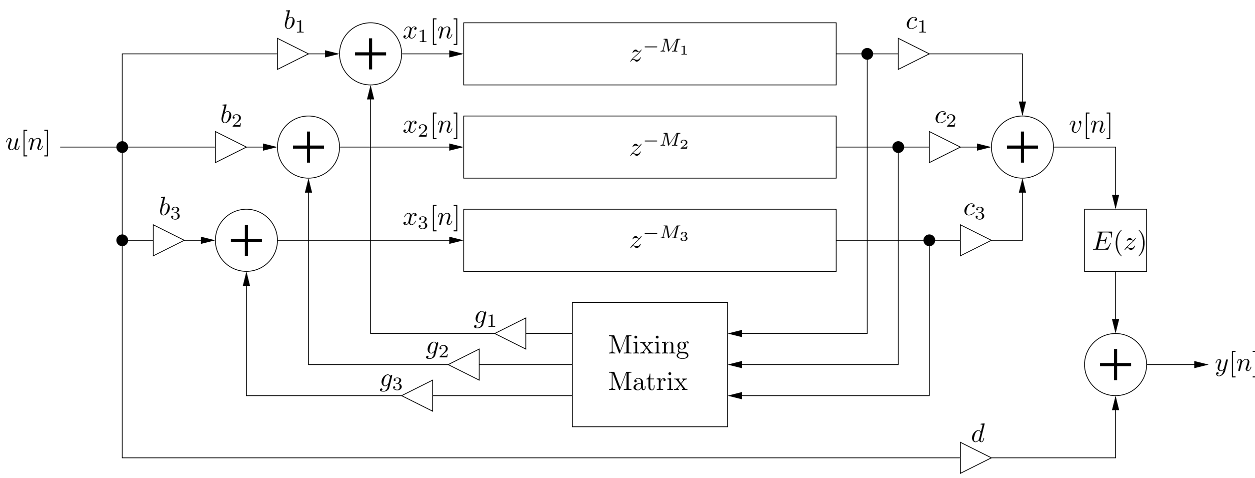

Figure 12: A feedback delay network structure proposed for artificial reverberation by Jot [1992].