

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 a variety of methods and criteria for the evaluation and design of an acoustic space. Acoustic spaces can have a variety of uses that may or may not share common desired acoustic characteristics.

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.
- Shortly after the arrival of the direct sound, a series of semi-distinct reflections from various reflecting surfaces (walls and ceiling) will reach the listener. These *early reflections* typically will occur within about 50 – 80 milliseconds.
- 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, the reverberant sound begins to decay immediately.

1.2 Direct Sound and Early Reflections

- Direct sound will decrease by 6 dB for each doubling of propagated distance.
- Our auditory system will determine the direction of a sound source from the direct sounds reaching the ear.
- Early reflections that arrive within about 50 – 80 milliseconds are not heard as separate from the direct sounds. Rather, they tend to reinforce the direct sound. For rapidly varying sound, such as speech, the limit is around 50 ms while for slowly varying music, the limit is closer to 80 ms.
- The source is perceived to be in the direction from which the first sound arrives provided that (1) successive sounds arrive within about 35 milliseconds, (2) the successive sounds have spectra and time envelopes reasonably similar to the first sound, and (3) the successive sounds are not too much louder than the first. This is referred to as the *precedence effect*.
- From a study by Leo Beranek (1962), a concert hall is considered “intimate” if the delay time between the direct and first reflected sound is less than 20 milliseconds.
- First reflections usually arrive from the nearest side wall (lateral reflections) or from the ceiling for those seated in the center.
- Reflections from the ceiling or overhead reflectors are not as perceptually desirable as those from side walls.

1.3 Late Reflections

- During a continuous sound, the reverberant sound level is reached when the rate at which energy is supplied by the source is equal to the rate at which sound is absorbed by the room and its contents.
- Too much reverberant sound will result in loss of clarity.
- In a bare room, where all surfaces absorb the same fraction of the sound that reaches them, the theoretical reverberation time (RT) is proportional to the ratio of volume to surface area.
- One definition for *reverberation time* is the time required for the reverberant sound level to decrease by 60 dB (or t_{60}). It normally varies with frequency.

1.4 Calculating Reverberation Time

- When expressed in units of cubic and square meters, the reverberation time can be estimated by $RT = 0.161V/A$, where V is the volume of the room and A is the effective “total absorption” area (Sabine, 1922).
- 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.
- Typical frequencies for which absorption coefficients are determined are 125, 250, 500, 1000, 2000, and 4000 Hz.
- A low absorption coefficient indicates a more reflective material.

1.5 Air Absorption

- Air contributes a substantial amount to the absorption of high frequency sound.
- Taking account of air absorption, $RT = 0.161 V/(A + mV)$, where m is a constant that varies with air temperature, humidity, and frequency.

1.6 Criteria for Good Acoustics

- 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).
- The optimum reverberation time of an auditorium is dependent on the use for which it is designed (speaking, chamber music, opera, orchestral music, etc.).
- Common requirements for good acoustics include *adequate loudness*, *uniformity*, *clarity*, *reverberance*, or *liveness* (the listener should be bathed in sound from all sounds, though still able to localize the sound source), *freedom from echoes*, and *minimal background noise*.
- Reflected sound arriving from the sides (lateral reflections) seems to be important to the overall reverberance of a room.

1.7 Concert Halls

- Important subjective attributes of concert hall acoustics include *intimacy or presence*, *reverberation or liveness*, *spaciousness: apparent width*, *spaciousness: listener envelopment*, *clarity*, *warmth*, *loudness*, *balance*, *blend*, and *ensemble*.
- Intimacy is the subjective impression of the size of a hall. Smaller, more intimate halls are generally preferred over wider, more remote halls.
- The acoustical measure of intimacy is called the *initial time delay gap* (ITDG), which is defined as the interval between the arrival of the direct sound and the first reflection at the listener.

- The *early decay time* (EDT), or initial rate of sound decay in a room, is perceptually most important to our impression of reverberance. The EDT consists of relatively few isolated early reflections.
- *Clarity* (C_{80}) is defined as 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). 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.
- 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.
- Spaciousness results from the presence of lateral reflections, which give the impression of a wider source width, as well as a diffuse sound field, which gives the impression of listener envelopment. Textured room surfaces help sound diffusion.
- Reverse fan-shaped halls provide better laterality of sound.
- *Warmth* is provided to a music hall via a slight increase in low frequency reverberation.
- Excessive high frequency damping can lead to a lack of *brilliance*. Brilliance is 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.
- Echoes, flutter echoes, sound focusing, sound shadows, and background noise should be avoided in an auditorium design.
- The greater the early decay time (up to two seconds), the greater the preference for the concert hall. Above two seconds, the trend it reversed.
- Narrow halls are generally preferred to wide ones.
- Preference is shown for halls having a high “binaural dissimilarity”.

1.8 Room Resonances

- For very simple geometries, it is sometimes possible to obtain an analytic expression describing the resonance frequencies of a room.
- The resonances in a rectangular box with dimensions a, b , and c and without damping are $f_{lmn} = \frac{v}{2} \sqrt{(l/a)^2 + (m/b)^2 + (n/c)^2}$, where v is the speed of sound and l, m , and n are integers.

2 Artificial Reverberation

There are a variety of approaches to synthesizing the effect of a reverberant space. Approaches based on direct measurement of a particular room response (convolution techniques) tend to be less extensible and computationally expensive, though possible using special purpose hardware. The use of three-dimensional physical modeling techniques is also limited by computational requirements. Most current work in simulating reverberation is based 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.

2.1 Transfer-Function Models

- The simulation of room reverberation ideally involves two transfer functions per sound source per listener (one for each ear). The transfer functions or filter representations will change if anything in the room changes.

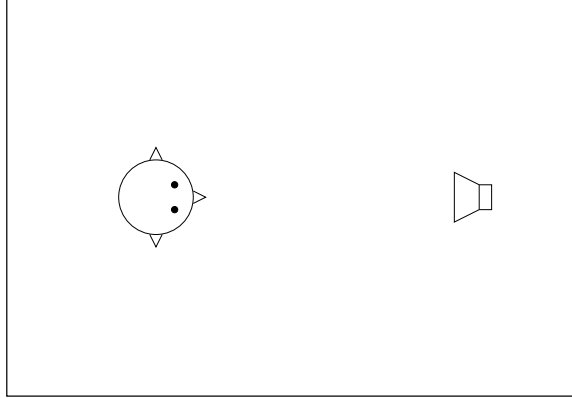


Figure 1: A listener-source setup in a room.

- For the three source, one listener setup depicted in Fig. 2, the output signals would be computed via six convolutions:

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

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.
- In addition to being very computationally demanding, this approach requires new filter representations whenever the room setup changes. In general, it is difficult to implement a flexible reverberation control scheme using convolution-based approaches.

2.2 A Physical Modeling Approach

- A “distributed” physical modeling approach to a reverberant space would allow for dynamic modifications of listener and source positions during an acoustic simulation.
- However, a brute force acoustic simulation of a room response using three-dimensional physical modeling techniques would require nearly 150 million “mesh” grid points to simulate a room of only 4 x 4 x 3 meters at a sample rate of 50 kHz.
- In addition, current three-dimensional modeling techniques are plagued by dispersion errors that would limit the quality of this approach. It is possible to use warping techniques to minimize the dispersion errors but this would significantly increase the already prohibitive computational burden.

2.3 Perceptual Aspects of Reverberation

- Above some frequency, the mode density (which increases as f^2) of a reverberation response becomes so high that it can be approximated by a random frequency distribution.
- Beyond some time, the echo density (which increases as t^2) of a reverberation response becomes so high that it can be approximated by a random time distribution.
- Based on perceptual limits, the impulse response of a reverberant space can be divided into two segments:

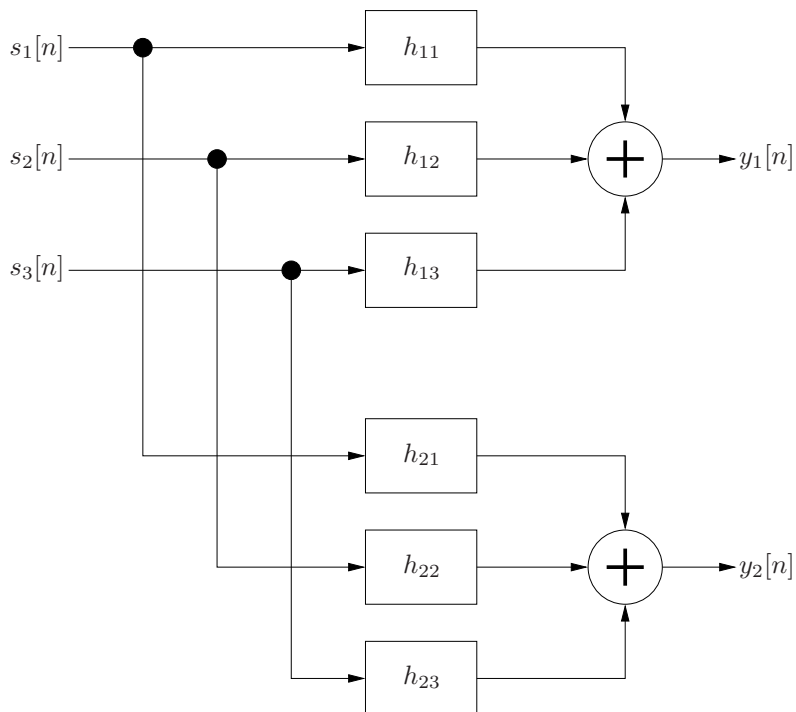


Figure 2: Transfer function approach to reverberation simulation for three sources and one listener.

- 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.
- 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 should 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

2.4 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.

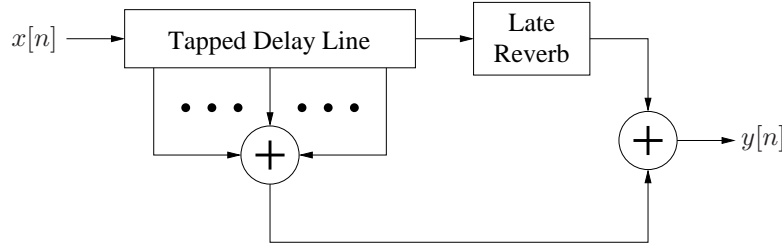


Figure 3: Early reverberation implemented with a tapped delay line, followed by a late reverberation processing block.

2.5 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.

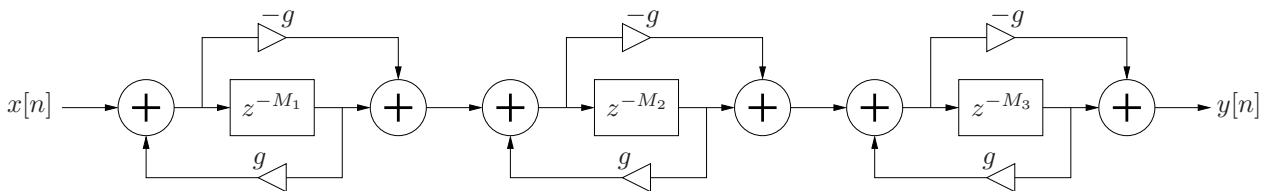


Figure 4: Cascaded Schroeder allpass sections.

- 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. 5.
- The feedforward 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.

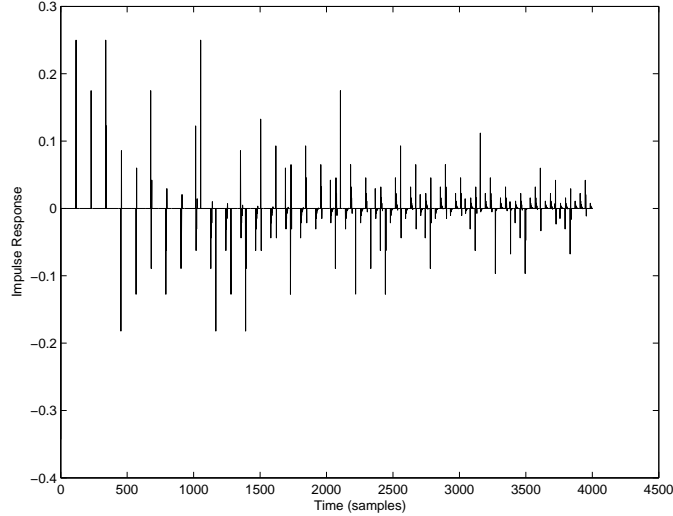


Figure 5: Impulse response of three cascaded Schroeder allpass sections ($g = 0.7$ and $M_i = [113, 337, 1051]$).

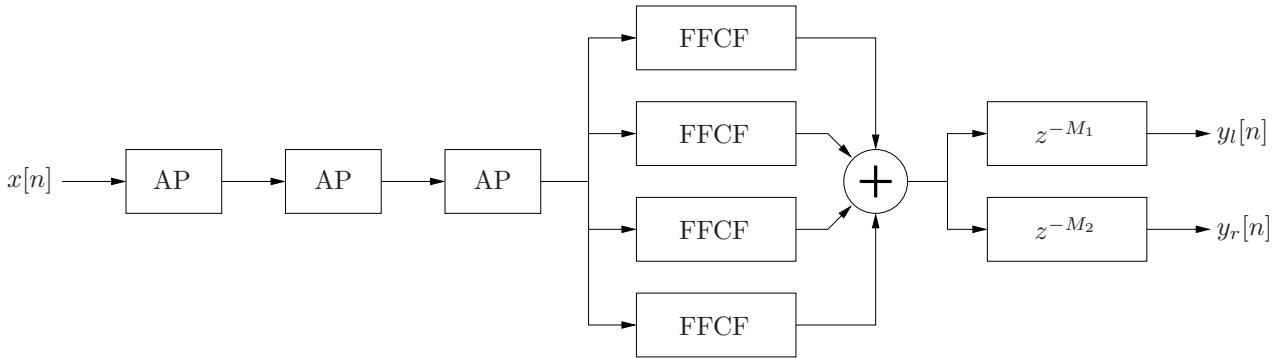


Figure 6: The JCRRev reverberator from CCRMA (based on Schroeder/Moorer).

2.6 Feedback Delay Networks (FDN)

- Figure 7 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. 7).
- 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 can be replaced by low-order digital filters.
- The “tonal correction” filter $E(z)$ is a low-order filter that serves to equalize modal energy amongst the three bands.
- The delay-line lengths are generally chosen to be mutually prime. System “tuning” remains a manual, trial and error process.

2.7 Feedback (“Mixing”) Matrix Stability

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

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

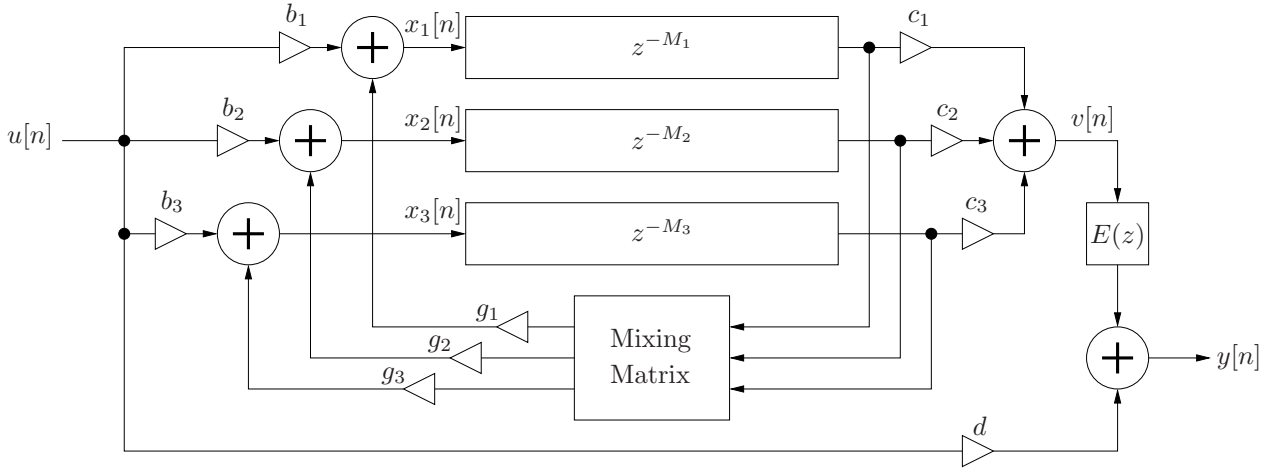


Figure 7: A feedback delay network structure proposed for artificial reverberation by Jot [1992].

- The inner loop calculations of the FDN shown in Fig. 7 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 $\mathbf{A} = \mathbf{GM}$, 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.

2.8 Householder Feedback Matrices

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

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

where $\underline{\mathbf{u}}_N^T = [1, 1, \dots, 1]$ 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 $\underline{\mathbf{u}}_N^T$ times the input vector, applying the scale factor $2/N$, and subtracting the result from the input vector).

2.9 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. Schroeder suggests a modal density of 0.15 modes per Hz for a 1 second t_{60} . This can be generalized to $M \geq 0.15t_{60}f_s$.

2.10 Reverberation Time

- 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

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

- 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

$$g_i = 10^{-3M_i T/t_{60}(0)}$$

and from [Jot and Chaigne, 1991]

$$a_i = \frac{\ln(10)}{4} \log_{10}(g_i) \left(1 - \frac{1}{\alpha^2}\right),$$

where

$$\alpha \triangleq \frac{t_{60}(\pi/T)}{t_{60}(0)}.$$

2.11 Other Techniques: Image Method

- The *image method* is based on a *ray tracing* model for room reflections.
- This technique is used to determine “virtual sources” at mirror image locations with respect to a reflecting surface.
- Once the virtual sources are determined, propagation distances can be easily calculated from two- or three-dimensional Euclidean geometry.
- This method assumes *specular* reflections from large, smooth surfaces. It may be useful for determining a set of early reflections in spaces with large, flat walls.
- However, this technique does not account for *diffuse* scattering.

2.12 Other Techniques: Convolution

- Because an acoustic space is by and large a linear, time-invariant system, one can “simply” measure its impulse response and use convolution to reproduce the effect of playing a given audio signal in that space.
- Problems:
 - Convolution is an expensive computation ... need efficient techniques.
 - It is not a “simple” process to measure the impulse response of a space.
 - A measured impulse response corresponds to a single source-listener configuration.
 - A measured impulse response is inflexible to modifications such as shortening the reverberation time.

References

- B. Blesser. Interdisciplinary synthesis of reverberation viewpoint. *J. Audio Eng. Soc.*, 49:867–903, 2001.
- J. Dattorro. Effect design. *J. Audio Eng. Soc.*, 45:660–684, 1997.
- J.-M. Jot. *Etude et Réalisation 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 J.*, 3(2):13–18, 1979.
- M. R. Schroeder. Digital simulation of sound transmission in reverberant spaces (part 1). *J. Acoust. Soc. Am.*, 47(2):424–431, 1970.
- M. R. Schroeder. Natural-sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3):219–223, 1962.