UNIVERSITÀ DEGLI STUDI
DI MILANO

Spatial Sound Rendering in Python and Pyo

Omid Torkan
Universita degli Studi di Milano
Department of Computer Science
Associate Professor
Federico Avanzini
Email:omid.torkan@studenti.unimi.it
Matriculation Number:930932

November 2020

Abstract:

The reverberation[1] effect plays a very important role in the spatial and peripheral processing of sound in closed environments, and various methods have been proposed to produce this effect, which aim to model the actual reverberation of sound from an audio source in the environment. Because the analysis of the acoustic properties of the environment is very complex and difficult, the process of producing reverberation is usually done by simpler and more approximate methods. In this project, the aim is to investigate some methods of reverberation production by pyo library in Python software, which according to the values of their input parameters, create the effect of reverberation in the input signal. These methods include WGverb, Convolve, Freeverb, HRTF, Binaural. For each method, several input sources are used and their audio effect is analyzed.

[1] Reverberation

Introduction and State of Art:

From the beginning of the introduction of artificial methods of reverberation production as an important part of the spatial processing of sounds in different environments, these methods are used to reproduce different types of reverberation in different environments. In the music, film, video game, and acoustics industries, these techniques have been very effective. The main advantage of these methods is the ability to control its variables by computer based on the needs of the person or system. In this field, many algorithms have been proposed that seek to model a wide range of environments and spaces, but no method is fully capable of performing all modeling.

In general, sound reverberation is the result of the propagation of sound waves in the room or any other space that is reflected from various walls, objects and obstacles, and by repeating this process of propagation and reflection, it finally reaches the listener or microphone. The approximate speed of sound at normal temperature is about 344 meters per second, which makes the return samples from the environment to be heard at times that are different from the original sound. Therefore, the human ear can perceive the effect of these reflections. With each reflection of a surface or object, part of the sound energy is absorbed and the other part is reflected, which means that the sound level or amplitude decreases in each reflection. On the other hand, after a few reflections, the level of sound amplitude decreases so much that it can no longer be heard, which depends on the environment, distance and type of surfaces. Incoming sounds from an audio source are divided into three categories: 1- Directly received sound 2- Early reflections 3- Late or final reverberation

The following figure shows this:

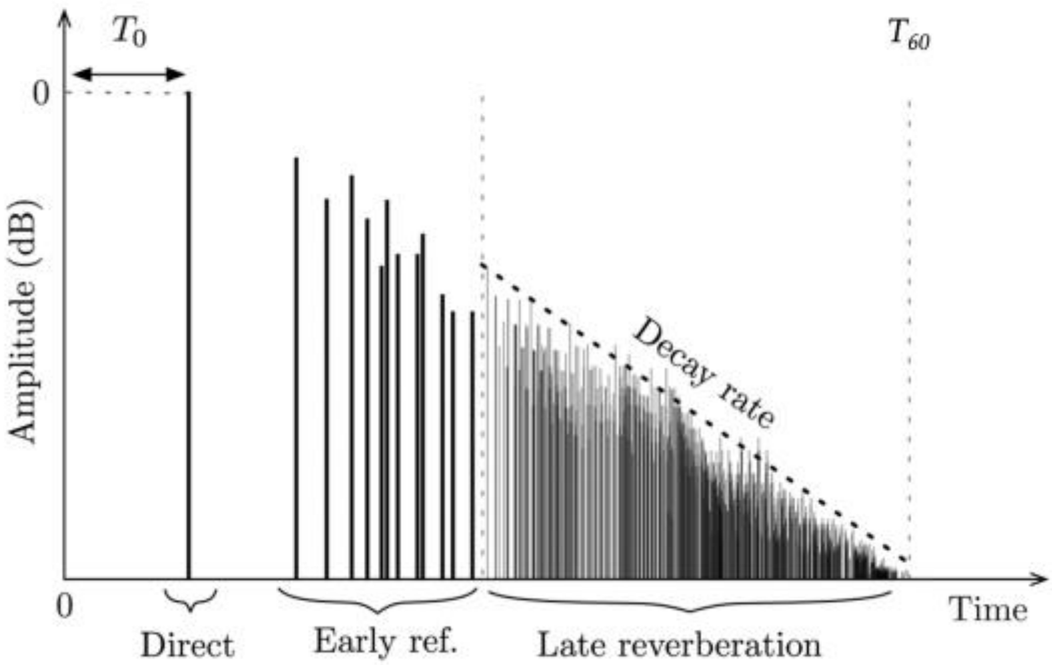

Figure1- Audio source reflections

It is clear that a copy is received directly without reflection. Then gradually the initial reflections are received at decreasing intervals and with smaller amplitudes. Eventually, reverberation versions with even lower amplitudes are received in terms of time density, so that their amplitude is almost negligible. Usually a drop of 60 decibels is considered to reach the lowest level of reverberation. In other words, reflections with an amplitude drop of more than 60 decibels compared to the direct version are not considered.

Primary reflections are usually the first and second, and in some cases the third. These samples differ in amplitude, arrival time, *direction and frequency content. The time limit between these reflections and the final echoes is usually 80 to 100 milliseconds. The third part is the reverberation phase of the return signals, which are so compressed that they are heard as an integrated structure. The power drop of this section is usually modeled as a damping. Since this part is received from almost all directions, it is not assumed to have a specific direction.*

It should be noted that the higher the frequency of sound, the more it is absorbed by the environment and the lower the reflection. Also, the ambient air has a significant effect on the decrease in the power of the received sound in the place of the listener.

1. Methods of producing reverberation:

We briefly review the artificial methods of producing reverberation in sound processing.

1.1 Reverberation through convolution:

Probably the clearest way to generate different reflections of a signal is to apply convolution to it by responding to the impact of a space or room. This impact response can be calculated by measuring the acoustic models of the environment for which various methods have been proposed.

In the method of producing reverberation through convolution, the assumption is on an invariant line with the time of reverberation system, the output of which follows the law of convolution. In other words, the model is completely determined by the impact response.
So the output is as follows:

$$y[n] = x[n] * h[n] = \sum_{m=0}^{N-1} h[k]x[n-k]$$

Where x [n] is the input signal, h [n] is impact response. However, due to the delay in calculating the output, this process is usually done in blocks.

One advantage of this method is that all the details and features of a room can be implemented by it, assuming that the impact response is measured correctly. But its main drawback is the limitation in changing the model parameters to produce an artificial reverberation. For this reason, another category of methods is presented, which is explained below.

1.2 Feedback delay network[2](FDN):

Schroeder and Logan[3] first introduced digital reverberation production. They proposed a method consisting of return delay filters that included comb filters[4].

Because these return filters produce a repetitive sequence of input signals with damping amplitudes over time, if a certain number of these filters are placed in parallel, their output results in a good approximation of the many reflections in real space in Occurs over time. This is Schroeder idea of how to use comb filters in parallel to increase the density of reflections, followed by the use of full-pass filters in series to distribute the signal evenly throughout the space. Of course, this method was later developed by others and different versions of it have been offered. An example of Schroeder structure is shown below:

[2] Feedback delay network
[3] Schroeder, Logan
[4] Comb filter

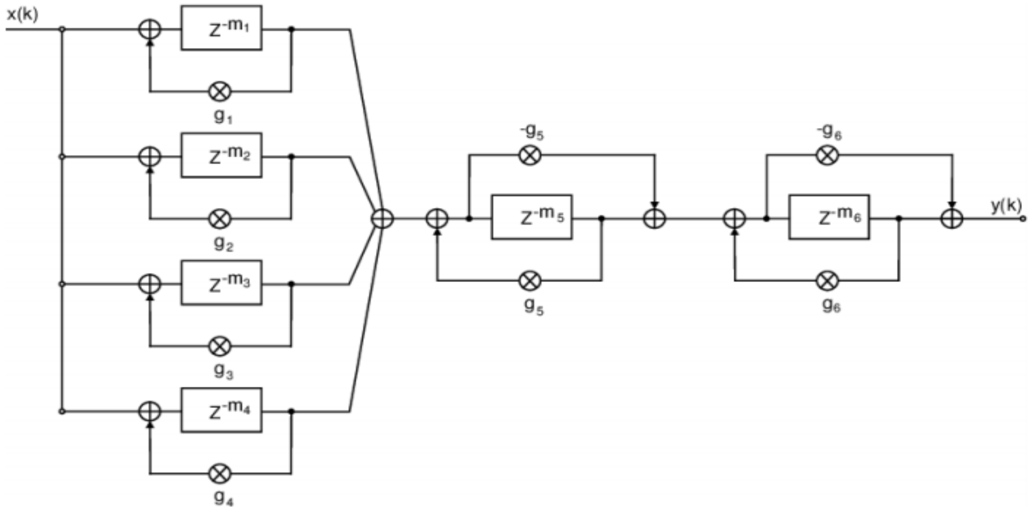

Figure 2- Schroeder reverberation production plan

It is clear that this structure includes 4 comb filters in parallel and two full-pass filters in series.

2. ***Binaural technology or two-way ear[5]:***

   Reproduction of sound signals at the listener's site should have the same effect as the actual experience of the event. This is the basis of binaural technology for spatial processing of sound in the listener's ear, which is one of the most important techniques for spatial sound processing. This is very difficult in spaces where different reverberations have to be intricately combined to produce a near-reality listening experience. The term binaural originates in such a way that distinct signals must be generated in the ears and their combination, for example through a headphone, produces the correct overall sound.

   There are several applications for this method such as virtual environment simulation, virtual reality, entertainment and computer games, music and movies and so on.

   The term binaural hearing[6] includes a set of processes in which different information received by two human ears in the brain is

[5] Binaural

[6] Binaural Hearing

analyzed in such a way as to create a unique experience of the sounds received. This helps a lot in locating the sound source in different spaces. Binaural technology seeks to generate distinct signals that can enable the location of different sound sources or sound sources through hearing with headphones. The propagation of sounds in the return from the human body, especially the head and ears, plays an important role in the auditory experience. This propagation model can be represented by a time-invariant linear transfer function measured at the listener's ear. This function is usually called a Head Related Transfer function[7]. Equivalent to the time domain Head Related Impulse Response[8]. This transfer function changes based on the environmental and spatial structure of the audio source and listener. The following figure shows an example of this function where the source is located at a surface angle (azimuth) of 45 degrees and an air angle (elution) of 0 degrees relative to the listener.

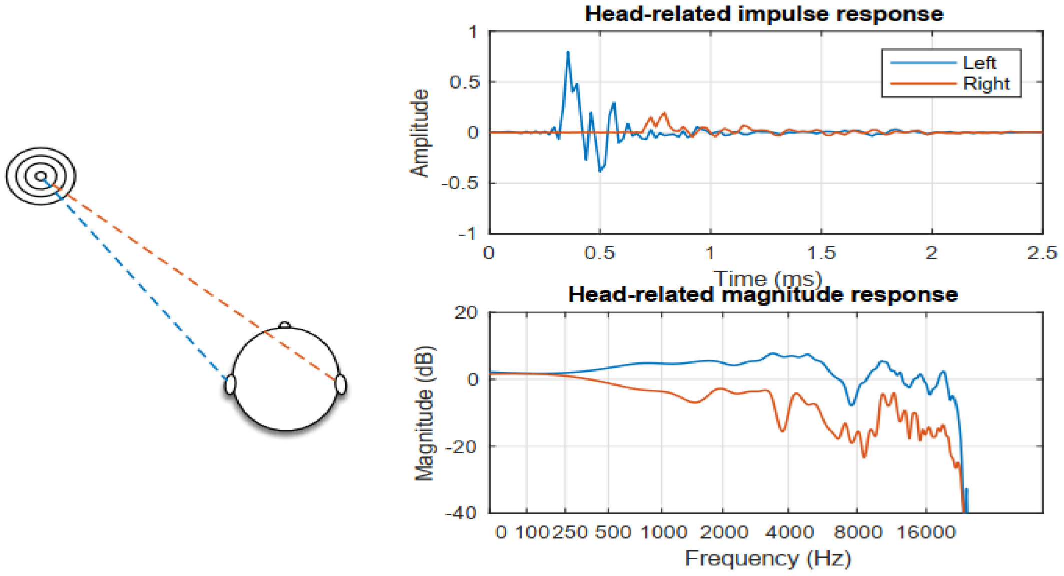

Figure 5- Head Related Transfer Function and Impulse Response

[7] Head Related Transfer Function(HRTF)
[8] Head Related Impulse Response(HRIR)

Binaural Rendering[9] *is the processing of input audio data that creates a truly deep audio experience by simulating binaural signals in the listener's ear. The input signal is processed by filters that model the location-dependent acoustic conversion function relative to the listener. This processed signal is commonly known as a virtual audio source. Such an audio spatial processing system can also have a dynamic structure in which, based on the relative motion of the audio source and the listener, it updates itself instantly and reproduces near-real virtual sound.*

One binaural rendering method is to use HRTF transfer function described earlier. To do this, the input signal is analyzed by two separate HRTF filters and transmits the output to two headphones.

The parameters of the HRTF filter are the surface and spatial angle of the listener relative to the source (θ and φ), the distance from the source, the position of the deviation of the head from its original position and the operating frequency. However, in some cases a simpler model uses this conversion function, which reduces the number of parameters.

3. *Spatial sound rendering to produce reverberation in Python:*

The functions that will be used in Python software are:

-Freeverb: The method of producing reverberation with the technique presented by Schroeder.

-WGverb: Reverberation generation method with feedback delay network technique.

-Convolve: Produces reverberation with circular convolution

-HRTF: 3D spatial analysis of sound with head related transfer function.

-Binaural: 3D spatial analysis of sound by combining the head related transfer function and Vector Base Amplitude Panning[10].

[9] Binaural Rendering
[10] Vector Base Amplitude Panning

1.3 Freeverb Method:

This function implements the Schroeder method described earlier. Includes 8 parallel comb filters followed by 4 series full-pass filters. In this function, you can select the reverberation length, which is between 0 and 1. The high frequency attenuation rate is also adjustable between 0 and 1, which determines how fast the high frequencies fall. There is another parameter for the balance between full reverberation and zero reverberation, which also varies between 0 and 1.

2.3 WGverb Method:

This function is based on the feedback reverberation model of the feedback delay network, which consists of 8 FDN lines. The rate of return feedback is adjustable between 0 and 1. The cut-off frequency of the feedback path filter can also be determined. Here, as in the previous function, there is a reverberation control parameter between 0 (no reverberation) and 1 (full reverberation).

3.3 Convolve Method:

It filters the input sound using the circular convolution method, which receives a table to determine the impulse response at the input. It also receives the resulting length of convolution. It should be noted that the impulse response table can be extracted for each audio file via the SndTable command.

4.3 HRTF Method:

This function implements the model of the head related transfer function on the audio signal. Its parameters include the surface angle and space previously described and the output signal gain. The optimal performance of this function is due to its internal impulse response when the output signal is listened to with headphones.

5.3 Binaural Method:

This function performs another type of Spatial Sound Rendering on the signal, which is a combination of HRTF and VBAP methods, in which the VBAP[11] method is to move the sound source on an array of

[11] Vector Base Amplitude Panning

16 virtual speakers around the listener, which eventually passes through the HRTF filter. And their combination creates a virtual sound in the space around the listener with the desired specifications. Binaural function parameters, like HRTF, include both surface and spatial angles and gain. In addition to these two parameters, there is a range of angle changes that determine the range of surface and spatial angle changes.

The SfPlayer function is also a tool for reading and executing audio file content, which includes file read speed parameters, the ability to run the file repeatedly in a loop, time offset at the beginning of the file to bypass part of the input, and interpolation to change speed. Of course, in the code we use the same default values that are the standard for voice processing.

4. *Results:*

1.4 Freeverb Function:

By changing the various parameters of this function, the music and speech signal is observed that the first parameter or the same size changes the amount of reflections. This parameter varies from 0 to 1, which the closer it gets to 1, the louder reverberation listening experience. At values close to 1, the sound reflections are too persistent and not very pleasant. The bal parameter, which determines the dryness of the sound, the closer it is to 1, the more it increases the volume of reflection or the environment, which creates the experience of hearing sound in a larger space. The damp parameter did not have much effect on the frequency content, but at high reverberation values, it can be effective in reducing the noise generated. The mul parameter, as mentioned earlier, raises and lowers the volume.

2.4 WGverb Function:

A new and important parameter in this function is the amount of feedback on the return path or the same feedback that more and more evokes the experience of hearing the sound reverbed in a larger environment. Also, the cutoff frequency is the cut-off frequency of the low-pass filter in the return path, which reduces noise by reducing it. Other parameters such as mul, bal have the same function as the previous function.

3.4 Convolve Function:

Here the two parameters of impulse function and convolution length are important. The percussion function is defined via the SndTable function of the audio file itself. The convolution length also produces noise for values less than 512 and more than 1024, indicating that the convolution action causes excessive interference between the delayed samples. There is also a slight noise at 512. The optimum value is 1024, which produces a better sound.

4.4 HRTF Function:

The two most important parameters in this function are the amount of surface angle that varies between -180 to +180 and the spatial angle that varies between -40 to +90. They are denoted by azi and ele, respectively. Here variable values are used for these two angles. The two functions Phasor and Sine, which produce a linear phase change in its defined range and a change in the value of the output sinusoidally with the frequency and range defined for it, respectively. For the surface angle, which ranges from -180 to +180, the Phasor function is as follows:

Phasor(freq=0.2, mul=360)

Which shows that every 0.2 seconds a complete phase change occurs in the range of 0 to 360 degrees, which is equivalent to -180 to +180.

For a spatial angle ranging from -40 to +90, the Sine function is as follows:

Sine(0.1).range(0,90)

Which shows a sine wave with a frequency of 0.1 Hz varies in the range of -40 to +90.

Of course, other values can be considered for these two angles. For example, you can set the same Sine function for the surface angle.

By examining the output, the motion of the audio source is fully felt in the area of the surface angle where the two sides of the head are recognizable.

As the frequencies increase, the periodic motion of the sound source becomes more intense than that of the ear.

5.4 Binaural Function:

The function of this function is similar to HRTF, and in addition, the parameter of the sound source amplitude can be adjusted in both surface and spatial angles. We define the surface and spatial angle as a variable as the HRTF function, and the amplitude as a variable in the form of constant and sinusoidal with amplitude 0.3. In terms of movement at the surface and height, the audio quality is the same as HRTF, but the output intensity or volume is higher here and a stronger sound is heard.

In terms of amplitude parameter in the steady state, the closer the value is to 0, the less reverberation the audio source is heard, and the closer to one, the more reverberation elements are heard in the sense of reverberation, which of course also creates some noise. In the case of variable amplitude, a sound with variable resonance is heard, which shows a different amount of reverberation in different places, which is quite expected because with the change of time, both the location of the sound source and its reverberation change according to the amount of spn amplitude.

5. Discussion and Conclusions:

In this project, the aim is to investigate artificial methods of reverberation production and spatial processing of sound in the computer environment to simulate real reflections in closed spaces with different dimensions. reverberation generation theoretically has different methods, the most important of which are convolution method and feedback delay network. Also, the method of simulating the head related transfer function is very important in implementing the real model that the human ear receives from different reflections of its body in the auditory experience close to reality. In this regard, different Python functions were tested and performed on different music files and the final processing result was analyzed in terms of listening experience, sound depth, sound source location and overall ambient sound quality. In this software, by adjusting the various parameters of each function, the expected quality can be created for the desired ambient space in which the audio signal must be heard.