Virtual Reality with Convolution

Mridul Choudhary (Report + Inbuilt Convolution)
Jayesh Verma (Report + Manual Convolution)

August 30, 2024

1 Links

Google Drive folder containing .sce files and output files: Folder

2 Introduction

The goal of this assignment was to create a nice audio experience by converting a mono audio recording into spatial sound using specific binaural room impulse responses (BRIR) through the process of discrete-time convolution(DTI). This involved simulating how sound would be perceived in different environments using stereo output, with each channel corresponding to the left and right ears. The implementation was carried out in Scilab using the provided skeletal framework `generate_reverb_data.sce`.

3 Inputs

- **inp**: Mono audio file (`BheegiRegular.wav`) sampled at 16000 Hz.
- **fs_inp**: Sampling frequency of the input audio file.
- **rir**: Two-channel BRIR representing the room impulse response measured in different environments (e.g., long hall, parking garage). The BRIR must have the same sampling rate as the input audio file.

4 Output

out: Two-channel (stereo) convolved signal designed to be played over headphones to create a virtual reality audio experience.

5 Background and Concept

5.1 Methods for Simulating Surrounding Sound

For this assignment, we are applying BRIR to create the illusion of spatial sound. BRIRs capture the impulse response of a room, allowing us to recreate how a sound would reflect and attenuate in that environment.

5.2 Binaural Room Impulse Responses (BRIRs)

Binaural Room Impulse Responses (BRIRs) are audio recordings that capture the unique acoustic characteristics of a specific environment. They provide a detailed snapshot of how sound behaves in a particular space, including factors like reverberation, echoes, and the overall sense of spaciousness.

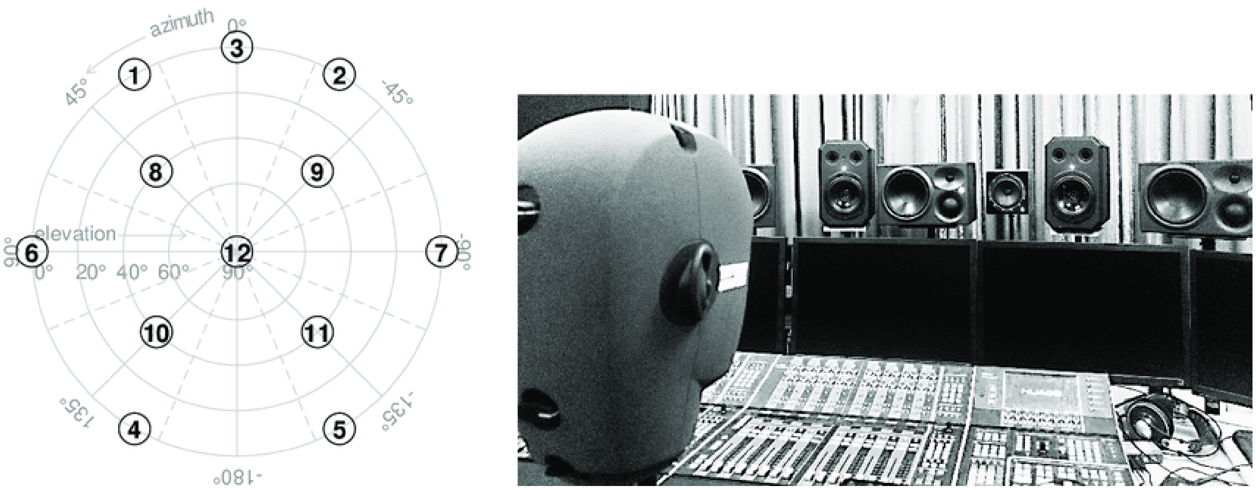

Figure 1: Binaural room impulse response measurement setup

5.2.1 *How BRIRs Are Created*

- **Sound Source**: A short, sharp sound, often referred to as an impulse, is emitted from a specific point in the room.
- **Microphone Pair**: A pair of microphones, typically placed at the listener's ears, records the sound as it travels through the room and interacts with surfaces.
- **Recording**: The recorded signal is the BRIR. It contains a wealth of information about the room's acoustics.

5.2.2 *Key Components of a BRIR*

- **Direct Sound**: The sound that travels directly from the source to the listener's ears without reflecting off any surfaces.

- **Early Reflections**: The first few reflections of the sound that reach the listener's ears. These reflections create the perception of space and distance.
- **Reverberation**: The prolonged decay of sound in the room due to multiple reflections. Reverberation can add depth and richness to the sound, but too much can make it difficult to understand speech or music.

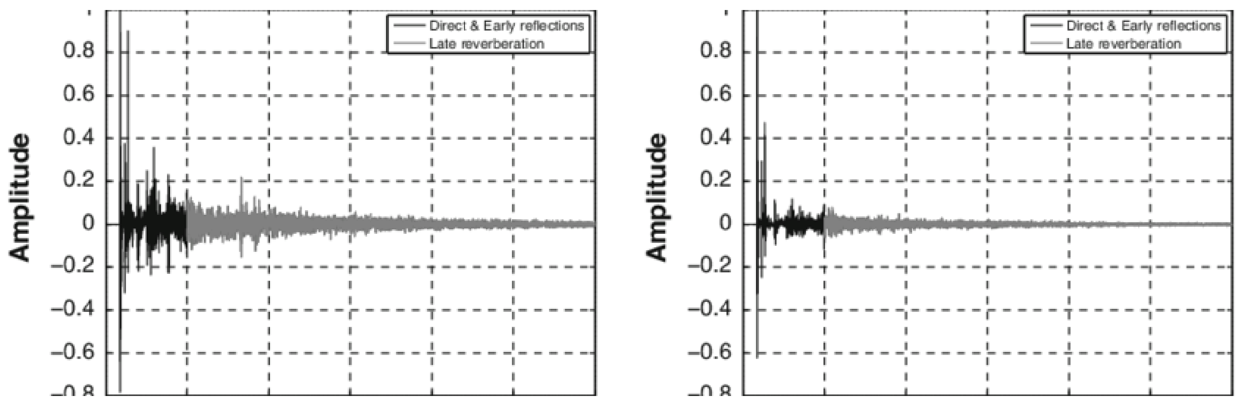

Figure 2: Binaural room impulse responses in a stairway having a reverberation time of approximately 0.86 s. Left left-ear room impulse response. Right right-ear room impulse response

5.2.3 Factors Affecting BRIRs

- **Room Geometry**: The size, shape, and materials of the room (walls, ceiling, floor, furniture).
- **Sound Source Position**: The location of the sound source within the room.
- **Listener Position**: The location of the listener within the room.
- **Sound Characteristics**: The frequency content and amplitude of the sound source.

6 Implementation Steps

6.1 Loading the Input Audio and BRIR Files

We loaded the mono audio (`BheegiRegular.wav`) and the BRIR file into Scilab using the `wavread` function. It was crucial to ensure that both files had the same sampling rate to maintain synchronization during the convolution process.

7 Implementation Steps

7.1 Loading the Input Audio and BRIR Files

We loaded the mono audio (`BheegiRegular.wav`) and the BRIR file into Scilab using the `wavread` function. It was crucial to ensure that both files had the same sampling rate to maintain synchronization during the convolution process.

7.2 Manual Convolution Implementation

The core of this assignment was implementing the convolution process manually, without relying on built-in functions. Convolution was performed separately for each channel of the BRIR:

- **Initialisation**:
  N_x: Stores the length of the input signal inp1.
  N_h: Stores the length of the impulse response rir1.
  N: Calculates the length of the output signal, which is equal to the sum of the lengths of the input and impulse response minus 1.
  Output: Creates a row vector of zeros with length N to store the output signal.

- **Convolution Loop**:
  The outer loop iterates from n = 1 to N, representing the indices of the output signal. The inner loop iterates from k = 1 to N_h, representing the indices of the impulse response. The conditional statements if n - k + 1 > 0 and n - k + 1 <= N_x ensures that the index n - k + 1 is within the valid range of the input signal. If the condition is true, the current element of the output signal output(n) is updated by adding the product of the corresponding elements of the input signal and the impulse response.

```
function [output]=fun_conv(inp1, rir1)
Nx = length(inp1);
Nh = length(rir1);
N = Nx + Nh - 1;

output = zeros(1, N);

for n = 1:N
    for k = 1:Nh
        if n - k + 1 > 0 && n - k + 1 <= Nx
            output(n) = output(n) + inp1(n - k + 1) * rir1(k);
        end
    end
end
endfunction
```

Convolution is mathematically defined as:

$$y[n] = \sum_{k=0}^{N_{rir}-1} x[k] \cdot h[n-k]$$

where $x[k]$ is the input audio, and $h[k]$ is the BRIR.

7.3 Saving the Output

The final convolved stereo output was saved as a WAV file (`output_stereo.wav`) using the `wavwrite` function. This file represents the spatially enhanced audio output that can be played over standard media players or headphones.

8 Challenges Faced

8.1 Computational Intensity

Convolution, especially with large BRIRs and longer audio samples, is computationally demanding. The processing time was significant, particularly when working with larger data sets.

8.2 Ensuring Consistent Sampling Rates

It was essential to ensure that both the input audio and BRIR files had matching sampling rates to prevent timing errors during convolution.