3D Audio

Mona Wang

For the recording session, I grouped with Yulu and Devyani. We chose to record at the Bobst library lobby, which has an excellent soundscape environment. We put the Neumann KU-100 binaural head ("Fritz") on a tripod in the center of the front lobby. Because we were all scared of playing sine sweep out loud in the lobby to calculate the impulse response, we chose to do a clap recording instead. After recording the clap, we recorded the three-minute sounds for the first time. It is not very ideal because the gain was low, and the sounds were not captured at all. We then raised the gain and started our second recording. The second recording was great, and we each listened to the recording through the headphones once to make sure the quality of the recording. We also tried to record some singing vocals from Devyani in one of the rooms on the fourth floor. We also recorded the impulse response by playing the sine sweep in that room. However, I will only use the Bobst lobby recording for my project, so I only calculated the impulse response for our first recording.

As I wanted to create a project in Bobst library with some usual sounds I heard in New York City and some beautiful music, I downloaded most of the virtual sounds related to street siren sounds, water dripping sounds, church clock, and three piano improvisation pieces on both the BBC free sound effect library and the freesound website. I loaded both virtual and recorded sounds using Python in a Jupyter Notebook through Visual Studio Codes. I also used a virtual environment through Anaconda for coding. The Python system that I used is 3.9.20. After loading the data, I coded to apply binaural processing with HRTF. I installed and utilized the pysofaconventions library to load the HRTF dataset. The HRTF dataset I chose is the SADIE II Database from York University. I only used the D1 SOFA data because it uses KU100 exactly, and within D1, I use the HRIR 44K_16bit_256tap_FIR_SOFA dataset. I used the SOFA file containing HRTF data to assign each virtual sound a degree of azimuth and elevation. If the degree I wanted was not included in the dataset, I then used the SOFA file to find the closest degree of azimuth and elevation. The virtual sound was convolved with the corresponding HRTF data for both the left and right ears to simulate spatial positioning. Then, I applied room simulation by convolving the HRTF-processed sound with a stereo impulse response (IR) recorded in the Bobst Library (the clapping recording), representing the room's acoustics. Since I wanted the sound sources coming from different distances, I adjusted the gain to make the sound source come farther or closer. Moreover, in order to prevent clipping, I also normalized the result after the HRTF process, room simulation, and gain adjustment. However, with the application of room simulation with the current IR, the reverb is too loud in the processed audio. I then made adjustments by trimming the length of the impulse response to avoid long reverberation. I also changed the convolution mode to improve the reverb level further. After doing this, the reverb level is greatly enhanced. The final output audios were saved as a stereo WAV file. After finishing processing the virtual audios, I blended these sounds into the binaural Bobst lobby recording. I first trim the recordings to three minutes. Then, I blended them in specific timestamps, ensuring the total length of these superimposed sounds would not exceed the original size of the binaural recordings. After that, I got my final mix.