BINAURAL ROOM IMPULSE RESPONSE (BRIR) SURVEY

Report for Module EEMT17 Spatial Audio

Haohan Zhu, Student ID 22308057
Trinity College Dublin
Lecturer: Dr. Enda Bates
zhuha@tcd.ie

February 2023

This report is submitted in part fulfilment for the assessment required in EEMT17 Spatial Audio. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year. These are found in Parts II and III at http://www.tcd.ie/calendar.

1 Introduction

Spatial audio aims to perceive and render sound in three dimensions to provide listeners with an immersive listening experience. This technology is used in virtual reality, augmented reality, and other gaming and film industries for sound localization widely. It is also used for creating surround sound field to make sure that listeners perceive sound coming from a certain direction and experience a sense of sound motion, and creating an immersive experience for the user all the time by pairing it with realistic graphics technology.

In the field of spatial audio, the impulse response corresponds to the acoustic response of three-dimensional space to an impulsive sound source. It is widely used to capture the spatial characteristics of the space in which the sound source is located. The impulse response mainly describes the spatial information expressed by the sound waves in the room, it contains information about the transmission time of the sound waves, frequency, information on the reflection path of the sound waves in the room, and reverberation.

This project aims to investigate and study the effect of impulse response in the same space at both time and space scales. In this project, I will use the binaural room impulse response (BRIR) as a test tool to test the impulse response. The binaural room impulse response is a specific type of impulse response that captures the acoustic properties of a room through a microphone that mimics the human binaural structure, simulating how humans perceive the sound signals received in the room. The researcher will use a sine sweep signal and anechoic samples as inputs to the BRIR, and subsequently parse the information and representations contained therein by analyzing the captured BRIR to relate different features to relevant perceptual parameters, such as directional cues (ITD and ILD) and distance cues (time and level of direct signals, early reflections, and their relationship to room geometry and surfaces).

In addition, I will use Reaper software and Matlab to analyze the environmental features contained in the captured BRIR.

In this introductory module, the authors provide an overview of the concepts and related to impulse response and binaural chamber impulse response, as well as their application to spatial audio. The authors also briefly discuss the tasks and goals carried out in the project and briefly outline the design and implementation methods of the study.

2 Methodology

2.1 Initialization

At the beginning of the study, in order to avoid the adverse effects on the recording quality of the KU-100 binaural microphone due to the voltage imbalance between the left and right channels. For example, regarding distortion and excess noise, I first ensured that the peak voltages of the two channels of the KU-100 were kept at the same level and not disturbed by excess noise.

This study used a good quality over-ear headset over the KU-100 binaural microphone. This measure is to ensure that the audio sensed in the initial environment of the microphone does not receive white noise from the natural environment that is outside the experimental expectations. White noise may cause deviations in voltage initialization (white noise may come from seemingly insignificant noise, such as the sound of a vent working).

As shown in Figure 1, I adjusted the microphone input knob on Presonus' AudioBox soundcard to adjust the voltage between the two channels of the microphone. The research expects to minimize interference with the binaural microphone in the real room.

2.2 Research Environment

The MMT lab room in Trinity College Dublin's Stack B building was the experimental location for the research, which the recordings were recorded at night and during unoccupied hours to ensure that the environment was as free as possible from distractions other than natural noise. The only

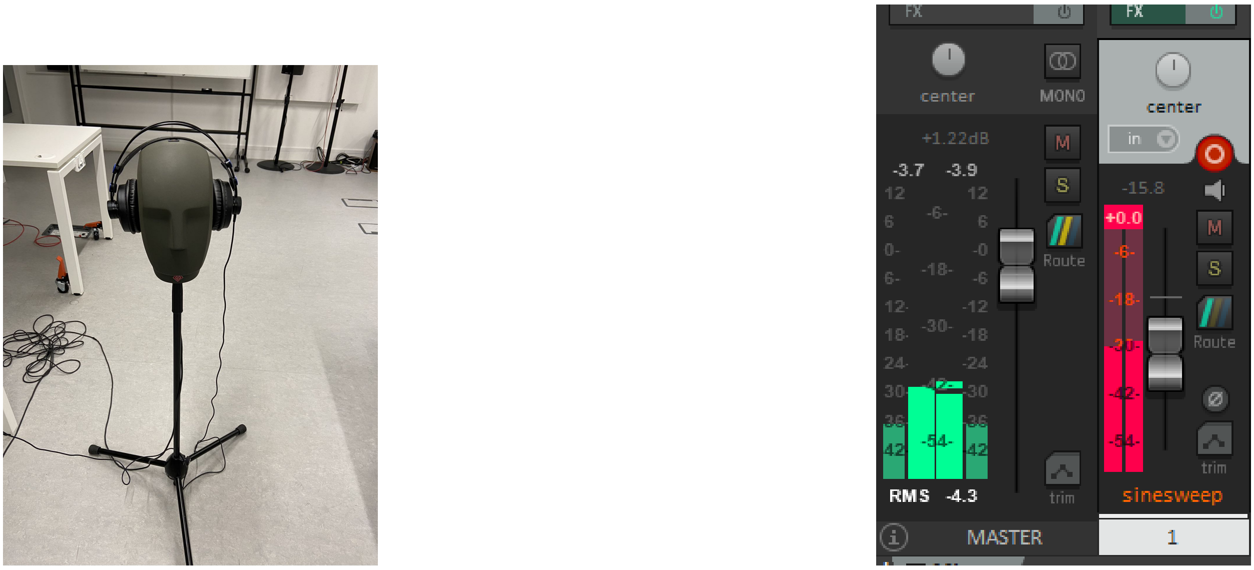


Figure 1: *The KU-100 in real environment (Left) and the peak of 2 channels in Reaper (Right)*

noise that affected the experiments came from the unclosable vents in the MMT lab. The room plan of MMT Lab is as follows:

Where the circle in the upper left corner represents the binaural microphone, whose position remains fixed. The remaining circles represent the speaker placement, and the individual circles represent the location of the recorded audio at one time. In this study, I recorded a total of five different versions of audio in two different directions.

The specific dimensions of the MMT lab can be observed in Figure 2, which is a rectangle of 8.1M in width and 14.4M in length. And the room contains a hexagonal recording room, and the single side of the recording room measures about 2 meters.

The large number of tables, chairs, and equipment placed in the MMT lab resulted in possible interference with the sound propagation path. When setting up the experimental environment, I minimized the number of obstructions or other potential interferences between the microphone and the speaker to avoid as much as possible the influence of the environment on the binaural stereo audio. Specific placement images will be given in other sections.

2.3 Devices Introduction

The Neumann KU-100 binaural microphone and a laboratory-supplied Eurolive B205D loudspeake were used to record binaural stereo audio and to play back the source audio. In addition, we connected the above equipment via a Presonus' AudioBox soundcard to ensure that the external system ran the equipment in good quality.

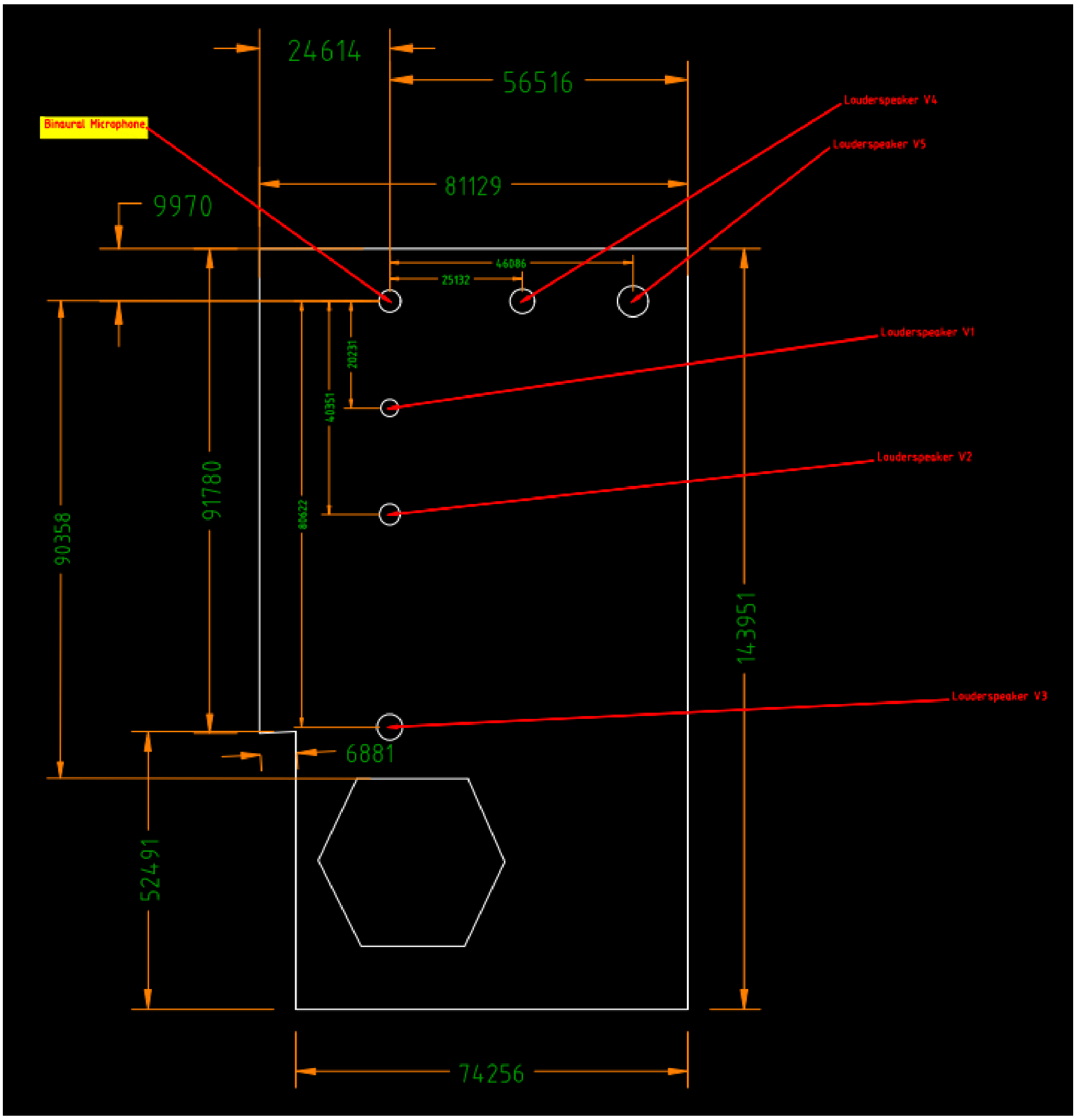


Figure 2: *Room structure plan, Loudspeaker and Microphone Setup*

2.4 Recording source location

I chose two directions with fewer obstacles to record stereo audio, one directly in front of the binaural microphone at 0 degrees and one in a straight line to the left ear of the microphone (90 degrees to the left ear or 270 degrees with 0 degrees directly in front), a brief diagram of the setup structure can be traced back to Figure 2.

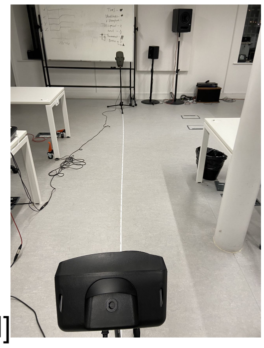
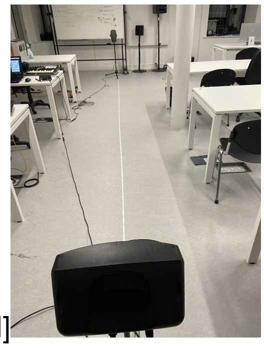
[2M]  [4M]  [8M] 

Figure 3: *2M, 4M and 8M at the 0 dgrees direction*

The first recording location was 2 meters directly in front of the binaural microphone KU-100 (I assume this direction is 0 degrees). I also recorded stereo audio at 4m and 8m respectively. The exact location is shown in the figure 3.

In addition to recording three sets of audio in the 0-degree direction, I also recorded two sets of audio in the 270-degree direction at distance intervals of 2.5 m and 4.6 m, respectively, as shown in the figure4.

2.5 Recording Processing

Two audio clips were used to generate the binaural chamber impulse response for this study to test the audio for realistic sounds heard by the human ear in the room. The first is a dryDrums.wav file provided by Dr. Enda Bates, a single channel sound file and an anechoic sample containing only the dry signal with no reverberation. The second audio is a sine test tone called sinesweep generated by the Reaper plugin. The sample rate of 2 audio files are 48000Hz while the drums file bit depth is 16 and all the others are 24.

The sine sweep provides a frequency range with a high signal-to-noise ratio, which excites the signal in the room to resonate and thus more accurately measures the true response of the human ear in the room.

I then recorded binaural audio of the stereo versions of each of these two audio clips at the locations mentioned in section 2.5 and all the files are shown in the figure 6

Figure 4: 2.5M and 4.6M in the 270 degrees(90 degrees to the left)

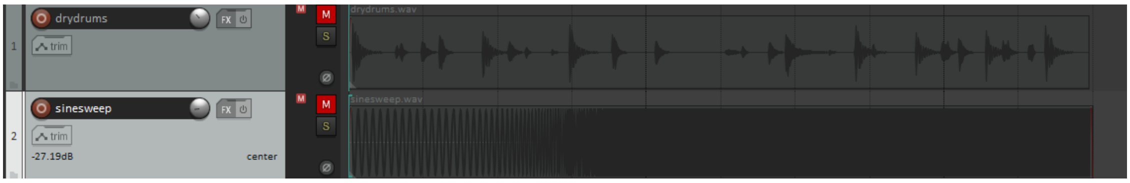


Figure 5: Waveform of Sine sweep and drydrums audio

2.6 BRIR generation and Rendering audio

BRIR generates binaural recordings which the information contains room directs signals, early reflections and reverberations by mathematically deconstructing the two channels of a two-channel stereo. This function can be easily implemented in Reaper via a plug-in to obtain BRIR files. In addition, I convolve the BRIR file with the audio of the anechoic samples within Reaper to finally generate an audio file that characterizes the features in the room as shown in figure7.

3 Result

In this section, the study will analyze the captured binaural room impulse responses through the correlation of different features with perceptual parameters. Examples include directional cues (interaural time difference (ITD) and interaural level difference (ILD)) and distance cues (time and level of direct signals, time and level of significant early reflections and their relation to room geometry and surface) [1].

Directional cues are auditory cues that help us determine the direction or location of a sound

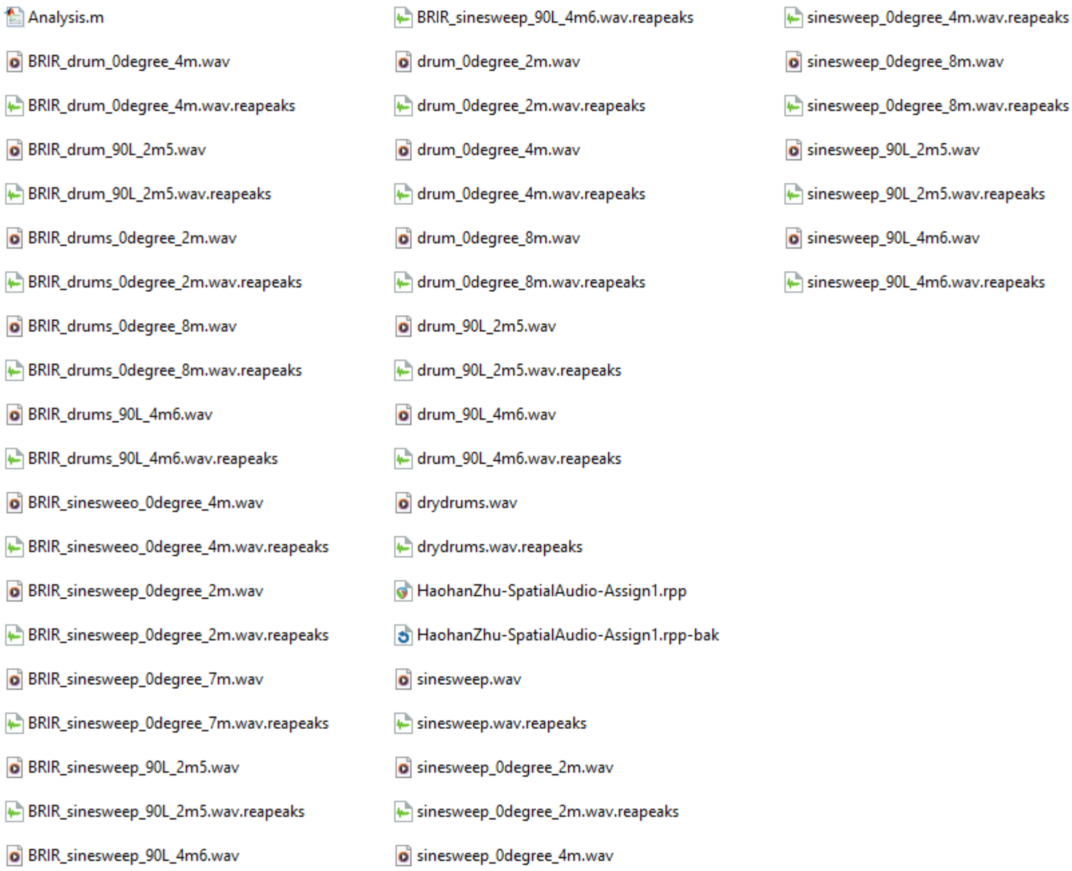


Figure 6: *Recording Files*

source. There are two main types of directional cues: interaural time difference (ITD) and interaural level difference (ILD). ITD is the difference in the time it takes for a sound to reach each ear, while ILD is the difference in the level or intensity of a sound as it reaches each ear. These cues are important because they allow us to determine the horizontal location of the sound source.

Distance cues, on the other hand, provide information about the distance or proximity of the sound source. There are several different types of distance cues, including the timing and level of direct signals, the timing and level of significant early reflections, and their relationship to room geometry and surfaces. The direct signal is the sound wave that travels directly from the source to the listener, while the early reflections are the first reflections of the sound bouncing off the ambient surface before it reaches the listener. These cues are important because, in addition to horizontal position, they allow us to perceive the depth or distance of the sound source.

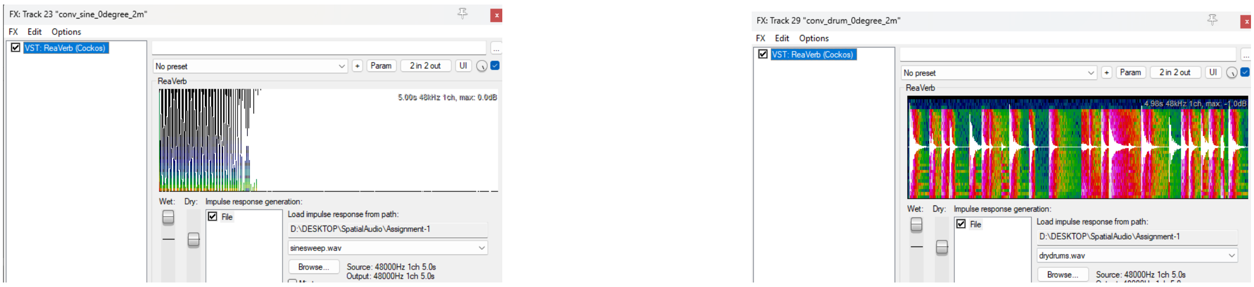


Figure 7: *Rendered Convoluted Sample Audio*

3.1 Sine sweep at 0 degree direction and distance of 2 meters

In the case of the relative distance between the audio source and the microphone at an angle of 0 degrees and a distance of 2 meters, it can be concluded from the graph8 that the delay between the left and right channels is negligible.

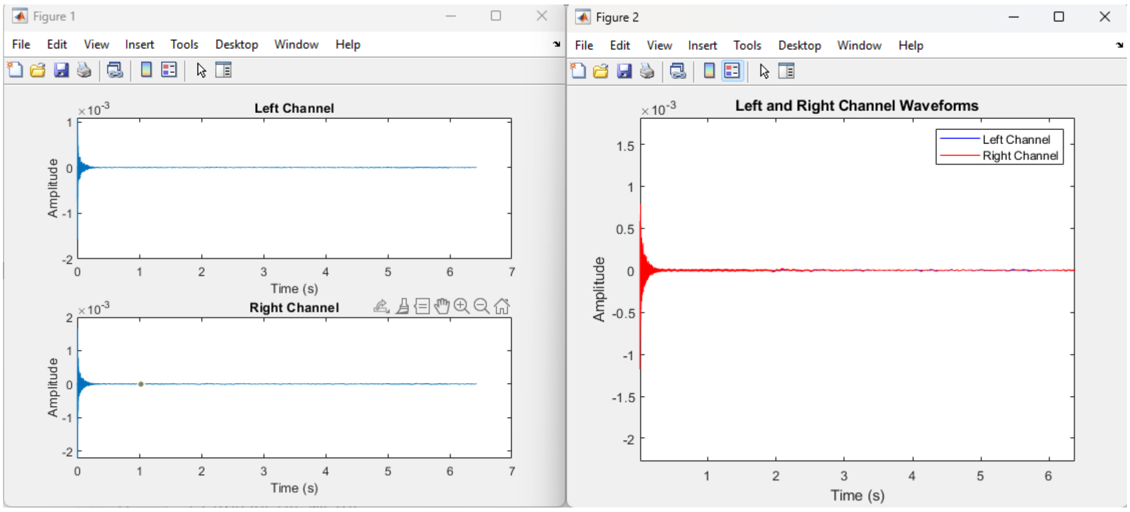


Figure 8: *Sine sweep at 0 degree direction and distance of 2 meters two channels*

In this case, this means that the left and right binaural room impulse responses (BRIRs) are very similar or that there is no significant delay between the two signals as shown in the figure 9.

By using distance cues we can find that the sound is reflected several times in the room due to too many obstacles in the room. This happened in almost every one of my audio files **??**.

3.2 Sine sweep at 0 degree direction and distance of 4 meters

The result of Sine sweep at 0 degree direction and distance of 4 meters as shown follow in figure 11 and figure 12

```
>> Analysis
ITD = 0 seconds
ILD = -3.4289 dB
```

Figure 9: *Sine sweep at 0 degree direction and distance of 2 meters ITD and ILD*

```
Direct signal time: 0.0061875 s
Direct signal level: 0.001574 Pa
Early reflection times:
  0.0061875 s
  0.006125 s
  0.0062917 s
Early reflection levels:
  0.001574 Pa
  0.0014762 Pa
  0.0010891 Pa
```

Figure 10: *Sine sweep at 0 degree direction and distance of 2 meters distance cues*

3.3 Sine sweep at 0 degree direction and distance of 8 meters

The result of Sine sweep at 0 degree direction and distance of 8 meters as shown follow in figure 13 and figure 14

3.4 Sine sweep at 270 degree direction and distance of 2.5 meters

The result of Sine sweep at 270 degree direction and distance of 2.5 meters as shown follow in figure 15 and figure 16

3.5 Sine sweep at 270 degree direction and distance of 4.6 meters

The result of Sine sweep at 270 degree direction and distance of 4.6 meters as shown follow in figure 17 and figure 18

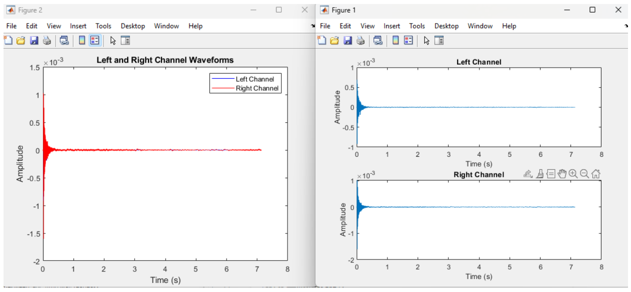


Figure 11: *Sine sweep at 0 degree direction and distance of 4 meters two channels*

```
>> Analysis
ITD = 0 us
ILD = -3.8979 dB
Direct signal time: 0.014646 s
Direct signal level: 0.00091839 Pa
Early reflection times:
  0.014646 s
  0.018437 s
  0.014583 s
Early reflection levels:
  0.00091839 Pa
  0.00091827 Pa
  0.00083995 Pa
```

Figure 12: *Sine sweep at 0 degree direction and distance of 4 meters directional cues and distance cues*

3.6 daydrums at 0 degree direction and distance of 2 meters

The result of daydrums at 0 degree direction and distance of 2 meters as shown follow in figure 19 and figure 20

3.7 daydrums at 0 degree direction and distance of 4 meters

The result of daydrums at 0 degree direction and distance of 4 meters as shown follow in figure 21 and figure 22

3.8 daydrums at 0 degree direction and distance of 8 meters

The result of daydrums at 0 degree direction and distance of 8 meters as shown follow in figure 23 and figure 24

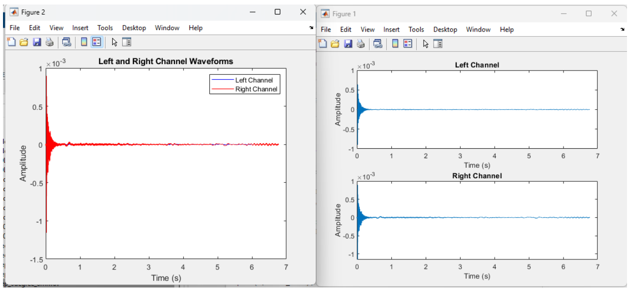


Figure 13: *Sine sweep at 0 degree direction and distance of 8 meters two channels*

```
>> Analysis
ITD = 2.0833e-11 us
ILD = -3.4743 dB
Direct signal time: 0.023583 s
Direct signal level: 0.00089955 Pa
Early reflection times:
  0.023583 s
  0.023979 s
  0.025958 s
Early reflection levels:
  0.00089955 Pa
  0.00063825 Pa
  0.00062764 Pa
```

Figure 14: *Sine sweep at 0 degree direction and distance of 8 meters directional cues and distance cues*

3.9 daydrums at 270 degree direction and distance of 2.5 meters

The result of daydrums at 270 degree direction and distance of 2.5 meters as shown follow in figure 25 and figure 26

3.10 daydrums at 270 degree direction and distance of 4.6 meters

The result of daydrums at 270 degree direction and distance of 4.6 meters as shown follow in figure 27 and figure 28

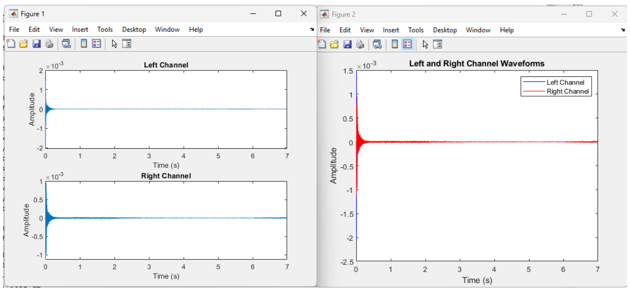


Figure 15: *Sine sweep at 270 degree direction and distance of 2.5 meters two channels*

```
>> Analysis
ITD = -7.7083e-10 us
ILD = -0.79651 dB
Direct signal time: 0.0072917 s
Direct signal level: 0.0020516 Pa
Early reflection times:
  0.0072917 s
  0.0076667 s
  0.0073542 s
Early reflection levels:
  0.0020516 Pa
  0.0014684 Pa
  0.0013528 Pa
```

Figure 16: *Sine sweep at 270 degree direction and distance of 2.5 meters directional cues and distance cues*

4 Conclusion

From the above experimental results, the following conclusions were drawn from this study. In a confined environment with many obstructing objects, sound tends to produce multiple reflection points on the propagation path and thus interferes with the reception of sound in the room by both ears. The difference between the signals received by the two ears is smaller when the direction of the sound source is directly opposite and the line formed by the two ears is parallel to the sound source. However, when the binaural deviation from the source is large, the signal difference is proportional to the distance.

In addition, in a room clutter environment, the sound path is reflected many times on an irregular plane and the time of reflection to the human ear varies from path to path. For example, in this experiment with a distance of 2 and 4 meters in front of 0 degrees, the 4-meter experimental

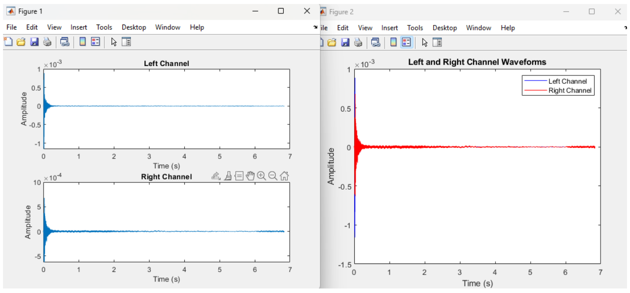


Figure 17: *Sine sweep at 270 degree direction and distance of 4.6 meters two channels*

```
>> Analysis
ITD = -6.875e-10 us
ILD = -0.52497 dB
Direct signal time: 0.017229 s
Direct signal level: 0.001159 Pa
Early reflection times:
  0.017229 s
  0.013104 s
  0.021792 s
Early reflection levels:
  0.001159 Pa
  0.0010797 Pa
  0.00088525 Pa
```

Figure 18: *Sine sweep at 270 degree direction and distance of 4.6 meters directional cues and distance cues*

group had a load-bearing column in front of the right side of the sound source, resulting in a greater degree of interference with the binaural recordings of the 4-meter experimental group comparing with 2-meter.

Bibliography

[1] LOISELLE, L. H., DORMAN, M. F., YOST, W. A., COOK, S. J., AND GIFFORD, R. H. Using ild or itd cues for sound source localization and speech understanding in a complex listening environment by listeners with bilateral and with hearing-preservation cochlear implants. *Journal of Speech, Language, and Hearing Research 59*, 4 (2016), 810–818.

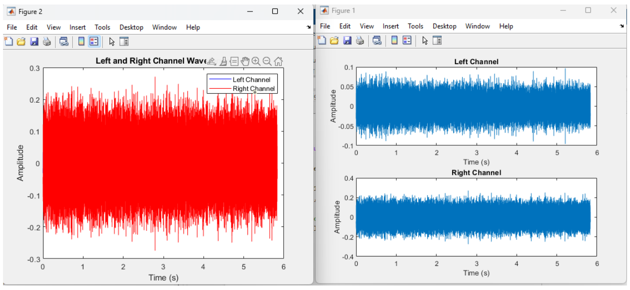


Figure 19: *daydrums at 0 degree direction and distance of 2 meters two channels*

```
ITD = 4.1667e-11 us
ILD = -10.2478 dB
Direct signal time: 5.2099 s
Direct signal level: 0.096692 Pa
Early reflection times:
  5.2099 s
  4.4128 s
  0.7439 s
Early reflection levels:
  0.096692 Pa
  0.089532 Pa
  0.088734 Pa
```

Figure 20: *daydrums at 0 degree direction and distance of 2 meters directional cues and distance cues*

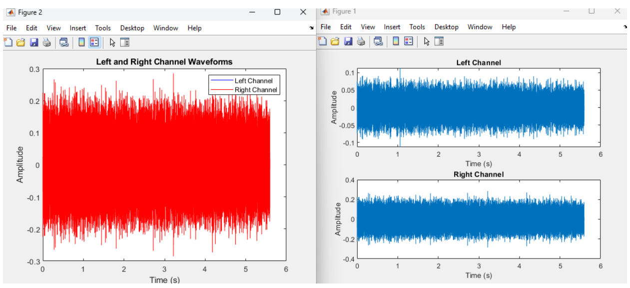


Figure 21: *daydrums at 0 degree direction and distance of 4 meters two channels*

```
ITD = -1.4917e-08 us
ILD = -9.6698 dB
Direct signal time: 1.0595 s
Direct signal level: 0.11301 Pa
Early reflection times:
  1.0595 s
  0.13737 s
  0.52375 s
Early reflection levels:
  0.11301 Pa
  0.090905 Pa
  0.09028 Pa
```

Figure 22: *daydrums at 0 degree direction and distance of 4 meters directional cues and distance cues*

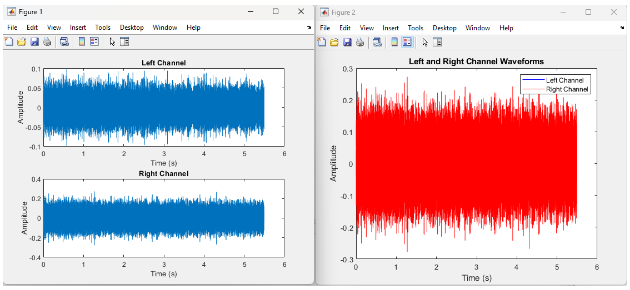


Figure 23: *daydrums at 0 degree direction and distance of 8 meters two channels*

```
>> Analysis
ITD = 4.5208e-08 us
ILD = -9.4558 dB
Direct signal time: 0.582 s
Direct signal level: 0.10129 Pa
Early reflection times:
  0.582 s
  2.5213 s
  3.0906 s
Early reflection levels:
  0.10129 Pa
  0.095327 Pa
  0.094734 Pa
```

Figure 24: *daydrums at 0 degree direction and distance of 8 meters directional cues and distance cues*

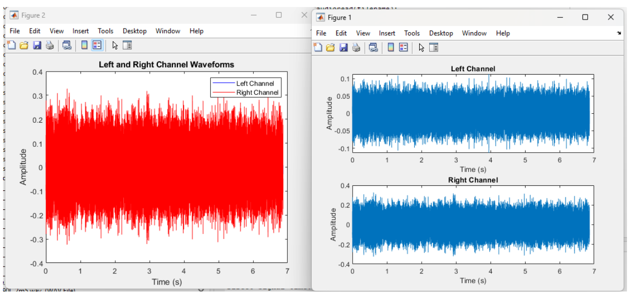


Figure 25: *daydrums at 270 degree direction and distance of 2.5 meters two channels*

```
>> Analysis
ITD = 2.1667e-08 us
ILD = -9.4795 dB
Direct signal time: 3.9523 s
Direct signal level: 0.1137 Pa
Early reflection times:
  3.9523 s
  1.3201 s
  1.3201 s
Early reflection levels:
  0.1137 Pa
  0.10929 Pa
  0.10713 Pa
```

Figure 26: *daydrums at 270 degree direction and distance of 2.5 meters directional cues and distance cues*

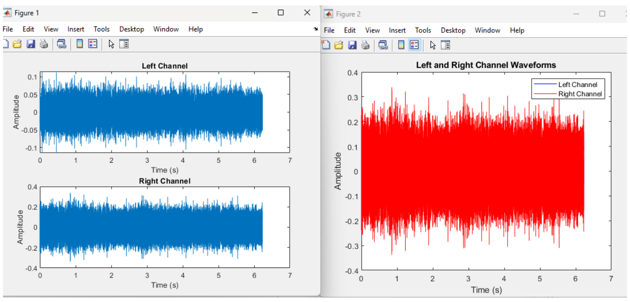


Figure 27: *daydrums at 270 degree direction and distance of 4.6 meters two channels*

```
>> Analysis
ITD = -2.1958e-08 us
ILD = -10.5272 dB
Direct signal time: 0.063229 s
Direct signal level: 0.11516 Pa
Early reflection times:
  0.063229 s
  0.4616 s
  4.0857 s
Early reflection levels:
  0.11516 Pa
  0.11266 Pa
  0.10636 Pa
```

Figure 28: *daydrums at 270 degree direction and distance of 4.6 meters directional cues and distance cues*