INDIAN INSTITUTE OF TECHNOLOGY KANPUR
Department of Electrical Engineering
2024-25

EE496: Undergraduate Project - I

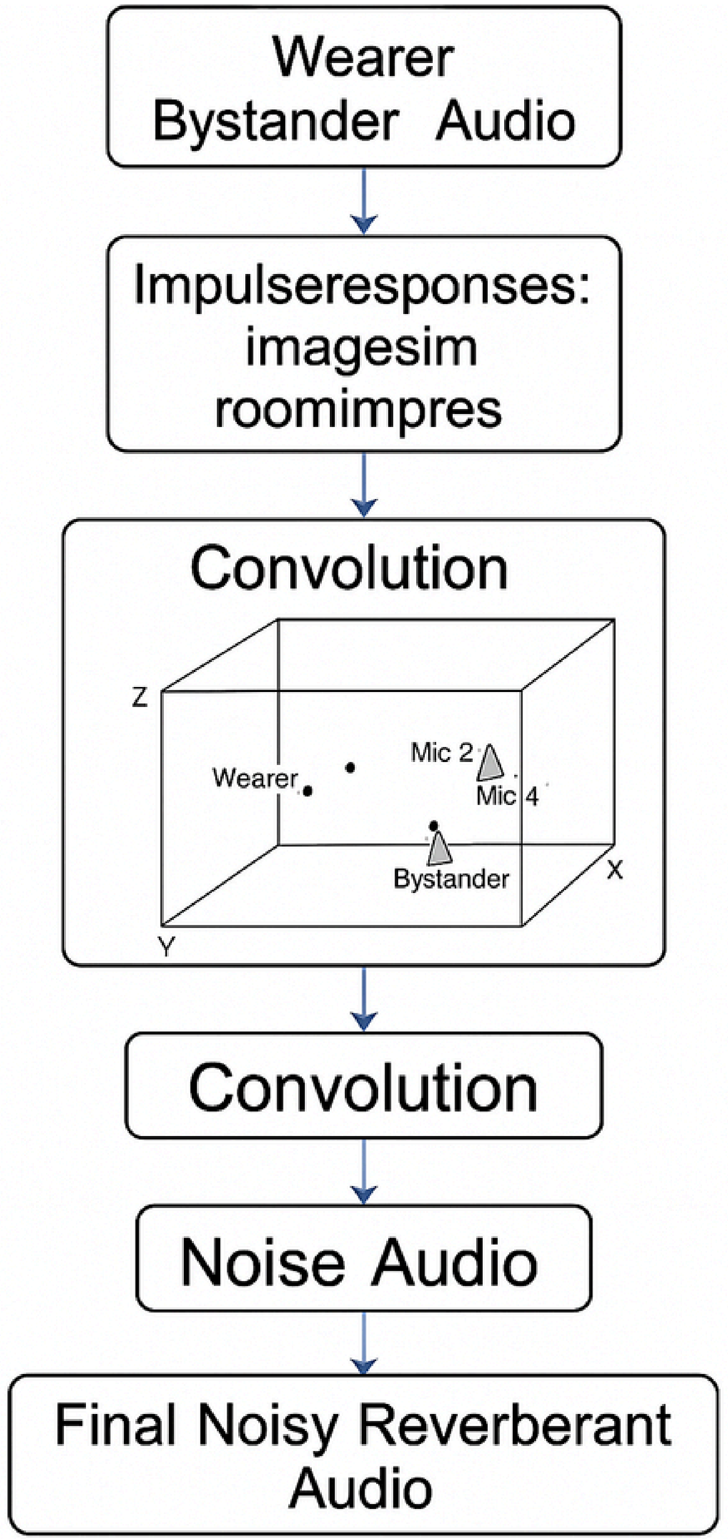
SIMULATING MULTI-CHANNEL SPATIAL AUDIO IN REVERBERANT ROOMS

Presented By:
Ayush Dixit (220262)   Ashwani Kumar Singh (220244)

Submitted To:
Prof. Rajesh M. Hegde

TABLE OF CONTENTS

INTRODUCTION



01 Project Overview

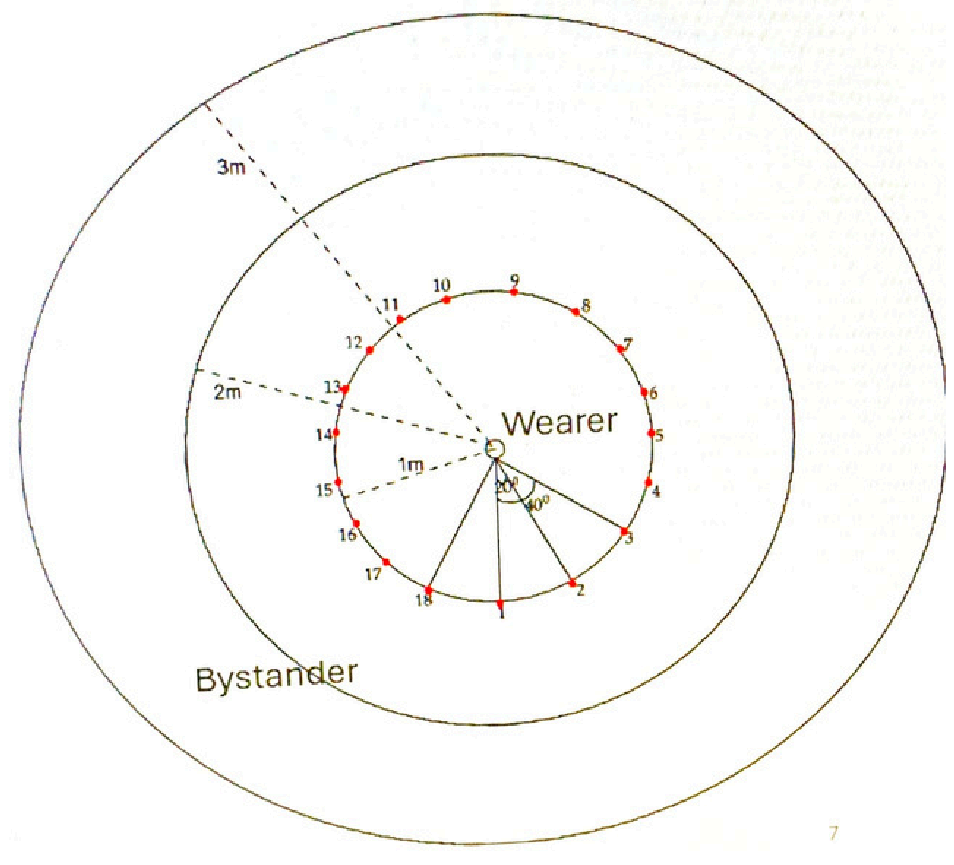
- Simulated multi-channel audio recordings in a 3D virtual room using Python and MATLAB tools.
- Modeled both wearable and external audio sources with varying positions to capture spatial acoustic effects.
- Generated high-quality, reverberant multi-channel datasets for testing beamforming and source localization techniques.
- Focused on replicating realistic room acoustics to enhance audio processing and localization research.

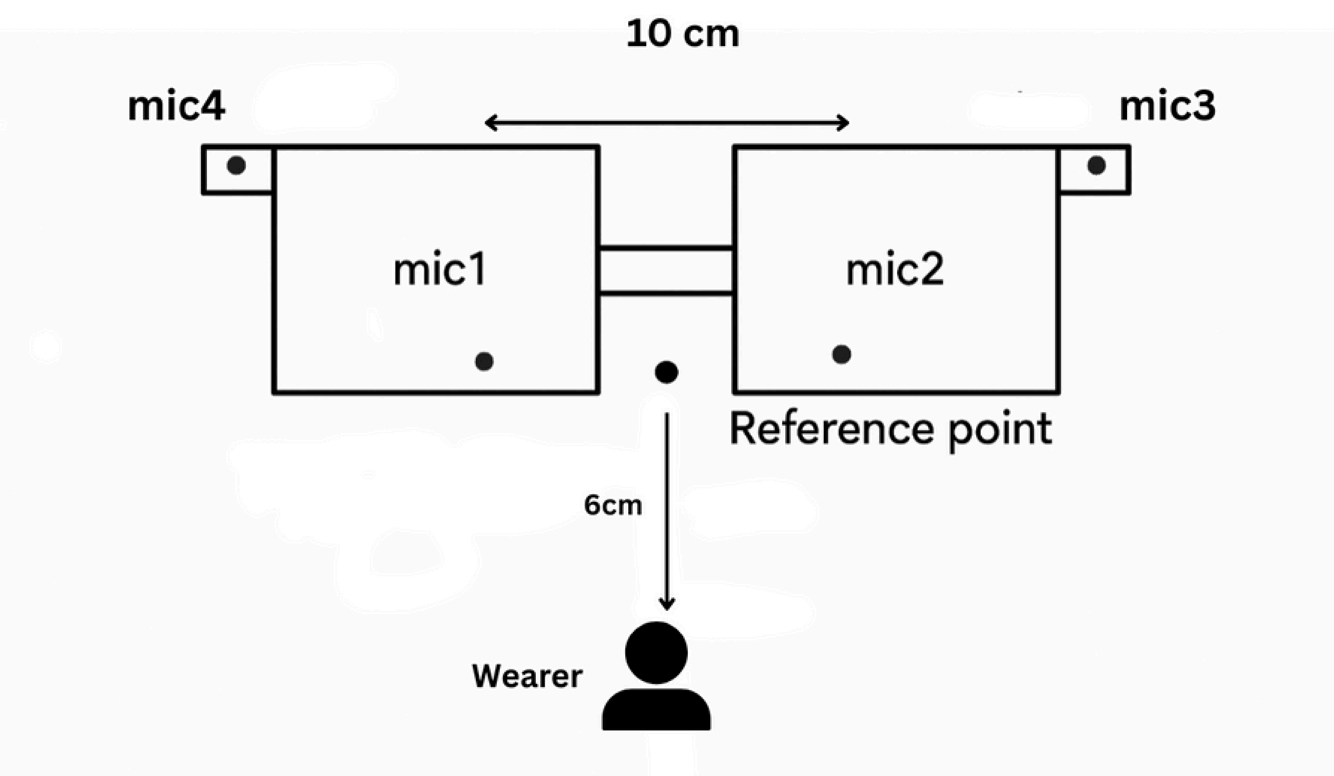
INTRODUCTION

02 Significance of the Study

- Developed a flexible audio simulation framework for replicating real-world multi-microphone recordings.
- Enabled controlled variation of source distances and angles to create diverse spatial configurations.
- Supported accurate modeling of room acoustics essential for advancing multi-channel audio processing research.

INTRODUCTION





03 Objectives

- Simulate a 3D acoustic room with wearable and external audio sources.
- Generate Room Impulse Responses (RIRs) using the image-source method.
- Perform convolution of clean speech signals with RIRs to simulate reverberation.
- Mix audio streams and add environmental noise to enhance realism.
- Save multi-channel outputs for further analysis, modeling, and machine learning applications.

TOOLS AND LIBRARIES

MATLAB:

- **Array Toolbox:** For simulating room acoustics using functions like imagesim and roomimpres.
- **Audio Toolbox:** For audio file handling and processing.
- **Custom Scripts:** For audio normalization, convolution, mixing, and file management.

.

TOOLS AND LIBRARIES

PYTHON:

- **Python (Pyroomacoustics):** For simulating room acoustics and generating Room Impulse Responses (RIRs).
- **Librosa:** For loading and processing clean speech audio signals.
- **Custom Scripts:** For signal convolution, gain control, file output, and noise mixing.

LibriSpeech Dataset: To take clean audios and use them to generate convolved audios.

MICROPHONES & SOURCES CONFIGURATIONS

ROOM CONFIGURATIONS:

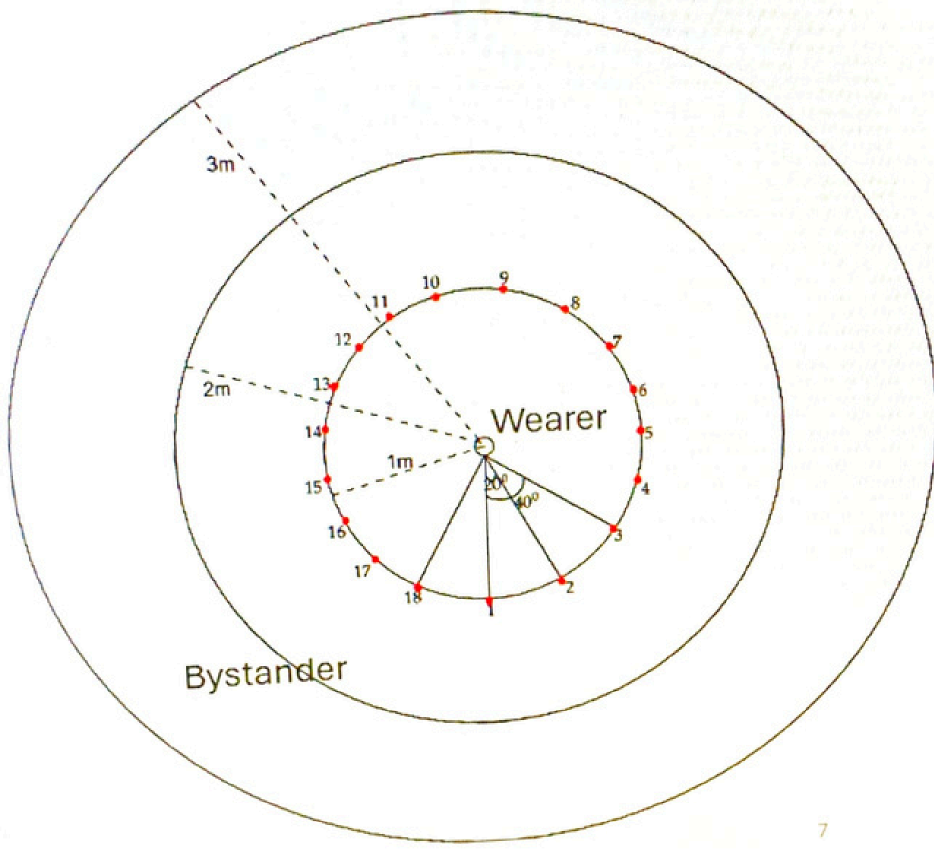
The room dimensions are set as 5m (length) x 4m (width) x 6m (height), centered at the origin to simplify coordinate calculations.

MICROPHONE CONFIGURATIONS:

The microphone array consists of four microphones placed as follows:

- **Mic 1:** 5 cm to the left of reference point (-0.05, 0, 0)
- **Mic 2**: 5 cm to the right of reference point (0.05, 0, 0)
- **Mic 3:** Positioned at (0.08, 0.45, 0.04)
- **Mic 4:** Positioned symmetrically at (-0.08, 0.45, 0.04)

The wearer (audio source) is slightly offset from Mic 1 to simulate real-world wearable conditions.



SOURCES:

- Wearer audio: primary speaker
- Bystander audio: secondary speaker
- Noise samples: from FSDnoisy18k dataset

SPATIAL AUDIO GENERATION USING MATLAB

Impulse Response (IR) Generation:

- For each mic & source (wearer/bystander), IRs generated using:
  imagesim: Computes delays and scaling using image-source method.
  roomimpres: Converts these to time-domain IRs.
- IRs normalized, then padded/truncated to 640 samples for uniformity
- Bystander source moved across 3 radii × 19 azimuth angles

Output Generation:

- For each (radius, angle) config:
- 4 WAV files (one per mic)
- 1 MAT file containing the full 4-channel output
- Filenames indicate config:
- Audio10_output_r{r}_phi{φ}_mic{n}.wav
- Total: 50+ hrs of spatial, reverberant, noisy audio

Audio Convolution & Mixing:

- Wearer and bystander audios convolved per mic using 1D conv
- Convolved signals summed:
  Outputmic=(W ⋆ IRW)+(B ⋆ IRB)
- Added randomly selected environmental noise (scaled by 0.0002)
- Applied fixed gain factor = 10 to maintain audio loudness
- All signals zero-padded to match max length

SPATIAL AUDIO GENERATION USING PYTHON

Room Definition – pra.ShoeBox()

Creates a 3D rectangular room with specified dimensions and wall absorption.

Microphone Array Setup – pra.MicrophoneArray()

- Defines location(s) of microphone(s) within the room.
- Can simulate mono or stereo setups.

Source Placement – room.add_source()

- Places sound source(s) at desired coordinates.
- Multiple sources can simulate real-world scenarios.

Impulse Response Computation – room.compute_rir()

- Uses the Image Source Method (ISM) to compute Room Impulse Responses (RIRs).
- Captures echo paths, reflections, and reverberations in the room.

Convolution with Clean Audio – fftconvolve()

- Convolves dry audio with RIRs to simulate reverberated output.
- Performed for each mic-channel (mono or stereo).

Output Generation – sf.write()

Saves the processed audio (with reverberation) as a WAV file.

OUTCOMES

Simulation Success

- Realistic multi-channel room environments were simulated using virtual room modeling.
- Room impulse response (RIR) were generated for both wearer and bystander positions.

Dynamic Source Modeling

- Simulated dynamic source movement by varying radius and azimuth angles.
- Capturing spatial variation across different microphone placements in a controlled acoustic environment.

Audio Realism & Processing

- Mixed clean speech with reverberation + real-world noise
- Applied energy normalization for consistent loudness and perceptual clarity.
- Ensured all signals are properly aligned and padded.

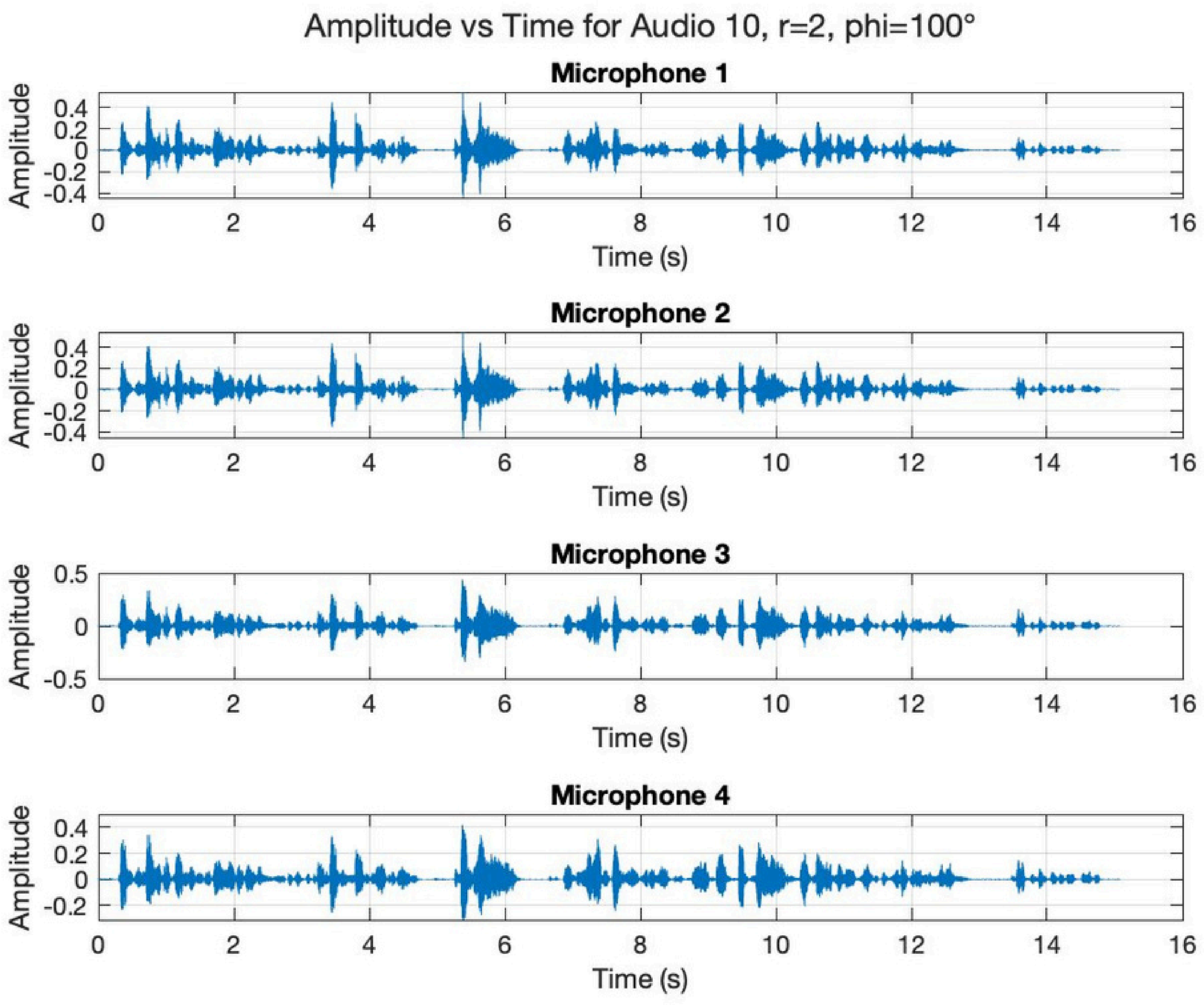
Dataset Generation

- Produced a large, labeled multi-channel dataset.
- Compatible outputs for both MATLAB and Python pipelines.
- Ready for tasks like source localization and beamforming.

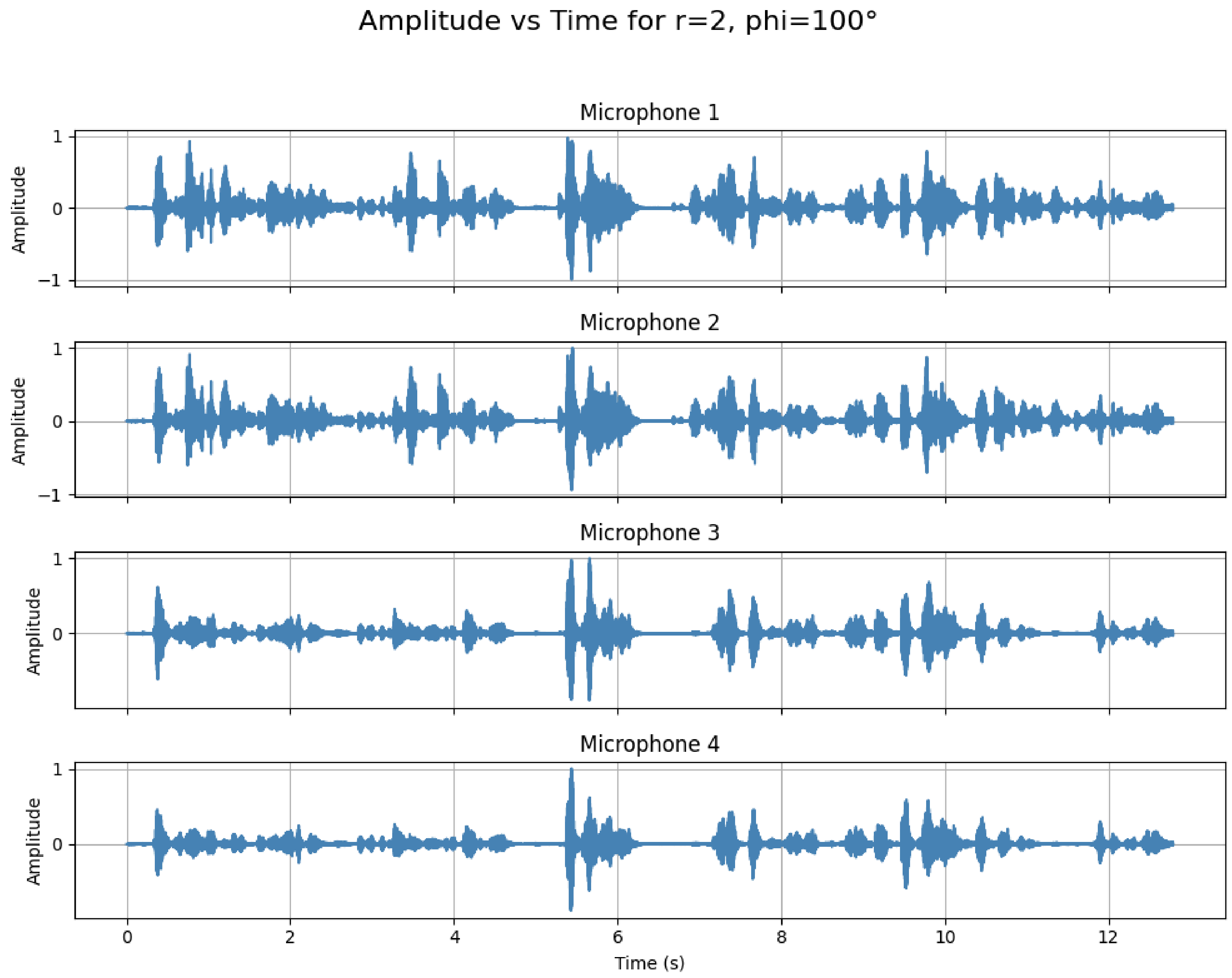
COMPARING MATLAB AND PYTHON OUTPUTS

We created amplitude plots to compare the output audio signals received at the four microphone positions. The plots showed largely similar shapes, with primary differences in amplitude levels. This variation is attributed to the shift of the room's origin to the center, which significantly reduced the loudness

- **MATLAB**



- **Python**



CONCLUSION

- Successfully utilized both Python and MATLAB for simulating realistic multi-channel audio environments.
- In Python, used libraries like pyroomacoustics, librosa, and numpy to generate Room Impulse Responses (RIRs), process clean audio signals, and create multi-channel spatial audio datasets.
- In MATLAB, employed the Array Toolbox to streamline the simulation process by efficiently generating impulse responses and allowing flexible placement of sources and microphones.
- Visualized and validated the simulated outputs through waveform plotting and analysis.
- The combined work in Python and MATLAB demonstrates a comprehensive approach to spatial audio simulation, resulting in high-quality datasets for advancing research in beamforming, speech enhancement, and source localization.

🔗 IMPORTANT LINKS :

- **Report -** https://docs.google.com/document/d/1XTBYBJc7H0ChtZNyHxCPQuREz5mz6lXPivrc4kKfOMw/edit?tab=t.0#heading=h.xr4cp7rumklw
- **MATLAB and Python code -** https://docs.google.com/document/d/1wTK2ZzMb_zRqbhElA3NixAZSNK0AzcGN1yUQNSFOzC0/edit?tab=t.0
- **Input Audio Files -** https://docs.google.com/spreadsheets/d/1zKVOFhMgzVvH5nOMFsRdMvYvvKgufhiGCdtnzWVZQ9I/edit?gid=861775566#gid=861775566
- **Generated Data -** https://iitk-my.sharepoint.com/:f:/g/personal/jsuraj_iitk_ac_in/EuD50DRfx2ZDvxOI8tMKnbwBN21QUvkg4Z41uwbunL2YPQ?e=NxUY63
- **Array tool box functions -** https://docs.google.com/spreadsheets/d/1rpVMOPcqYDynCNNqSlesEpcA1rgvePTLnEkTSx1RhqI/edit?gid=0#gid=0

THANK YOU

INDIAN INSTITUTE OF TECHNOLOGY KANPUR
Department of Electrical Engineering
2024-25