Enhancing Auralization and Sonification through Decay Envelope Manipulation of Room Impulse Responses

Project-II (EE47004) report submitted to

Indian Institute of Technology Kharagpur

in partial fulfilment for the award of the degree of

Bachelor of Technology

in

Electrical Engineering

by

Rudransh Gupta

(20EE38023)

Under the supervision of

Professor Anirban Mukherjee

Department of Electrical Engineering

Indian Institute of Technology Kharagpur

Spring Semester, 2023-24

May 03, 2024

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to certify that the project report entitled "**Enhancing Auralization and Sonification through Decay Envelope Manipulation of Room Impulse Responses**" submitted by **Rudransh Gupta** (Roll No. 20EE38023) to Indian Institute of Technology Kharagpur towards partial fulfilment of requirements for the award of degree of Bachelor of Technology in Electrical Engineering is a record of bona fide work carried out by him under my supervision and guidance during Spring Semester, 2023-24.

Date: May 03, 2024
Place: Kharagpur

Professor Anirban Mukherjee
Department of Electrical Engineering
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Abstract

Name of the student: **Rudransh Gupta** Roll No: **20EE38023**

Degree for which submitted: **Bachelor of Technology**

Department: **Department of Electrical Engineering**

Thesis title: **Enhancing Auralization and Sonification through Decay Envelope Manipulation of Room Impulse Responses**

Thesis supervisor: **Professor Anirban Mukherjee**

Month and year of thesis submission: **May 03, 2024**

Room Impulse Responses (RIRs) are essential in the field of acoustics, as they are extensively employed to assess the acoustic characteristics of settings accurately and to create auditory simulations of a room's acoustics, a process known as auralization. This paper presents novel signal processing strategies to optimise the efficiency of Room Impulse Responses (RIRs) in convolution-based room simulations, a specific type of auralization. These techniques utilise ambient background noise to lengthen the decay phases of room impulse responses (RIRs), enhancing the auditory experience with a longer echo effect.

Moreover, the study investigates modifications to the decay rate in Room Impulse Responses (RIRs) to enhance the detectability of tiny acoustic characteristics. This allows for better control over the acoustic properties of a room and improves the ability to perceive small differences. The progress made in this field is especially important for RIR sonification, which aims to convert acoustic data into sound representations that are both more analytically valuable and simpler to understand.

Acknowledgements

I would like to express my sincere gratitude to all those who have contributed to the completion of this work. Special thanks go to Professor Anirban Mukherjee for their invaluable guidance, support, and insightful feedback throughout the entire process.

I am also thankful for the Indian Institute of Technology Kharagpur for providing the necessary resources and environment for this research. I extend my appreciation to my colleagues and peers who have shared their expertise and perspectives, enriching the depth of this project.

I am grateful to my friends and family for their unwavering encouragement and understanding during the challenges and demands of this endeavor. Their support has been a constant source of motivation.

Lastly, I want to express my appreciation to the broader academic and research community whose work has laid the foundation for this study. The collaborative spirit and shared knowledge within this community have significantly contributed to the advancement of our understanding in this field.

Thank you all for being an integral part of this journey.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Room Impulse Responses (RIRs) are a crucial tool in representing the acoustic properties of a space, detailing how sound travels from a source to a receiver within a room. These RIRs can be utilized to compute key acoustic metrics such as reverberation time, clarity index, and speech transmission index. Additionally, binaural acoustic parameters, like the interaural cross-correlation coefficient, can be extracted from binaural RIRs, and spatial parameters from multi-channel RIRs that capture responses from various directions. RIRs are integral in the process of auralization, which involves the convolution of these impulse responses with dry audio signals for playback or real-time applications.

Moreover, RIRs serve a significant role in sonification, a method similar to visualization but for auditory data, allowing listeners to perceive data characteristics through sound. This paper delves into advanced techniques for sonifying RIRs beyond merely playing back the unaltered responses. Methods explored include time reversal to enhance the detectability of early reflections, time-stretching to improve the perceptibility of discrete echoes through resampling or using phase vocoder techniques, and autoconvolution to amplify spectral features.

Our study specifically examines the decay envelope of RIRs, investigating innovative strategies to alter or entirely remove this feature, thereby expanding the potential applications and understanding of RIRs in acoustic analysis.

Chapter 2

Literature Review

2.1 Decay Manipulation

2.1.1 Extrapolating Reverberant decay through the noise floor

Numerous strategies exist to optimize the signal-to-noise ratio (SNR) in Room Impulse Response (RIR) measurements, but some level of background noise is often unavoidable. If the noise level is low enough, these RIRs can be suitable for auditory playback through convolution with unprocessed signals. However, in some cases, the presence of a noise floor may lead to discernible auditory artifacts, as observed in some of our listening experiments. Instead of simply cutting off or gradually fading out the noise floor, we engage in a more sophisticated technique called multi-band extrapolation, extending the reverberation into the noise-dominant regions.

2.1.2 Experimental Setup

The technique of segmenting the frequency spectrum into multiple bands proves essential due to the variability in decay rates and the signal-to-noise ratio (SNR) across different frequencies in measured Room Impulse Responses (RIRs). Commonly, the spectrum is divided into octave bands, employing a high-pass filter for the highest and a low-pass filter for the lowest frequencies. Typically, eight bands ranging from 125 Hz to 16 kHz suffice, but adjustments to nine or ten bands may be necessary if the low-frequency SNR is robust. To maintain phase consistency across these bands, zero-phase filtering via Python's Scipy 'filtfilt' function and 6th order Butterworth filters are utilized.

Identifying and applying decay slopes and noise floors involve a streamlined method where the decay envelope combines exponential decay with a constant noise level. By squaring octave band RIR values and smoothing these using a 4 Hz low-pass filter, we can prevent time shifts with zero-phase filtering. The filter's order adjusts according to the band, with lower frequencies typically requiring a higher order.

This method's advantage lies in fitting the function on a logarithmic (decibel) scale, preventing issues with extremely low sample values which could approach negative infinity. By identifying and normalizing the peak of the envelope to 0 dB, we then fit a curve described by the following equation:

$$\mathbf{L}(\mathbf{t}) = [10 \log_{10} \left(10^{\frac{at}{10}} + b\right)], \quad (2.1)$$

In this formula, t represents time in seconds, $L(t)$ models the decay curve, a determines the decay slope, and b indicates the noise floor level, set relative to the envelope's peak. The process smooths the initial peak, raising the relative noise floor.

To correct for discrepancies between a noise-free decay envelope and the modeled one, a gain compensation function $g(t)$ is formulated:

$$g(t) = \frac{10^{\frac{at}{10}} + b}{10^{\frac{at}{10}}} \quad (2.2)$$

This function $g(t)$ is not applied uniformly across the RIR to prevent undue smoothing of unaffected decay phases. Instead, it starts where the envelope is 10 dB above the noise floor, targeting only the impacted signal sections.

Despite the simplicity of the assumptions behind this function, it effectively analyzes typical recorded RIRs. While narrower or wider bands are possible, octave bands generally yield the best results. Issues might arise when dealing with exceptionally low SNR, non-exponential decay patterns, such as in coupled rooms, or non-uniform noise floors.

Addressing these issues requires selecting appropriate spectral bands and possibly modifying the curve-fitting approach to incorporate additional parameters for adjusting a secondary decay slope's rate and gain. In automated systems, differentiating between single and double-slope decays is crucial; adjustments should respect the underlying decay model. In cases of tonal background noise, substituting the real noise with synthetic noise that follows the derived exponential decay curve may be more effective.

2.2 Adjusting reverberation time

Adjusting the reverberation time in a Room Impulse Response (RIR) can be efficiently accomplished by applying an exponential modulation to the RIR. Increasing the reverberation time is achieved using a positive exponential factor, while decreasing it involves a negative factor. The exponential damping constant, δ, associated with a specific RIR band, is mathematically related to the reverberation time, T, as shown in the following equation:

$$\delta = \frac{\ln(10^6)}{2T} \approx \frac{6.91}{T} \quad (2.3)$$

This formula underpins the relationship by considering a perfect exponential decay model of the reverberation envelope, where x is the envelope amplitude, t is time in seconds, and x_0 is the amplitude at $t = 0$:

$$x^2(t) = x_0^2 e^{-2t\delta} \quad (2.4)$$

If a new reverberation time is desired, this relationship adjusts accordingly with a new damping constant δ_1:

$$y^2(t) = x_0^2 e^{-2t\delta_1} \quad (2.5)$$

To align the original RIR to this new reverberation time, the relationship between the squared envelopes before and after adjustment is used to determine $a(t)$:

$$x^2(t)e^{-2t\delta_1} = a^2(t)x^2(t)e^{-2t\delta_0} \quad (2.6)$$

From which, the necessary adjustment factor $a(t)$ is derived:

$$a^2(t) = e^{-2t(\delta_1 - \delta_0)} \quad (2.7)$$

Applying this factor adjusts the original RIR, represented by $x(t)$, to match the new target reverberation time:

$$y(t) = x(t)e^{-t(\delta_1 - \delta_0)} \quad (2.8)$$

In this formula, $x(t)$ is the initial RIR, and $y(t)$ is the RIR modified to have the adjusted reverberation time. This operation modifies the pressure amplitude directly, hence the exponent is not squared.

Decay rate adjustments are typically made using octave bands through the use of 6th order Butterworth zero-phase filters, enabling precise control over the reverberation time of each band without undue interaction. This ensures a realistic acoustic response without abrupt changes across frequency bands. Both finer and coarser spectral resolutions can be utilized depending on the application requirements.

It is particularly crucial to manage the extension of the reverberation decay past the noise floor when the reverberation time is increased. Failure to do so might result in the amplification of noise at the tail end of the RIR, leading to audible and unwanted distortions.

Chapter 3

Implementation

3.1 Extrapolating reverberant decay through the noise floor

The Python code below simulates the generation of a synthetic Room Impulse Response (RIR) with distinct reflections at specified times and amplitudes, incorporating an initial decay and adding Gaussian noise to mimic real-world acoustic scenarios. The RIR is then processed through octave band filters ranging from 125 Hz to 16 kHz to analyze its frequency components. For each band, the RIR is filtered, squared, and converted to a logarithmic scale (decibels) to assess the intensity of the response. To manage the influence of noise, especially in lower signal areas, the noise floor is defined and applied to ensure that no part of the RIR falls below this threshold. The results for each frequency band are visualized in subplots, comparing the original and adjusted RIRs, thus illustrating how the RIR behaves across different frequency bands and how it is affected by the noise floor settings. Additionally, the code adjusts the reverberation characteristics by applying specific high-pass or low-pass filters for the extreme frequency bands, allowing for more detailed control and analysis of the RIR's behavior within each octave band, thereby ensuring accurate acoustic analysis and simulation.

3.1.1 Pseudo-Code

Algorithm 1 Octave Band Filtering

$rir \leftarrow$ zeros_like(t)
$initial\_decay \leftarrow -0.005$
$reflections \leftarrow [0.05, 0.1, 0.15, 0.3, 0.8]$ ▷ Reflection times in seconds
$reflections\_amplitudes \leftarrow [1, 0.5, 0.25, 0.125, 0.06]$ ▷ Amplitudes of reflections
for $r, amp \in$ zip$(reflections, reflections_amplitudes)$ **do**
  $rir \mathrel{+}= amp \cdot e^{initial\_decay \cdot (t-r) \cdot fs} \cdot (t > r)$
end for
$rir \mathrel{+}=$ np.random.normal$(0, 0.0005,$ len$(rir))$ ▷ Adding some noise
function OCTAVE_BAND_FILTER$(data, center_freq, fs, bandwidth_oct)$
  $b, a \leftarrow$ butter$(1, [((2^{(bandwidth_oct/2)})^{-1}) \cdot center_freq, ((2^{(bandwidth_oct/2)}) \cdot center_freq]$, 'band', $fs)$
  return filtfilt$(b, a, data)$
end function
$octave\_bands \leftarrow [125, 250, 500, 1000, 2000, 4000, 8000, 16000]$
$fig, axs \leftarrow$ plt.subplots$(4, 2,$ figsize $= (12, 24))$
for $i, ax \in$ enumerate$(axs.$flatten$())$ **do**
  if $octave\_bands[i] = 125$ **then**
    $b, a \leftarrow$ butter$(1, octave_bands[i]/(fs/2),' low')$
  else if $octave\_bands[i] = 16000$ **then**
    $b, a \leftarrow$ butter$(1, octave_bands[i]/(fs/2),' high')$
  else
    $b, a \leftarrow$ butter$(1, [((2^{(1/2)})^{-1}) \cdot octave_bands[i], ((2^{(1/2)}) \cdot octave_bands[i]]$, 'band', $fs)$
  end if
  $filtered\_rir \leftarrow$ filtfilt(b, a, rir)
  $rir\_squared \leftarrow filtered\_rir^2$
  $rir\_db \leftarrow 10 \cdot \log_{10}(rir\_squared/\max(rir\_squared))$
  $noise\_floor\_db \leftarrow -80$
  $noise\_floor \leftarrow 10^{(noise\_floor\_db/10)}$
  $rir\_db\_adjusted \leftarrow 10 \cdot \log_{10}(\max(rir\_squared, noise\_floor))$
end for

3.2 Adjusting reverberation time

The code below performs several key operations to adjust the decay characteristics of a Room Impulse Response (RIR). Initially, it calculates the maximum amplitude of a smoothed envelope and establishes an initial decay rate based on a hypothetical reverberation time of 1.9 seconds. The code then measures the noise floor by averaging the envelope values towards the end of the signal. Using a target reverberation time of 4.0 seconds, it adjusts the decay rate of the RIR. This adjusted decay rate is applied to the filtered RIR to extend or modify its reverberation characteristics. Lastly, the adjusted RIR is converted to decibels, and a minimum threshold is set by extending the noise floor, ensuring that the decay does not fall below this level, which helps in maintaining auditory consistency and preventing any potential loss in the perceptible quality of the sound due to too low signal levels.

3.2.1 Pseudo-Code

Algorithm 2 Adjusting Decay Rate and Extending Noise Floor in RIR

$A \leftarrow \max(smoothed\_envelope)$

$initial\_decay\_rate \leftarrow -1/(fs \times 1.9)$ ▷ Initial estimated decay rate

$noise\_floor \leftarrow \mathrm{mean}(smoothed\_envelope[-\mathrm{int}(0.1 \times fs):])$

$adjusted\_decay\_rate \leftarrow -1/(fs \times 4.0)$ ▷ Target decay rate for 4.0s RT

$adjusted\_rir \leftarrow filtered\_rir\_octave \times e^{(adjusted\_decay\_rate - popt[0]) \times t}$

$adjusted\_rir\_db \leftarrow 10 \times \log_{10}(adjusted\_rir^2)$

$adjusted\_rir\_db\_extended \leftarrow \mathrm{where}(adjusted\_rir\_db < 20 \times \log_{10}(noise\_floor),$
$20 \times \log_{10}(noise\_floor), adjusted\_rir\_db)$

Chapter 4

Simulation and Results

In the subsequent analysis, the plots show how sound decays in different frequency bands within a simulated room environment, both with and without adjustment for a baseline noise floor. This analysis is crucial in fields like acoustics and audio engineering, where understanding and controlling the reverberation characteristics of a space are essential for achieving desired audio qualities.

The plot shows a series of comparisons between original Room Impulse Responses (RIRs) and their adjusted counterparts across several frequency bands. Each subplot appears to represent an octave band (or similar frequency-specific band), displaying the waveform of the RIR before and after processing.

Original vs Adjusted RIRs

Original RIR: The pink line represents the original room impulse response filtered to specific frequency bands. It captures how sound decays naturally within the frequency band of interest in a room. Adjusted RIR: The black line is the adjusted RIR where a noise floor is applied. This means that for very low sound levels that fall below a set threshold (in this case, -80 dB), the response is artificially maintained at this noise floor level rather than allowing it to continue decaying into the noise of the measurement system. This is useful for acoustic analysis as it prevents the influence of low-level noise from distorting the true decay characteristics of the room.

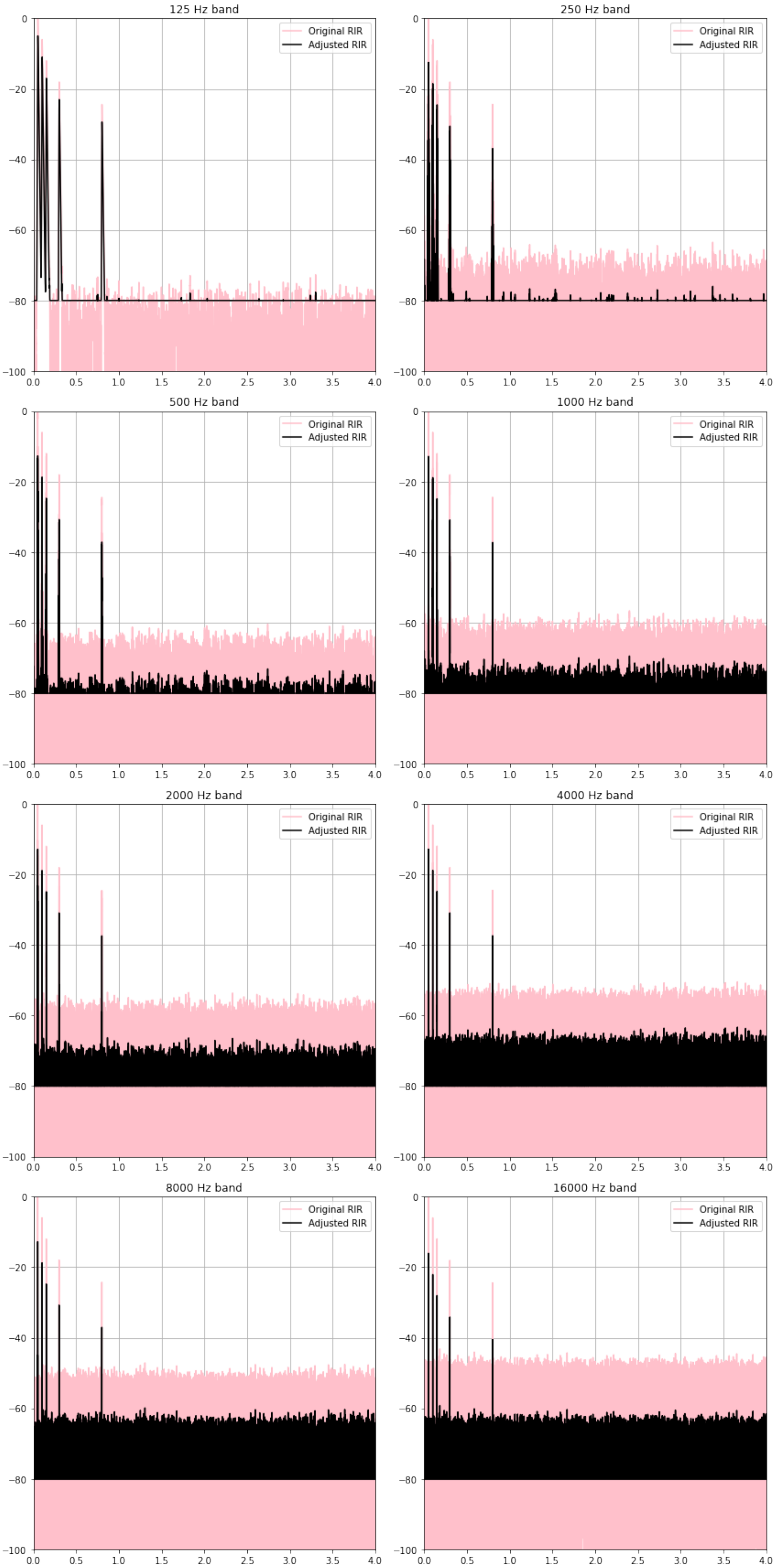

FIGURE 4.1: Original and Adjusted RIRs

4.0.1 Adjusting reverberation time

The first plot is the raw RIR filtered into one octave band (centered on 4 kHz), squared, and expressed in decibels (dB).

The first large peaks at the very beginning of the plot represent the direct sound—the sound that travels straight from the source to the listener or microphone without any interaction with the environment. This is typically the loudest and most distinct part of an RIR. Following the initial peak, the subsequent spikes are the early reflections. These are sounds that have bounced off one or more surfaces (walls, floor, ceiling) before reaching the listener. These reflections are important for characterizing the acoustic properties of the room, such as size and shape. As the plot progresses, the spikes diminish and the waveform becomes a more uniform, dense area known as the reverberant tail. This part of the RIR represents the sound energy being scattered numerous times, leading to a complex overlay of many reflections that gradually lose intensity due to absorption by surfaces and air.

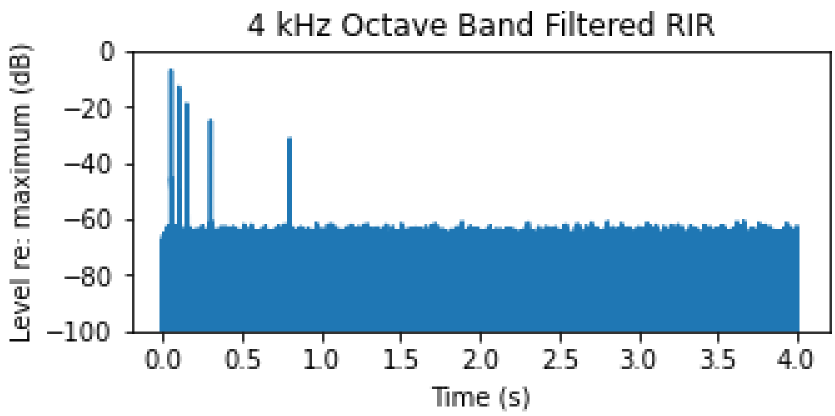

FIGURE 4.2: 4 kHz Octave Band Filtered RIR

The second plot shows the smoothed envelope of the octave band RIR with a fitted exponential decay curve.

The plot illustrates two key components typically used in acoustic analysis:

1. Smoothed Envelope (Blue Line): This line represents the envelope of a Room Impulse Response (RIR) that has been smoothed. The envelope is a curve that outlines the upper boundary of a sound wave, showing the gradual decrease in amplitude over time. The smoothing process helps in reducing the fine details and fluctuations to focus on the general trend of decay in the sound energy. This is useful for clearly observing the overall decay characteristics without being distracted by the minor variations due to specific reflections or other transient sounds.

2. Fitted Curve (Orange Dashed Line): This dashed line represents a mathematical curve that has been fitted to the smoothed envelope. The purpose of this curve is to model the decay behavior of the sound in the room. Commonly, an exponential decay function is used because it closely represents the physical process of sound energy dissipation in an enclosed space. The parameters of this curve (such as decay rate) are critical for determining acoustic properties like reverberation time, which is a key descriptor of how sound behaves in a room.

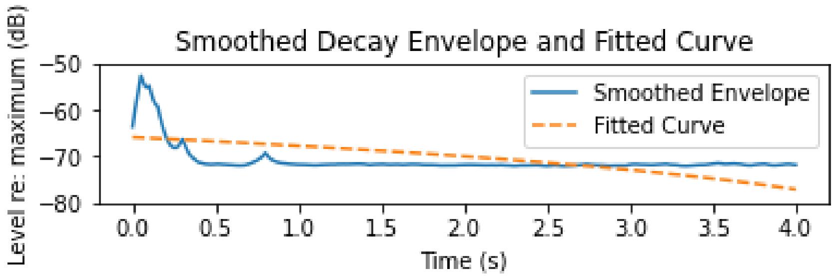

FIGURE 4.3: Smoothed Decay Envelope and Fitted Curve

The third plot illustrates the modified RIR, where the noise floor is extended as part of the decay, and the reverberation time has been changed.

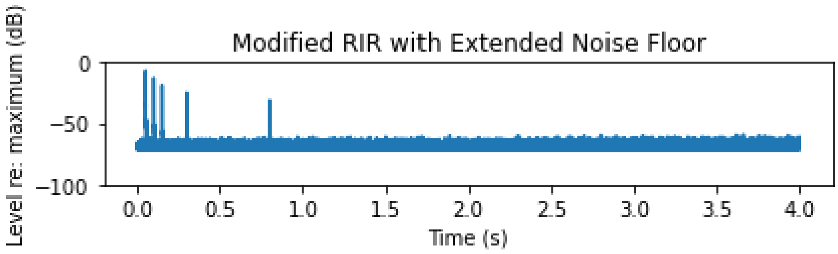

FIGURE 4.4: Modified RIR

Chapter 5

Future Plan

The outcomes of our study reveal that employing such techniques can help in manipulation of RIRs effectively. The methods discussed are primarily used for preparing audio content for auralization and enhancing the audibility of Room Impulse Response (RIR) features for sonification purposes. The decay slope is a critical characteristic of RIRs, and its manipulation is essential either to highlight its significance or to diminish its impact, thereby allowing other features of the RIR to be more discernible.

Combining our results with the outcomes achieved in the RAIR Dimensionality Reduction study we could develop more efficient and accurate methods for auralization and sonification. Dimensionality reduction could streamline the process of decay envelope manipulation by focusing efforts on the most significant acoustic features, reducing computational load and potentially improving the quality of the simulated audio. Applying the combined techniques to system identification could enhance the precision of identifying and modeling various audio systems, such as echo cancellation modules, hearing aids, or even complex virtual reality audio environments.

Few of the major applications can be improvement in architectural acoustics, enhanced audio forensics and optimization of hearing aids.

Bibliography

Cabrera, D, S. H. M. W. . L. D. (2009). Binaural measurement and simulation of the room acoustical response from a person's mouth to their ears. *Acoustics Australia*, 37.

Carbera, D. and Ferguson (2007). 'sonification of sound: tools for teaching acoustics and audio. *Proceedings of the 13th international conference on auditory display.*

Densil Cabrera, Doheon Lee, M. Y. and Martens, W. L. (2011). Decay envelope manipulation of room impulse responses: Techniques for auralization and sonification. *Proceedings of ACOUSTICS 2011.*

Gongping Huang, Jacob Benesty, J. C. (2023). Dimensionality reduction of room acoustic impulse responses and applications to system identification. *IEEE SIGNAL PROCESSING LETTERS*, 30.