A Dynamic Reverberation Algorithm for Virtual Acoustic Rendering

Huan Mi

Master of Philosophy

University of York

Physics, Engineering and Technology

April 2023

Abstract

Reverberation is crucial in virtual acoustic rendering, shaping spatial perception and immersion. This thesis explores the applicability of feedback-delay-network-based artificial reverberation algorithms for real-time virtual environments. To optimise reverberator design, perceptual thresholds of key binaural room impulse response (BRIR) parameters—initial time delay gap (ITDG), early reflection strength, and late reverberation strength—are analysed across different room sizes.

A new binaural artificial reverberation algorithm, enhancing the Schroeder and Moorer models, is developed with high-frequency noise filtering, air absorption simulation, and refined binaural synthesis. Comparative evaluations show that it improves reverberation rendering, particularly for small and medium-sized rooms.

To further optimise virtual acoustic simulation, the proposed algorithm (for small rooms), Schroeder (for medium rooms), and Gardner (for large rooms) are integrated into a hybrid model, implemented as an audio plug-in for Digital Audio Workstation (DAW) etc. virtual environments. Numerical evaluation confirms the model's effectiveness in simulating real-world BRIRs, contributing to the development of computationally efficient and perceptually accurate artificial reverberation algorithms for immersive virtual audio applications.

Acknowledgements

I would like to express my deepest appreciation and gratitude to all those who have supported and guided me throughout my research journey.

First and foremost, I would like to express my heartfelt gratitude to my first supervisor, Prof. Gavin Kearney, for your patient guidance and encouragement when I first entered this field, and continuous support throughout my entire Ph.D. journey. Your expertise, encouragement, and patience have been instrumental in shaping this thesis and my overall academic growth. I am truly grateful for your dedication and commitment to my research.

I would also like to thank my second supervisor, Prof. Helena Daffern. You provided constructive suggestions for my entire research process and carefully helped me correct the errors in my thesis, discussing the best solution with me. Your insightful comments and expertise have been invaluable in shaping my research.

Dr. Jez Wells, as an excellent thesis advisor, has consistently provided me with invaluable feedback, constructive criticism in the first two years. Although you have now switched jobs and no longer serve as my thesis advisor, your insightful suggestions had significantly improved the quality of my thesis.

I am grateful to my colleagues and friends who have provided me with a supportive and stimulating research environment. Their discussions, ideas, and ca-

maraderie have been instrumental in shaping my research and making this Ph.D. journey an enriching experience. Especially my friend Hongbo Hu, who works tirelessly to help me solve all the problems I encounter in life, so that my research path can avoid being influenced by trivial things in life.

I extend my heartfelt thanks to my family for their unwavering love, encouragement, and support throughout my academic journey. Their belief in me and constant motivation have been a driving force behind my achievements, and I am deeply grateful for their presence in my life.

Last but not least, I would like to express my sincere appreciation to all the research participants who contributed their time and efforts to my study. Their willingness to participate and share their insights have been invaluable in advancing my research.

In conclusion, I am deeply grateful to all those who have supported me on this research journey. Your encouragement, guidance, and unwavering support have been invaluable, and I am honoured to have had you by my side. Thank you for your contributions and for making this thesis a reality.

Declaration of Authorship

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. I also declare that parts of this research have been presented in previous conference and journal publications, which are listed as follows:

Chapter 3 is based on the following journal paper:

> H. Mi, G. Kearney, and H. Daffern. Impact Thresholds of Parameters of Binaural Room Impulse Responses (BRIRs) on Perceptual Reverberation. *Applied Sciences*, 12(6):2823, 2022.

Chapter 4 is based on the following journal paper:

> H. Mi, G. Kearney, and H. Daffern. Perceptual Similarities between Artificial Reverberation Algorithms and Real Reverberation. *Applied Sciences*, 13(2):840, 2023.

Contents

List of Tables

List of Figures

List of Acronyms

3D	3-dimensional
3DoF	three degrees of freedom
6DoF	six degrees of freedom
ANOVA	analysis of variance
AR	Augmented Reality
BAQ	Basic Audio Quality
BRIR	binaural room impulse response
CIPIC	Center for Image Processing and Integrated Computing
CPU	central processing unit
D/R	direct-to-reverberant
DAW	Digital Audio Workstation
dB	decibel scale
DFDN	Directional Feedback Delay Network
DS	direct sound
DSP	digital signal processor
DWN	Digital Waveguide Network
EDC	energy decay curve
EDR	energy decay relief
EDT	early decay time
ER	early reflection
ERB	Equivalent Rectangular Bandwidth
ESS	exponential sine sweep

FDN	Feedback Delay Network
FER	forward early reflection
FFT	fast Fourier transform
FIR	finite impulse response
FPGA	field-programmable gate array
GA	Geometrical Acoustics
GUI	graphical user interface
HMS	Hybrid Moorer-Schroeder
HMSG	Hybrid Moorer-Schroeder-Gardner
HRIR	head-related impulse response
HRTF	head-related transfer function
IACC	interaural cross-correlation coefficient
IC	interaural coherence
IIR	infinite impulse response
ILD	interaural level difference
IQR	interquartile range
ISM	Image Source Method
ITD	interaural time difference
ITDG	initial time delay gap
ITU	International Telecommunication Union
JND	just noticeable difference
LKFS	Loudness, K-weighted, relative to Full Scale
LPF	low-pass-feedback
LR	late reverberation
LSS	linear sine sweep
LUFS	Loudness Units relative to Full Scale
MLS	Maximum Length Sequence
MUSHRA	Multiple Stimuli with Hidden Reference and Anchor
Pa	pascal
RER	reverse early reflection

RIR	room impulse response
RMS	root mean square
SDN	Scattering Delay Network
SH	spherical harmonic
SHD	spherical harmonic domain
SHT	Spherical Harmonic Transform
SPL	sound pressure level
SRIR	spatial room impulse response
UI	user interface
VR	Virtual Reality
VST	Virtual Studio Technology
WARA	Wearable AR Audio

List of Symbols

$*$	convolution	
α	probability of making the type I error	
α_s	individual physiological parameters set	
β	probability of making the type II error	
$\boldsymbol{A}$	orthogonal recirculation matrix	
$\boldsymbol{Y}_{\text{uni}}$	spherical harmonic matrix	
Δd	travel difference between direct and first reflected sound	m
Δd_1	travel difference between direct and first reflected sound at position 1	m
Δd_2	travel difference between direct and first reflected sound at position 2	m
Δ	random delay parameter	
∞	infty	
λ	wavelength	m
$\lceil\rceil$	ceil function	
$[]$	rounding function	
μ	mean value	
ω_1	start frequency	rad/s
ω_2	end frequency	rad/s
ϕ	elevation	$^\circ$

σ	population parameter standard deviation	
σ^2	population parameter for the variance	
θ	azimuth	$^\circ$
A	feedback matrix	
a	sound absorption coefficient	
b_i	input gain	
C	correction parameter of 2.5 ms sample value	
c	speed of sound	m/s
c_i	output gain	
CL	confidence level	
d	direct sound gain	
d_i	non-prime-number delay length	
d_{D1}	direct sound travel	
d_{R1}	first reflected sound travel	
f	frequency	Hz
Fs	sample rate	Hz
g	gain	
g_i	target gain of each delay line	
g_{dB}	logarithmic scale gain	
g_{lin}	linear scale gain	
H_0	null hypothesis	
H_a	alternative hypothesis	
H_L	left ear head-related transfer function	
H_R	right ear head-related transfer function	

m	sound attenuation constant of the air	m^{-1}
m_i	delay length	
M_{min}	total minimum delay length	
N	population data size	
n	sample size	
P	sound pressure	Pa
P_0	complex-valued free field sound pressure	Pa
P_L	left ear complex-valued sound pressure	Pa
P_R	right ear complex-valued sound pressure	Pa
P_{ref}	reference sound pressure	Pa
$ppwr_i$	prime number position index	
pri_i	prime number set	
Q	orthogonal recirculation matrix	
r	source-to-head distance	m
RT_{60}	reverberation time	s
S	area	m^2
s	sample standard deviation	
s^2	sample variance	
SE	standard error	
$T0$	direct impulse sample value	
$T1$	first reflection impulse sample value	
T	time duration of sweep	s
t	time	s
T_{20}	reverberation time calculated by using a 20 dB interval	

T_{30}	reverberation time calculated by using a 30 dB interval	
T_c	center time	ms
T_d	filter delay time	s
t_e	early reflection time	ms
t_e	early time	
V	volume	m^3
x_i	the value of the ith data point	

1 Introduction

Sound is all around us all the time. In addition to visual perception, humans need to perceive the world through their sense of hearing. Even with our eyes closed, we can sense where a sound is coming from, whether it is far away or close to us, and we can tell by sound whether the space we are in is closed or open. Humans localise sound and identify their environment mainly through the time difference between the arrival of the sound signal at the right and left ears, the level difference, the spectral cues created by the head, ear and torso and the perceived reverberation. This is known as binaural hearing.

The binaural mechanisms by which humans locate sounds and identify acoustic environments make it possible to reproduce plausible real-world audio in a virtual acoustic space. Virtual acoustic rendering involves simulating acoustic properties, with the potential to provide highly immersive and realistic auditory environments. Head tracking, binaural filtering using head-related transfer functions (HRTFs) and artificial reverberation are three factors that contribute to producing a realistic virtual acoustic spaces with accurate localisation and timbre [1]. The implementation of reverberation in particular is a pivotal area of research to create auditory experiences that closely mimic real-world scenarios. For example, the development of computationally efficient dynamic reverberation algorithms is essential to provide plausible virtual acoustic rendering on low-cost mobile de-

vices [2, 3]. Reverberation modelling involves understanding and replicating how sound reflects, diffuses, and decays in an environment. This has significant implications for enhancing the realism and spatial awareness within Virtual Reality (VR) or Augmented Reality (AR) and gaming, as well as in the practical assessment of architectural acoustics and in the study of psychoacoustics.

In real rooms, reverberation is caused by the coupling of a sound source to the acoustic space. After the sound is emitted, the sound wave will interact with its environment, undergo reflections, diffraction and absorption, so that the audience will receive filtered replicas (echoes) of the original wavefront arriving from different directions at different times. These filtered replicas are mixed with the direct sound and perceived by the listener as reverberation [4]. Figure 1.1 [5] is an illustrative diagram of reverberation in an auditorium. A room impulse response (RIR) is the transfer function between a microphone receiving point and an impulsive sound source at any location in a room [6]. The RIR is composed of three main components: the direct sound, which arrives first and provides the initial clarity; the early reflections, which follow closely and contribute to the spatial impression and perceived distance; and the late reverberation, which consists of numerous reflections that gradually diminish, creating the sense of reverberation. The illustrative diagram of a room impulse response is shown in Figure 1.2 [7].

The measurement of RIR typically involves placing a sound source, such as a loudspeaker emitting an impulsive sound (e.g., a balloon pop or an applause), and a microphone to record the response at various positions within the space. However, applying RIRs measured in real spaces in virtual acoustics is impractical because it requires extraction from a large database of impulse responses The measurement of impulse responses is complex and therefore difficult to apply to real-time

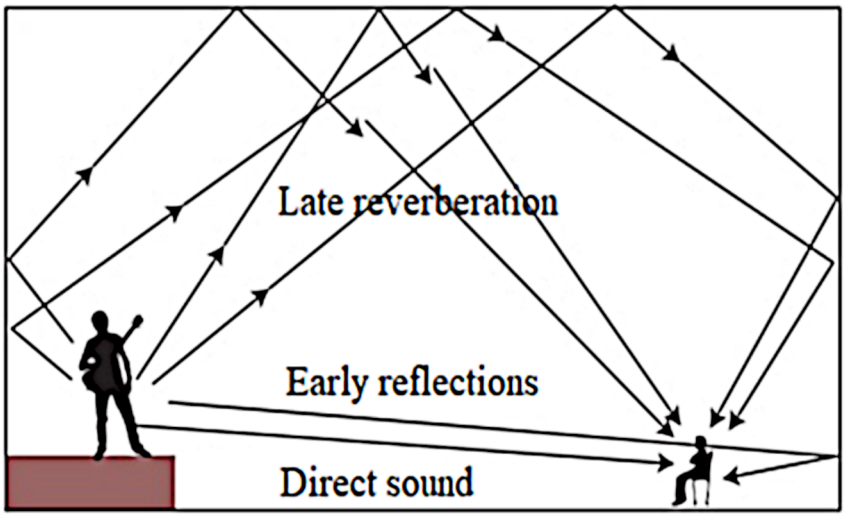


Figure 1.1: Illustrative diagram of reverberation in an auditorium [5].

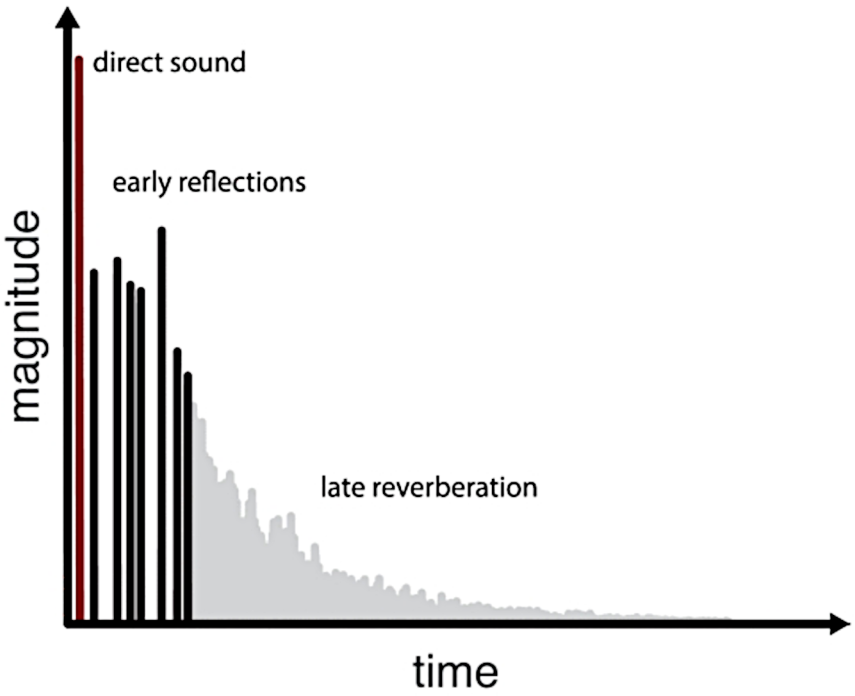


Figure 1.2: Illustrative diagram of a room impulse response [7].

dynamic reverberation processing. Therefore artificially developed methods such as wave-based modelling of room acoustics, geometrical acoustics and recursive reverberation algorithms based on feedback delay networks which also simulate room impulse responses are also potential candidates for this purpose [8, 9].

Wave-based modelling can accurately represent sound wave phenomena, but its computational cost rapidly increases with the increase of frequency and spatial complexity, which makes it occupy a large amount of memory [10]. Geometric acoustic rendering methods such as the Image Source Method [9] or Ray-tracing [9] have high computational efficiency for high-frequency sounds, but their accuracy is lower in low-frequency simulations [9]. This is because in geometric acoustics, sound is treated as rays, neglecting all wave properties. This approach works well at high frequencies, where the sound wavelength is short relative to surface and room dimensions, but at lower frequencies, approximation errors increase because wave phenomena, such as diffraction, become more significant [9]. Recursive reverberation algorithms based on feedback delay networks [11] cannot fully and accurately replicate the specific acoustic characteristics of a given room, but they are well suited to real-time audio applications in virtual rendering due to their low computational cost and flexible parameterisation characteristics [12]. A method of binaural acoustic simulation proposed by Agus et al. [13] that models first-order reflections and late reverberation within one feedback delay network has been shown to achieve good accuracy, with the added benefits of simplicity, efficiency, and real-time processing, as demonstrated in an iOS application. Therefore, the recursive reverberation algorithm based on feedback delay network is adopted as the core algorithm in this thesis.

Current trends in the field are directed towards more computationally efficient

models for simulating realistic and plausible reverberation, with approaches such as feedback delay networks and hybrid models combining different acoustic principles to optimise perceptual characteristics and reduce computational load [14–18]. Feedback delay networks are composed of recursive digital filters with extremely low computational complexity, making them suitable for real-time applications. Significant work has extended these models to enhance realism without compromising efficiency. For example, Alary et al. [19] extended FDNs to model non-uniform, direction-dependent decay times. Schlecht and Habets [18] proposed optimised feedback matrix designs for spatialised FDNs, focusing on enhancing spatial accuracy while maintaining efficiency. Das and Abel [20] modified FDNs to simulate coupled volumes efficiently, providing a more accurate room response without increasing processing complexity.

The future of virtual reverberation rendering promises further integration with emerging technologies like artificial intelligence and more advanced VR or AR systems, potentially revolutionising how we interact with and perceive sound in virtual environments, as demonstrated by recent developments in machine-learning-driven sound design and spatial audio in gaming and immersive experiences [21–23].

Harma et al. proposed a technique and application of Wearable AR Audio (WARA) [24] in 2003. The model's direct sound and 14 early reflections (six first-order and eight lateral planar second-order early reflections) were calculated from a simple shoebox room model with user-adjustable wall, floor, and ceiling positions. To match the length of the reverberation to that of the pseudo-acoustic environment, diffuse late reverberations were added using a variant of a FDN. The parameters of the reverberation algorithm (e.g. reverberation time) were set manually by analysing the reverberation of the environment or estimating

parameters automatically from the binaural microphone signals. The additional information about the user's orientation and position required for room modelling and binaural synthesis was estimated using a number of head-tracking devices, but a limitation of many current technologies is the need to place external devices in the environment, so these can only be used at close range and are therefore largely limited to applications where the augmented sound source is positioned on some real-world object that can also host the transmitter of a head-tracking system.

A real-time binaural room modelling proposal for AR applications based on Scattering Delay Network (SDN) reverberation was presented by Yeoward, et al. in 2021 [25]. SDN integrates Digital Waveguide Network (DWN) and ray-tracing Image Source Method (ISM) to achieve a more physically accurate simulation to room models, while meeting the lower processing requirements like other reverberators that use delay networks [25–27]. A study by Djordjevic et al. [28] indicated that SDNs are considered more natural than binaural room impulse responses, ray tracing, and feedback delay networks in evaluations based on non-interactive simulations of two listening rooms. However, the suitability of SDNs for AR applications needs to be determined, as binaural SDN models have not been proposed in real-time architectures suitable for mobile or wearable computing devices [25].

Above researches represent the ongoing efforts towards integrating realistic acoustic modelling in virtual applications. These models offer a more physically accurate simulation while still meeting low processing requirements, which is pivotal for applications that operate on mobile or wearable devices. However, developing a real-time dynamic reverberation algorithm capable of accurately reproducing real-world perceptual reverberation in virtual acoustic renderings is a complex task. It requires careful consideration of computational complexity, especially in the pur-

suit of high realism, the dynamic properties of sound, such as adapting to variables such as room size, shape, material and listener motion, as well as psychoacoustic complexity in response to subjective human perception. Combining the advantages and disadvantages of the above reverberation simulation technologies and current trends in virtual reverberation rendering, the aim of the work presented in this thesis is to implement a real-time dynamic binaural reverberation algorithm that can reproduce plausible real-world perceptual reverberation in virtual acoustic rendering. The work will focus on enhancing user perceptual experience and ensuring computational efficiency. In order to produce a better accuracy and realism of sound and sense of immersion, virtual reverberation needs immersive, real-time, interactive and dynamic features. This thesis defines these four features as follows.

- **Immersive** refers to an experience that deeply engages the senses and creates a feeling of being fully surrounded by or involved in a virtual or simulated environment. It typically involves the use of technologies such as virtual reality or augmented reality to create a sense of immersion.

- **Real-time** refers to activities or processes that occur instantly or with minimal delay. In this thesis, real-time features require the algorithms to have high computational efficiency and thus low latency when converted to audio streaming implementations to make sure they are consistent with the perception of the human auditory system.

- **Interactive** describes an experience or system that allows users to actively participate, engage, or manipulate elements within it. It involves two-way communication or feedback between the user and the system, enabling the user to have control and influence over the experience.

- **Dynamic** refers to adaptability, responsiveness, or the ability to modify or update components in real-time based on changing conditions or inputs.

This thesis implements a binaural reverberation algorithm that can simulate plausible real-world reverberation by evaluating the effect of the parameters of binaural room impulse responses (BRIRs) on perceptual reverberation, and develops a real-time dynamic binaural reverberation audio plug-in by combining it with other reverberation algorithms for computationally efficient reproduction of plausible real-world perceptual reverberations in virtual acoustic rendering.

1.1 Aims and Research Questions

The primary aim of this thesis is to implement a real-time dynamic binaural reverberation algorithm that can reproduce plausible real-world perceptual reverberation in virtual acoustic rendering, with a focus on enhancing user's perceived experience whilst ensuring computational efficiency. The key purpose of this work is to address the challenge of creating a digital reverberator that not only mimics the acoustics of real environments convincingly but also operates efficiently to be used in dynamic, real-time applications.

To achieve this aim, the question that motivates the research is presented: *How can a plausible real-world reverberation effect be simulated in real-time in virtual acoustic rendering, using a computationally efficient reverberation algorithm?*

The key terms of the research question and how they relate to this thesis are explained specifically below:

- **Reverberation:** As a physical phenomenon, reverberation refers to the

echoes that the human ear hears when sound waves are launched from a source and then reflected by obstacles.

- **Artificial reverberation algorithms:** A filter structure for the reconstruction of a room impulse response by methods such as signal processing to achieve reverberation by convolution with a dry source or by directly filtering the dry source to capture the reverberant output.

- **Virtual acoustic rendering:** Virtual acoustic rendering is the use of spatial audio technology to synthesise virtual sound sources that are used to mimic the process of real sound sources reaching human ears, allowing the listener to perceive, for example, information about the location of the virtual sound source in space. It involves the use of advanced algorithms and mathematical models to generate sound waves that can mimic the behaviour of sound in a real environment.

- **Computationally efficient:** Computationally efficient means that reverberation algorithms must accurately process data and provide responses within a budgeted time frame and efficient use of computational resources to enable real-time application of the algorithms.

- **Plausible:** Plausible means producing sound that convincingly mimics the acoustics of the real environment. The plausibility of a reverberation algorithm is determined by the similarity of the algorithm's output to real-world acoustics, that is, the degree to which the output reverberation matches the real-world acoustic properties. Plausibility is regulated by a combination of objective acoustic properties and subjective human perception.

The main question is divided into three sub-questions for further study:

1. The room impulse response consists of different components and parameters that all contribute to the perceptual reverberation. In order to implement a dynamic reverberation algorithm, some important components or parameters could be more perceptually relevant to manipulate than others. Therefore it needs to be clarified what the most perceptually important reverb parameters are to manipulate in digital reverberation. Chapter 3 investigates this question. For convenience, these components and parameters are referred to uniformly as 'parameters' in this thesis.

2. By evaluating the impact of room impulse response parameters on perceptual reverberation and the realisation of the manipulability of these parameters, can we create a perceptually plausible and efficient digital reverberator that sounds as good as real-world acoustic measurements? This question is addressed in Chapter 4.

3. With the implementation of a reverberation algorithm that can simulate plausible real-world reverberation, it is also important to ensure that the algorithm is interactive and computationally efficient in order to satisfy the real-time dynamic rendering of the algorithm in a virtual acoustic environment. What is the effectiveness of such a design in a real-time dynamic (changing reverb parameters) rendering scenario when compared to real-world measurements? Chapter 5 synthesises the three reverberation algorithms into a hybrid algorithm and verifies its real-time dynamics and interactive properties by generating it as a reverb plug-in.

1.2 Novel Contributions

The research presented in this thesis provides the following novel contributions to the field:

- **Threshold evaluation of BRIR parameters:** Parameters of binaural room impulse responses to auditory perception and the perceptual thresholds for different sized rooms are assessed through a staircase method. The generic thresholds for perceptual reverberation are defined by evaluating the average perceptual thresholds for four BRIR parameters to determine whether they have a significant effect on perceptual reverberation. These include the initial time delay gap (ITDG), the forward and reverse early reflection strengths, and the late reverberation strength. These threshold measurements can provide the basis for accurate manipulation of reverberation parameters to suit different virtual environments, thus enhancing the realism and plausibility of reverberation effects in various virtual acoustic setups.

- **A new binaural reverberation algorithm and evaluation:** A new reverberation algorithm - Hybrid Moorer-Schroeder (HMS) is proposed based on the evaluation of binaural room impulse response parameters and by combining and improving the Schroeder and Moorer reverberation algorithms. This algorithm can simulate the reverberation effects in large, medium and small rooms respectively. In this thesis, large, medium and small rooms are used to denote long, medium and short reverberation times respectively. The algorithm includes high frequency noise filtering and air absorption simulation, fully combines direct sound, early reflections and late reverberation, and separates the left and right channels by using delay constants to pro-

duce complete binaural synthesis. It also achieves matching of correlation between the left and right channels, direct-to-reverberant (D/R) energy ratio and timbre to the real reverberation. The perceptual similarity and objective acoustic properties of this algorithm and six other well-known reverberation algorithms to measured reverberation is also evaluated. It is demonstrated that this enhanced reverberation algorithm is able to more closely mimic the perception of auditory reverberation in real-world environments.

- **Development and verification of a real-time dynamic reverb audio plugin:** By verifying the real-time dynamic properties of reverberation algorithms, a dynamic reverb audio plug-in, named Hybrid Moorer-Schroeder-Gardner (HMSG) plug-in, has been implemented by combining different reverberation algorithms that can be adapted to different reverberation requirements in virtual acoustic rendering in real time. The HMSG plug-in is implemented by combining the reverberation structure of the HMS reverberation algorithm for simulating small rooms, the reverberation structure of the Schroeder reverberation algorithm for simulating medium rooms, and the reverberation structure of the Gardner reverberation algorithm for simulating large rooms. An objective numerical evaluation between the parameters of the binaural room impulse responses generated by the plug-in and the parameters of the real-world measured impulse response is also presented verifying the objective plausibility of the plug-in.

1.3 Statement of Ethical Approval and Publications

The protocols for perceptual tests using human participants presented in this thesis, and the management of corresponding data, were approved by the University of York Physical Sciences Ethics Committee with reference Mi111120 for Chapter 3 and Mi070821 for Chapter 4. These protocols are provided as accompanying materials, and described in Appendix C.

The relevant publications in this thesis are also listed:

Chapter 3 is based on the following journal paper:

> H. Mi, G. Kearney, and H. Daffern. Impact Thresholds of Parameters of Binaural Room Impulse Responses (BRIRs) on Perceptual Reverberation. *Applied Sciences*, 12(6):2823, 2022.

Chapter 4 is based on the following journal paper:

> H. Mi, G. Kearney, and H. Daffern. Perceptual Similarities between Artificial Reverberation Algorithms and Real Reverberation. *Applied Sciences*, 13(2):840, 2023.

1.4 Structure of the Thesis

The thesis is organised as follows:

Chapter 2 provides a comprehensive literature review. It begins with an explanation of sound fields and sound propagation, providing the foundational understanding necessary for virtual acoustic rendering. The chapter then explores

spatial audio, covering key concepts such as HRTFs, head-tracking and artificial reverberation simulation, which are essential for accurate spatialisation in virtual environments. The chapter also delves into the composition, parameters, measurement, and synthesis of room impulse responses, including binaural reverberation and its application through convolution and auralisation. Next, it reviews existing artificial reverberation algorithms and discusses their application in scenarios of virtual acoustic rendering. Finally, the methods of audio perception evaluation and how to statistically analyse their results were reviewed, establishing a framework for the perceptual evaluation of reverberation algorithms.

Chapter 3 presents the motivation for assessing the impact thresholds of four specific parameters of BRIRs. The chapter details the design and implementation of listening tests to evaluate these parameters, including a thorough description of the participants, stimuli, procedures, and apparatus used. The results section provides an average threshold analysis for each parameter type, followed by statistical analysis and a discussion of the threshold distribution. The chapter concludes with a discussion of the findings, considering their implications for the development of a new reverberation algorithm.

Chapter 4 introduces and evaluates the new reverberation algorithm developed in this thesis, known as the HMS algorithm. The chapter begins by discussing the motivation for the algorithm. It then details the implementation of traditional reverberation algorithms before introducing the new algorithm. The basic mono structure and binaural optimised structure of the proposed algorithm are explained, followed by an objective analysis of its performance. The chapter includes a listening test, describing the participants, stimuli, procedure, and apparatus, and presents the results, including a mean value analysis, statistical analysis, and com-

putational cost analysis. A discussion of these results is provided, highlighting the strengths and limitations of the algorithm, and the chapter concludes with a summary of the findings.

Chapter 5 focuses on the implementation of reverberation algorithms as an audio plug-in designed for real-time use in virtual environments, known as the HMSG plug-in. The chapter begins with a discussion of the motivation for this work and reviews relevant published studies. It then details the implementation of the plug-in, including the adjustable parameters and the integration of the reverberation component. An objective analysis of the plug-in's output is conducted, comparing it to measured impulse response grids to evaluate its accuracy. The chapter concludes with a discussion of the results and a summary of the findings, considering the implications for the use of the plug-in in real-time virtual acoustic rendering.

Chapter 6 summarises the main findings and contributions of the research, reiterating the research question and explores areas for future work. The future work includes threshold measurements for other parameters of the BRIR and different stimuli, optimising the computational efficiency of the HMS reverberation algorithm, improving the match between the adjustable parameters of the HMSG plug-in and the measured impulse response, and implementing perceptual evaluations for the HMSG plug-in in Virtual Reality (VR) and Augmented Reality (AR) environments. The chapter concludes with final thoughts, reflecting on the broader implications of the research and its potential impact on the field of virtual acoustic rendering.

2 Literature Review: Sound and Reverberation

To reproduce and render plausible real-world reverberation in a virtual environment, it is essential to understand the basic principles and properties of sound and reverberation. This chapter will review the fundamentals of acoustics and reverberation, in particular the application of digital artificial reverberation algorithms in virtual acoustic rendering. The basic principles of sound propagation and sound fields are first introduced in Section 2.1. Section 2.2 then reviews the spatial audio. The fundamentals of reverberation are then described in Section 2.3. Section 2.4 provides a comprehensive overview of current artificial reverberation algorithms, then the application scenarios of virtual reverberation is discussed in Section 2.5. Section 2.6 reviews the methods of audio perception evaluation.

2.1 Sound Field and Sound Propagation

A sound wave is the vibration of particles in the direction of propagation near its equilibrium position as shown in Figure 2.1 [29], which means that sound propagation is the transfer of energy in a medium [30]. The speed of sound propagation is related to the type of medium. Sound travels fastest in solids, followed by liquids and finally gases, while a vacuum cannot transmit sound [31]. This is because

sound needs a medium to travel in, and the denser and more elastic the medium, the faster the sound travels [32]. The speed of sound propagation is also influenced by temperature, the higher the temperature the faster it travels [33]. This is particularly the case in gases, because temperature affects the kinetic energy of particles in the medium. Higher temperatures mean more energetic and faster-moving particles, which results in quicker sound propagation [33]. This relationship is less direct in liquids and solids due to their different molecular structures and the different ways in which temperature affects their properties, but in general, the tendency for the speed of sound to increase with temperature has been observed to varying degrees in different media [34]. At 20 °C, in dry, windless air, the speed of sound is approximately 343 m/s [35]. The speed of sound is the distance a sound wave travels per unit of time, given by:

$$c = \lambda \times f, \tag{2.1}$$

where c is the speed of sound, λ is the wavelength (unit: m), and f is the frequency (unit: Hz).

Sounds can be distinguished by their perceptual attributes: pitch, loudness and timbre [36]. Each of these attributes is determined by a different physical aspect of the sound wave. Pitch is primarily determined by the frequency of the sound wave, where the higher the frequency the higher the pitch of the sound. The human ear can distinguish sound frequencies between 20 Hz and 20 kHz. Fluctuations above this range are called ultrasound, while those below this range are called infrasound.

Loudness is related to the amplitude of the sound wave, or in other words, the volume of the sound, which is the perceptual correlate of sound intensity. The intensity of a sound is usually expressed by the sound pressure level (SPL)

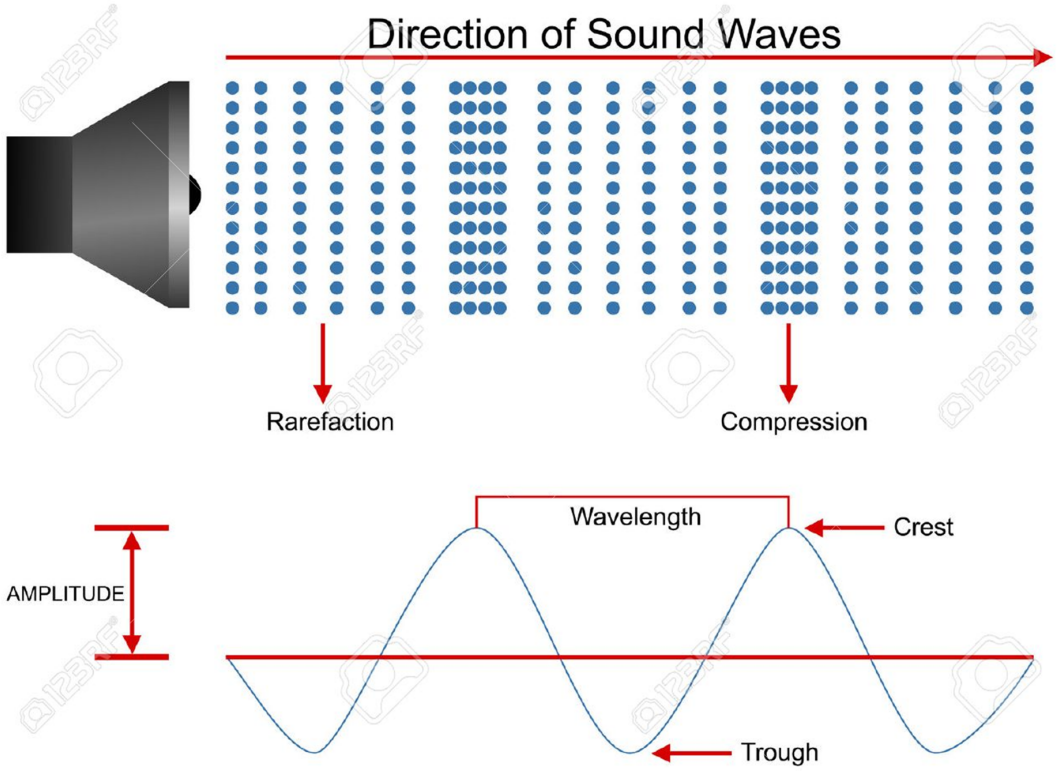


Figure 2.1: Schematic diagram of sound wave propagation from [29].

in decibel scale (dB) [37]. Sound pressure is the increment of pressure due to the presence of sound waves, typically measured as the root mean square (RMS) pressure in pascal (Pa). The RMS pressure represents the effective pressure of the sound wave over time [38]. The sound pressure level is calculated by multiplying the logarithm of the ratio of the RMS sound pressure to the reference sound pressure by 20, as shown in Equation 2.2:

$$\mathrm{SPL} = 20 \log \left(\frac{P}{P_{\mathrm{ref}}} \right), \tag{2.2}$$

where SPL is the sound pressure level (unit: dB), P is the RMS sound pressure

(unit: Pa), and P_{ref} is the reference sound pressure (2×10^{-5} Pa), which is the lowest sound pressure just audible to the human ear, occuring normally between 1000 and 4000 Hz [39]. The normal range of loudness for human hearing is 0 to 130 dB, but sounds above 90 dB can damage the inner ear, and even above 120 dB can already cause irreparable damage to human hearing [40]. Loudness is also influenced by the sensitivity of the human ear, which varies with frequency. Like Figure 2.2, the human auditory field is limited by the threshold of quiet and the threshold of pain, but when the SPL is above the limit of damage risk, the sound can damage the inner ear.

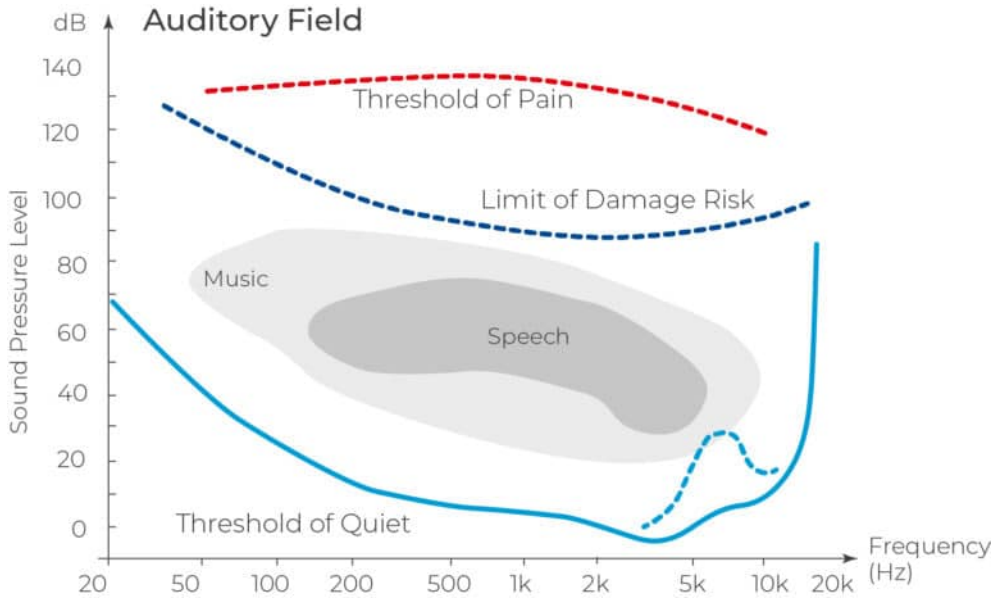


Figure 2.2: Variation of loudness threshold with frequency from [41].

Timbre can also be described as tone color or tone qulity of the sound that encompasses a very complex set of auditory attributes, as well as a large number of complex psychoacoustic issues [42]. Timbre is determined by the complex combination of frequencies in a sound wave, including fundamental frequencies and overtones, and how these frequencies change over time. Humans can distinguish

different sound sources that produce sound at the same pitch and loudness by their timbre [42]. For example, a violin and a flute playing the same note at the same loudness can still sound different.

A concept that is closely related to sound propagation is the sound field. Sound field refers to the area in which sound waves exist. Four categories of sound fields are the near field, far field, free field and diffuse field. Of these, the area covered by the near field is usually within a distance of about one wavelength from the sound source [43], as shown in Figure 2.3 [44]. In the near field, the sound energy circulates back and forth with the vibrating surface of the source, and has not yet reached a state of fully developed propagation, so it can be heavily influenced by factors such as the shape and size of the sound source, the frequency of the sound, and the acoustic properties of the surrounding environment [45]. Due to the complex nature of sound in the near field and as there is no specific relationship between sound pressure and distance, it is often difficult to accurately measure or predict its behavior. Within this part of the sound field, the sound pressure does not obey the inverse square law and the particle velocity is not in phase with the sound pressure [46]. The inverse square law states that the intensity of sound is inversely proportional to the square of the distance from the source [47].

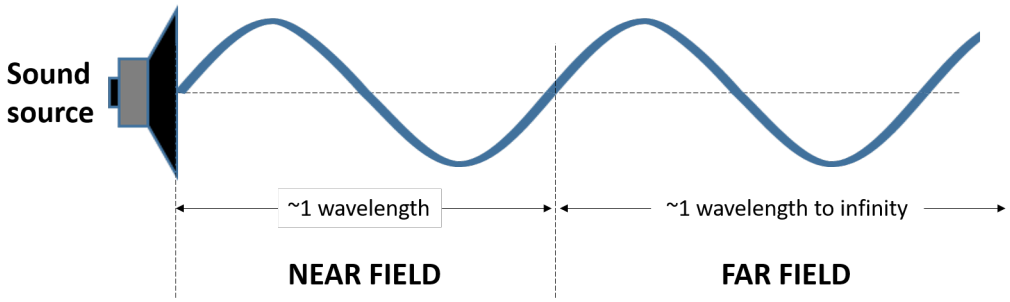


Figure 2.3: Sound field: near field vs. far field from [44].

The far field begins from about one wavelength from the sound source, where

the near field ends and extends to infinity [48], as also shown in Figure 2.3 [44]. When far enough away from the source, the source can be approximated as a point source and the sound waves can be approximated as plane waves. In the far field, the sound pressure and the velocity of the sound particles are essentially isotropic, which refers to the uniformity of sound pressure and particle velocity in all directions. This means that the sound waves radiate equally in every direction from the source, resulting in a spherically symmetric wavefront [49]. In such conditions, the sound pressure level decreases according to the inverse square law, i.e. the sound pressure level decreases by 6 dB for every doubling of the distance from the source [50] as shown in Figure 2.4 [47]. Due to this it can be more easily measured, predicted, and modeled using mathematical equations, and it is therefore common in many acoustic standards to specify that measurements are made in the far field at a distance of at least one metre from the sound source to ensure that the measurement is taken in the far field for the most critical frequencies [51]. The most critical frequencies are usually the frequency range in which the human ear is most sensitive and detects most of the acoustic energy in speech and music. This range is usually about 1 kHz to 4 kHz [39].

Within the far field, there are also free fields and diffuse fields. A free field is an ideal environment without any reflective surfaces, which means there is no echo or reverberation [43]. In a free field, the sound source propagates in a homogeneous isotropic medium [51]. The ideal free field is usually difficult to obtain, but an approximate free field that satisfies certain measurement error requirements can be considered a free field. A typical approximate free field is a specially designed anechoic chamber. In the anechoic chamber, all incident sound waves are effectively absorbed by fully covered glass fibre wedges so that they are not reflected [52].

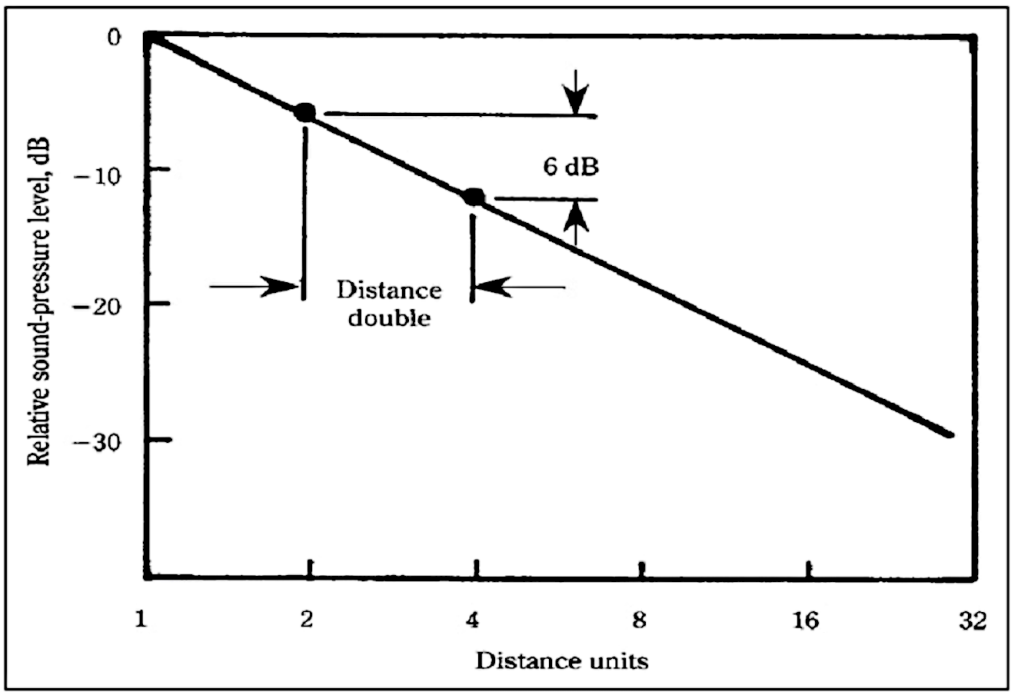


Figure 2.4: The sound pressure inverse square law from [47].

The distance at which the sound pressure of the direct sound is equal to the sound pressure of the reverberant sound in the room where the source is located is called the critical distance [53]. A sound field with a distance greater than the critical distance is called a diffuse field [43]. The diffuse field is the area of the room where the sound pressure level is uniform. In other words, in a diffuse field, the sound pressure intensity is almost equal at all points in space, the probability of the sound energy flowing from each direction to a point is the same, and the phase of the sound waves arriving in each direction is irregular. In a diffuse field, the sound pressure and particle velocity are out of phase, so that the net sound intensity is zero and there appears to be no single source of sound [53].

2.2 Spatial Audio

Understanding the fundamentals of sound propagation and sound fields is necessary to more accurately and realistically simulate the phenomena and behaviour of sound in a virtual rendered scenario. Examples include room reflections and reverberation, as well as modelling how listeners perceive these phenomena. Spatial audio techniques are used in virtual acoustic rendering to create the illusion of sound from a specific location within 3-dimensional (3D) space. This is achieved by manipulating the sound signal to include spatial cues used by the human auditory system to localise the sound source [54]. It allows the listener to perceive sounds from different directions and distances, mimicking the way we experience sound in the real world. Spatial audio can be delivered through loudspeakers or headphones and various audio processing techniques are used to manipulate the sound cues. These techniques include head-related transfer functions (HRTFs), head tracking and artificial reverberation simulation.

Binaural sound is a type of spatial audio that focuses on creating 3D sound experiences through headphones. It uses two audio channel exclusively to create the sensation of sound source localisation for listeners using headphones [55]. It closely mimics the natural listening experience by providing audio with spatial characteristics that indicate directionality and distance. Binaural audio relies on recording or rendering techniques. Binaural recording is accomplished using two microphones that are positioned to mimic the position of the human ears on a dummy head, capturing the audio as a human would hear it. This includes the use of HRTFs to filter the audio signal in a way that mimics the acoustic shadows and time delays caused by our head or ears.

The processing techniques used to achieve spatial audio will be described in detail next.

2.2.1 Head-Related Transfer Functions (HRTFs)

A HRTF is a frequency response that describes the pressure transition from a specific free field source position to an eardrum [56]. As a critical element for the spatialisation of sound, the HRTF contains information on interaural time difference (ITD), interaural level difference (ILD) and spectral cues [57].

ITD describes the time difference between the sound waves arriving at the left and right ears as shown in Figure 2.5 [58]. It is one of the main cues used by the auditory system to localise sound in space, especially in the horizontal plane. Horizontal plane, as shown in Figure 2.6 [59], is the flat plane that divides the head into upper and lower parts and is parallel to the ground when a person is standing upright [60]. The angle of sound incidence on the horizontal plane is represented by the azimuth angle θ. When the sound source is located in the median plane, that is, when $\theta = 0$, ITD is approximately zero because the path lengths from the sound source to both ears are identical. Median plane, divides the head into left and right two symmetrical halves as shown in Figure 2.6 [59]. The incident angle of sound on the median plane is represented by the elevation angle ϕ. When a sound source is off to one side, the sound reaches the closer ear before reaching the farther ear. The brain uses this time difference to determine the direction of the sound source in the horizontal plane [61]. ITD is most effective for low-frequency sounds because their wavelengths are long enough to create a significant time difference between the ears. However, high-frequency sounds are more easily disturbed by the presence of the head to create significant sound shadowing because of their

shorter wavelengths [62] as shown in Figure 2.7 [60].

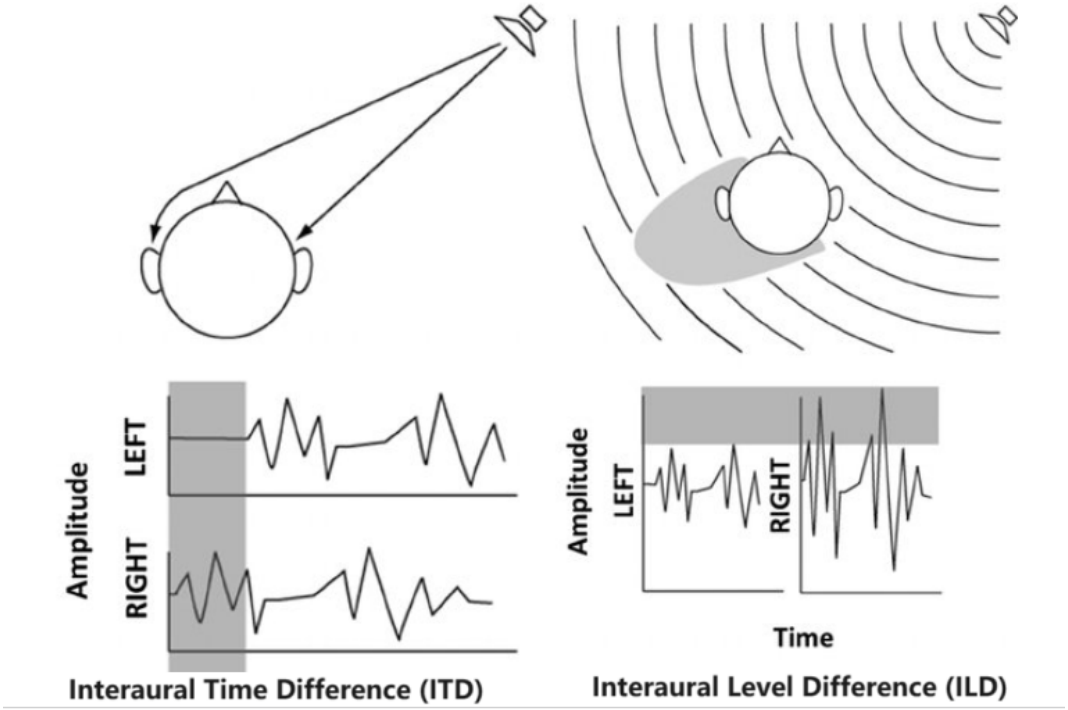


Figure 2.5: Illustration of the interaural time difference and interaural level difference from [58].

ILD describes the difference of sound pressure intensity between the left and right ears as also shown in Figure 2.5 [58]. It is another important cue to localise sound in space. Due to the shadow effect of the head, the intensity of the sound is diminished at the ears farther away from the sound source. High-frequency sounds have shorter wavelengths and more likely to be obstructed by the head. This obstruction creates a shadowing effect, which results in differences in the level (loudness) of sound perceived by each ear [62] as shown in Figure 2.7 [60]. When the sound source deviates from the median plane, the sound pressure level at the ear away from the sound source (contraalteral ear) is lower, whereas the sound pressure level at the ear near the sound source (ipsilateral ear) is higher [61].

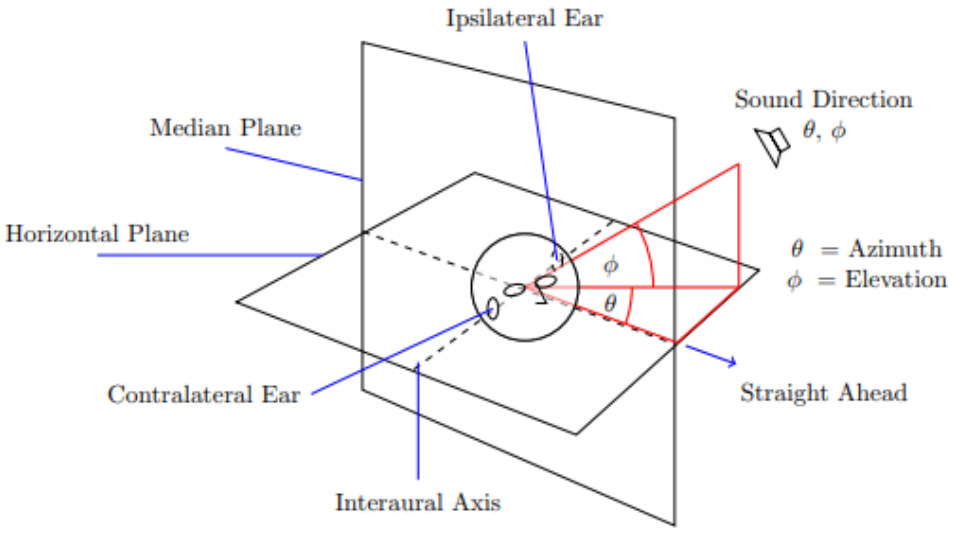
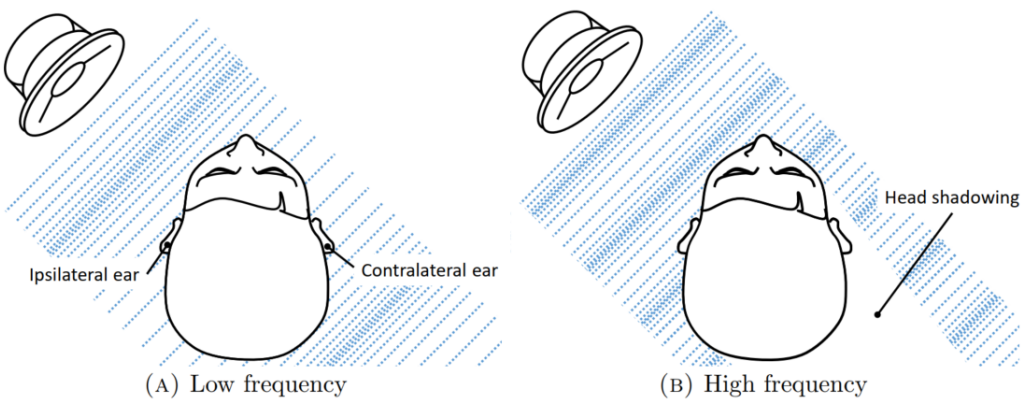


Figure 2.6: The head-related coordinate system for sound direction from [63].



(A) Low frequency (B) High frequency

Figure 2.7: Illustration of the effect of head shadows on low and high frequency sounds from [60].

ILD complements ITD, which is more effective for localising low-frequency sounds, while ILD is more effective for high-frequency sounds. The auditory system uses both cues to accurately localise sounds across the entire audible frequency range on the horizontal plane.

ITD and ILD cues primarily contribute to the perception of sound localisation in the horizontal plane. However, they provide limited information for determining the elevation of a sound source, because when the sound source is located parallel to the cone region extending from the median plane, which results in multiple locations along the equidistant curve where both ITD and ILD have the same value, these locations are known as the cone of confusion [64]. This may lead to errors in judgement of the front-back or up-down location of sounds [65]. The cone of confusion is an imaginary cone extending outwards from each ear along the interaural axis that indicates in Figure 2.8 [66], representing the location of the sound source that produces the same interaural differences [67]. Listeners often unconsciously move their head, altering the sound path to reduce the confusion caused by the limitations of ITD and ILD cues. This is because when the head is rotated, the direction of the sound relative to the ear changes, and so do the localisation cues [68].

The process of hearing involves the ear converting sound waves from the environment into electrical signals that the brain can interpret. The ear consists of three main parts: the outer ear, the middle ear and the inner ear [69] as shown in Figure 2.9 [70]. Sound waves enter the ear through the outer ear, which consists of the pinna and the ear canal. The pinna helps to collect and direct sound waves into the ear canal. The sound waves travel along the ear canal and vibrate the eardrum, or tympanic membrane. The vibrations from the eardrum are transmitted to the

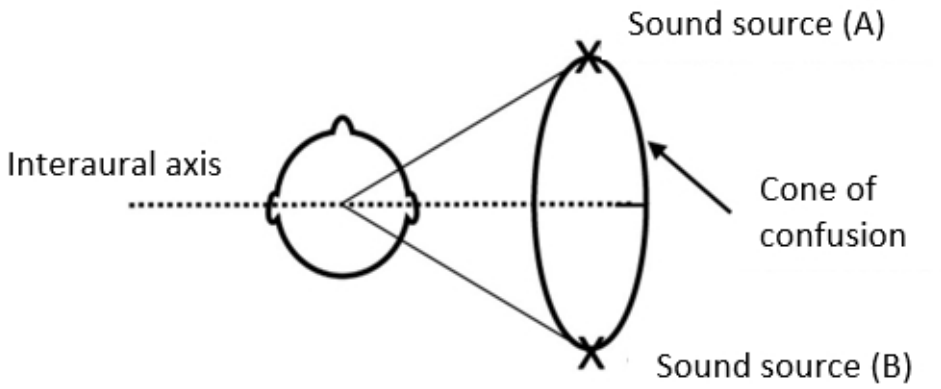


Figure 2.8: Illustration of the cone of confusion from [66].

middle ear, which contains three small bones (malleus, incus, and stapes) called ossicles [71]. These bones amplify the vibrations and transmit them to the oval window, an opening in the inner ear covered by a membrane [72]. The vibrations then enter the cochlea, a fluid-filled, spiral-shaped organ in the inner ear. Inside the cochlea are thousands of tiny hair cells called cilia. The vibrations cause the fluid in the cochlea to move, which in turn causes the cilia to move [71]. The movement of the cilia produces electrical signals. Different frequencies of sound cause different cilia to move, enabling us to distinguish different sounds [71]. These electrical signals are then transmitted to the brain via the auditory nerve. Once in the brain, particularly in the auditory cortex, these signals are interpreted as sound [71].

The spectrum of the sound is modified according to the listener's ear anatomy, primarily the shape and form of the outer ear (pinna). Thus spectral cues provide more useful localisation information for vertical localisation, allowing the auditory system to discriminate whether a sound is coming from up or down and front or behind. Incident sound from different spatial directions may be reflected by different parts of the pinna, and different parts of the pinna enhance or attenuate

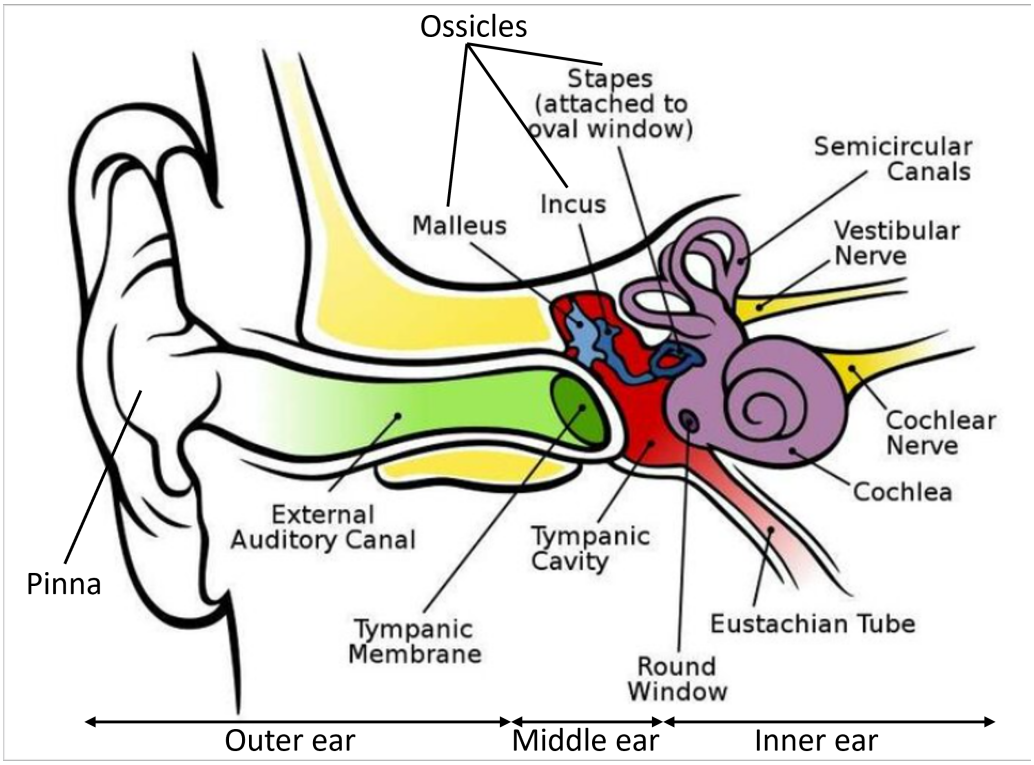


Figure 2.9: The anatomy of the ear from [70].

particular frequencies. As sound waves enter the ear, they interact with the folds and contours of the pinna. These structures reflect and diffract sound waves, altering the frequency spectrum of sound and thus helping the auditory system to localise directional information [61].

At low frequencies ($f < 1.5$ kHz), ITD is the main contributor to localisation. At medium frequencies ($1.5 < f < 2.5$ kHz), ITD and ILD work together. And in high frequencies ($f > 2.5$ kHz), ILD plays a major role [73]. When ITD, ILD and spectral cues cooperate with each other, HRTFs can help realise accurate location of sound source in virtual acoustics.

HRTFs are defined as left and right ear functions, H_L and H_R respectively [74].

Their expressions are shown in Equation 2.3:

$$\begin{aligned} H_L &= H_L(r, \theta, \phi, f, \alpha_s) = \frac{P_L(r, \theta, \phi, f, \alpha_s)}{P_0(r, f)} \\ H_R &= H_R(r, \theta, \phi, f, \alpha_s) = \frac{P_R(r, \theta, \phi, f, \alpha_s)}{P_0(r, f)}, \end{aligned} \tag{2.3}$$

where

- P_L and P_R are the complex-valued sound pressure in the frequency domain at the left and right ears respectively.
- P_0 is the complex-valued free field sound pressure in the frequency domain at the centre of the head with the head absent.
- Binaural HRTFs are the functions of source-to-head distance r, source position azimuth θ, elevation ϕ, frequency f and individual physiological parameters set α_s.

Therefore, HRTFs vary significantly between different people, so when a person listens to sound spatialised with a non-individualised HRTFs, dramatic distortion of sound location may be perceived [75]. For instance, obvious intracranial hearing of sound, front-back reversals and misjudgement of elevation may occur. Due to differences in individual ear and head anatomy, non-individualised HRTFs may not provide accurate spatial localisation for all users, with a consequent potential reduction in overall immersion and realism. Some users may experience auditory discomfort or disorientation if the non-individualised HRTFs are significantly different from their own. However, non-individualised HRTFs are more readily available and can be used immediately in a variety of applications without the need for individualised measurements, and are designed to work well for a large

part of the population, making them suitable for mass-market audio products.

In contrast, individualised HRTFs can provide a more accurate, natural spatial listening experience and improve intracranially hearing sound, front-back reversals and misjudgement of elevation to make virtual sound more plausible [76], because they match the listener's unique ear and head anatomy. For spatial audio auditory applications, immersion and realism can be greatly enhanced by using individualised HRTFs. However, measuring individualised HRTFs requires specialised equipment and an anechoic chamber, and the process of the measurement can be time-consuming, as it involves detailed measurements at multiple angles and frequencies, making it a complex and expensive process. It is therefore impossible to measure HRTFs for all users, so measuring individualised HRTFs in a simple way is a great challenge now or may become a future research direction.

Given these challenges, individualised HRTFs were not used in this work. The focus of this research is on developing a computationally efficient and broadly applicable reverberation algorithm that can be utilised in real-time audio applications. Using non-individualised HRTFs allows the findings and methods developed in this research to be more widely applicable and easier to implement across different platforms and for a broader audience. The complexity, time, and cost associated with individualised HRTFs measurement make it impractical for the scope of this study, where the priority is on accessibility and efficiency rather than personalised spatial accuracy.

In summary, while individualised HRTFs can provide the best personalised spatial audio experience, they can be limited in complexity and cost. Non-individualised HRTFs, while less accurate, provide a practical and cost-effective solution for a wider range of applications. The choice between the two usually depends on the

specific requirements and limitations of the application and the target audience. In this work, non-individualised HRTFs were chosen to ensure broader applicability and to focus on the efficiency and accessibility of the reverberation algorithm.

2.2.2 Head Tracking

Head tracking technology enables audio renderers to capture and recognise a user's head movements to significantly enhance the immersive quality of the listening experience, especially in applications such as virtual reality, gaming and spatial audio [77]. In virtual acoustic rendering applications, head tracking technology can track a user's head movements by capturing raw data from a camera or through the calculation of accelerometers and gyroscopes on a device, which can give the position of the wearer's head so that each simulated audio channel can be positioned as the user turns their head [1]. By tracking the listener's head movement in real time, audio playback can be dynamically adjusted to maintain a consistent and realistic soundstage. Thus, the listener's head movement is tracked and compensated for in order to keep the sound source stationary during head rotation [78], so that the sound field can dynamically change as the listener moves, similar to how we experience sound in the real world. Begault et al. [1] showed that head tracking is primarily used for eliminating reversals, as head movement resolved the cone of confusion, and head tracking improved the azimuth error by about 3 degrees.

2.2.3 Artificial Reverberation Simulation

Artificial reverberation simulation is a technique in audio processing and sound engineering that uses digital modelling or digital algorithms to create the effect of

sound reflecting off the surfaces of a room or space, thus adding a sense of spatiality to the audio for better listening enjoyment. It involves the use of algorithms and acoustic software to simulate the reflection, diffusion and attenuation of sound in space after the original source has stopped. Artificial reverberation simulation can be used to create realistic audio environments that match the visual elements of a virtual space, enhancing the user's sense of presence and significantly affecting the overall acoustic quality and functionality of the space. Reverberation in real spaces is created by complex patterns of sound reflected off walls and other objects that define the space. When the sound source becomes quiet, sound waves continue to reflect off walls, floors, and ceiling surfaces until the waves lose energy and dissipate. The prolongation of the reflected sound is regarded as reverberation [79]. The reflections modify the perception of the sound, changing the loudness, timbre, and most importantly, the spatial characteristics of the sound [80]. In virtual acoustics, if the sound is not reverberated or the reverberation time is too short, it sounds dry, and may not be aesthetically pleasing to listen to. Instead, if the reverberation time is too long, the sound may become unintelligible and music will sound muddy. For example, the ideal reverberation time for a concert hall is 2-2.25 s, while the ideal reverberation time for a classroom is 0.5-0.7 s [81, 82]. Only an appropriate reverberation can create a more realistic natural reverberation and a sense of space. Therefore, proper artificial reverberation simulation that can simulate room acoustic characteristics is critical for producing a plausible immersive experience in a virtual context.

The research and literature related to the perception of headset-delivered three-dimensional sound indicates that optimal spatialisation and localisation of virtual sound result from the combination of head tracking, individualised HRTFs and re-

verberation [1, 80, 83–86]. By comparing these three factors, head tracking is used to update the position of the stimuli in real time in response to head movement to reduce front-back reversals significantly, but neither reverberation nor individualised HRTFs have significant effect on the reversal rates [1]. Individualised HRTFs play an important role in localisation and help improve intracranially hearing sound, and to a large extent, reverberation helps sound externalisation and reduces azimuth errors and misjudgement of elevation [1, 74, 87].

As the aim of this thesis is to investigate a dynamic reverberation algorithm adapted to virtual acoustic rendering, the basics of reverberation, the artificial reverberation algorithms and how the dynamic reverberation algorithms enhance the virtual acoustic rendering experience are reviewed in detail below.

2.3 Reverberation and Room Impulse Responses

As expressed above, reverberation is the echo heard when a sound wave interacts with the environment and is reflected. Reverberation is usually described by RIRs in acoustics research. The reverberated sound can be seen as the result of convolving the source sound with RIRs. A RIR is a mathematical expression of reverberation, typically represented by a time-domain function $h(t)$, which describes the sound pressure at the listener's position as a function of time t, as shown in Figure 2.10. A RIR is the resultant pressure fluctuation measured at a receiving point due to an impulsive sound source at an arbitrary location in a room [6]. It represents the transfer function between the sound source and microphone, which contains all the acoustic characteristics of the indoor sound field. The RIRs can be measured in physical environments, and in addition the synthesis of RIRs can be achieved through various computational methods. Techniques such as delay

networks, computational acoustics and virtual analog models are employed to reconstruct RIRs effectively [88].

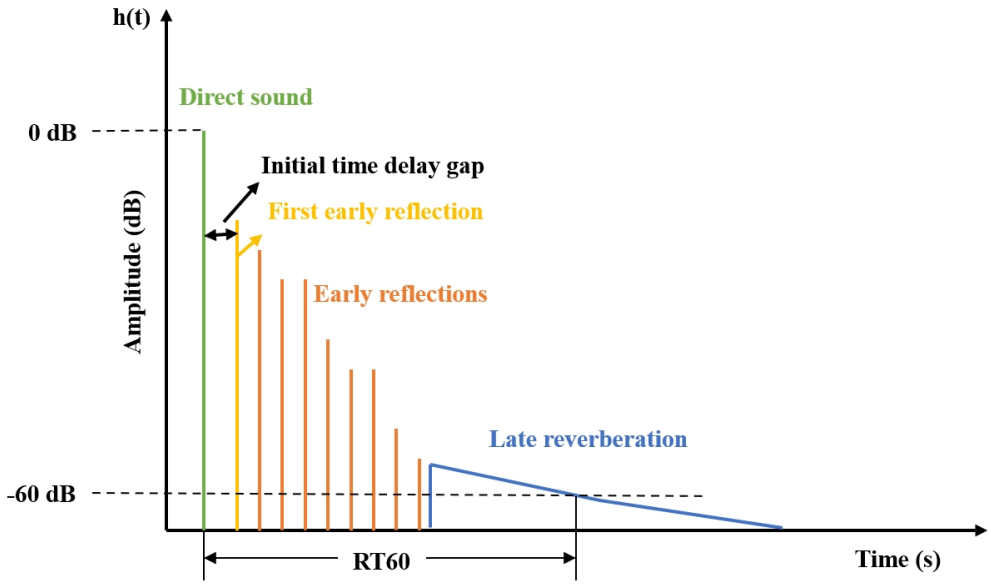


Figure 2.10: Representation of a room impulse response.

2.3.1 Composition of Room Impulse Responses

As the room impulse responses is a mathematical representation of reverberation, the study of reverberation presupposes an understanding of the composition of the room impulse responses.

In an enclosed space a proportion of the radiated sound source will be reflected by the room boundaries and eventually decay due to absorption by the room surfaces or air. Any impulsive stimulus will rapidly change the nature of the sound field from being coherent to partially coherent to non-coherent [89]. Initially, the sound field is coherent, meaning that the sound waves are in phase and exhibit a uniform pattern emanating from the source. When these waves interact with

the room, they begin to reflect, refract and scatter, leading to a transition to a partially coherent state. In this state, due to these interactions, the uniformity of the sound waves begins to be destroyed. Eventually, the sound field becomes incoherent. Here, the sound waves undergo multiple interactions, leading to complex and unpredictable wave patterns. A room is modelled as a linear time invariant system in acoustics [90]. Linear means that the response of the room (in terms of sound propagation) is proportional to the input (impulsive sound). Time-invariant means that the response does not change over time. For example, assuming the characteristics of the room do not change, the way the room responds to a sound today will be the same tomorrow. This linear time invariant system is characterised by a room impulse response [89], consisting of the following components (shown in Figure 2.10):

- **The direct sound (DS):** The DS reaches a listeners' ears directly from the source before being reflected from the boundaries of the enclosure [6]. Its amplitude is large with less energy loss relative to the reflections because of the shorter propagation path. Here Figure 1.1 and Figure 2.10 can help understand it. Its function is to transmit sound information and provide the direction of source.

- **The early reflections (ERs):** These are the sound waves that arrive in a temporal order after being reflected from at least one boundary of the enclosure [6]. They arrive within typically 10 to 80 ms after the direct sound, and typically constitute up to fourth order reflections before the sound field becomes stochastic. This stochastic sound field usually occurs after the early reflections. This is due to the fact that the reflections tend to scatter the sound rather than the ideal specular reflection. In addition, the density of re-

flections increases, making individual reflections less important. This means that although the sound field starts as a superposition of distinguishable individual reflections, it gradually becomes more and more diffuse and stochastic. As a result, any reflection path eventually becomes a statistical path rather than representing an ideal specular reflection path [9, 91]. Their energy is reduced by absorption and scattering. The point at which the sound field transitions from being dominated by distinguishable reflections to a diffuse reverberation is known as the mixing time [92, 93]. This is a key moment in room acoustics because it marks the critical transition point from clear, directional echoes to a rich reverberant field where individual reflections are no longer discernible [94]. In the period before mixing time, the sound field is dominated by early reflections, which are few and can be traced back to their origins. These early reflections are vital for providing cues about the size of the room and the location of the sound source. They enhance sound clarity and help in localising sounds in space. Once the mixing time is reached, the sound field transitions to a diffuse state where the reflections become so numerous and scattered that they no longer provide distinct directional information. The sound energy is distributed more evenly across the room, leading to what is known as the reverberant field.

Furthermore, early reflections are critical in shaping the perception of spatial characteristics of sound, including the size and nature of the environment, as well as for sound localisation [95, 96]. They help the listener to recognise the direction and distance of a sound source, thus enhancing the sense of space and dimension in the environment. Early reflections arriving with the first 50 ms after the direct sound are integrated for directional cues rather

than perceived separately [97]. Early reflections can increase the perceived overall sound pressure level and sound clarity, and contribute to the overall acoustic quality of a space [95,98]. When an RIR has appropriate strong late reflections, under certain conditions, early reflections have a conducive effect on recognition accuracy and a sense of space [97]. However, too many early reflections can lead to a comb-filtering effect, causing certain frequencies to be amplified or attenuated, which may affect the quality of the sound. Thus balancing early reflections is essential for room acoustics to ensure a natural and clear auditory experience. Golzer and Kleinschmidt [97] investigated the importance of certain portions of the impulse response in different contexts. They evaluated the importance of early and late reflections for the accuracy of automatic speech recognition and determined the effective time cutoff between conducive and detrimental portions of the impulse response. They found that when a detrimental late portion is removed, early reflections up to a certain critical delay time can carry useful information and contribute to the automatic speech recognition accuracy, and for room impulse responses measured or simulated in various types of rooms or spaces that have distinct acoustic properties, the cutoff time is in the range of 25 to 50 ms [97]. Christensen's study investigated the importance of early reflections in creating an ideal acoustic environment [99]. The authors found that the presence of early reflections was essential in creating a sense of envelopment and spaciousness and that the number of early reflections relative to the total reverberation time of the sound field influenced the perceived clarity of the sound. Lokki and Pätynen's research [100] has demonstrated that early reflections play a key role in sound quality, perceived dynamics and timbre. In particular, the

direction in which these reflections reach the listener is important for human spatial hearing.

- **The late reverberation (LR):** It is a stochastic sound field that consists of diffuse reflections [6]. It is an exponentially attenuated dense collection of echoes diffusing in all directions. The echo density is proportional to the square of time. As sound waves continue to reflect and diffuse around, they lose energy and their paths become more complex and scattered, resulting in a diffuse sound field. It is thus the end result of multiple sound reflections from various surfaces in a room. This diffusion causes sound waves to arrive from multiple directions, creating the reverberant environment and enveloping quality typical of enclosed spaces. Paulus et al. [101] showed that the shape of the reverberation tail has an important effect on the perceived level for an equal average loudness level: a long reverberation tail has a higher perceived level although it has a lower absolute instantaneous level. An appropriate amount of late reverberation can contribute to a sense of spatialisation and fullness, but too much reverberation can result in reduced intelligibility of speech [102].

The intensity ratio of the direct sound to early reflections and late reverberation is the main factor for depth location, so when adding appropriate reverberation sound to the original dry sound, the sound image has a certain depth sense [103].

2.3.2 Parameters of Room Impulse Responses

The acoustic parameters of the room impulse responses are important factors in assessing reverberation. These parameters can be used to clarify the attributes

of RIRs and as tools for objective analysis of reverberation.

Initial time delay gap (ITDG): This is the time period between the direct sound and the first arriving reflection as shown in Figure 2.10. ITDG is the main contributor towards the perception of 'presence' [104], an attribute that is recognised as the perceptual sense of feeling boundaries of an enclosed space [105]. It is the hearing-equivalent of 'seeing' the walls of a room [106]. Hyde [107] observed that a short ITDG generally indicates an important contributing factor to acoustical quality in a hall through discussing its relation to acoustical intimacy. Beranek [103] reported that the listener's impression of the size of a hall is determined by the time delay of the first major reflection after the direct sound. He also observed that halls that have intimate acoustics had ITDG values at or shorter than 20 ms, and that the shorter the ITDG, the more intimate the experience [103]. He also stated that with a short ITDG, more reflections can occur in the first 80 ms after the arrival of the direct sound, and more early reflections contribute to a greater feeling of intimacy [108]. ITDG has been shown to have a significant impact on the perceived space and envelope of the sound field, as it provides important cues for source localisation and distance perception [109]. A study by Jeong et al. [110] showed that ITDG can affect the perceptual reverberation of a sound field by presenting listeners with simulated sound fields with different ITDGs and asking them to rate the amount of reverberation perceived in the sound. The results showed that as the ITDG increased, the perceived amount of reverberation increased even if the total reverberation time of the sound field remained constant. ITDG typically decreases with increasing distance between the source and receiver, as the travel difference Δd between direct and first re-

flected sound is reduced. Figure 2.11 [109] is given as an example. When the receiver is in position 1, d_{D1} represents the direct sound travel and d_{R1} represents the first reflected sound travel. Δd_1 (d_{R1}-d_{D1}) represents the travel difference at position 1. Similarly, Δd_2 (d_{R2}-d_{D2}) represents the travel difference for position 2. The receiver at position 2 is significantly further from the source than the receiver at position 1, and the travel difference Δd_2 is significantly smaller than Δd_1.

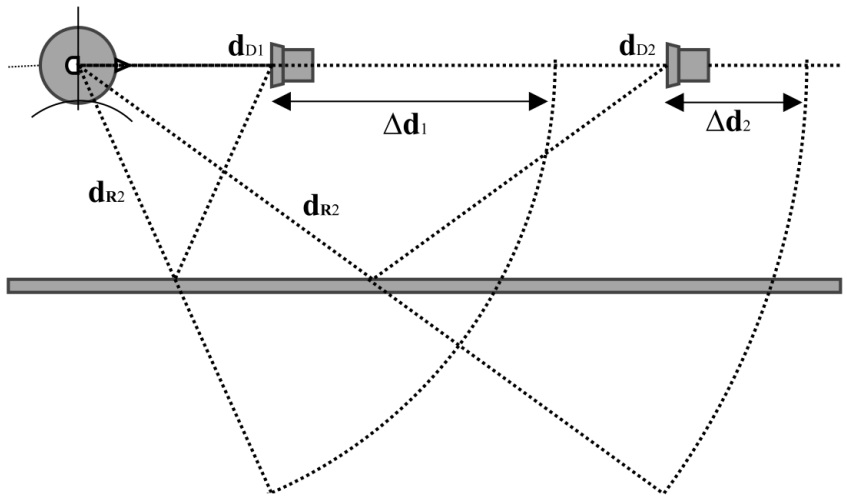


Figure 2.11: Schematic diagram of ITDG's travel difference variation from [109].

Energy decay curve (EDC): The fading of sound energy as a function of time is defined as the EDC, which indicates how the sound energy decreases with time after the source stops emitting as shown in Figure 2.12. It is the tail integral of the squared impulse response at time t and can be calculated with the Schroeder integral [111, 112] as shown in Equation 2.4:

$$\mathrm{EDC}(t) = \int_t^\infty h^2(\tau)\mathrm{d}\tau, \tag{2.4}$$

where $h(k)$ is the measured impulse response. EDC(t) describes the remaining energy of the impulse response after the time t. The decay of the EDC is smoother than the impulse response itself, so it is more useful for estimating reverberation times than the normal amplitude envelope.

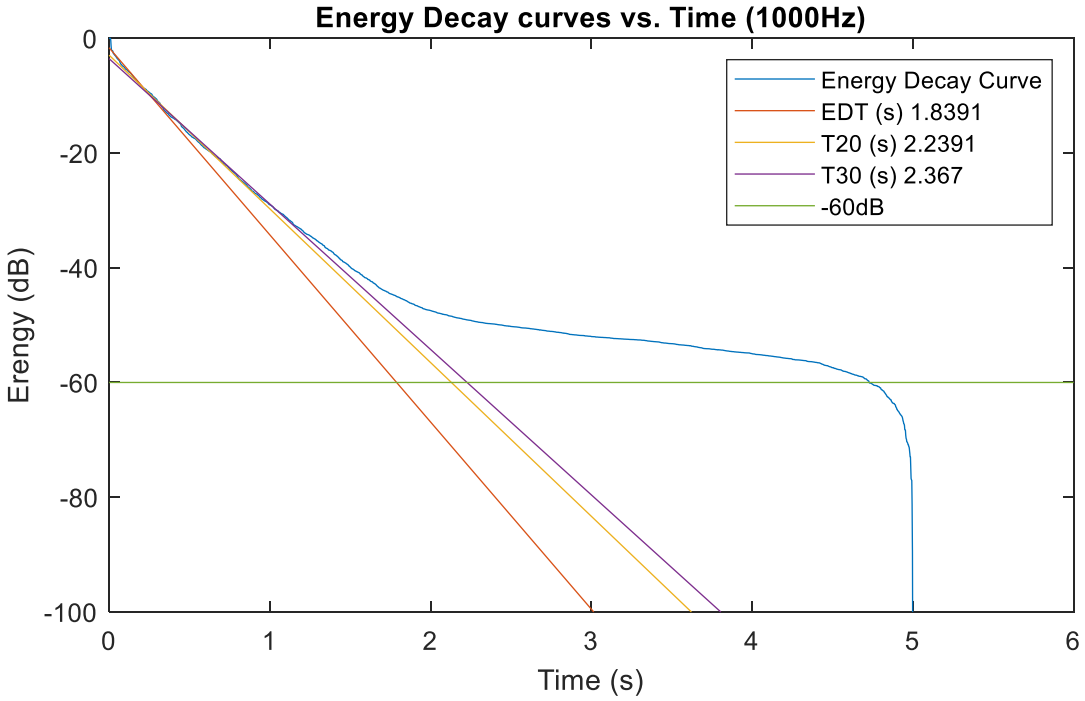


Figure 2.12: Schematic representation of the energy decay curve.

EDC allows the derivation of important acoustic parameters such as early decay time (EDT), T_{20}, T_{30} and RT_{60}, which are essential for assessing the acoustic properties of a room.

Energy decay relief (EDR): The reverberation energy can be expressed as a time-frequency representation, which is called EDR [111–113]. It can be calculated as:

$$\mathrm{EDR}_h(t, f) = \int_t^\infty p_h(\tau, f)\mathrm{d}\tau, \tag{2.5}$$

where $p_h(t, f)$ is the energetic time-frequency representation of the signal $h(t)$. $\mathrm{EDR}_h(t, f)$ describes the remaining energy of the impulse response after the time t in a frequency band centered about f.

It provides a detailed analysis of how sound energy decays in a room over frequency bands as shown in Figure 2.13 [114]. The plot clearly illustrates that higher frequencies tend to decay faster than lower frequencies. This could be due to various factors including the absorption characteristics of the air and room surfaces which are generally more effective at absorbing higher frequencies. Low frequencies show a more gradual decay. This is typical in enclosed spaces where low frequencies have longer wavelengths and are not as easily absorbed by surfaces. Unlike conventional energy decay curves that show decay over time, EDR illustrates the rate of decay at different frequencies. It is useful for understanding the frequency-dependent behaviour of room acoustics, especially with regard to frequency-specific reverberation times. By analysing the EDR, acoustic problems in specific frequency ranges can be more accurately assessed and solved, leading to more targeted and effective acoustic studies.

Reverberation time: Reverberation time is defined as the time taken for the sound pressure level in a room to decay to 60 dB below source level after the source is muted.

Reverberation time can be calculated by the Sabine formula and the Eyring formula [115]. Sabine is the first researcher who began to predict reverberation time in 1900 [116]. He directly related reverberation time to the volume of the room, and created a simple formula inversely proportional to the absorptivity of the surfaces of walls and other objects. The Sabine formula at normal room

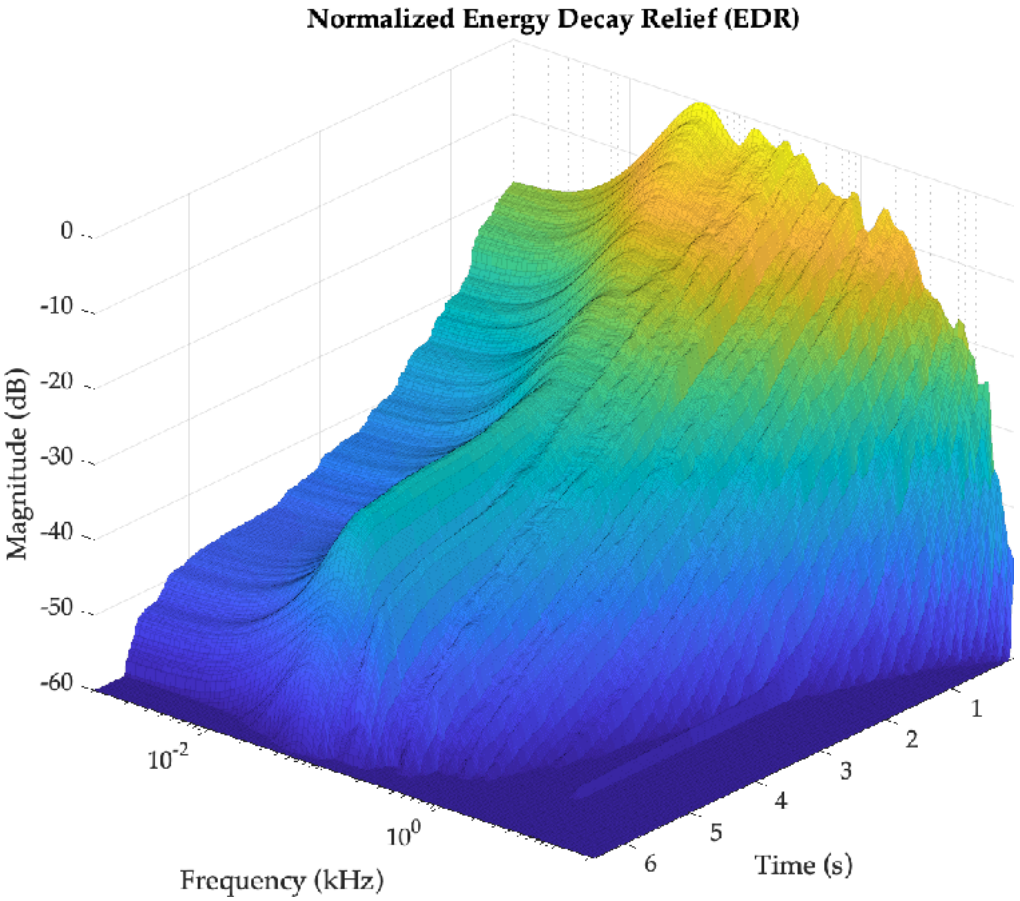


Figure 2.13: Schematic representation of the energy decay relief from [114].

temperature (22 ℃) is described as:

$$T_{60} = \frac{0.161V}{Sa + 4mV}, \tag{2.6}$$

where V is the volume of a room in cubic meters and S is the sound absorption area of the room, m is the sound attenuation constant of the air in units of m^{-1} and a is the sound absorption coefficient. In fact, when there is a room covered with material that has a fully absorption coefficient, in other words, the surface of the material does not have reflection sound, the sound should be absorbed fully,

resulting in the reverberation time should be zero. However, according to the Sabine formula (Equation 2.6), when the sound absorption coefficient a is 1.0, the calculated reverberation time is not zero. As a modification, Eyring presented an alternative formula that can calculate a reverberation time infinitely close to zero with a material infinitely close to the full absorption coefficient of 1.0 in 1930 [116]. The Eyring formula is described as:

$$T_{60} = \frac{-0.161V}{S\ln(1-a) + 4mV}, \tag{2.7}$$

where the parameters V, S, a and m have the identical meaning with them in the Sabine formula.

The Sabine formula was originally derived from average sound energy loss of a room by calculating differential equations, but actually, it could also be derived from the Eyring formula. The natural logarithm $\ln(1-x)$ has Maclaurin series [117],

$$\ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n} = -x - \frac{x^2}{2} - \cdots - \frac{x^n}{n} - \cdots \tag{2.8}$$

It converges for $|x| < 1$ and $x = -1$. According to the Equation 2.8, the Equation 2.7 can be expressed as Equation 2.9:

$$T_{60} = \frac{-0.161V}{S(-a - \frac{a^2}{2} - \frac{a^3}{3} - \cdots - \frac{a^n}{n} - \cdots) + 4mV} \tag{2.9}$$

The absorption coefficient a is always less than or infinitely close to 1, so the series $(-a - \frac{a^2}{2} - \frac{a^3}{3} - \cdots - \frac{a^n}{n} - \cdots)$ is always convergent. When a tends to 0, the series can be approximated by $-a$, so Equation 2.9 can be approximated by Equation 2.6. Thus the Sabine formula can be deduced from the Eyring formula.

Experiments [115] indicated that the Sabine formula can underpredict rever-

beration time. The reverberation time calculated by the Sabine formula is more accurate when the absorption coefficient is lower because it provides longer reverberation time. When the average absorption coefficient a is lower than 0.2, the Sabine formula can provide the most accurate results [118]. On the contrary, the Eyring formula is more accurate than the Sabine formula when the average absorption coefficient is higher than 0.2 [118, 119]. In practical applications, the choice of reverberation formula can be determined by the absorption coefficient.

In addition, the energy decay curve was also introduced to measure and define the reverberation time. The measurement of reverberation time is difficult to achieve a full 60 dB attenuation due to background noise etc., so the derivation of the reverberation time is usually done using an attenuation of 10 dB, 20 dB or 30 dB of the sound pressure level. The different RT_{60} measurements (e.g. EDT, T_{20}, T_{30}) are derived by calculating the slope of the best-fit line for the energy decay curve over different ranges. Figure 2.12 is an example to illustrate the derivation of EDT, T_{20}, T_{30}. According to ISO 3382-2:2008 standard [120], T_{20} is derived by multiplying the time taken to reduce the sound pressure level of the source from -5 dB to -25 dB by 3. Similarly, T_{30} is obtained by multiplying the time taken to reduce the sound pressure level from -5 dB to -35 dB by 2. Since the initial decay of the SPL is most important to the subjective perception of reverberation in a room, the EDT is used to characterise the rate of decay of the initial SPL. Although it is similar to the calculation of T_{20} and T_{30}, which is derived by multiplying the reduction in sound pressure level of the source from 0 to -10 dB by 6, it is not generally used to characterise RT_{60}.

Definition and Clarity: Both Definition and Clarity are measures of the

intelligibility and clarity of speech and music [121]. Definition is the ratio of early-to-total sound energy. It is defined as [122]:

$$D_{t_e} = \frac{\int_0^{t_e} h^2(t)\mathrm{d}(t)}{\int_0^{\infty} h^2(t)\mathrm{d}(t)} \cdot 100\% \tag{2.10}$$

Clarity is the logarithmic ratio of early-to-late sound energy, which is expressed in dB. It is an objective measure of the clarity or intelligibility of sound. It is defined as [122]:

$$C_{t_e} = 10 \log \frac{\int_0^{t_e} h^2(t)\mathrm{d}(t)}{\int_{t_e}^{\infty} h^2(t)\mathrm{d}(t)}, \tag{2.11}$$

where $h(t)$ is the impulse response, t_e equals 50 or 80 ms. When Definition or Clarity is related to the musical perception, the early time t_e is limited to 80 ms, whereas if Definition or Clarity is related to speech, t_e is limited to 50 ms. Clarity (C50, C80) is the ratio between the 'useful energy' received in the first 50 or 80 ms and the 'detrimental energy' received afterwards [123]. High values of Definition and Clarity indicate a large amount of early energy, which coincides with the subjective perception of clarity. Conversely, low values of Definition and Clarity indicate unclear and excessively reverberant sound [121]. However, it is common to use D50 to measure the ratio of early to total sound energy in the first 50 milliseconds. D50 focuses on speech intelligibility, with higher D50 values indicating a clearer speech. C80 evaluates the ratio of early to late sound energy in the first 80 milliseconds. C80 focuses on musical balance, with higher values indicating a clearer sound, which is critical for music that requires clarity and reverberant mixing.

The relationship between Clarity and Definition is

$$C_{t_e} = 10 \log \frac{D_{t_e}}{1 - D_{t_e}}, \tag{2.12}$$

so, knowing either of these two parameters, it is possible to derive the other.

Centre Time: Centre Time is defined as the centre of gravity of the squared impulse response [124] as Equation 2.13:

$$T_c = \frac{\int_0^\infty t \cdot h^2(t) \mathrm{d}(t)}{\int_0^\infty h^2(t) \mathrm{d}(t)} \tag{2.13}$$

T_c is related to the perceived definition or the balance between clarity and reverberance, and avoids the discrete division of the impulse response into early and late reflections or energy, like with for instance the clarity C80 [124]. A low T_c indicates that most of the energy arrives early with sensation of clarity, whereas a large T_c suggests the energy arrives late with a reverberant sound. The specific values for T_c in acoustics can vary depending on the type of room and its intended use. For instance, in concert halls designed for classical music, a slightly higher T_c might be preferred to enhance the richness and fullness of sound. On the other hand, in spaces like lecture halls or theaters where speech clarity is paramount, a lower T_c is desirable to ensure early arrival of sound energy and clearer speech perception. There isn't a fixed numerical range that applies universally, as the optimal Tc value is context-dependent, balancing between clarity and reverberation for the specific acoustic requirements of each space.

These parameters collectively describe the balance and interaction between early reflections and late reverberation in a room. Early reflections are crucial for

clarity and speech intelligibility, as measured by parameters like ITDG, Clarity, and Definition. On the other hand, late reverberation contributes to the sense of space and envelopment, as characterised by parameters like RT_{60}, EDR, and Centre Time.

Direct-to-reverberant (D/R) energy ratio: The D/R energy ratio is defined as the ratio of the energy of the direct sound arriving at a listener's position to the energy of the reverberant sound in the same environment [125]. Mathematically, the D/R energy ratio can be expressed as

$$\mathrm{DRR} = \frac{E_D}{E_R} \tag{2.14}$$

where E_D is the energy of the direct sound and E_R is the energy of the reverberant sound. This ratio can also be measured in decibel scale (dB) by taking 10 times the logarithm of the energy ratio:

$$\mathrm{DRR_{dB}} = 10 \log \frac{E_D}{E_R} \tag{2.15}$$

A higher D/R ratio typically enhances speech clarity, making it essential in spaces like classrooms and conference rooms, while a well-balanced D/R ratio contributes to the richness and clarity of music in concert halls and recording studios. The D/R energy ratio can be used as an acoustic cue in a reflective environment in relation to a one-to-one distance, avoiding confusion with source characteristics [126]. In a typical listening room, the direct sound field energy decays proportionally to the (logarithmic) distance, while the energy of the reverberant sound field is independent of the distance. Therefore, D/R energy ratio can in principle be used

to estimate the distance of a sound source [127].

2.3.3 Measurement of Room Impulse Responses

A room impulse response can be measured with conventional and digital sound sources. Conventional sources are natural sound sources such as balloon explosions, gun firing or applause. These types of sources present difficulties such as an uncontrolled spectral response and poor reproducibility [128]. Digital sources include noise signals such as Maximum Length Sequence (MLS), linear sine sweep (LSS) or exponential sine sweep (ESS) signals, etc [129, 130]. These computer-generated digital signals are converted to analogue signals, amplified by a power amplifier and then emitted by a loudspeaker. They are then captured by a microphone at the measurement point and converted into a digital signal back to the computer, which is then processed to obtain the RIR of the sound field. Digital sound sources are the preferred stimuli for room impulse responses measurements due to the reduced price of faster processors and the availability of high quality audio equipment [131].

Maximum Length Sequence noise is a periodic pseudo-random binary sequence consisting of random sequences of zeros and ones [132]. The generated MLS is used as the excitation signal for the acoustic system to be measured. The impulse response of the system to be measured is obtained by sampling this output signal at the output and performing a cross-correlation operation with the input MLS signal under the same sampling. This method is highly influenced by time variation when measuring the room impulse responses and is poorly resistant to non-linear distortions, so the room impulse responses obtained during deconvolution may cause artificially distorted peaks [130, 133].

A linear sine sweep increases or decreases frequency at a constant linear rate over time [133] and it is presented by

$$S(t) = \sin(\omega_1 t + \frac{\omega_2 - \omega_1}{T} \cdot \frac{t^2}{2}), \tag{2.16}$$

where ω_1 and ω_2 are the start and end frequencies of the sweep (in radians) and T is the total time duration in seconds. The impulse response can be obtained by convolution of the measured signal at the microphone with the inverse filter (time-reversed input sine sweep). This method is sensitive to distortion artefacts introduced by the loudspeaker, and impulsive noise during recording [134].

An ESS increases or decreases the frequency at an exponential rate [130, 133] and it is presented by

$$S(t) = \sin[K \cdot (e^{\frac{t}{L}} - 1)], \tag{2.17}$$

where

$$K = \frac{\omega_1 \cdot T}{\ln(\frac{\omega_2}{\omega_1})}, L = \frac{T}{\ln(\frac{\omega_2}{\omega_1})}, \tag{2.18}$$

and ω_1 and ω_2 are the start and end frequencies of the sweep (in radians) and T is the time duration of the sweep in seconds. Similar with the LSS method, the impulse response can be obtained by convolution of recorded signal with time reversed version of input sine sweep signal. Compared with LSS with the flat spectrum, the energy of the ESS decreases with the increasing frequency, so compensation is required for the increased low frequency energy in the exponential sweep. ESS utilises different harmonic nonlinear properties to better distinguish impulse response from other harmonic distortions and is therefore widely used in room acoustic measurements [133].

2.3.4 Synthesis of Artificial Room Impulse Responses

In addition to impulse responses that can be obtained from measurements in real space, synthetic methods such as wave-based modelling of room acoustics, geometrical acoustics and recursive reverberation algorithms based on feedback delay networks can also simulate room impulse responses.

Wave-based modelling of room acoustics is achieved through numerically solving the wave equations. Usually the space or its boundary surfaces are discretised into small elements and the interaction between them is modelled [135]. The wave-based approach takes into account wave interference and diffraction effects, which means that it can accurately model the low-frequency components of the room impulse response, where constructive and destructive wave interference forms the room modes [136]. Room modes have the effect of amplifying and attenuating specific frequencies in the room impulse response and produce many subjective sound 'colours' or 'charactersitocs' of the room. The reproduction of these room modes is therefore essential for assessing the acoustics of rooms such as concert halls and recording studios or for producing pleasing reverberation. Wave-based modelling is able to provide the most accurate results, but the computational cost of the method increases rapidly with frequency, making it computationally very expensive [135]. Therefore, this method is suitable for low frequency conditions.

Computational acoustics, also known as Geometrical Acoustics (GA), is a method that simulates the propagation of sound in a specified geometric space [88]. GA simulate sound propagation by assuming the sound paths as simple rays. When the ray hits an object such as a wall, it is reflected to a new direction. This means that the wave properties of the sound are ignored. GA rendering therefore does not cover the diffraction of sound waves. This leads to a limitation of the GA

rendering method. Because the sound wavelength is shorter in comparison to the surface dimensions and overall dimensions of the room at high frequencies, the smaller the role played by wave phenomena, the smaller the errors from GA rendering [9]. At lower frequencies, however, the approximation error increases as wave phenomena play a greater role. Research in geometric room acoustic simulation therefore focuses on more accurate modelling of propagation effects, such as diffraction, scattering and other wave effects in a room [137], so that a reasonable auralisation of room acoustics can be achieved. The computational cost of GA is lower compared to wave-based rendering, but at the same time the accuracy is also lower [9].

Delay networks synthesise RIRs to simulate the reverberation process using delay lines, filters, and feedback connections [88]. It is the method of filtering unit impulse function (also know as Dirac delta function) [138] with a reverberator constructed by filter structures. Dirac delta function, theoretically defined as an idealized impulse that occurs at a single point in time with infinite amplitude and zero duration, yet integrates to one, as shown in Figure 2.14 [139], is conceptually important but not directly used in digital system. Instead, the unit impulse response, which consists of a single non-zero value (typically one) at the origin and zeros elsewhere, is used in digital implementations. As the unit impulse is processed through the reverberator, it is modulated by the filter structures to produce an output that reflects the time it takes for sound to bounce around the room and reach the listener from various angles. This output, known as the room impulse response, characterises the acoustic properties of the space. Reverberation research based on delay networks pays more attention to selection of appropriate gain, delay, decay and feedback matrix values, which is crucial to produce a more authentic

reverberation and good sound quality. It is useful to examine the output of such algorithms by looking at their response to a discrete unit sample function [140] passing through various filters and delay lines and feedback connections. This is equivalent to visualization of RIRs, and is readily achieved within the MATLAB environment [141], which has a rich filter structure database. These make artificial reverberation algorithms based on delay network structures have flexible parameterisation characteristics and better real-time performance [142]. In addition to recursive filters based on delay networks, simple finite impulse response (FIR) filters can be used to implement early reflection simulations. As a filter with no feedback support, FIR filters can be designed with arbitrary amplitude-frequency characteristics while maintaining strict linear phase characteristics, and are widely used in speech and data transmission [143]. However, because the filter has no feedback, more coefficients are required in the system equations to accurately simulate the behavior of physical systems, such as the multiple reflections in a room, which require detailed and precise filter responses [144]. For the purposes of virtual rendering, simplicity, real-time performance and computational efficiency must be considered. Therefore, on the basis of ensuring the effectiveness of the reverberation algorithm, the number of coefficients of the FIR filter should be set properly in order to ensure the speed of the algorithm and to reduce its memory requirements.

2.3.5 Binaural Reverberation

Humans use both ears to perceive the spatial properties of sound, so binauralisation is a more realistic simulation of the human auditory system than mono reverberation. The use of two channels to simulate a realistic auditory experience

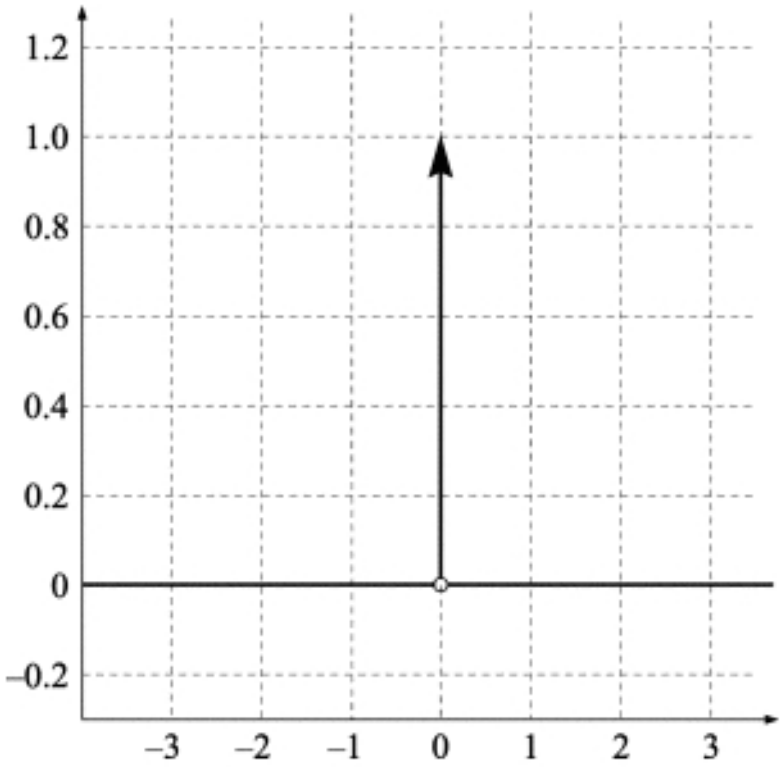


Figure 2.14: Schematic representation of the Dirac delta function [139].

is known as binauralisation. Binaural reverberation is a technique used to reproduce a realistic reverberation experience in an enclosed acoustic space and has a wide range of applications in music production and virtual acoustic rendering. By understanding how binaural reverberation works, it is possible to design and construct acoustic environments that are tailored to a particular application. This can help to reduce the negative effects of noise, as well as improve the clarity and intelligibility of speech. This type of acoustic design can also be used to create sound environments that are more enjoyable for the listener. It involves the creation of a simulated acoustic environment within headphones that simulates the sound of being in the same room as the source. Binaural reverberation is achieved by convolving the source audio with computed or physically measured binaural

room impulse responses. The binaural head-related impulse responses recorded in the room are often referred to as binaural room impulse responses [145], which consists not only of reverberation, but also cues of human source localisation, including interaural time and level differences and spectral cues induced by the torso, head and pinnae.

As with RIR measurements, high-resolution binaural room impulse response signals are also rarely measured with impulsive sound sources, such as balloon explosions, firing guns and applause. Instead, computer-generated broadband stimuli, like MLS noise, LSS and ESS, etc. [129] [130] are used to excite the room and a deconvolution process is used to extract the resultant BRIRs.

In order to measure a BRIR, in addition to a suitable stimuli, it is most important to have a dummy head with in-ear microphones. A dummy head is a binaural recording array that closely replicates the shape, size, and acoustic properties of the human head and torso, and is commonly used for acoustic measurements such as binaural room impulse response. The purpose of the dummy head is to capture audio identical to that heard by a human listener, taking into account the natural acoustic shadows, reflections, and diffraction caused by the shape of the head, ears, and torso. These devices are widely used in spatial audio recording and research to measure how sound interacts with the listener's environment, providing a detailed representation of the sound field from the listener's perspective. Suitable dummy heads (Shown in Figure 2.15) include the widely used Neumann KU100 [146] or G.R.A.S. KEMAR head and torso [147]. Alternatively, BRIRs can be simulated in room acoustics software such as ODEON [148]. ODEON is a room acoustics simulation software. It is designed to model and analyse sound propagation and distribution in virtual spaces. By simulating various acoustic parameters

and conditions, ODEON allows the user to predict the behaviour of sound in real or hypothetical spaces without the need for physical measurements.

(a) (b)

Figure 2.15: Dummy heads for BRIR measurement. (**a**) Neumann KU 100, reproduced from Neumann [149]. (**b**) G.R.A.S. KEMAR, reproduced from G.R.A.S. Sound & Vibration [150].

When reproduced binaurally, reverberation increases the sense of externalisation, i.e. the illusion that the virtual sound source is outside the head [1, 151]. Previous research has investigated the contribution of monaural and binaural cues to externalisation of the reverberant binaural signal. The limited importance of monaural cues has been demonstrated by Hassager et al. [152] and Jiang et al. [153], who argue that spectral detail is not as important in reverberation as it is in direct sound. Studies have demonstrated that binaural cues are crucial in increasing externalisation and other subjective properties. Leclere et al. [1] suggest that reverberation increases externalisation of the binaural signal whenever interaural differences are introduced. This was supported by Catic et al. [154], who reported that externalisation was significantly reduced when the reverberant portion of au-

ralised speech was presented monaurally. This effect is related to specific binaural cues such as ILD and interaural coherence (IC) [155].

The interaural cross-correlation coefficient (IACC) is a measure used in acoustics to quantify the similarity between the signals reaching the two ears. It is a critical parameter for assessing spatial properties of sound, particularly in terms of how sound is perceived in a binaural or three-dimensional space [156]. Higher IACC values indicate a high degree of similarity between the signals at the two ears, which corresponds to a more central or less spatially dispersed sound source. Lower IACC values suggest greater differences between the ear signals, indicating a more spatially dispersed or enveloping sound field. The IACC is used to assess spatial impression and envelopment in concert halls, virtual acoustic environments, and other acoustic spaces.

2.3.6 Convolution and Auralisation

Convolution and auralisation of reverberation are two important concepts in acoustic research. Convolution is the process of combining two signals to produce a new signal. It is done by multiplying two functions' overlapping values at each position as one function shifting over the other, and summing the resulting values to produce a new value for the resulting function at that position [157] as shown in Equation 2.19.

$$y[n] = x[n] * h[n] = \sum_{m=-\infty}^{\infty} x[m]h[n-m], \tag{2.19}$$

where $x[n]$ presents the input signal, $h[n]$ presents the impulse response, $y[n]$ presents the convolved signal, and * denotes convolution [158]. If the length of

input signals $x[n]$ and $h[n]$ are M and N respectively, the length (L) of the convolved signal ($y[n]$) will be $L = M + N - 1$. This is because the first value of $y[n]$ occurs when the first value of $x[n]$ overlaps with the first value of $x[n]$, and the last value of $y[n]$ occurs when the last value of $x[n]$ overlaps with the last value of $h[n]$.

Convolution reverberation is a technique used to simulate the reverberation of an acoustic space, it involves applying an impulse response to a sound source, which can be used to recreate the acoustics of a particular room or environment [159]. Convolution is a powerful tool for creating complex sounds and manipulating existing audio signals. By using convolution, arbitrary reverberation impulse responses can be applied to audio data.

Auralisation can be used to create realistic virtual soundscapes for video games, films, and other applications. It involves the synthesis of a sound source within a simulated acoustic environment, taking into account the characteristics of the environment, such as reverberation, absorption, and diffusion. Room auralisation is typically achieved through convolution of source audio with a computationally derived or physically measured impulse response [160].

Whether RIR or BRIR, once the impulse response has been obtained from measurement or in simulation modelling, it can be convolved with an anechoic audio sample using signal processing to achieve auralisation [161]. This auralisation includes loudness, timbre and spatial characteristics of the measured or modelled space. The difference is that using RIR typically results in mono or non-binaural stereo sound production, while BRIR represents the reproduction of binaural sound [161].

Research has consistently shown that auralised sounds closely mimic actual

acoustics, supporting their use in virtual environments. Kearney [162] compared the perceptual differences between actual acoustic recordings and 'virtual' recordings. The actual recording signals here refer to the recording signals measured with a dummy head in the test environment, and the virtual convolution-based recording signals refer to the direct source signal rendered with the BRIRs measured with a dummy head. He assessed the differences between real and virtual violin and female speech recordings on five subjective attributes (source width, reverberance, clarity, naturalness and source movement). The source width refers to the auditory width or spatial extent of the soundfield created by a source as perceived at a particular listener position. Reverberance describes the duration of the reflected sound. Clarity denotes how well-defined and intelligible the sound is perceived to be. Naturalness is accurate and realistic tonal quality. Source movement represents small movements of the sound source within the acoustic space. It was found that, the virtual acoustic recordings of violin samples had a smaller source width compared to the actual recordings. For the reverberance attribute, the samples did not differ statistically. For the clarity attribute, in the case of female singing, the virtual recording showed a significant improvement in clarity over the actual recording. No significant differences were found between the actual and virtual recordings in terms of natural timbre. For both source types, no variation in source movement was perceived overall.

Blau et al. [163] compared the various perceptual properties of real and auralised rooms in 2018. Their results showed that when using measured BRIRs, even non-individual BRIRs, highly convincing speech auralisation is possible. Small defects in the simulations can be reliably detected. Blau et al. [164] later compared the consistency of the auralisation of measurement-based binaural room impulse

response with the real source (speech signal) for five attributes (reverberation, source width, source distance, source direction, and overall quality). The results showed that, although the measured BRIR set for 'reverberance' was rated slightly lower than the other attributes with a median score of around 7 on a nine-point scale, for all other attributes, the agreement with the real room was rated with medians of 7.5 or higher on the nine-point scale based on the measured BRIR. The fact that the measured BRIR sets were capable of eliciting such high ratings suggests that the measured impulse response can be made realistically audible for speech. Therefore, it was concluded that close-to-real binaural auralisations of speech are possible if all modalities (auditory, visual, etc.) are appropriately reproduced [163].

Lokki et al. [165] conducted a comprehensive evaluation of algorithmic simulations of binaural auralisation compared to real head recordings. They modeled direct sound and early reflections via the image-source method and included edge diffraction modelling. Their objective analysis revealed that while the spatial properties of the auralizations were nearly identical to those of real head recordings, some timbral differences, particularly at lower frequencies. Subjective listening tests involving thirteen participants confirmed that the algorithmic auralizations, especially for clarinet sounds, were almost indistinguishable from real recordings. However, differences were more noticeable with transient-like signals such as guitar and snare drum. Despite these timbral variations, the study concluded that the auralization algorithm generally achieved reasonable and natural binaural auralisation, affirming its efficacy for realistic virtual acoustic environments.

These results confirm that auralisation is an effective tool for crafting authentic auditory environments. The demonstrated perceptual alignment between real

and virtual recordings in prior research supports the use of anechoic audio samples processed with measured binaural room impulse responses to simulate 'real' reverberation. However, it is noted that most evaluations were conducted in static environments without interactive elements like head tracking or testing in Augmented Reality or Virtual Reality scenarios.

2.4 Artificial Reverberation Algorithms

Artificial reverberation algorithms are computer algorithms used to digitally simulate the sound of reverberation. These algorithms are used to create a sense of space for sound, and can range from simple echo effects to complex simulations of real-world acoustic environments. Artificial reverberation algorithms typically involve the processing of the input signal to create a series of delayed and attenuated versions of the signal, which are then mixed together to create the desired reverberation effect. Common parameters used to control the sound of the reverberation include the reverberation time, the decay time, and the pre-delay time.

Digital reverberation algorithms, such as artificial reverberation based on delay networks, are composed of digital filters. This traditional approach to synthesise reverberation involves combining feedforward paths to render early reflections, and feedback paths to synthesise late reverberation [166]. To implement these algorithms, basic filter structures such as delay filters, low-pass filters, comb filters, and all-pass filters, among others, are necessary [141]. A detailed description of these basic filter structures is provided below.

- **Delay filter:** This is used to create a delay effect in audio signals. They work by delaying the signal by a certain amount of time. The purpose of

a delay filter is to shift the entire signal in time without altering its shape or other characteristics. In stereo systems, delay filters are used to create depth or spatial effects by slightly delaying the audio signal in one channel relative to another. In addition, delay filters are an integral part of the construction of more complex filters, such as comb filters or all-pass filter, where the original signal is mixed with its delayed version to create periodic cancellations or enhancements. The structure of the delay filter is shown as Figure 2.16, and its transfer function is given by

$$H(z) = z^{-m} \tag{2.20}$$

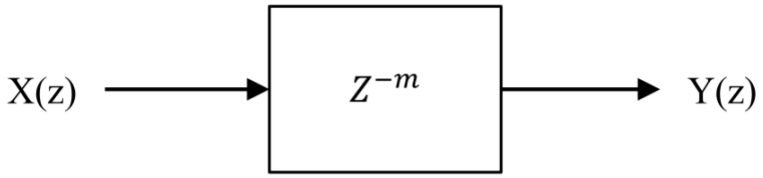


Figure 2.16: The structure of a delay filter.

- **Low-pass filter:** This is a type of filter that is used to allow low frequencies to pass through while blocking higher frequencies. It passes signals with frequencies lower than a cut-off frequency and attenuates signals with frequencies higher than the cut-off frequency [167]. The cutoff frequency is the frequency at which the filter begins to block signals. Low-pass filters are used to achieve a variety of effects in audio, such as removing high-frequency noise, creating a smoother sound, and creating a warmer sound. They can also be used to boost bass frequencies, creating a fuller sound. Low-pass filters are also used to reduce harmonic distortion in electrical signals. An example of a first-order low-pass filter structure is given in Figure 2.17, and

the transfer function of the low-pass filter is

$$H(z) = \frac{1-g}{1-gz^{-1}} \tag{2.21}$$

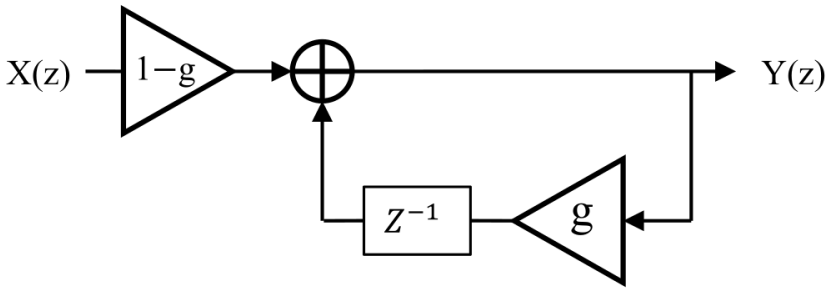


Figure 2.17: The structure of a low-pass filter.

The denominator polynomial of the first-order low-pass filter includes a term with z^{-1}, which corresponds to a single reactive component (a pole in the z-domain). The z-domain is a mathematical framework used in digital signal processing to analyse the behavior of discrete-time systems, such as digital filters [168]. It involves transforming signals from the time domain into the complex frequency domain using the z-transform. The transfer function of the first-order low-pass filter exhibits a single pole at $z = g$. For the filter to be stable, this pole must lie inside the unit circle in the z-plane, which corresponds to the condition $|\alpha| < 1$. Stability ensures that the filter's output remains bounded when a bounded input is applied, making this a fundamental design requirement. Since the filter has only one pole, it is also commonly referred to as a single-pole low-pass filter [169]. The simplicity of this structure makes it computationally efficient while still providing effective attenuation of higher frequencies, making it a widely used design in digital signal processing applications.

- **Comb filter:** A comb filter is a signal processing filter that adds a delayed version of a signal to itself, causing a series of constructive and destructive interferences [170]. The frequency response of a comb filter is characterised by a series of peaks and valleys, in the shape of a 'comb'. Its time domain impulse response is an exponential decaying pulse train [171]. An example of the impulse response in the time domain and the corresponding frequency response of a comb filter is shown in Figure 2.18. Comb filters are generally used to reduce harmonic distortion in audio signals. The peaks and valleys of the frequency response of a comb filter create an effect similar to that of a phaser or chorus effect, but without the need for additional processing. In equalization or signal processing applications, a comb filter can help shape the tone or timbre of an audio signal. By adjusting the spacing and depth of the filter's peaks and troughs, specific elements of the sound can be highlight or suppressed, such as making a voice clearer or a musical instrument more resonant. Comb filters can also be used to reduce feedback in sound systems. Feedback occurs when sounds from speakers are picked up by microphones and re-amplified, creating a loop that can cause high-pitched squealing noises. The comb filter can attenuate specific frequencies that cause feedback to break the feedback loop and stabilise the audio output. The structure of the comb filter is shown in Figure 2.19, and the transfer function of the comb filter is expressed as the Equation 2.22:

$$H(z) = \frac{z^{-m}}{1 - gz^{-m}} \tag{2.22}$$

- **All-pass filter:** An all-pass filter is a type of filter which passes all frequencies with equal gain while changing the phase relationship between the

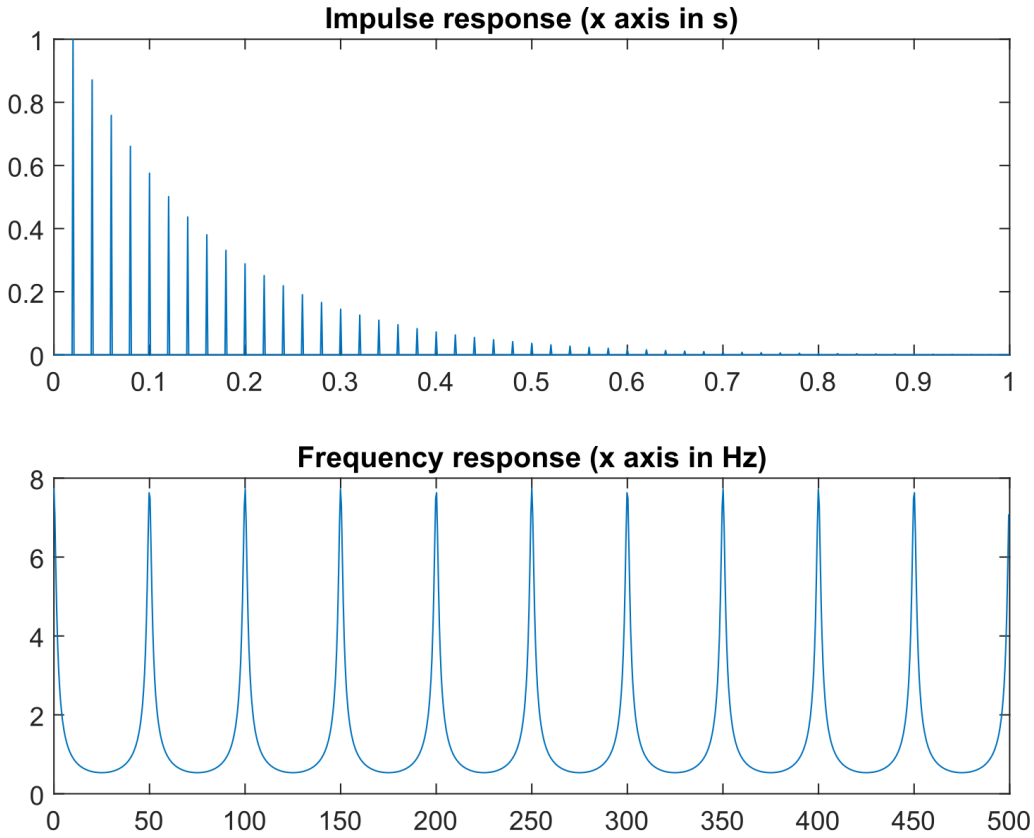


Figure 2.18: An example of the time domain impulse response and corresponding frequency response of a comb filter.

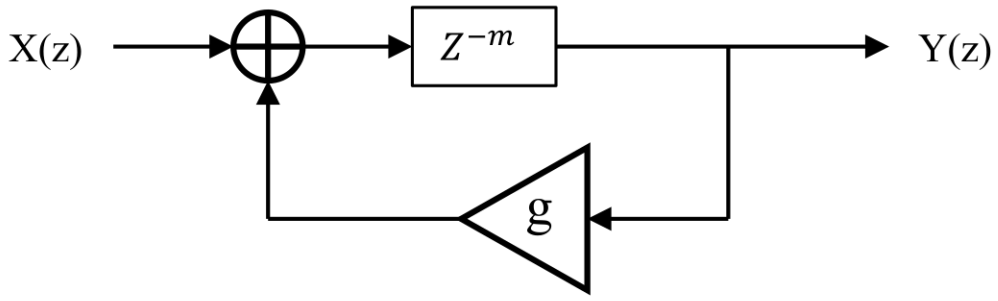


Figure 2.19: The structure of a comb filter.

various frequencies [172]. In other words, the amplitude frequency response of the all-pass filter is 1, while the phase frequency response can be arbitrary. This means that the phase response can be designed to vary with frequency according to specific needs. The amplitude frequency response of an all-pass filter being 1 means that this type of filter allows all frequencies of the input signal to pass through with the same gain, or amplitude level, effectively maintaining the original strength of the signal across its frequency spectrum. The phase shift introduced by an all-pass filter can vary across the frequency spectrum. This variation can be designed based on specific needs, such as compensating for unwanted phase shifts introduced by other parts of a signal processing chain, or creating specific phase effects required for applications like audio spatialisation or digital signal alignment. The all-pass filter is often used to achieve a constant group delay, which is a characteristic of a filter that is capable of keeping the time delay between input and output signals constant over a range of frequencies [173]. This is important for many audio applications, such as equalization and crossovers. Constant group delay ensures that all frequencies are delayed by the same amount of time as they pass through the filter. This uniform delay prevents phase issues from causing audible artifacts such as pre-echoes or smearing, which can degrade the clarity and quality of audio signals [174]. Crossovers are used in multispeaker audio systems to divide the audio signal into separate frequency bands that are then sent to different speakers optimized for those bands (e.g., woofers, mid-range speakers, and tweeters). Constant group delay in this context ensures that all divided signals are synchronized in time, maintaining coherent wavefronts and accurate soundstage reproduc-

tion. Without this, different frequency components might reach the listener's ears at different times, leading to a confused or blurred audio image [174]. All-pass filters can also be used to create notch filters, band-pass filters, and other types of filters. The structure of the all-pass filter is shown as Figure 2.20, and the transfer function of the all-pass filter is

$$H(z) = \frac{-g + z^{-m}}{1 - gz^{-m}} \tag{2.23}$$

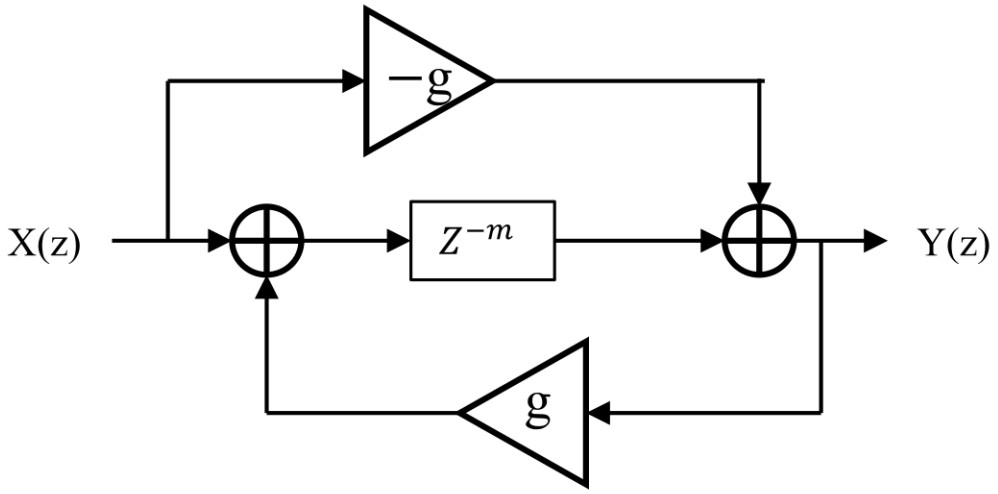


Figure 2.20: The structure of an all-pass filter.

The transfer functions of these basic filters can be derived from their structure diagrams. Their transfer functions provide the exact filter coefficients for their implementation in some signal processing softwares, such as MATLAB [141].

Some more complex common filters can be constructed from these basic filters, such as single nested all-pass filter, double nested all-pass filter, and low-pass-feedback comb filter. The nested all-pass filter can be used to construct Gardner

reverberator [171] and the low-pass-feedback comb filter can be used in Moorer reverberator [175]. The main advantage of these nested filter structures is that there is no repetition of the same echo pattern. Instead, the echo density increases with pattern size, just like in a real room [141]. This is because every signal impulse from the inner all-pass filter is recirculated to their inputs via the out feedback path [112]. This way allows each filter to impact the signal independently from a phase perspective. This independence helps in creating a variety of echo patterns rather than repeating the same pattern. Each filter can adjust the phase of different frequency components of the signal, resulting in diverse timing and characteristics of echoes. The low-pass-feedback comb filter simulates the absorption of sound by air [112] and it also simulates the exponential attenuation characteristics of impulse responses. The structures and transfer functions of these nested systems and the low-pass-feedback comb filter are described below.

- **Single nested all-pass filter:** This is implemented by surrounding one all-pass filter by another [176]. In the single nested structure, the inner structure is a delay followed by an all-pass filter. A single nested all-pass filter can create a more complex frequency response than a single all-pass filter. when two all-pass filters are nested, the phase shifts introduced by the inner filter are further modified by the outer one. This interaction can create a more varied phase response curve because the outer filter acts on the already phase-shifted signal from the inner filter. Besides, the complexity also depends on the specific coefficients and the design of the filters involved. Each all-pass filter in the nest can be designed with different parameters (such as delay coefficients and feedback coefficients), which further diversify the combined frequency response. This type of filter is often used in audio

processing to shape the more complex frequency response of a signal to create a more pleasing sound. The structure of the single nested all-pass filter is shown as Figure 2.21, and the transfer function of the single nested all-pass filter is

$$H(z) = \frac{Y(z)}{X(z)} = \frac{-g_1 + F(z)}{1 - g_1 F(z)} = \frac{-g_1 F_d(z) + F_n(z)}{F_d(z) - g_1 F_n(z)}, \tag{2.24}$$

where

$$F(z) = \frac{F_n(z)}{F_d(z)} \tag{2.25}$$

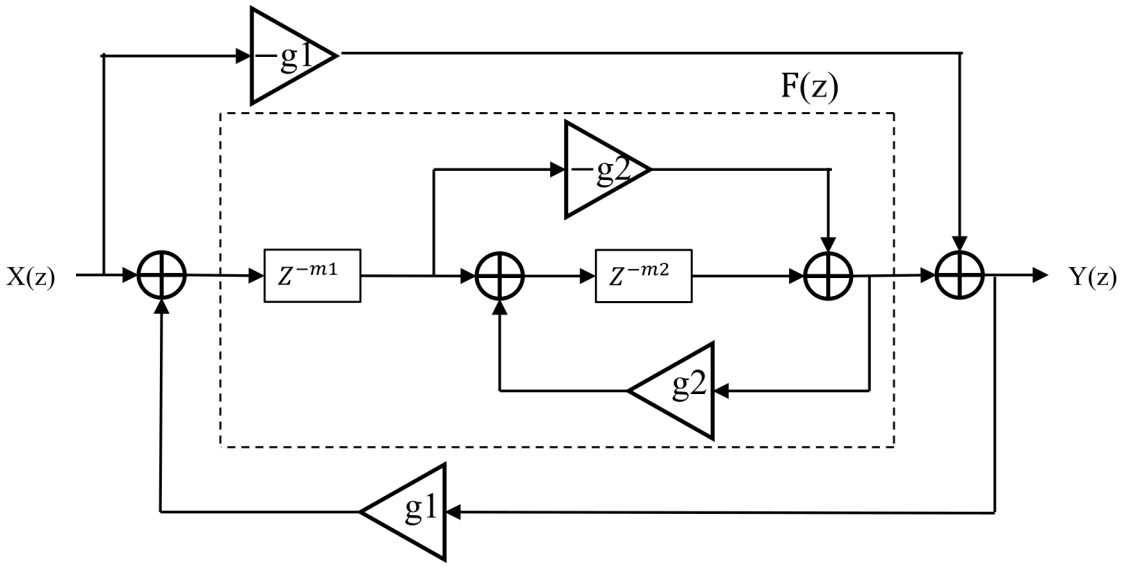


Figure 2.21: The structure of a single nested all-pass filter.

- **Double nested all-pass filter:** A Double nested all-pass filter is implemented by surrounding two all-pass filters in series by the third filter [176]. In a double nested structure, the inner structure is made up of a delay followed by two all-pass filters in series. This arrangement is used to create a very smooth reverberation effect with a long decay time. The structure of the double nested all-pass filter is shown as Figure 2.22, and the trans-

fer function of the double nested all-pass filter is shown in Equation 2.26. Its function structure is the same as that of the single nested all-pass filter, except the inner function is different.

$$H(z) = \frac{Y(z)}{X(z)} = \frac{-g_1 + F(z)}{1 - g_1 F(z)} = \frac{-g_1 F_d(z) + F_n(z)}{F_d(z) - g_1 F_n(z)}, \tag{2.26}$$

where

$$F(z) = \frac{F_n(z)}{F_d(z)} \tag{2.27}$$

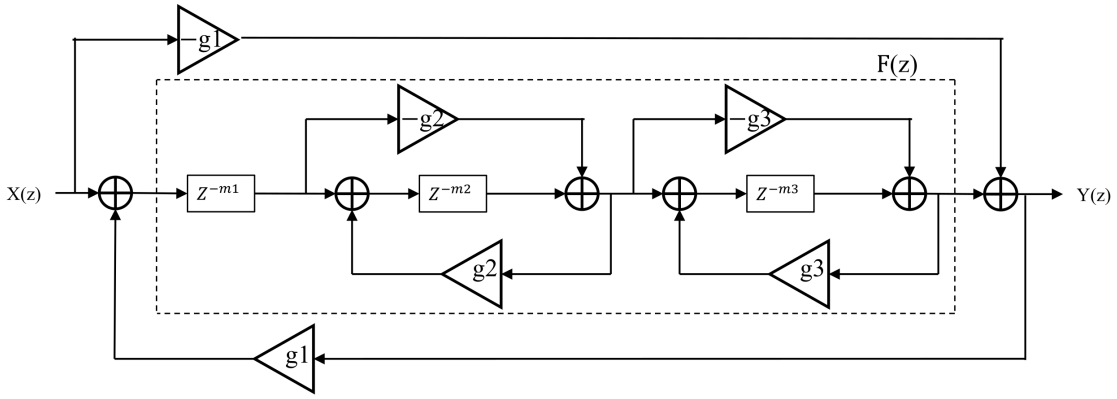


Figure 2.22: The structure of a double nested all-pass filter.

- **Low-pass-feedback comb filter:** This is implemented by inserting a one-pole low-pass filter into the feedback loop of a comb filter [112]. In the low-pass-feedback comb filter, the inner structure is a one-pole low-pass filter, designed to cut out high frequencies and allow only lower frequencies to pass through, while the feedback loop adds a decaying echo-like effect, enhancing the depth and spatial perception of the sound. This filter type is commonly used to create the illusion of a larger space by simulating natural reverberations that occur in such environments, where sounds reflect off multiple

surfaces at varying distances and angles. The periodic reinforcements and cancellations (echoes) introduced by the comb filtering effect, combined with the attenuation of high frequencies by the low-pass filter, mimic the natural absorption by air and surfaces. High frequencies, which are quickly absorbed when interacting with soft, porous materials like curtains and carpets, and scattered by minor obstructions and air currents due to their shorter wavelengths, decay more rapidly [177]. This rapid decay of high frequencies in larger rooms enhances the depth of the sound, as the longer-lasting lower frequencies dominate. The low-pass filtering of the feedback signal effectively mimics the natural absorption of high frequencies by air and surfaces in large rooms or other spaces. The combined effect of repetitive echoes (comb filtering) and the attenuation of high frequencies (low-pass filtering) not only adds depth but also gives the listener the impression that the sound is emanating from a larger space. This is because in larger spaces, sound reflections arrive at the listener at different times and frequencies, similar to the echo pattern created by the filter. The low-pass-feedback comb filter can also be used to create a chorus effect if the feedback loop is set to a very short delay time. The structure of the low-pass-feedback comb filter is indicated in Figure 2.23, and the transfer function of the low-pass-feedback comb filter is

$$H(z) = \frac{Y(z)}{X(z)} = \frac{z^{-m}}{1 - gF(z)z^{-m}} = \frac{z^{-m} - g_L z^{-(m+1)}}{1 - g_L z^{-1} - (g - gg_L)z^{-m}}, \qquad (2.28)$$

where

$$F(z) = \frac{1 - g_L}{1 - g_L z^{-1}} \qquad (2.29)$$

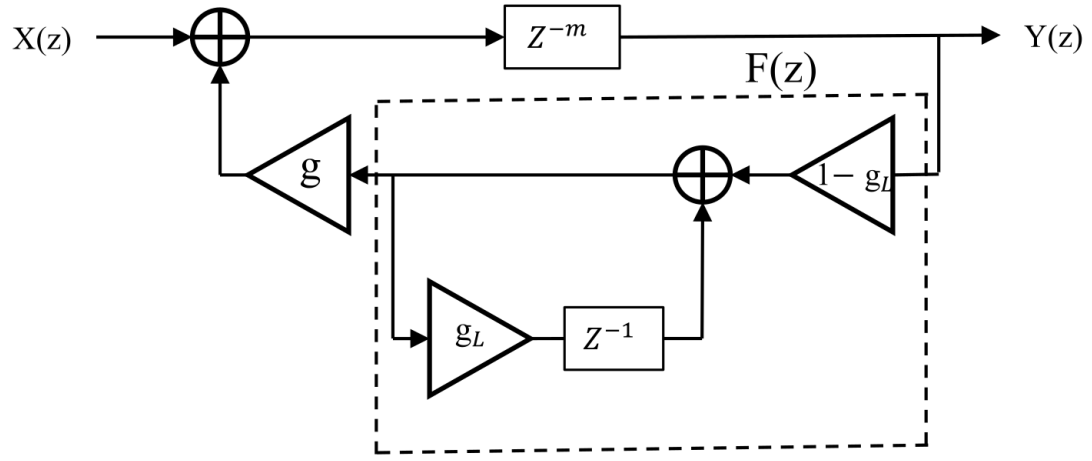


Figure 2.23: The structure of the low-pass-feedback comb filter.

When these filter structures are properly combined, the corresponding reverberation algorithms can be obtained. Since the 1960s, many artificial reverberation algorithms have been presented [15, 19, 171, 175, 178, 179]. Besides the earliest and most famous reverberation algorithm presented by Schroeder, five popular reverberation algorithms also have been presented by Moorer, Gardner, Jot, Dattorro and Alary.

The first digital reverberation algorithm, one of the earliest and most famous, was presented by Schroeder in 1961 [178]. The Schroeder reverberator is widely used in music production and sound design, as it can be used to create a wide range of acoustic effects without the need for expensive hardware [180]. This flexibility makes it an essential tool in studios looking to enhance spatial audio effects economically. It is also used in audio processing, such as voice recognition, speech synthesis, and noise reduction [181]. In addition, the algorithm is often used in video games and virtual reality applications to create a realistic 3D soundscape [182].

It has the simplest reverberation structure which consists of four parallel comb filters and two series all-pass filters as shown in Figure 2.24 [141]. The comb filter in the algorithm acts as an acoustic reflector, producing an echo-like effect, as if the sound is bouncing around the room. The all-pass filters are used to introduce a frequency-dependent delay, effectively simulating the natural reverberation of an acoustic space. The comb filters produce the long reverberant decay, and the all-pass filters multiply the number of echoes output [141]. Parallel comb filters can ensure their comb-like frequency responses to be evened out, and serial all-pass filters can ensure their phase characteristics cause minimum disturbance.

Natural reverberation consists of random closely-spaced impulse responses of exponentially decaying amplitude [183]. The 'random' aspect refers to the variability in the time and amplitude of the echoes. As sound bounces off irregular and varied surfaces in an environment, reflections occur at different times and with different intensities, depending on the distance and nature of the surfaces. 'Closely-spaced' indicates that these reflections often occur in rapid succession, especially in environments with many reflective surfaces close to each other. This rapid succession of echoes contributes to the characteristic texture of the reverberation. As sound waves reflect and scatter, they lose energy. Each subsequent echo is generally quieter than the last, leading to an exponential decay in amplitude. The first two conditions can be achieved by ensuring that the delay of each comb filter has no common divisor, and exponential decay can be achieved by making their reverberation times equal [15, 178, 184–187]. By choosing delay lengths that have no common divisor, each comb filter operates independently in terms of timing, with the echoes from different filters not aligning predictably. This randomness helps to distribute the echoes more evenly and more irregularly across the time

spectrum, much like the natural reverberations in different types of room. When the reverberation times of all comb filters are set to the same value, the decay rates of the echoes produced by each filter are synchronized, contributing to a coherent and uniform exponential decay of the overall sound. This uniformity is important because it prevents the sound from having irregular or disjointed fading, which can sound unnatural.

Taking Figure 2.24 as a reference, according to Schroeder's recommendation [141, 178], the delay values (*delay*1 to *delay*4) of the four comb filters should be distributed between approximately 30 to 45 ms, and the ratio of the maximum delay value (*delay*4) to the minimum delay value (*delay*1) should be approximately 1.5. The gains (*gain*1 to *gain*4) of the comb filters can be adjusted trough desired reverberation time (RT_{60}), as shown in Equation 2.30 [178].

$$g = 0.001^{(T_d/\mathrm{RT}_{60})}, \tag{2.30}$$

where g represents the gain of the comb filters, and T_d is the delay time of the comb filters. The delay (*delay*5 and *delay*6) of two all-pass filters is recommended to be set to 5 ms and 1.7 ms, respectively, and their gains (*gain*5 and *gain*6) are both recommended to be set to 0.7 [141, 178].

Of course, as the earliest reverberation algorithm proposed, the Schroeder reverberation algorithm is not perfect, so it inevitably has some disadvantages. For example, the echo density is insufficient and does not increase with time, and for large reverberation time the sound is metallic [185]. It also does not simulate air absorption. However, these shortcomings do not prevent it from being a good reverberation algorithm, according to Gardner [80], Zölzer [180], Ballou [188] et al.

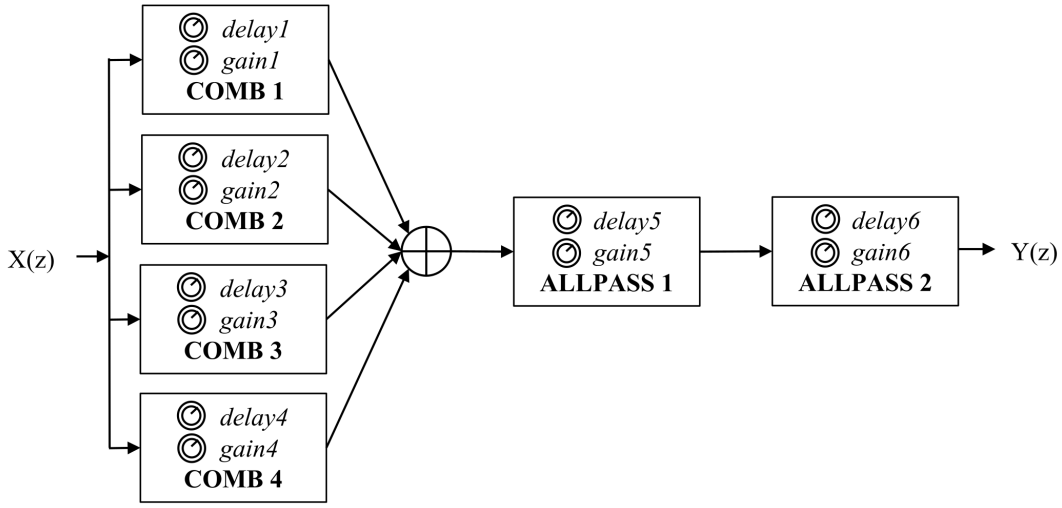


Figure 2.24: The structure of the Schroeder reverberation algorithm [141].

Moorer enhanced the Schroeder reverberation algorithm in 1979 [175] through four critical modifications, as shown in Figure 2.25 [189]. The first improvement was a FIR delay line inserted in order to simulate the early reflections of the RIR. The FIR delay line consists of a series of delay taps. Each tap in the delay line is associated with a coefficient that multiplies the input signal. The outputs of these multiplications are then summed together to produce the final output of the filter. Another modification was inserting a one-pole low-pass filter into the feedback loop of each comb filter to simulate the absorption of sound by air to decrease the reverberation time at high frequencies, making the sound appear more real. Here it is called the low-pass-feedback (LPF) comb filter. The third improvement was to increase the number of comb filters from four to six to get higher echo and modal density, especially for longer reverberation times. Finally, Moorer replaced the all-pass filter near the output with a delay filter to ensure that the late reflections arrive at the output a little later than the early reflections.

The part A in Figure 2.25 is the FIR delay line for simulating early reflections, and the part B is composed of six parallel low-pass-feedback comb filters, an all-pass filter and a delay filter. Parallel low-pass-feedback comb filters are used to simulate a smooth decay with high frequency roll off as time progresses, and the all-pass filter is to increase echo density without adding colouring to the magnitude frequency response [189]. The replaced delay filter ensures that late reflections arrive at the output just a little later than early reflections [80]. The overall improvements of Moorer reverberator are that it provides a frequency-dependent reverberation time and a higher echo density [112].

Moorer proposed parameters for all filters in his study [175]. As in Schroeder's reverberation algorithm, the random and closely-spaced characteristics of impulses can be realised by ensuring that the delay of each comb filter has no common divisor, while the exponential attenuation can be realised by making its reverberation time (RT_{60}) equal.

Gardner proposed a set of reverberation algorithms in 1992 aimed at simulating the reverberation effects of different sized rooms. These reverberation algorithms share three similar structures for different sized rooms (small, medium, and large). These algorithms were designed to accommodate the reverberation characteristics of small, medium, and large rooms within specific reverberation time ranges, and their corresponding ranges are shown in Table 2.1 [171]. These categorizations are based on the acoustic concept, where small rooms have quicker sound decay, medium rooms have moderate decay, and large rooms have prolonged sound decay.

These three reverberator structures are shown in Figure 2.26, Figure 2.27 and Figure 2.28 respectively [141] [171]. The structures of these three algorithms mainly include all-pass filters (simple and nested ones), delay filters and first-order

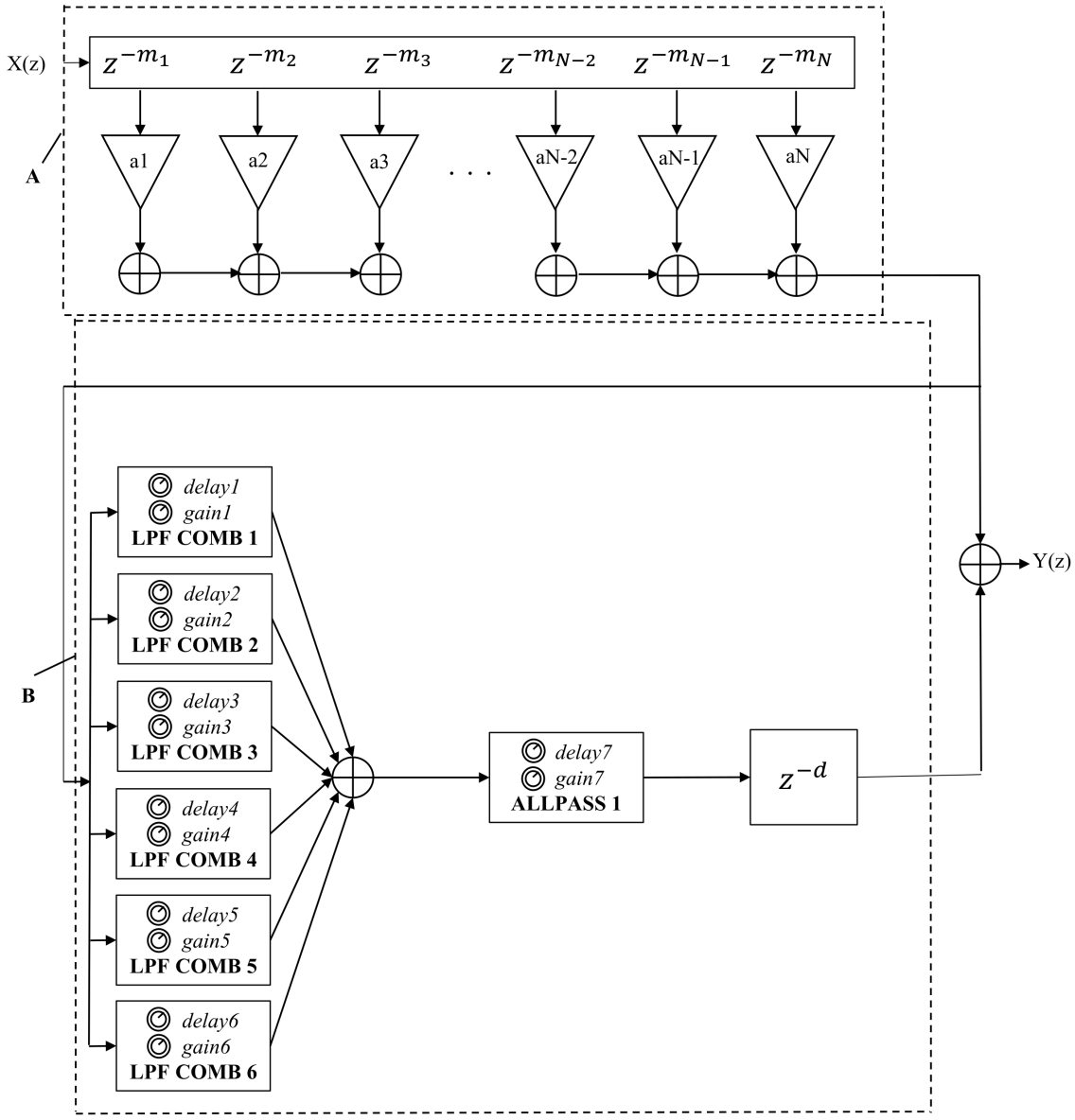


Figure 2.25: The structure of the Moorer reverberation algorithm [189], where part A is the FIR delay line for simulating early reflections and part B that is composed of six parallel low-pass-feedback comb filters, an all-pass filter and a delay filter is to simulate late reverberation.

Table 2.1: The corresponding reverberation time ranges for each Gardner reverberator [171].

Gardner reverberator structures	Reverberation time ranges (s)
Small-size-room reverberator structure	$0.38 \leq RT_{60} \leq 0.57$
Medium-size-room reverberator structure	$0.58 \leq RT_{60} \leq 1.29$
Large-size-room reverberator structure	$1.30 \leq RT_{60} < \infty$

low-pass filters. In these reverberators, input signals pass through these cascaded all-pass filtering structures, and are recirculated through a low-pass filter and a weighted gain which can affect the reverberation time. The advantages of these algorithms are that they are adjustable in reverberation time and high frequency damping [112] to get high quality results according to Gardner's study [190]. He thought the independence of reverb time can help realise a wide range of reverb effects. The damping frequency parameter controls the absorption of high frequencies in the late reverberation: a room with a high damping frequency sounds bright, a room with a low damping frequency sounds warm [190].

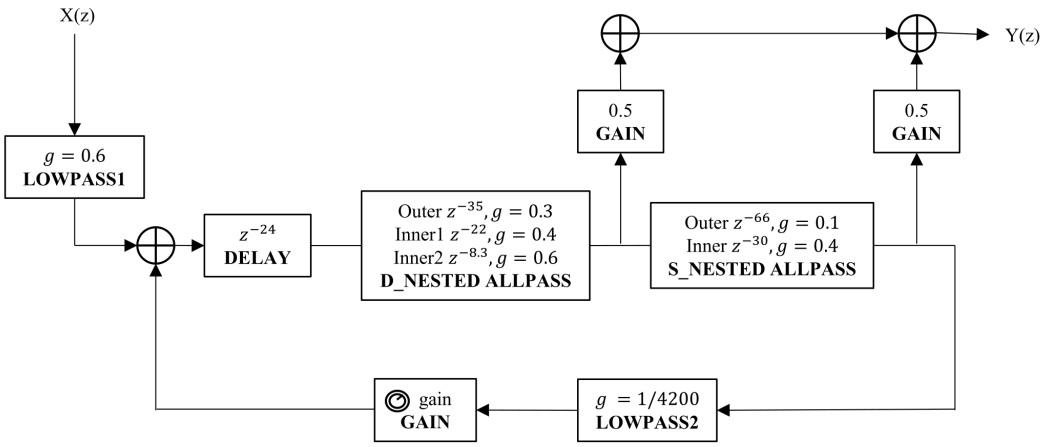


Figure 2.26: The structure of the Gardner reverberation algorithm for small size room [141, 171].

The transfer functions of three Gardner reverberator structures are easily cal-

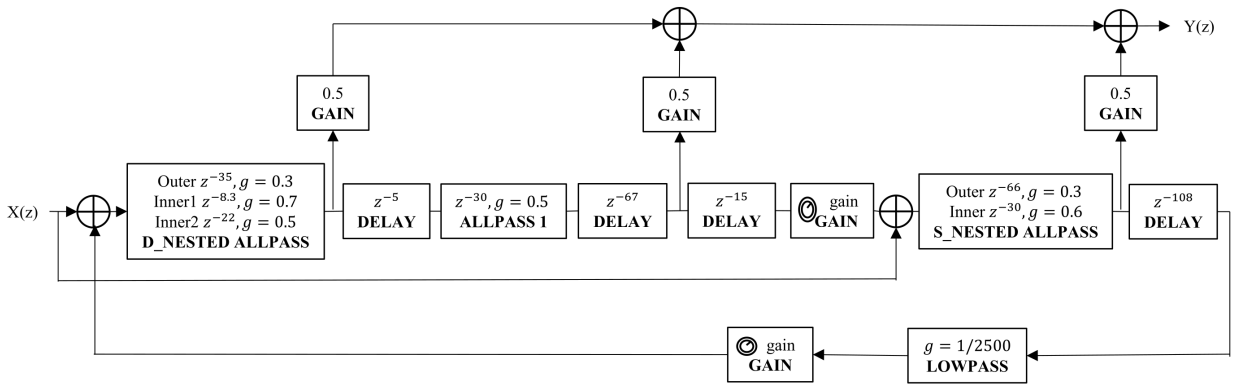


Figure 2.27: The structure of the Gardner reverberation algorithm for medium size room [141, 171].

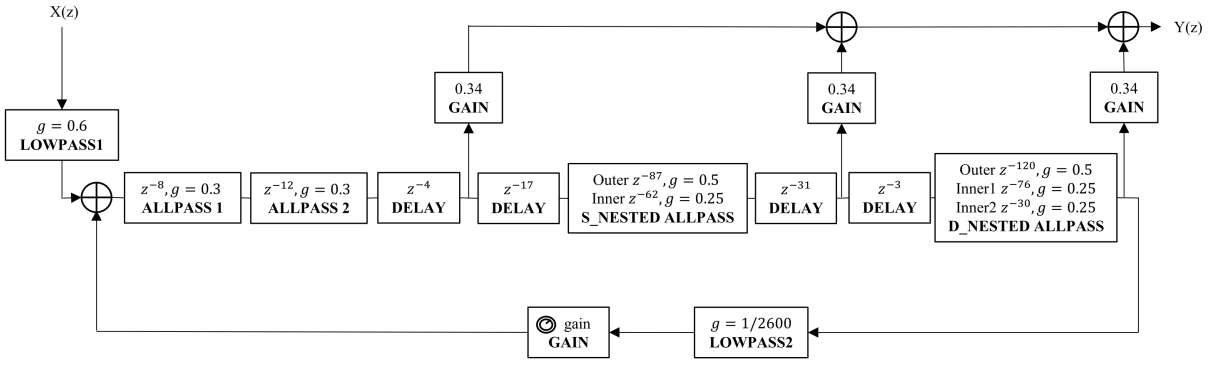


Figure 2.28: The structure of the Gardner reverberation algorithm for large size room [141, 171].

culated by simplifying their structures. Their simplified structures are shown in Figure 2.29 and Figure 2.30, where the small room and large room share the same simplified structure (Figure 2.29), and the medium room structure is simplified as Figure 2.30.

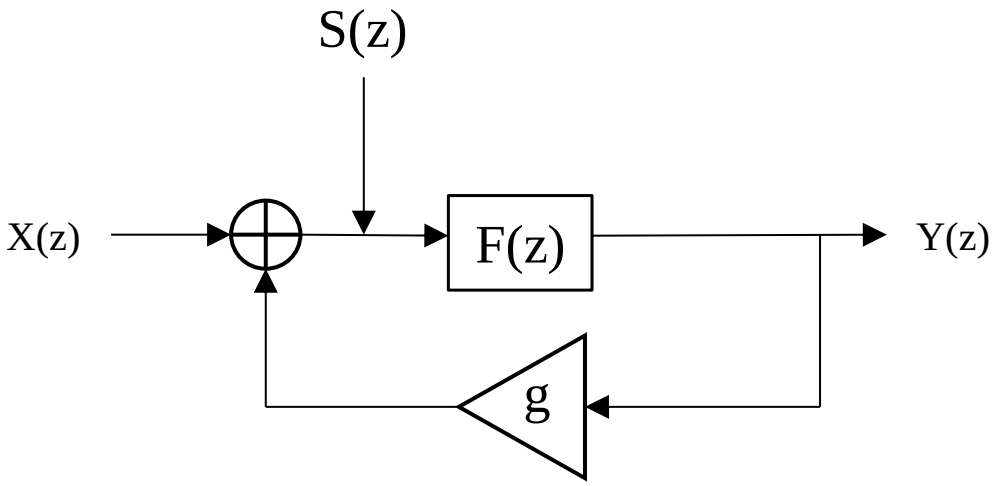


Figure 2.29: The simplified structure of the Gardner reverberator for small and large size rooms [141].

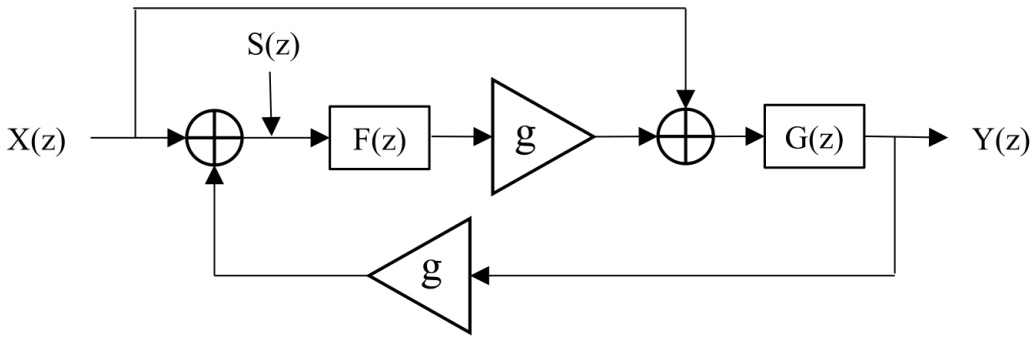


Figure 2.30: The simplified structure of the Gardner reverberator for medium size room [171].

When Figure 2.29 is used as a simplified structure of the small room, $F(z)$ is the transfer function of the filter structure composed of a delay, a double nested all-

pass filter, a single nested all-pass filter and a low-pass filter in series. g indicates an adjustable gain that can affect reverberation time. According to Figure 2.29, the transfer function $H(z)$ of the input signal $X(z)$ to the reference signal $S(z)$ is

$$H(z) = \frac{S(z)}{X(z)} = \frac{1}{1 - gF(z)} = \frac{F_d(z)}{F_d(z) - gF_n(z)}, \tag{2.31}$$

where

$$F(z) = \frac{F_n(z)}{F_d(z)}, \tag{2.32}$$

where $F_n(z)$ indicates the numerator polynomial of $F(z)$, and $F_d(z)$ represents the denominator polynomial of $F(z)$.

For medium room structure (see Figure 2.27), there is a forward input feedback in its structure, so the inner transfer function is simplified to two parts, as shown in Figure 2.30. $F(z)$ is the transfer function of the filter structure composed of a double nested all-pass filter, three delays and an all-pass filter in series. $G(z)$ is the transfer function of the filter structure consisting of a single nested all-pass filter, a delay and a low-pass filter in series. g indicates an adjustable gain that can affect reverberation time. According to Figure 2.30, the transfer function $H(z)$ of the input signal $X(z)$ to the reference signal $S(z)$ is

$$H(z) = \frac{S(z)}{X(z)} = \frac{1 + gG(z)}{1 - g^2G(z)F(z)} = \frac{G_d(z)G_d(z)F_d(z) + gG_n(z)G_d(z)F_d(z)}{G_d(z)G_d(z)F_d(z) - g^2G_d(z)G_n(z)F_n(z)}, \tag{2.33}$$

where

$$F(z) = \frac{F_n(z)}{F_d(z)}, G(z) = \frac{G_n(z)}{G_d(z)}, \tag{2.34}$$

where $F_n(z)$ and $G_n(z)$ indicate the numerator polynomial of $F(z)$ and $G(z)$, and $F_d(z)$ and $G_d(z)$ represent the denominator polynomial of $F(z)$ and $G(z)$,

respectively.

The simplified structure of the large size room reverberator is the same as that of the small size room (shown in Figure 2.29), so the transfer function $F(z)$ of the input signal $X(z)$ to the reference signal $S(z)$ is the same. However, the transfer function $F(z)$ of the filter structure is composed of two all-pass filters, four delays, a single nested all-pass filter, a double nested all-pass filter and a low-pass filter in series, which is different from the small size room reverberator. g indicates an adjustable gain that can affect reverberation time.

Once the reference signal $S(z)$ is obtained, the output signal of each individual filter can be obtained successively so that the output signals can be built up by adding up the required weighted output taps.

Jot proposed a multiple feedback system reverberation algorithm that controls decay characteristics and frequency response in 1991 [15]. It is named Feedback Delay Network (FDN). This reverberation algorithm provides frequency-dependent attenuation for each delay to simulate absorption in the air, and increases the echo density in the response of the reverberator through the interaction of multiple delay lines interconnected by a feedback matrix [15]. As a novel and versatile reverberator design structure, it provides separate and independent control of the energy storage, damping, and diffusion components of the reverberator [14].

It consists of N-dimensional delay lines and a feedback matrix as shown in Figure 2.31 [19, 141, 191]. The feedback matrix can recirculate the output of each delay line to each corresponding input. The coefficient d indicates the ratio between the direct signal and reverberation sound. The feedback matrix A contains all the feedback coefficients a_{ij} which control the feedback level for the jth delay output to the ith delay input. Compared with parallel comb filters, this feedback structure

can generate much higher echo density when given a sufficient number of non-zero feedback coefficients and delay lengths [191].

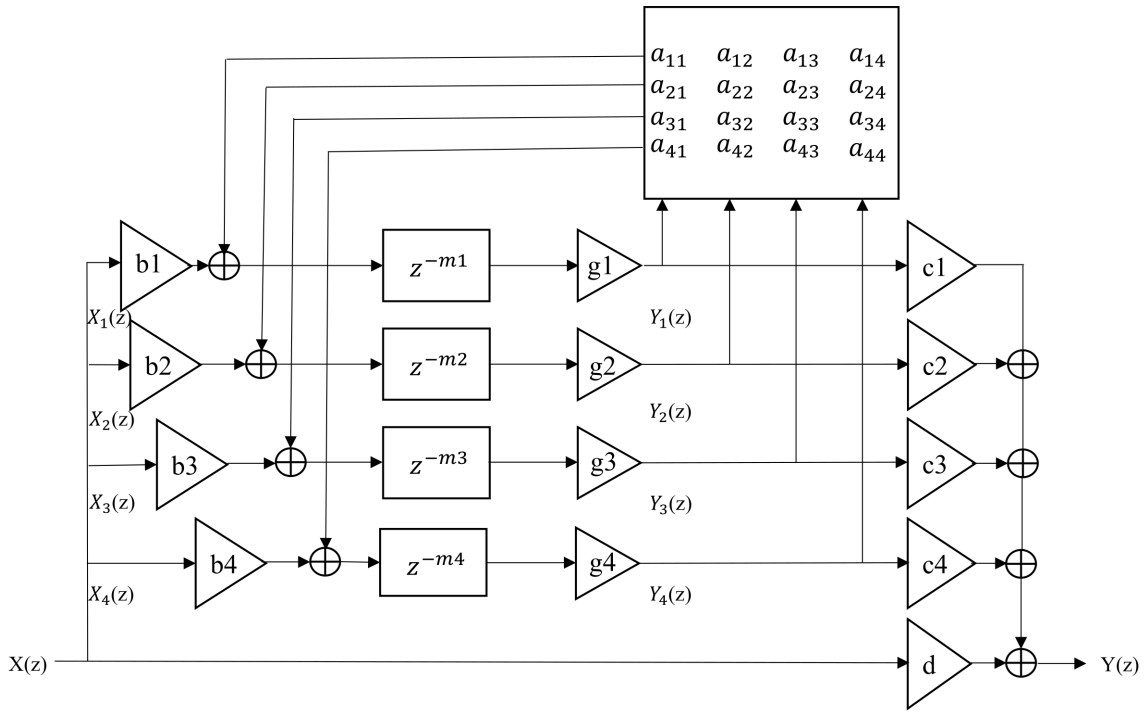


Figure 2.31: The structure of the FDN, where the part A represents the feedback coefficient matrix [19, 141, 191].

Developing this algorithm using filter functions in MATLAB programming presents more challenges compared to other reverberation algorithms. The core component of the FDN is its feedback matrix, which is crucial for determining the manner in which delayed signals are mixed and fed back into the system. Unlike typical MATLAB filter functions, which do not support the use of matrices as the numerator or denominator in the transfer function, the FDN requires such a capability for its implementation. This necessitates alternative approaches or custom functions to handle matrix-based operations in the feedback network of the reverberation algorithm. It is a possible way to implement the FDN algorithm using

a for loop, in which the output is computed sample by sample. However, because of the constant feedback processing, it requires a substantial amount of calculations. This can be computationally intensive, particularly if the algorithm is not optimized for performance. However, this reverberation algorithm can have any reverberation time and frequency density without undesirable tonal coloration [185]. Furthermore, it can reach the highest echo density [112].

In an FDN reverberator, the output signal is the sum of the direct signal $X(z)$ and reverberation signals. According to Figure 2.31 [19, 141, 191], the transfer function is

$$Y(z) = \sum_{i=1}^{N} c_i Y_i(z) + dX(z), \tag{2.35}$$

where the output signals $Y_i(z)$ of each delay line are given by

$$Y_i(z) = [a_{i,j} Y_j(z) + b_i X(z)] \cdot g_i z^{-m_i} \tag{2.36}$$

Dattorro published a topology structure for a reverberation algorithm in 1997 [179]. The implementation of this reverberation structure is composed of a predelay followed by a low-pass filter, a decorrelation stage, and a 'Tank implementation' as shown in Figure 2.32 [141]. The decorrelation stage includes four cascaded all-pass filters, which can perform a rapid build-up of echo density. The decorrelation stage does not have a feedback loop, so it is straightforward, but the 'Tank' is more complex because of the recursive structure [141]. The 'Tank implementation' consists of two cross-coupled symmetrical lines which cause the infinite recirculation of the input signal. Each line contains two all-pass filters, two delays, a low-pass filter and two decay coefficients for attenuation. These decay coefficients also help control reverberation time. The input signal passes through these filtering structures

accordingly, and the network produces a final decorrelated all-wet stereo output by summing six output signals extracted from the 'tank implementation' [141].

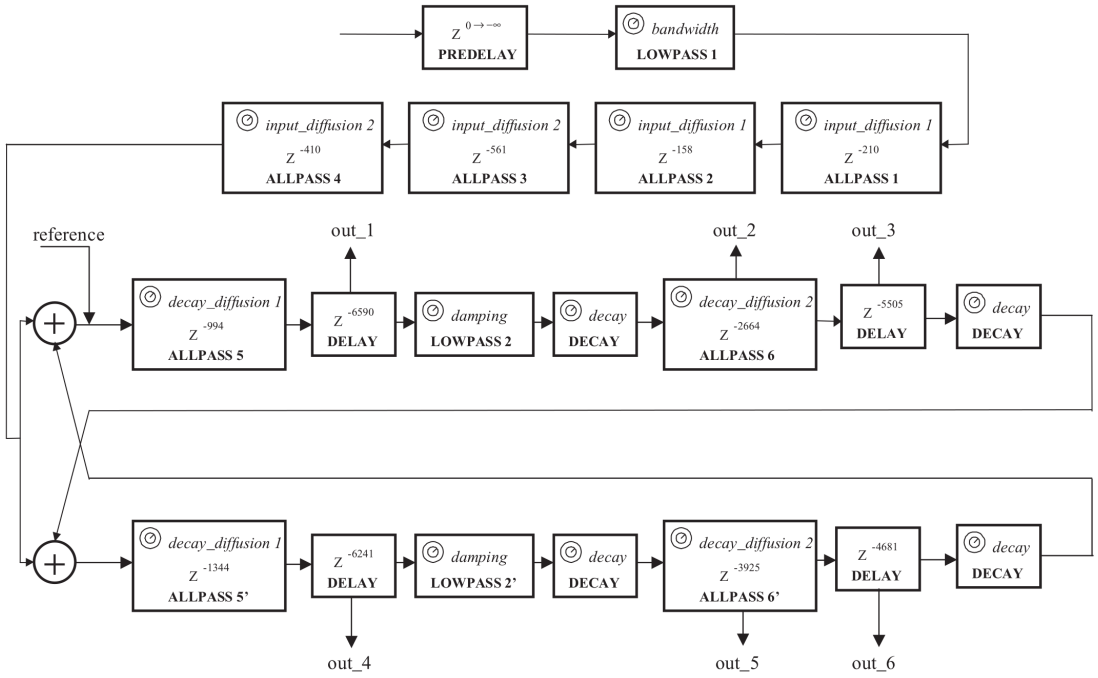


Figure 2.32: The structure of the Dattorro reverberation algorithm [141].

Similar to the Gardner reverberator structures, in order to calculate the transfer function of the 'Tank' structure, the 'Tank' structure needs to be simplified as shown in Figure 2.33.

$H_1(z)$ and $H_2(z)$ are both the transfer functions of a filtering structure composed of two all-pass filters, two delay filters and a low-pass filter in series. Besides, g_1 and g_2 both consist of two decays. According to Figure 2.33, the transfer function $H(z)$ of the input signal $X'(z)$ to the reference signal $S(z)$ is

$$H(z) = \frac{S(z)}{X'(z)} = \frac{1 + g_2 H_2(z)}{1 - g_1 g_2 H_1(z) H_2(z)} = \frac{H_{1,d}(z) H_{2,d}(z) + g_2 H_{1,d}(z) H_{2,n}(z)}{H_{1,d}(z) H_{2,d}(z) - g_1 g_2 H_{1,n}(z) H_{2,n}(z)}, \tag{2.37}$$

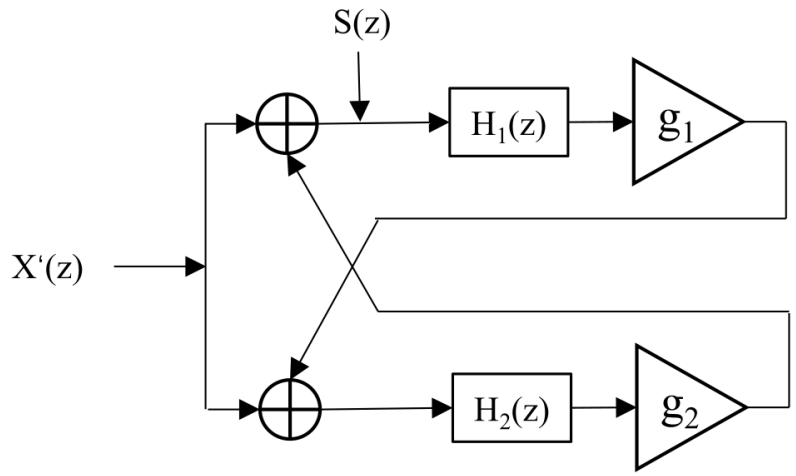


Figure 2.33: The simplified 'Tank' structure of the Dattorro reverberator [141]. Note that $g_1 = g_2 = \text{decay}^2$.

where

$$H_1(z) = \frac{H_{1,n}(z)}{H_{1,d}(z)}, H_2(z) = \frac{H_{2,n}(z)}{H_{2,d}(z)} \tag{2.38}$$

Once $S(z)$ is determined, the output signal of each individual filter can be derived successively, allowing for the extraction of the six output signals from the 'Tank'.

Alary designed a new reverberation algorithm named the Directional Feedback Delay Network (DFDN) in 2019, which is shown in Figure 2.34 [19]. The DFDN is an extension to a conventional FDN that can produce direction-dependent reverberation times and control the energy decay of a reverberant sound field, which is suitable for anisotropic decay reproduction on a loudspeaker array or in binaural playback through the use of Ambisonics [19]. The algorithm expands individual delay lines into a set of multichannel signals by encoding the mono input signal in the spherical harmonic (SH) domain, and modifies the direction-dependent energy decay by including a directional energy weighting function in the recirculating

path. In this algorithm, the discrete-time input signal $x(n)$ is encoded into an Ambisonic signal using the input gain vector $\boldsymbol{b}_i$, followed by a delay line for delay and attenuation. The delay lines within a delay line group have the same delay length m_i. Each delay line group is attenuated by a common gain factor g_i. Each matrix $\boldsymbol{T}_i$ is a directionally weighting matrix transform that takes a set of Ambisonic signals as the input and outputs the signals in the same format. $\boldsymbol{A}$ is an orthogonal recirculation matrix defining the recirculation gain between the delay lines. The vector $\boldsymbol{c}_i$ applies final attenuation to each delay line group, allowing direction-dependent gain control of the final output, and each delay line group is mixed with the direct sound to produce the final output $y(n)$. The output signal in the SHD can be decoded spatially to a loudspeaker array or binaurally to headphones using the appropriate decoding matrix [192].

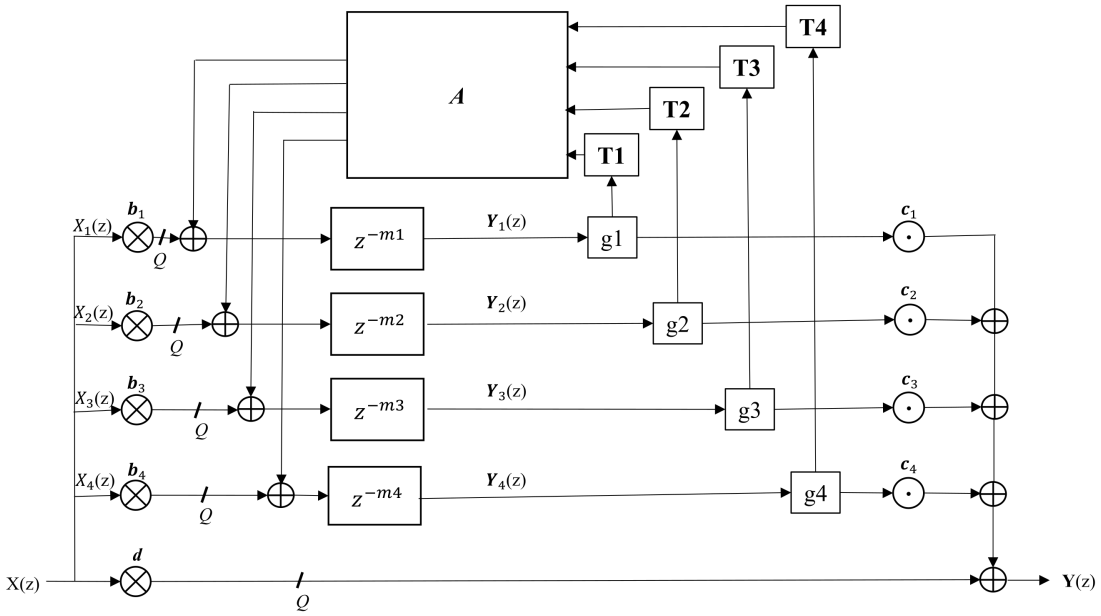


Figure 2.34: The structure of the Direction Feedback Delay Network (DFDN) [19].

Ambisonics is a full-sphere surround sound format, which encodes sound fields

or reproduces spatial sound by using a series of spatial basis functions known as spherical harmonics [19,193]. In scenarios where sound sources are in the far-field, the sound field can be modeled as a continuous distribution of plane waves with amplitudes $\boldsymbol{s}(n, \phi, \theta)$, each characterised by unique azimuth (ϕ) and elevation (θ) angles, and specific amplitudes. The core process in Ambisonics is the Spherical Harmonic Transform (SHT), which converts the sound field described in spatial terms into the spherical harmonic domain (SHD). The SHT is mathematically expressed as an integral over the sphere, involving the product of the amplitude density $\boldsymbol{s}(n, \phi, \theta)$ of the sound field and the spherical harmonic functions ($y(\theta, \phi)$), integrated over the appropriate angular domains:

$$\boldsymbol{s}(n) = \text{SHT}\left[(s(n, \phi, \theta)\right] = \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} s(n, \phi, \theta) y(\phi, \theta) cos(\theta) \mathrm{d}\theta \mathrm{d}\phi \tag{2.39}$$

This transform computes the Ambisonic signal vector $\boldsymbol{s}(n)$, up to the Ambisonic order L. $y(\theta, \phi)$ contains all spherical harmonics channels ($Q = (L+1)^2$). Practically, a discrete version of SHT is employed by using a set number of points that ensure accurate sampling without errors, appropriate for the transform order:

$$\boldsymbol{s}(n) = \frac{4\pi}{K} \boldsymbol{Y}_{\text{uni}}^T \boldsymbol{s}(n, \phi, \theta), \tag{2.40}$$

where $\boldsymbol{Y}_{\text{uni}}$ is defined as a matrix comprised of SHs, represented $[\boldsymbol{y}(\phi_1, \theta_1), \ldots, \boldsymbol{y}(\phi_K, \theta_K)]^T$. This matrix is constructed from SHs evaluated at $K \geq (L+1)^2$ uniformly distributed points on the sphere, which ensures the SHT can be performed accurately up to order L without sampling errors [194]. $\boldsymbol{s}(n, \phi, \theta)$ is represented as $[s(n, \phi_1, \theta_1), ; s(n, \phi_K, \theta_K)]^T$, indicating the sampled amplitude distribution at the same points.

B-format, an essential part of Ambisonics, is a method of representing a sound field used in first-order Ambisonics, which captures the spatial characteristics of sound within an environment [195]. Unlike traditional stereo or surround sound formats, B-format does not directly store the audio signals corresponding to individual speakers. Instead, it encodes the sound field in terms of its directional components, typically using four channels: W (omnidirectional), X (front-back), Y (left-right), and Z (up-down), as shown in Figure 2.35 [196]. This allows the sound field to be reproduced or manipulated with greater flexibility, enabling soundfield rotation, binaural decoding, and rendering to arbitrary speaker arrays, making it ideal for applications in immersive audio, where accurate spatial representation is crucial.

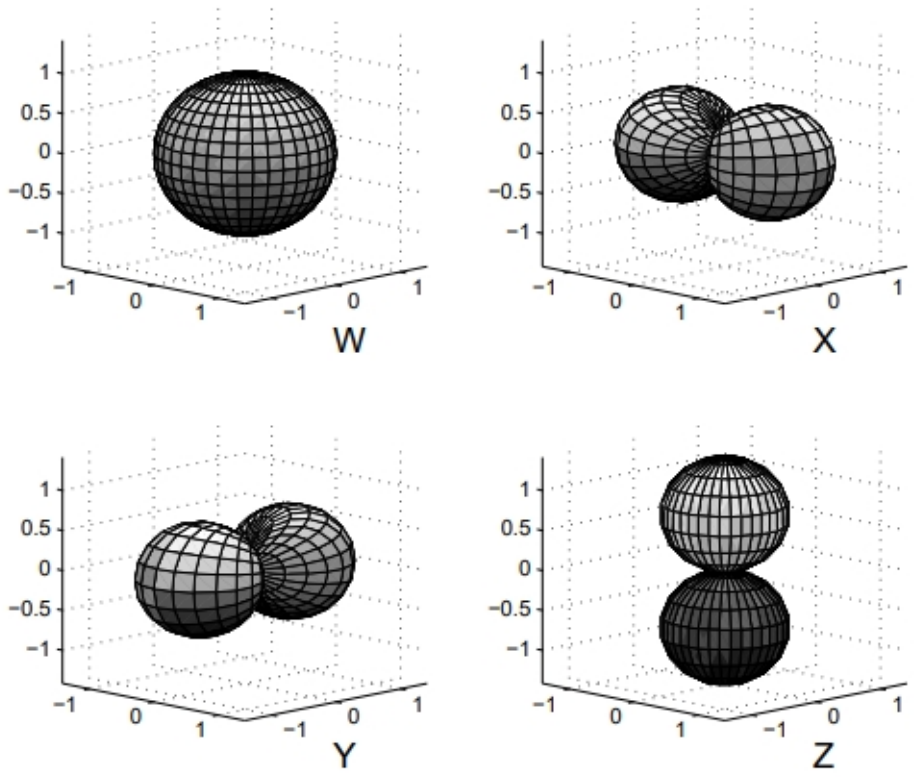


Figure 2.35: The directional patterns of the four channels in a B-format [196].

2.5 Application Scenarios of Virtual Reverberation

In virtual acoustic rendering, the accurate simulation of reverberation is essential to creating immersive and plausible audio experiences. Reverberation provides critical spatial and environmental cues that enhance the perception of depth, size, and texture of virtual spaces, making it a key component in various application scenarios, such as AR, VR, and music technology.

2.5.1 Reverberation in Application Scenarios of Virtual Acoustic Rendering

AR and VR are transformative technologies that have significantly impacted how we interact with digital content and the physical world. AR overlays digital information onto the real world, often through devices like smartphones, tablets, or AR glasses [197]. These digital overlays can include images, videos, sounds, or other sensory enhancements that enhance the user's perception and interaction with the environment. In contrast, VR immerses users in a completely virtual environment through VR headsets and other sensory equipment, disconnecting them from the real world and creating an entirely new sensory experience [198].

In AR scenarios, virtual reverberation plays a crucial role in blending digital sounds with the real-world environment. This blending is essential for maintaining the illusion that virtual objects are part of the physical space [199]. For instance, AR games can use virtual reverberation to enhance immersion by dynamically adjusting the acoustics based on the player's real-world surroundings. Sounds

from virtual objects can reflect off physical surfaces, creating a more engaging and believable gaming experience [200, 201].

In VR scenarios, creating a convincing and immersive audio environment is critical for user engagement and realism. Virtual reverberation enhances the realism of VR environments by simulating how sound interacts with the virtual space [3]. VR training programs for pilots, military personnel, or emergency responders benefit from realistic soundscapes that include appropriate reverberation [202]. Accurate audio cues help trainees develop better situational awareness and improve their responses in real-life scenarios [203].

Music technology encompasses a broad range of tools, techniques, and software used in the creation, recording, production, and performance of music [204]. This field includes everything from traditional musical instruments enhanced with electronic features to sophisticated Digital Audio Workstations (DAWs) and software plugins. One key component of music technology is the Virtual Studio Technology (VST) plugin. VST plugins are software modules that integrate with DAWs to enhance audio production capabilities. Developed by Steinberg in 1996, VST plugins can emulate various musical instruments or provide audio processing effects [205]. These plugins are highly versatile and widely used in modern music production for tasks such as synthesising sounds, adding effects like reverb and delay, and mastering tracks.

Virtual reverberation plays a crucial role in music technology, enhancing the spatial and acoustic qualities of recorded and synthesised sounds. VST plugins are integral components in modern music production, offering a wide range of audio effects and virtual instruments [205]. Virtual reverb plugins can replicate the natural reverberation of various physical spaces, such as concert halls, churches,

studios, and small rooms [206]. This ability allows producers to place audio tracks within a believable acoustic context, making them sound as if they were recorded in a particular environment. Users can also customise the parameters of the reverb to simulate specific environments, adjusting factors like room size, decay time, and diffusion to achieve the desired sound [207]. By adding reverb to different tracks in a mix, producers can create a sense of depth and space, making the music sound more three-dimensional. This technique is essential for creating a natural-sounding mix that is pleasing to the ear.

2.5.2 Multiple-degrees-of-freedom Reverberation Rendering

To achieve plausible reverberation simulations in virtual acoustic rendering, a very important technique is multiple-degree-of-freedom rendering, such as the typical three- and six-degree-of-freedom rendering. The six degrees of freedom (6DoF) in which an object can move through three dimensions of space are shown in Figure 2.36. In a three degrees of freedom (3DoF) virtual rendering system, the user's position is fixed, so they cannot walk freely around the environment, but can only perform head movements, i.e. yaw, pitch and roll [208]. In a 6DoF virtual rendering system, the user is able to move freely within the virtually recreated environment, thus experiencing all 6DoF, i.e. in addition to moving the head in 3DoF, there is also forward/backward, up/down, and left/right position movement [209]. This makes the virtual rendering of the reverberation closer to the perception of the real world.

Researchers are also working on the implementation of 6DoF virtual reverberation rendering, aiming to simulate plausible real-world reverberation. A compact

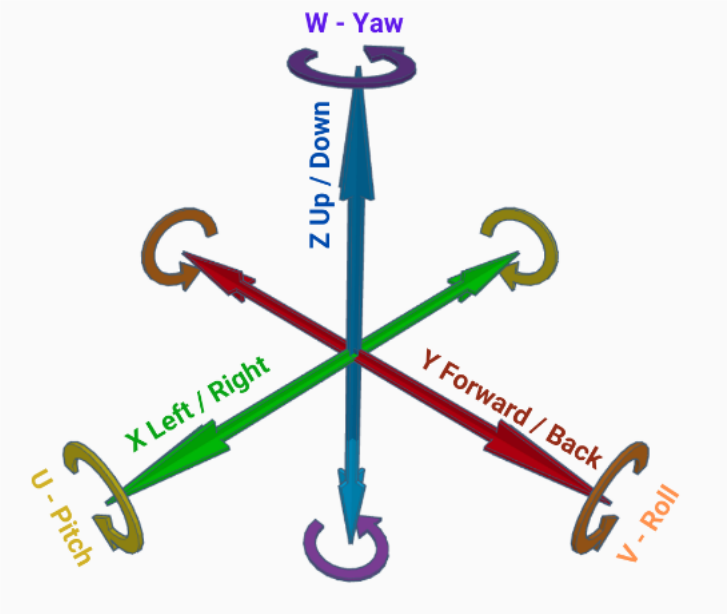


Figure 2.36: Example diagram of 6 degrees of freedom in three dimensions [210].

dataset of attributes that can specify parameters of a DFDN for auralisation was generated using a method of extracting key directional features from a set of captured spatial room impulse responses (SRIRs) by Araly et al. [211]. The DFDN reverberation algorithm provides 6DoF auralisation by interpolating between the values in the dataset to modulate the gain in the delay network. To achieve 6DoF auralization, the DFDN algorithm interpolates between recorded spatial room impulse responses (SRIRs). This interpolation allows the algorithm to simulate how sound would behave at any point within the space, not just where measurements were taken. The interpolation adjusts the gains in the delay network, which modifies the reverberation effect to match the expected real-life acoustic response based on the listener's position and orientation.

A spatial room impulse response dataset in a variable acoustic room measured by McKenzie et al. [212] includes 7 SRIRs recorded via a 6DoF recording application using a 19-capsule Zylia ZM-1 spherical microphone array model. This

dataset may be applied to analysis and rendering of 6DoF.

A 6DoF parametric spatial audio rendering method based on monophonic RIR has been proposed by Arend et al. [213]. Synthetic BRIRs are generated for any desired head orientation and position in a room by parametrically describing the spatial sound field through a scalable and perceptually motivated encoding of the listener's head orientation or spatial position. A 6DoF implementation framework is implemented to evaluate the synthesised BRIRs. The results show that the synthetic BRIRs can provide reasonable binaural reproduction for listeners moving around the room, but also reveal the technical challenges of generating artefacts with 6DoF systems.

2.6 Perceptual Evaluation of Audio

The study of binaural reverberation is aimed at producing plausible real-world reverberation effects in virtual acoustic rendering. In order to assess the acoustic quality and similarity of virtual binaural reverberation to real reverberation, it is often necessary to compare reverberation algorithms to reference acoustic measurements and to know how to measure the quality.

Research aimed at natural sounding audio is usually divided into two areas. One is the measurement of the physical characteristics of the audio signal, which provides an objective representation of the nature of the audio. The other is to assess the subjective perception of audio signals by humans. While some insight into the quality of a binaural audio system can be gained through numerical analysis methods by comparing measurements from two or more systems, the results are ultimately an estimate of human perception and a rigorous audio assessment

should always include some measurement of human perception [60]. Formal listening tests are the most common form of perceptual assessment [214]. Listening tests are experimental procedures used to assess the subjective perception of sound, i.e. to evaluate how the human auditory system interprets and quantifies stimuli [214,215]. They are designed to measure the subjective perception of a listener, often using a rating scale. The tests involve playing back a sound to a listener, who is asked to rate the sound's quality. The results of a listening test can then be used to compare and evaluate different sound systems or audio devices. Listening tests are widely used in acoustic research fields such as sound insulation [216–218], soundscapes [219–222], sound quality [223,224], and room acoustics [215,225–228].

In acoustical research, listening tests are used for three main purposes. Firstly, to describe just noticeable differences (JNDs) in room acoustic parameters [225, 229–236], i.e. to determine the smallest change in a parameter that must occur in order for that change to become apparent [215], also called threshold estimation.

Secondly, to assess whether differences between two or more stimuli are perceptible [94,227,228,237–241] (e.g., differences between real recordings and auralisations, differences between auralisations with different patterns of sound source directionality, etc.) [215]. From this, it can be evaluated whether auralisation can well simulate or even replace real recordings.

Finally, to address listeners' preferences for room acoustics [226,230,242–245] (e.g., preferred acoustic characteristics of venues for specific music genres or speech, etc.) [215]. This facilitates the simulation of more appropriate sound effects for the listener in a specific acoustic room.

A listening test should be designed very carefully to minimise possible effects that could affect the results. All aspects should be considered, from the test

methodology and choice of stimuli to the statistical analysis of the results. This is due to the complex procedures and demanding performance involved, which involve the interaction between different entities (e.g., experimenter, participant and test interface) and factors (e.g., experimental conditions).

The next section presents some commonly used methods of subjective perception evaluation as a basis for the methods chosen in the experiments covered in Chapter 3 and Chapter 4. Subsequent section describes methods of statistical analysis of the test results to analyse whether the listening test results are statistically significant.

2.6.1 Methodologies of Listening Test

In the field of indoor acoustics, the most commonly used hearing tests can be divided into:

- Integrative scaling tests [246], involve participants providing scaled responses to questions regarding their perception of the stimuli presented. These tests are primarily utilised to address issues related to preference and quality [215]. The primary goal of integrative scaling tests is to identify the optimal choice, meaning the most preferred or highest quality option as perceived by the participants based on the specific attributes being evaluated. This helps researchers determine which stimulus or product version best aligns with user preferences or delivers the desired sensory experience.

- Discrimination tests [246] require participants to determine if there are perceptible differences among various stimuli, or to discern which stimuli exhibit specific attributes. These tests are mainly used to assess the detectability of subtle differences [215]. Discriminatory tests can be subdivided into two cat-

egories: attribute-related tests and overall difference tests [247]. In attribute-related tests, participants are asked to identify stimuli based on specific characteristics. In overall difference tests, participants access whether any perceptible differences exist amiong the stimuli overall [215].

Depending on the requirements and objectives, the listening test can be carried out by choosing the most appropriate one from a variety of methods. Some of the most common test methods used in listening tests are described in the following content.

Paired Comparison: A paired comparison test is a psychophysical method used to determine the preference or difference between two sounds. In a paired comparison test, the subject is presented with two sounds and asked to choose their preferred sound or to judge whether difference exist between the two sounds based on some certain criteria [248]. Paired comparison testing can be used to compare two or more sound samples with different characteristics, such as pitch, loudness, and timbre. Paired comparison tests are suitable for detecting differences in very similar sounds. Before conducting a listening test, therefore, it is important to clearly define the objective. This involves deciding whether the test is intended to look for subtle differences between sounds or to resemble a real-life auditory experience. However, paired comparison test can be affected by the order in which the sound samples are presented. The preference or difference between two sounds may be influenced by the sounds that were presented immediately before. An AB blind test is a method that can compare two sensory stimuli to judge detectable differences between them [249]. The perception test performed in Chapter 3 of this thesis is a typical example of a paired comparison test.

Staircase Method: The staircase method is a psychophysical technique used

to measure perceptual thresholds, usually in the form of a listening test [250]. It is a widely used technique in the field of auditory research due to its ability to accurately measure the threshold of influence of a parameter on perceived hearing. This technique is suitable when the researcher wants to quantify a participant's ability to distinguish different sound levels. The staircase method is also used to measure the ability of a participant to detect small changes in auditory stimuli. The staircase method is advantageous in measuring thresholds of auditory perception accurately through a small number of trials. Staircases usually begin with a detectable difference between a reference and test stimulus. This difference is then reduced with predetermined repeat intervals until the participant provides a negative response. At this point, the staircases reverse and the difference increases with predetermined repeat intervals until the participant makes a different response again, triggering another reversal. Predetermined repeat intervals can be set to be the same or different, and this process can be repeated as needed, until the stimuli reach an asymptotic level. The asymptotic level is called a plateau and refers to the phase where changes in the stimulus no longer cause changes in the participant's responses, indicating that perception has reached a stable level. Participants then hover around the plateau as long as the conditions remain unchanged [251]. There are a number of ways to determine the value that represents the threshold. The simplest method is to calculate the average of a certain number of stimuli after the series of trials has reached its final level. This requires an arbitrary decision as to when the final level is reached. Since the staircase method always reaches the asymptotic level in the last few trials, it is common to discard the data from the first few trials and use the remaining data to estimate the threshold [251, 252]. Experiments in Chapter 3 on the effect of variations in BRIR

parameters on perceptual reverberation give a detailed example of the staircase method.

The staircase method has three predetermined conditions, the start point, the step sizes and the stop point.

- **Start point [251]:** The start point refers to the initial level of stimulus intensity presented to the subject. It's crucial to set this level close to the anticipated threshold to ensure efficiency and reduce the number of trials needed. The aim is to begin the test with a stimulus intensity that is detectable but not too far from the threshold, making it easier to pinpoint the exact point of sensory change as the test progresses.

- **Step size [251]:** Step sizes determine the increment or decrement in stimulus intensity after each trial based on the subject's response. These steps should be sized to create approximately equal perceptual intervals, allowing the subject to discern changes without making the shifts too abrupt or too subtle. The ideal step size prevents long sequences of consistent responses ('yes' or 'no'), which can occur if the step size is too large (leading to rapid oscillations across the threshold) or too small (resulting in negligible perceptual differences). A common strategy is to employ a '1-up-N-down' rule where the intensity is decreased after several consecutive correct responses and increased after an incorrect one.

- **Stop point [251]:** The stop point is the phase of the test where changes in stimulus levels tend towards an asymptotic level, and it typically occurs when the participant's responses have stabilised around a specific level, indicating a plateau. This stop criterion is essential for concluding the test without extending it unnecessarily, ensuring efficient use of time and resources while

still achieving reliable results. The stop point can be predefined either as a fixed number of trials, a fixed number of reversals (changes in the direction of stimulus intensity change), or after a certain number of responses is achieved following the plateau.

The staircase method allows for a more accurate measurement of the difference threshold between two sounds than conventional listening tests and it allows the difference threshold to be determined with fewer trials, due to the fact that the difficulty of the task is adjusted according to the listener's response [253]. In addition, it can be used to compare more than two sounds or sounds with different characteristics, such as pitch, loudness and timbre. However, the staircase method relies on the listener's subjective response. This can lead to individual differences in the results, and the staircase method has a limited range of applications. It is only used to determine the threshold of difference between two sounds. It cannot be used to assess the overall quality of a recording, music or speech. Despite its limitations, the staircase method has been widely used in various fields, including speech perception, and music perception.

Category Judgement: A category judgment test is a type of listening test that is used to evaluate the quality of sound recordings, music, and speech, with applications in fields such as audio engineering, speech perception, and music psychology. Category judgment tasks are often used to assess the perceived quality of audio signals, such as speech or music, and to compare different audio samples based on their acoustic properties. The Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) paradigm [254] is a typical category judgment test and is used by international standards organizations such as the International Telecommunication Union (ITU) [255] to evaluate the quality of voice and audio codecs.

It is recommended when there are medium to large differences between test conditions. Medium differences are noticeable but not obvious differences that can be detected without significant difficulty but are not immediately apparent in a casual listening scenario. Listeners may need to pay closer attention to notice these differences, especially in environments with minimal distractions. Large differences are easily perceptible differences that are readily noticeable and can be detected even by untrained listeners or in less controlled listening environments.

In MUSHRA, a reference signal is presented along with several other signals that are being evaluated. One of the evaluated signals is designated as an 'anchor' signal, which is used to establish a common reference point for all the other signals. The reference signal is not directly presented to the listener, but instead, it is used as a hidden reference for comparison with the other signals. The test signal usually includes one or two 'anchor' signals used to eliminate abnormal results. A low-pass filtered reference signal with cut-off frequency of 3.5 kHz is usually used for low 'anchor' signal, while low-pass filtered reference signal with cut-off frequency of 7 kHz is usually used for mid 'anchor' signal [255]. In a category judgement test [248], audio samples are presented to subjects one by one and subjects rate the respective sounds on several category scales according to certain criteria. For example, in an audio perception study, like the Basic Audio Quality (BAQ) of the audio sample as shown in Figure 2.37 [256], the listener might be presented with several audio samples that are classified into different categories. The listener would then rate the audio samples within each category on a scale of 0 to 100, with 0-20 indicating bad quality and 80-100 indicating excellent quality. In the judgment based on category scale, various bias effects, also called context effect, may occur [248]. That is, the bias or influence that previous sounds have on the

perception of the sound currently being judged [257]. For example, a relatively poor quality sound may be rated more favorably if it immediately follows a much poorer quality sound, or conversely, a good quality sound may seem less impressive following a superior quality sample. This effect can be improved by presenting each sound multiple times in a randomised order. An example of using a category judgment test are detailed in Chapter 4.

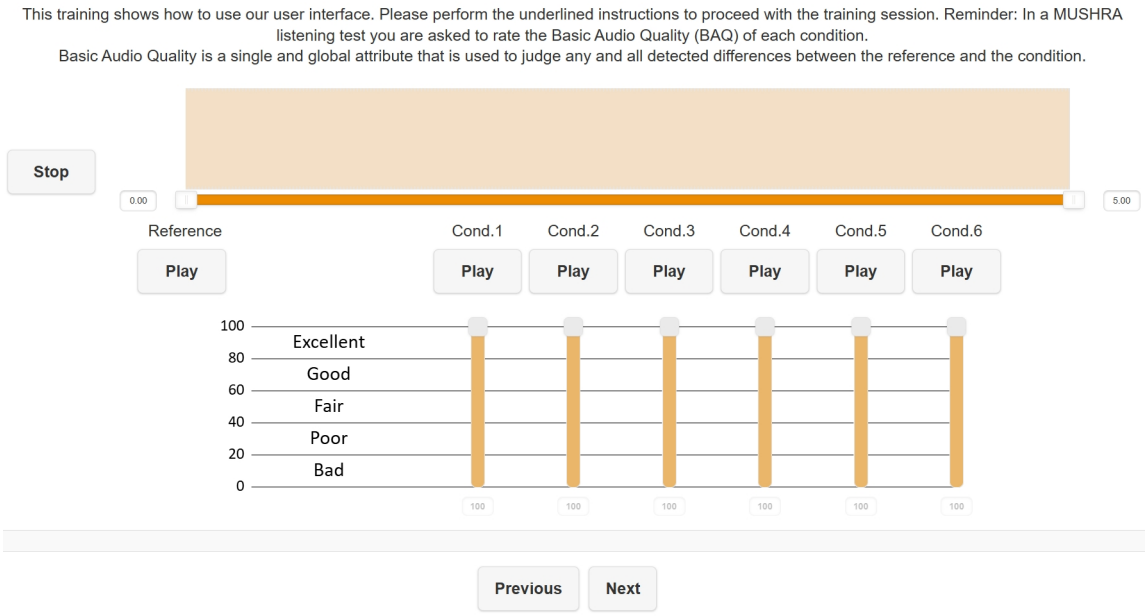


Figure 2.37: An example of the interface of the category judgement test [256].

Ranking: A ranking test is a type of evaluation method used to determine the order of preference or quality of a set of items based on specific criteria. In the case of audio samples, a ranking test involves presenting a set of sound samples to participants and asking them to rank them in order of preference or based on a specific criterion. For example, when assessing noise annoyance in buildings, there is no clear standard for how to quantify the level of annoyance. One possible solution is to perform a ranking test related to the estimation of annoyance instead of evaluating annoyance using an absolute scale [258]. The ranking can be done using

a variety of methods, such as a drag-and-drop interface and numerical rating. This task becomes more difficult because more sound samples have to be sorted. Therefore, this ranking test should not provide too many audio samples, and according to suggestion [248], the maximum number of audio samples provided should be six. This test method provides a simple, easy way to check the first impression of the audio samples; for example, using subject preference as an attribute. The disadvantage of this evaluation method is that the subject only specifies the order of individual audio samples, without providing information about the 'distance' between audio samples on that quality scale. Therefore, the results of the ranking test are not necessarily suitable for computational correlation with technical measurements or analytical results. Figure 2.38 [248] shows an example of a ranking test. By clicking on the sound button (Door01-Door06), the subject can play the corresponding sound and change its position in the ranking list using the arrow buttons.

Semantic Differential: The semantic differential test is a research method that is used to measure the meaning of a concept or a thing [259]. It is a questionnaire-based method that aims to capture the respondent's attitude or perception towards a particular topic or object. The application of the semantic differential test in listening tests is to measure a listener's perception of different aspects of sound, such as loudness, pitch, timbre, and rhythm. The test is typically administered after the participant has listened to a piece of music, a sound recording, or a speech segment. The subject assesses the presented sound according to several bipolar scales marked with an adjective and its antonym at each end of the scale. For example, for the attribute 'loudness', the opposing poles could be 'soft' and 'loud'. The listener is then asked to rate the sound they just heard based on

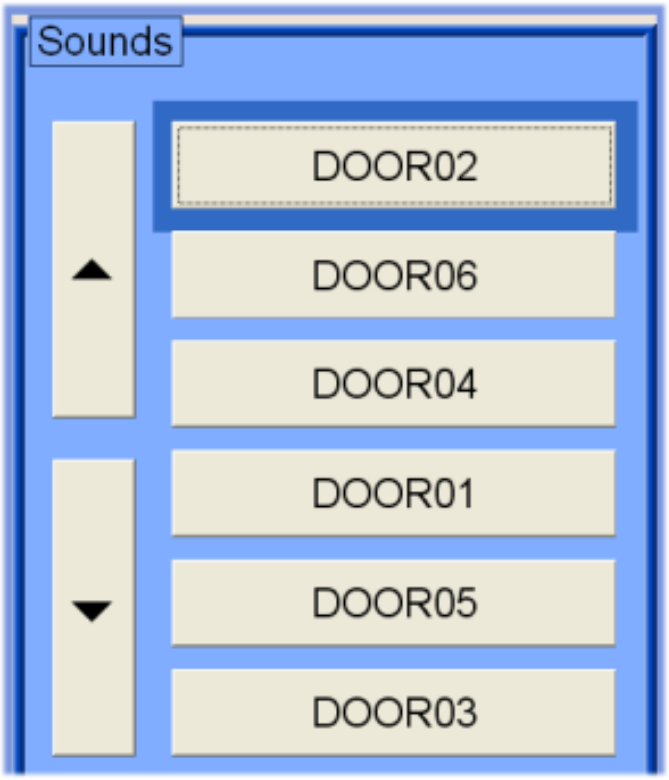


Figure 2.38: An example interface of a ranking test [248].

their perception of the attribute being measured, by placing a mark on a scale that lies between the two opposing adjectives. The scales usually have between seven or nine points [248]. This testing method provides more comprehensive insights into an audio sample than merely indicating which sample is preferred. Therefore, allow participants to express their feelings or perceptions about the audio sample in a nuanced way, capturing subtle attitudinal nuances. Judgements on several scales make it easier to find correlations with technical measurements and analysis results.

A typical example of the use of semantic differential test is the semantic differential analysis of the soundscape in urban open public spaces conducted by Kang

et al. [260]. This method allows researchers to discover why sounds are rated poorly and which aspects of the sound must be changed to improve the sound quality. However, the judgement of audio samples with semantic differential pairs is more time-consuming than other testing methods [248]. The number of audio samples and the scale of judgement should not be too large, otherwise the subject may lose concentration towards the end of the listening test. According to the recommendation [248], there should be no more than eight to twelve attribute pairs. Also, based on the purpose of listening test, in order to avoid the possibility that attribute correlation and adaptive effect may affect the results of listening test, the selection of attribute pairs needs to be very careful. Figure 2.39 [248] shows an example of a semantic differential test with a seven-point scale and four pairs of antonyms. This information can be used to understand how listeners respond to different aspects of sound and to identify which elements of sound are most important for different types of music or speech.

Figure 2.39: An example interface of a semantic differential test [248].

2.6.2 Statistical Analysis

Data acquired from listening tests must undergo rigorous analysis to derive meaningful conclusions. For this purpose, it's crucial to first determine the nature of the data distribution. The Lilliefors test [261] or the Kolmogorov-Smirnov test [262, 263] should be employed to verify normality, whereas homogeneity of variances can be tested using the Levene test [264] or the Bartlett test [265]. If the data conforms to the assumptions of normality and homogeneity of variance, parametric statistical tests such as analysis of variance (ANOVA) can be used. Conversely, for data that do not meet these criteria, non-parametric methods such as the Kruskal-Wallis test [266] are recommended, which do not require normality and are robust against heteroscedasticity [267].

Upon identifying significant differences using overarching tests like ANOVA, it becomes necessary to conduct post hoc pairwise comparisons [268] to pinpoint specific group differences. These subsequent analyses provide detailed insights into the significant distinctions between individual group pairs.

Statistical significance in academic research is generally inferred at a 95% confidence interval, with p-values less than 0.05 denoting significant differences [269]. In cases requiring greater statistical rigor, a 99% confidence level may be employed, demanding a more stringent p-value threshold.

2.7 Summary

This chapter has introduced the basic principles of sound, including the characteristics of different sound fields and sound propagation principles, which are

fundamental for creating realistic virtual environments. The basic requirements for the implementation of spatial audio were discussed, mainly HRTFs, head tracking and artificial reverberation simulation. The psychoacoustics of sound localisation was covered in the introduction of HRTFs: horizontal localisation through interaural time and level differences, and vertical localisation through spectral features interacting with the ear, head, and torso. These elements are crucial for achieving spatial audio realism, particularly in simulating how sound behaves in a three-dimensional space relative to listener orientation and position.

The composition of room impulse responses and their parameters involved in their evaluation were described, and discussed the measurement and synthesis of room impulse responses. These are central to understand how sound interacts within a space, and set the stage for designing plausible artificial reverberation algorithms. The importance of binaural reverberation, which is closely related to the human auditory system, in producing a realistic auditory experience was emphasised, and how reverberation can be auralised by convolution was presented.

Next, a review of the artificial reverberation algorithms based on delay network structures that have been proposed was presented, including the Schroeder [178], Moorer [175], Gardner [171], FDN [15], Dattorro [179] and DFDN [19] reverberation algorithms. This review provided an understanding of how these algorithms simulate the reverberation characteristics of the environment. By referring to these algorithms, further improvement and optimisation of artificial reverberation algorithms can be considered.

Application scenarios of virtual reverberation were reviewed and work exploring the implementation of reverberation in 6DoF rendering was discussed. This provides directions for exploring dynamic reverberation algorithms in virtual acoustic

rendering.

Finally, a perceptual evaluation of audio is presented. Subjective listening tests are crucial in the study of binaural reverberation. As subjective listening tests can be performed to assess the acoustic quality of virtual binaural reverberation and its similarity to real reverberation. This chapter describes the different methods of performing listening tests. The advantages and disadvantages as well as the scenarios in which they are applicable are compared by discussing the paired comparison method, the staircase method, the category judgement method, the ranking method and the semantic differential method, so as to select the appropriate test methods for the experiments covered in Chapter 3 and Chapter 4. In addition, statistical analysis methods of the test results are discussed. Through subjective listening tests on audio and statistical analysis of their results, the acoustic performance of the binaural reverberation model and its effectiveness in simulated real-world environments can be comprehensively assessed and validated. This not only enhances the reliability of the study, but also ensures that the results are scientifically sound and accurate.

While existing reverberation algorithms have proven effective in simulating reverberation, they still fall short of achieving the level of realism that fully mirrors true binaural hearing. The literature has identified several key challenges, including the computational inefficiencies that hinder real-time processing, especially in dynamic environments where the acoustic conditions change frequently. Traditional algorithms often struggle to balance accuracy with computational efficiency, leading to either high-quality reverberation that is too computationally expensive for real-time use, or faster algorithms that compromise on the realism and perceptual plausibility of the sound.

The need for a new reverberation algorithm arises from these gaps. To achieve a more realistic and dynamic reverberation experience in virtual acoustic rendering, it is essential to enhance both the realism and computational efficiency of artificial reverberation algorithms. A new algorithm must be capable of producing reverberation effects that are nearly indistinguishable from those in the real world while also being efficient enough to maintain real-time performance.

This research is informed by the gaps and limitations identified in the existing literature. By reviewing and analysing current methods, this work proposes a novel approach that aims to bridge the gap between realism and efficiency. The goal is to develop a dynamic reverberation algorithm that can deliver high-quality, realistic auditory experiences in virtual environments without sacrificing the computational performance required for real-time applications.

The background provided in this chapter forms the foundation for the study of binaural reverberation in virtual acoustic rendering that is carried throughout this thesis. The next chapter determines the effect of different parameters on the perceptual reverberation by evaluating the thresholds of the binaural room impulse response parameters, thus laying the foundation for the design of dynamic reverberation algorithms.

3

Evaluation of Parameters of Binaural Room Impulse Responses

This chapter undertakes an evaluation of binaural room impulse response parameters, essential for enhancing virtual acoustic rendering. The focal point of this study is a listening test designed to establish the influence and thresholds of critical BRIR parameters on perceptual reverberation. The outcomes from this test are important in refining the proposed reverberation algorithm, ensuring its performance closely aligns with human auditory perception within practical applications. As the efficacy of simulated reverberation largely depends on the accuracy and dynamic adaptability of these parameters, identifying their key influences enhances both the development of a real-time dynamic reverberation algorithm and its application in immersive audio environments.

The pursuit of realistic acoustic simulations compels us to analyse the mechanisms behind sound reverberation and propagation. As discussed in Section 2.3, reverberation is usually described by room impulse responses in acoustics research. Acoustic parameters that make up RIRs, including ERs, ITDG, and LR, influence the resultant perceived reverberation. These parameters form the backbone of our perception of virtual spaces, dictating the localisation of sound sources and the perceived quality of space. Their precise manipulation and understanding not only enhance user experience but also broaden the application of virtual acoustics in

various multimedia and entertainment technologies.

Many experts have conducted in-depth research on the perceptual properties of room acoustics and their relation to binaural reverberation parameters, as reviewed in Section 2.3.1 and 2.3.2. Their findings highlight how factors such as early reflections, ITDG, and late reverberation affect the localisation of sound sources and the perceived spatial quality, providing fundamental knowledge that supports the need for precise control of BRIR parameters in virtual acoustic systems.

These studies underscore that while ITDG, early, and late reflections influence auditory perception in a reverberant environment, but the contribution of each component of the BRIR and the thresholds of perceptibility under different reverberant conditions still remain uncertain. This chapter, therefore, delves into perceptual testing on ITDG, early reflections, and late reverberation to identify which parameters significantly impact listener experiences. Establishing perceptual thresholds for each parameter is crucial for determining which aspects of a BRIR are most influential and should be prioritised in the development of dynamic artificial reverberation algorithms.

By distinguishing between perceptually significant and insignificant changes, it is useful to create more efficient and effective acoustic simulations, ensuring that enhancements or modifications to BRIR parameters lead to meaningful improvements in auditory perception. The results of the test are analysed and discussed, ultimately guiding focused adjustments in the algorithm design to replicate the reverberation of real environments within virtual spaces effectively.

In practical terms, real-time virtual acoustics systems should operate on portable, low-power computing devices with minimal computational demand, emphasising the importance of optimising each parameter for maximum perceptual impact.

This chapter will explore these aspects, underpinned by empirical data from the listening test, to direct the development of more efficient and effective virtual acoustic systems.

3.1 Method

In the study of artificial reverberation, the RIR is is usually divided into two parts: one comprising direct sound and early reflections, and the other including late reverberation [270]. For this study, the ERs, ITDG and LR were evaluated, as the focus was primarily on understanding the effects of reflections, which can vary drastically based on room geometry and surface materials. Direct sound was kept constant, as a baseline for assessing the impact of variable acoustic parameters like ERs and LR [271].

Based on the fact that early reflections, late reverberation and ITDG all play a key role in the perception of reverberation in acoustic environments, affecting the spatial quality and clarity of the auditory signal, four specific parameters: reverse early reflection (RER) removal, forward early reflection (FER) removal, initial time delay gap (ITDG) extension and late reverberation (LR) removal were selected for a listening test to measure the perceptual thresholds of BRIR parameters. For comparison, BRIRs with three different reverberation times (calculated from T_{30} at 1000Hz octave), long (1.51 s), medium (0.91 s) and short (0.31 s), were used for testing. These three BRIRs were simulated in a same room model with only the materials affecting the reverberation time.

3.1.1 Participants

Twenty participants were recruited, all classified as expert listeners according to the ITU-R BS.1543-3 recommendation [255]. Each participant was paid to take part in the test which lasted about 1h. All participants were over 18 years old. Participants were not asked to report their gender. All of these participants come from the AudioLab (part of the Communication Technologies Research Group at the University of York) or music related majors at the University of York and Beijing Contemporary Music Academy.

3.1.2 Stimuli

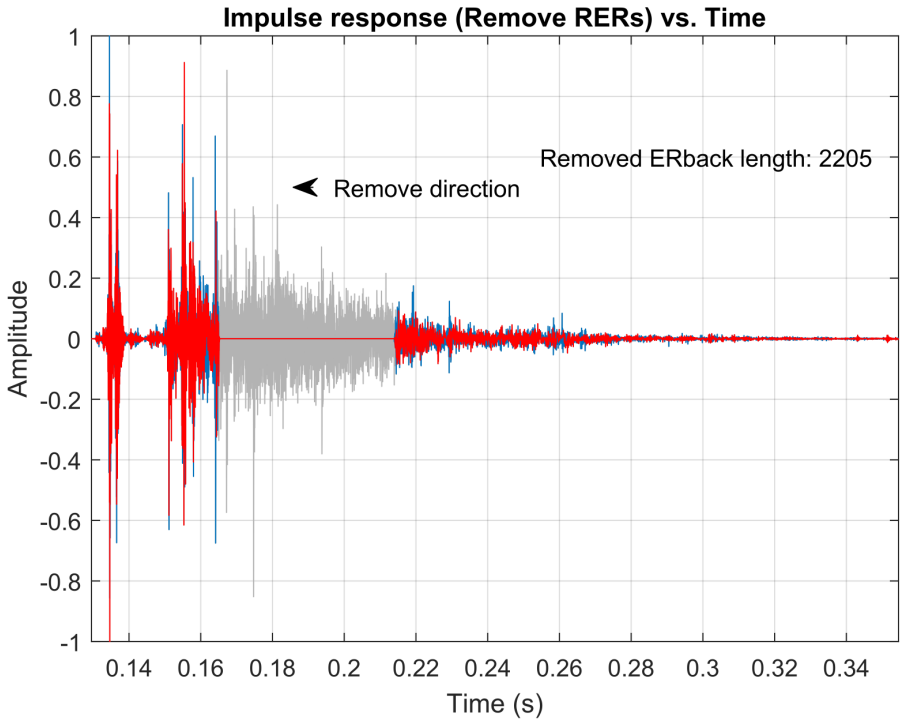
FER removal has been implemented through the removal of the initial reflections with subsequent tests removing further reflections forward from the initial reflections of the BRIRs. The opposite scenario, known as RER removal is achieved when the last early reflections (those just before the late reverberation) are removed first, and at each subsequent test render, the earlier reflections are removed through traversing further backwards towards the direct sound. LR removal is achieved by first removing the last late reverberation tail and then removing the earlier late reverberation by traversing further backwards towards the direct sound in subsequent tests. ITDG extension is increasing the time interval between the direct sound and the first early reflection.

The schematic figure referred to as Figure 3.1 illustrates how various parameters are manipulated within a BRIR that has a reverberation time of 0.31 s. The original impulse response, as a reference, is marked in gray, and the arrows in the figure indicate the direction of removal or extension of the parameters. Each part of the figure demonstrates a different modification to the BRIR to help explain

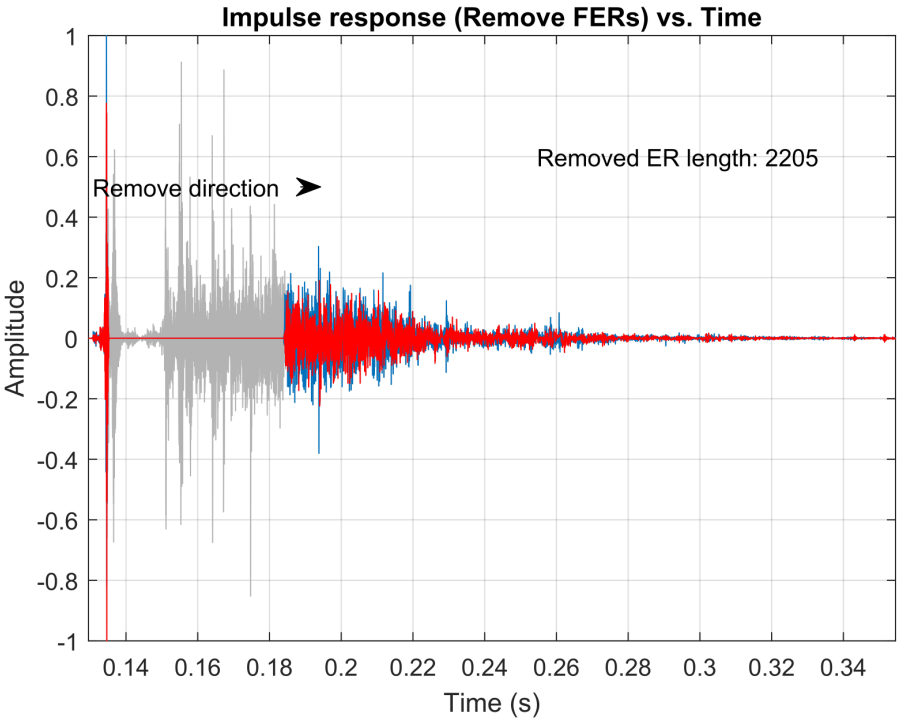
the impact of these changes. Figure 3.1 (a) shows the BRIR after ERs have been truncated reversely by 50 ms. This means that the early reflections occurring within the last 50 ms of the early reflection phase have been removed, starting from the point closest to the transition to late reflections and moving backward towards the direct sound. Figure 3.1 (b) displays the BRIR with early reflections forward truncated by 50 ms. In this adjustment, the first 50 ms of reflections following the direct sound are removed, which emphasizes the effect of removing the initial early reflections that immediately follow the direct sound. Figure 3.1 (c) presents the BRIR with the ITDG artificially extended by 50 ms. This modification delays the onset of all reflections relative to the direct sound by an additional 50 ms. Figure 3.1 (d) shows the BRIR with the LR cut off by 465 ms from the reverberation tail. This adjustment truncates the reverberation tail, removing the reflections that occur towards the end of the reverberation period.

In the process of removing RER, FER and LR from an impulse response, a technique involving a Hanning window is used to enhance the transitions. Specifically, half of a 64-point Hanning window is applied. This window, which translates to a 0.726 ms transition period at a sample rate of 44.1 kHz, is used to smooth out the abrupt edges that occur between the silent sections and the impulse response. This smoothing helps to reduce any artifacts or harshness that might result from cutting the impulse response abruptly, thereby ensuring a more natural-sounding fade in and fade out around the edited sections.

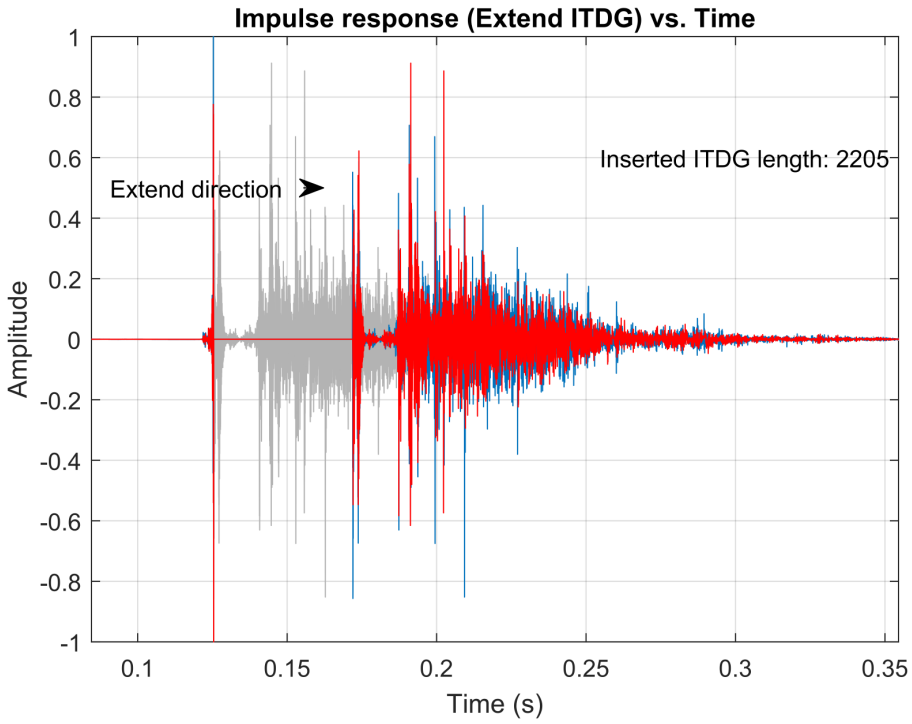
Three BRIRs with reverberation times of 0.31 s, 0.91 s, and 1.51 s, generated using the ODEON software, were employed as the reference impulse signals, as shown in Figure 3.2 (a)-(c). ODEON is regarded as a useful tool for research in objective and subjective room acoustics [148]. The BRIRs were generated using a


Impulse response (Remove RERs) vs. Time
Removed ERback length: 2205
Remove direction
Amplitude
Time (s)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0.14
0.16
0.18
0.2
0.22
0.24
0.26
0.28
0.3
0.32
0.34

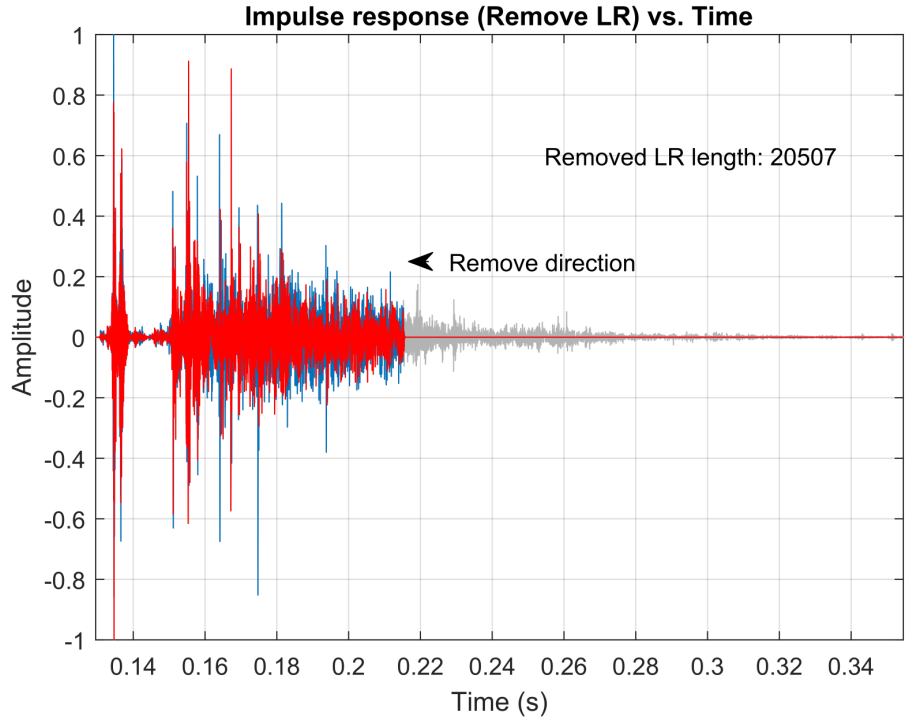
(a)


Impulse response (Remove FERs) vs. Time
Removed ER length: 2205
Remove direction
Amplitude
Time (s)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0.14
0.16
0.18
0.2
0.22
0.24
0.26
0.28
0.3
0.32
0.34

(b)



(c)



(d)

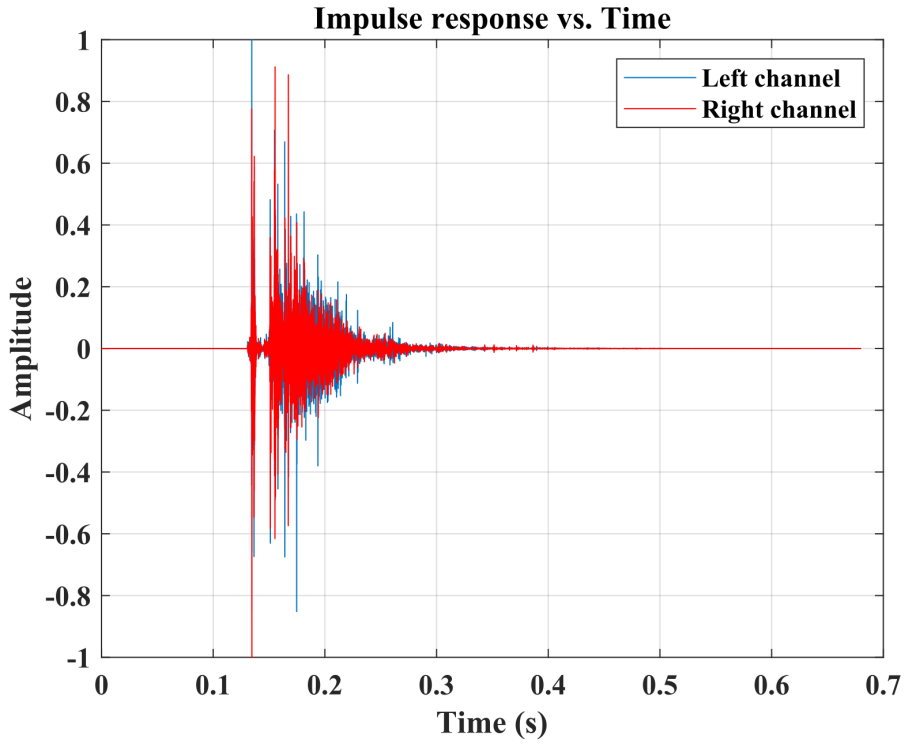
Figure 3.1: The explanation of changing different BRIR parameters, including RER removal, FER removal, LR removal and ITDG extension. (**a**) Cut off ERs reversely by 50 ms on the original BRIR. (**b**) Cut off ERs forward by 50 ms on the original BRIR. (**c**) Extend ITDG by 50 ms on the original BRIR. (**d**) Cut off LR by 465 ms on the original BRIR.

10 s, 44.1 kHz exponential sine-tone sweep [130], and the HRTF implemented in ODEON was Subject 21 from the Center for Image Processing and Integrated Computing (CIPIC) HRTF database. The CIPIC HRTF database is a public-domain database of high-spatial-resolution head-related transfer functions measured at the U. C. Davis CIPIC Interface Laboratory [272].

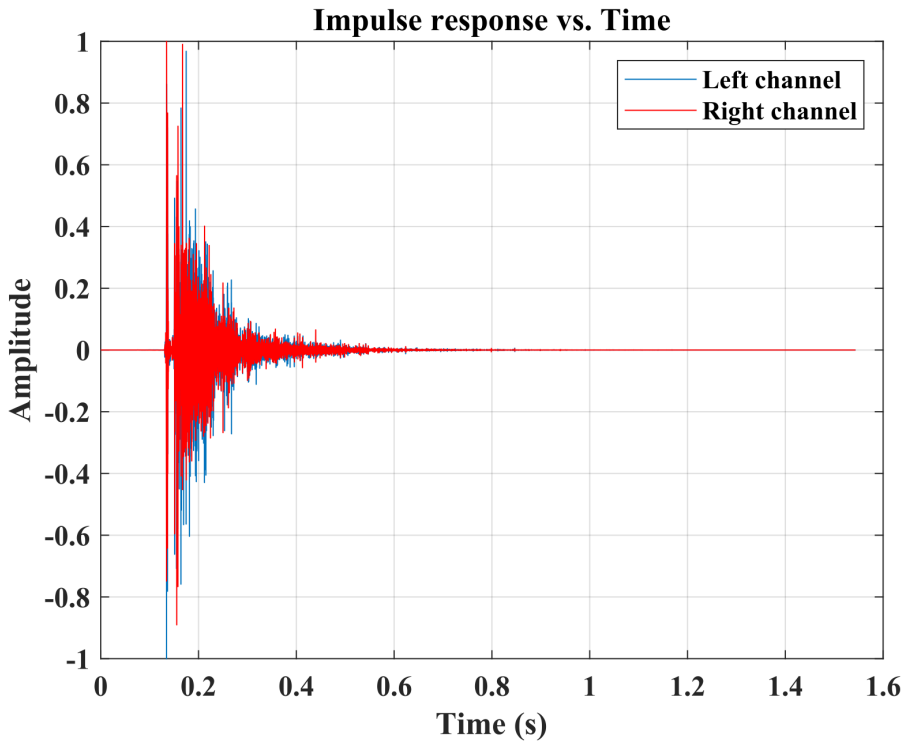
The setup involved placing the speaker 4 m from the front wall of a reverberant room, with the listening position at 13.5 m from the source at a 0-degree angle. The absorption coefficients were varied by altering the materials of the room surface, resulting in different reverberation times measured within the same room model. The materials used for these different reverberation times are detailed in Table A.1 in Appendix A. The room model was a cuboid with a length, width and height of approximately 22 m, 16 m and 10 m, respectively, as shown in Figure A.1 in Appendix A.

A brief segment of anechoic male speech audio was used as the test signal as shown in Figure 3.2 (d), as male speech is a common and recognised sound source with familiar timbre. The segment is a two-channel anechoic audio of 2.6 s length, with both channels being identical. Its sample rate is 44.1 kHz and bit depth is 24 bit. The listening test reference audio samples were generated by convolving these three BRIRs with the anechoic male speech audio, allowing for an evaluation of how different reverberation times affect the perceived acoustics of the test signal.

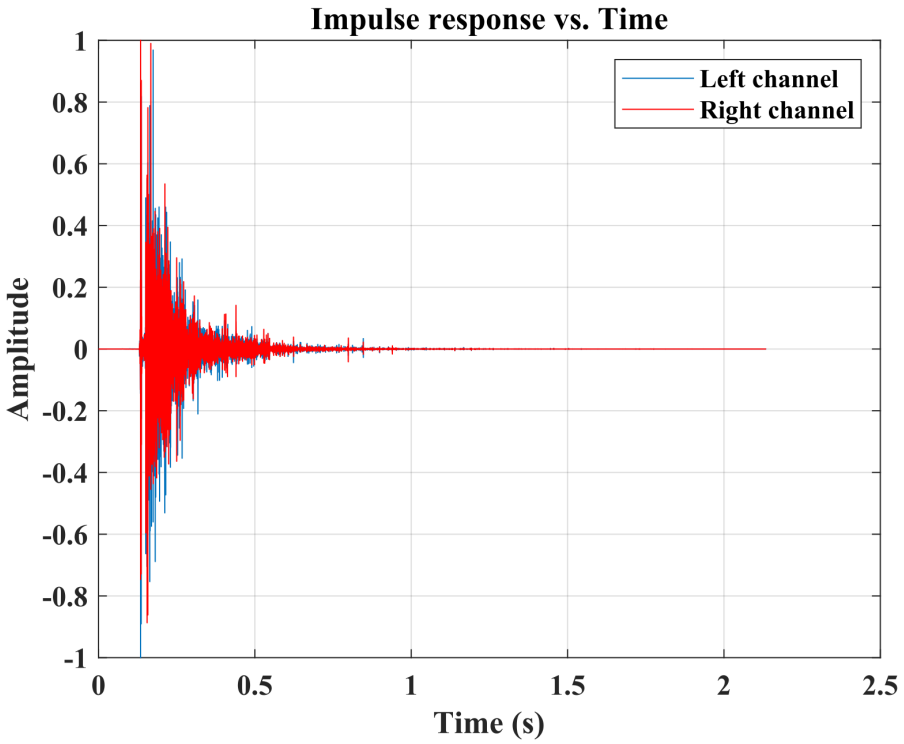
Contrast BRIRs were generated by changing one of the identified acoustic parameters of the reference BRIRs from above four variable acoustic parameters (RER removal, FER removal, LR removal and ITDG extension). These altered BRIRs were convolved with the same anechoic male speech audio to generate the contrast stimuli.


Impulse response vs. Time
Left channel
Right channel
Amplitude
Time (s)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

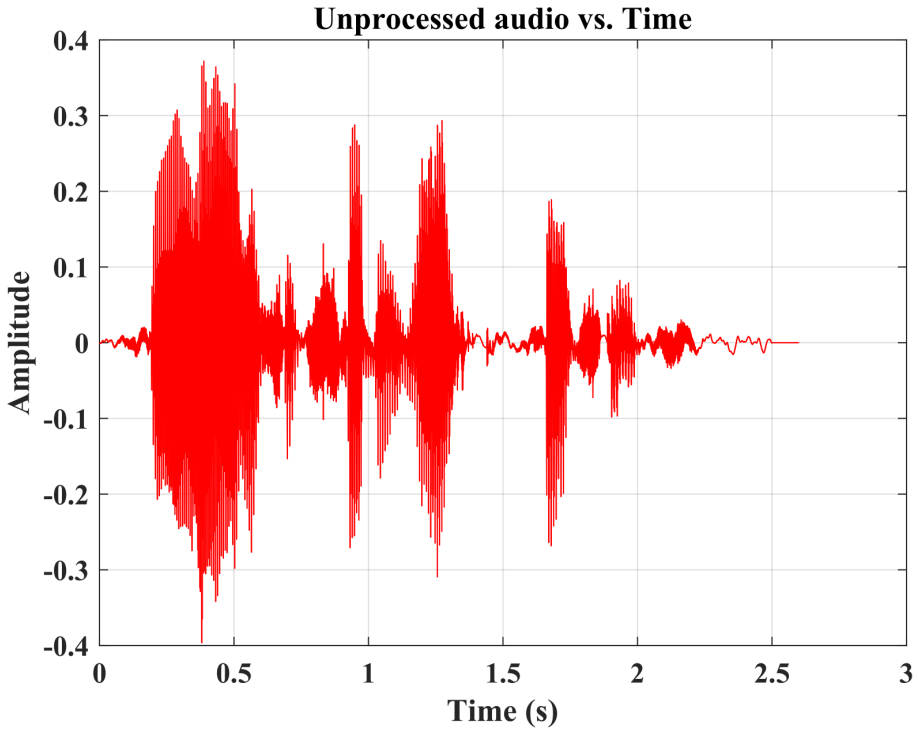
(a)


Impulse response vs. Time
Left channel
Right channel
Amplitude
Time (s)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

(b)



(c)



(d)

Figure 3.2: The BRIRs and the test signal used in the listening test. (**a**) The BRIR with 0.31 s reverberation time. (**b**) The BRIR with 0.91 s reverberation time. (**c**) The BRIR with 1.51 s reverberation time. (**d**) The dry male speech audio signal.

3.1.3 Procedure

The entire listening test is built on a graphical user interface (GUI) application self-developed by the author using MATLAB's Appdesigner toolbox, where Appdesigner is a MATLAB environment for application development [273]. The application was sent to the participants in the form of an installation package and the participants were guided through the instructions to install it. Prior to conducting the test on the application, participants were informed of the purpose of the experiment and the experimental protocol. After participants reviewed the information sheet and agreed to the consent form, they could access the entire listening test through the application. This study received approval from the University of York Physical Sciences Ethics Committee under the approval code Mi111120. For further details, please refer to Appendix C.

Three different reverberation times and four acoustic parameters were employed to evaluate changes in perceptual thresholds of reverberation, so the whole experiment is divided into three groups and each group includes four parts as shown in Table 3.1. All 12 parts were presented in a random order and a 30 s rest time is set between each part to reduce fatigue and experimental error caused by sustained concentrated listening.

Table 3.1: The experimental design and corresponding predetermined conditions of the staircase method (Same stop condition is 5 'Yes' add 10 trials or 30 trials).

Parts / Groups		1 (RER Removal)	2 (FER Removal)	3 (ITDG Extension)	4 (LR Removal)
1	start point:	50ms	35ms	40ms	465ms
(0.31s reverb time)	step size:	5ms to 3ms to 1ms	5ms to 3ms to 1ms	5ms to 3ms to 1ms	10ms to 5ms to 3ms
2	start point:	50ms	35ms	40ms	780ms
(0.91s reverb time)	step size:	5ms to 3ms to 1ms	5ms to 3ms to 1ms	5ms to 3ms to 1ms	10ms to 5ms to 3ms
3	start point:	50ms	35ms	40ms	1250ms
(1.51s reverb time)	step size:	5ms to 3ms to 1ms	5ms to 3ms to 1ms	5ms to 3ms to 1ms	10ms to 5ms to 3ms

The measurement of perceptual thresholds for each parameter was established through AB blind test [249] and the staircase method [251] discussed in Section 2.6.1. In this listening test, participants were asked to listen to and compare reference audio samples to audio samples rendered with a single variable acoustic parameter. They responded 'Yes' or 'No' to the question 'Are audio samples A and B the same?'.

This experiment involves the threshold detection for four BRIR parameters: RER, FER, ITDG, and LR. The starting points for ITDG, RER, and FER were established empirically through multiple preliminary listening tests conducted by the author. For ITDG, the initial extended interval was set at 40 ms, identified as the point where changes became clearly audible across all tested conditions. Similarly, the start points for removing RER and FER were determined based on the author's perceptual sensitivity, with RER and FER removal start points set at 50 ms and 35 ms, respectively. The preliminary tests utilised a method of adjustments. In this method, extension or removal time was incrementally adjusted until the alterations were clearly perceptible to the author. For instance, with ITDG, the extension began at 0 ms and was increased in 5 ms steps until the delay was distinctly noticeable. Similarly, for RER and FER, removal began from a baseline of 0 ms, increasing in 5 ms increments until the changes were perceptually evident to the author. The rationale behind conducting multiple preliminary tests was to ensure the reliability of the start points to be used in the main experiment. It was essential to determine a starting point that was close to the actual perceptual threshold but still discernible, as it minimises the number of trials needed and reduces potential biases that might arise from over or underestimation of sensory changes.

Unlike the process of determining the starting points of RER, FER, and ITDG, the process of determining the removal starting point of LR involves additional steps due to its significant relation to the reverberation time across different BRIRs. To minimise the number of preliminary tests and thus prevent auditory fatigue, the removal of LR starts from 0 ms but with increments of 100 ms until a perceptible change is noted by the author. Recognising the challenge in discerning changes in LR compared to the other parameters, a pilot test was conducted to identify a perceptually meaningful start point for LR removal. The author carried out a basic pilot test involving three expert listeners (in accordance with the recommendations of ITU-R BS.1543-3 [255]), including the author themselves. These experts determined a range of potential starting points using the same preliminary listening tests described above. From this pilot study, the author empirically established the starting points for LR removal. For BRIR with a short reverberation time of 0.31 s, the starting point was set at 465 ms. For a medium reverberation time of 0.91 s, it was set at 780 ms. For BRIR with a long reverberation time of 1.51 s, the removal begins at 1250 ms.

In the experiments for ITDG, RER and FER, the initial step sizes are set at 5 ms. These step sizes are dynamically adjusted based on the participants' responses: they are reduced to 3 ms after receiving three affirmative ('Yes') responses and further decreased to 1 ms following five affirmative responses. For the LR tests, which begin with larger initial differences due to the large time proportion of late reverberation in the entire BRIR, the starting step sizes are set at 10 ms. Similar to the other tests, these are adjusted downwards in stages—first to 5 ms after three 'Yes' responses, and then to 3 ms after five 'Yes' responses. This gradual adjustment allows for a more precise determination of the threshold as the test

progresses, refining the resolution based on the participant's sensitivity to changes.

There are two kinds of stop conditions. One is to decide the end point at several trials after a predetermined number of 'Yes' responses, as this experiment presumes an initial response of 'No'. The other is to decide a fixed end point through a predetermined number of trials. Theoretically, the greater the number of trials, the more reliable the results will be, but as the number of trials increases it can cause participant fatigue which affects the accuracy of the results, and the time taken to carry out the test increases. In this experiment, two types of stopping conditions are employed to balance both the reliability of the results and the time efficiency of the tests. First Stop Condition: After a participant provides five affirmative ('Yes') responses, this condition is activated if they complete an additional ten trials without reaching a total of thirty trials for that part of the test. Second Stop Condition: Alternatively, this condition automatically halts the test after thirty trials are conducted, regardless of the number of affirmative responses. If the total number of trials in a test part fail to reach thirty, the first stop condition takes precedence. Upon triggering either stop condition, the resolution threshold is calculated by averaging the last ten values obtained from the reversals. Figure 3.3 illustrates this process, showing an example of how the start point, step size, stop point, and threshold calculation are determined for clarity. Furthermore, Table 3.1 details the predetermined conditions for each part of the test, noting that since the above stopping conditions are uniformly applied across all 12 parts of the experiment, they are not redundantly listed in the table.

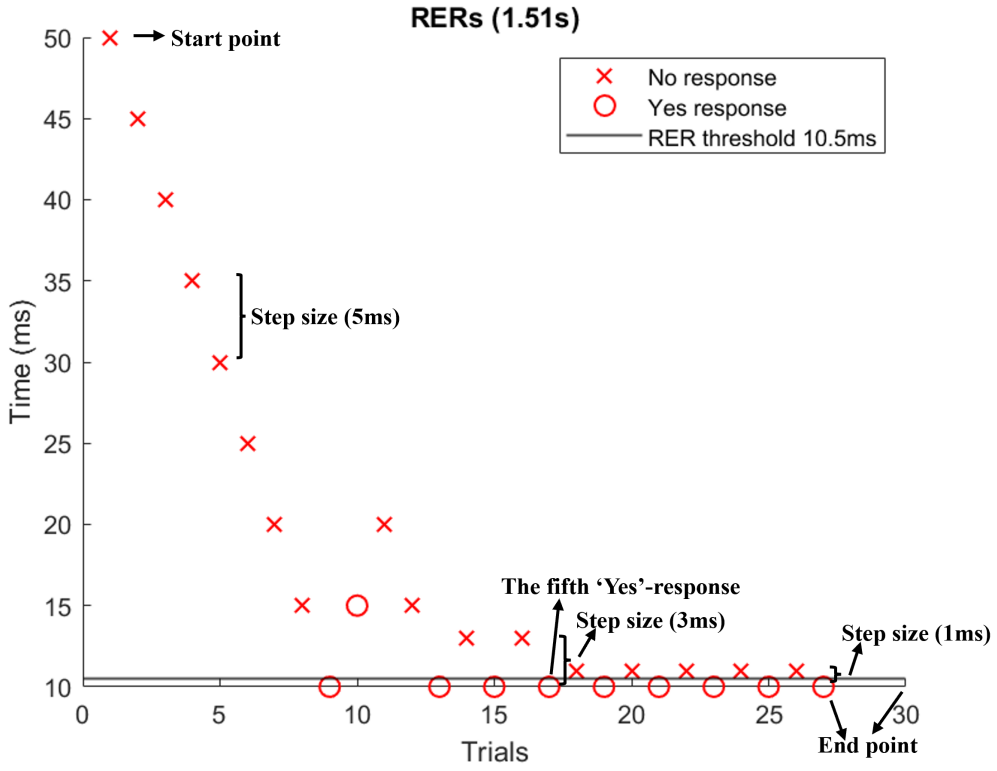


Figure 3.3: The diagram illustrates the predetermined conditions of the staircase method. The red circles represent affirmative responses ('Yes') and the red crosses represent negative responses ('No'). The initial step size for the test was set at 5 ms. Depending on the participant's response, these step sizes were reduced to 3 ms after three affirmative responses and further reduced to 1 ms after five affirmative responses. The test ended when the stop condition was triggered by the participant giving five affirmative responses followed by ten more trials, and the last ten values of reversals are averaged to obtain the resolution threshold.

3.1.4 Apparatus

The listening test was conducted online, utilising the test subjects' own personal computers or laptops along with a pair of 'Beyerdynamic DT990 Pro' headphones. Subjects were instructed to conduct the experiment in a quiet listening environment to ensure consistent and accurate results.

The experimental software used was a custom listening test application devel-

oped by the author using Appdesigner (a MATLAB environment for application development [273]). The application's operation interface is shown in Figure 3.4. Participants can click the 'Play A' or 'Play B' button to listen to reference or contrast audio samples, respectively, and then click the 'Yes' or 'No' button to record their responses. The 'Previous' button allow participants to return to a previous trial to review or change their responses, with any new answer overwriting the previous one. The 'Next' button is used to proceed to the next trial. The number on the interface shows the number of trials completed in the current part.

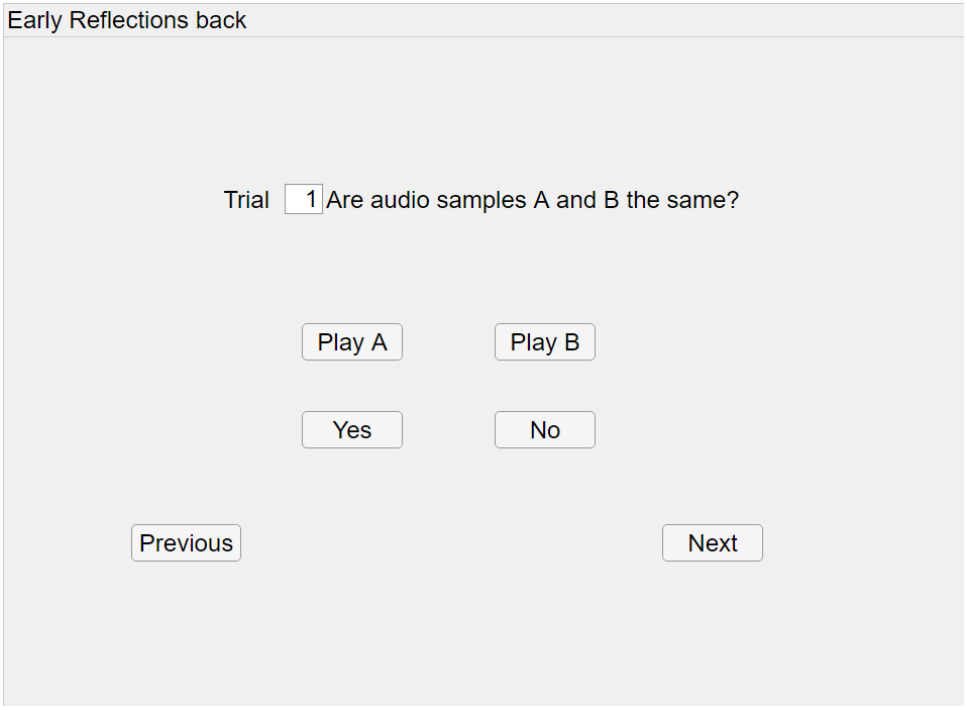


Figure 3.4: The operation interface of the listening test software.

It is important to note that participants may have completed more trials than what was recorded, as they could revisit and change previous responses without limit, though the exact number of such repeated trials is unknown. This factor could potentially introduce listener fatigue, especially in an online setting where

control over the environment is limited. Prolonged participation in the test could lead to decreased attention and increased errors, potentially influencing the accuracy and reliability of the responses.

3.2 Results

The test data are shown in Table 3.2 and 3.3. A higher threshold suggests a smaller impact of changes in that parameter on perceptual reverberation. This implies that only more significant alterations to this parameter enable participants to distinguish between the modified and the original reverberation. Data marked in red denote the maximum alterations that were applied to the parameters during the experiment. At these maximum values, participants were unable to perceive any differences between the reference and contrast audio samples, thus these are considered the upper thresholds for perceptual change. Conversely, data marked in blue represent the minimum changes that were applied to the parameters during the experiment. The smallest alterations to parameters lead to perceivable differences between the reference and contrast audio samples. These minimum values are hence considered the lower thresholds for perceptual sensitivity.

3.2.1 The Average Threshold Analysis of Each Parameter Type

In order to obtain generic impact thresholds of these parameters on perceptual reverberation, it is essential to compute the arithmetic mean of the collected individual thresholds. Impact thresholds refer to the specific points at which changes in certain parameters of a BRIRs start to produce a noticeable effect on perceptual

Table 3.2: The test threshold data of each participant (unit: ms). Data marked in red denote the upper thresholds for perceptual change, while data marked in blue represent the lower thresholds for perceptual sensitivity.

Parameter Types / Participants	0.31s RER Removal	0.31s ITDG Extension	0.31s FER Removal	0.31s LR Removal	0.91s RER Removal	0.91s ITDG Extension
1	27.5	17.9	20.7	380.1	36.3	24.5
2	2.1	8.9	17.3	415.3	14.7	27.7
3	38.5	34.7	39.3	464.5	26.9	37.1
4	20.7	65	35.7	450.5	21.7	33.7
5	13.9	17.1	23.9	435.1	15.9	23.3
6	12.3	9.3	22.5	268.3	18.3	24.7
7	3.3	19.3	20.5	400.5	1.3	7.7
8	7.3	12.1	22.7	380.1	10.5	6.1
9	18.7	14.3	30.1	437.7	18.1	48.1
10	2.7	2.7	5.1	409.5	15.1	12.9
11	16.3	2.5	31.3	451.9	18.9	40.5
12	2.7	14.5	28.7	423.7	15.5	31.3
13	29.9	23.1	40.1	455.3	21.3	25.9
14	13.7	15.1	29.1	464.1	19.1	16.9
15	12.3	16.5	29.1	429.7	9.1	30.5
16	51.5	51.7	36.7	464.5	61.7	45.1
17	12.5	31.5	35.3	440.5	18.7	1.5
18	25.3	27.1	29.9	464.1	21.1	17.9
19	28.1	30.3	41.1	463.2	44.3	43.3
20	12.5	0.5	14.5	444.5	14.9	5.5
Average	17.59	20.705	27.68	427.155	21.17	25.21
Standard Deviation	12.87	16.11	9.35	46.04	13.19	13.89
Standard Error	2.88	3.60	2.09	10.29	2.95	3.11

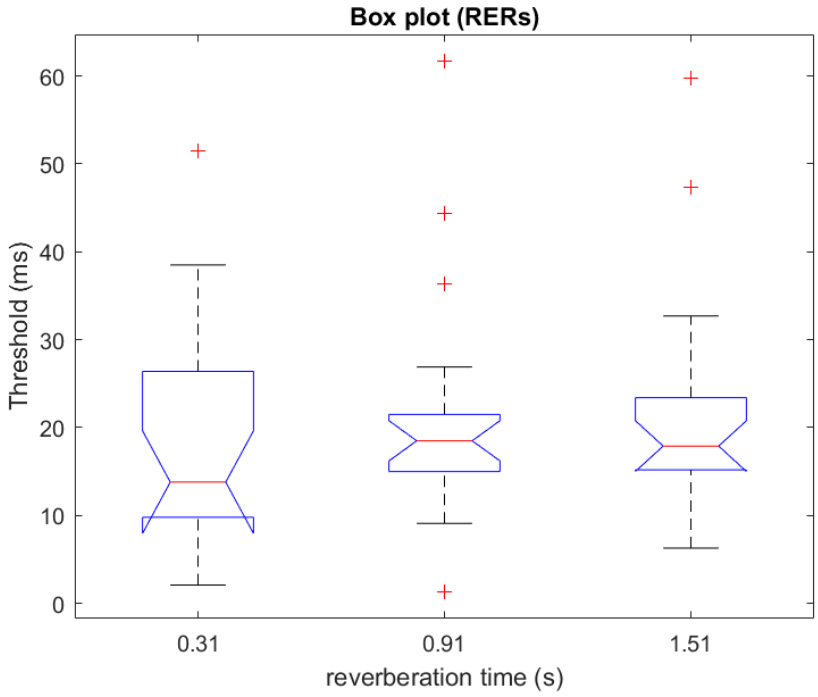
Table 3.3: The test threshold data of each participant (unit: ms). Data marked in red denote the upper thresholds for perceptual change, while data marked in blue represent the lower thresholds for perceptual sensitivity.

Parameter Types / Participants	0.91s FER Removal	0.91s LR Removal	1.51s RER Removal	1.51s ITDG Extension	1.51s FER Removal	1.51s LR Removal
1	32.7	731.9	24.7	65	25.9	1310
2	29.7	730.9	17.5	13.3	28.3	950
3	33.7	830	32.7	50.7	49.1	1310
4	45.3	830	22.3	65	46.7	1310
5	22.9	803.3	18.1	13.7	22.7	1195
6	27.3	696.5	17.7	18.9	40.5	950
7	32.9	804.9	6.3	40.7	40.5	1291.1
8	21.5	652.5	15.7	3.7	19.9	950
9	28.3	744.9	15.1	3.1	24.7	1282.1
10	20.3	756.7	17.7	20.5	9.5	1252.1
11	20.7	787.1	22.3	15.3	44.1	1310
12	34.9	820	14.1	46.7	30.7	1310
13	60	752.3	15.3	46.7	60	950
14	47.3	782.5	24.5	27.3	34.9	1259.3
15	44.9	830	10.7	31.3	49.1	1310
16	40.9	798.7	59.7	49.3	49.7	1275.1
17	39.3	818.1	19.9	26.1	50.1	1310
18	39.7	759.5	11.9	27.7	25.3	1260.5
19	46.3	796.9	47.3	44.5	22.7	1283.7
20	17.9	696.5	20.7	5.5	17.5	1161.5
Average	34.325	771.16	21.71	30.75	34.595	1211.52
Standard Deviation	11.14	50.64	12.43	19.39	13.72	139.76
Standard Error	2.49	11.32	2.78	4.34	3.07	31.25

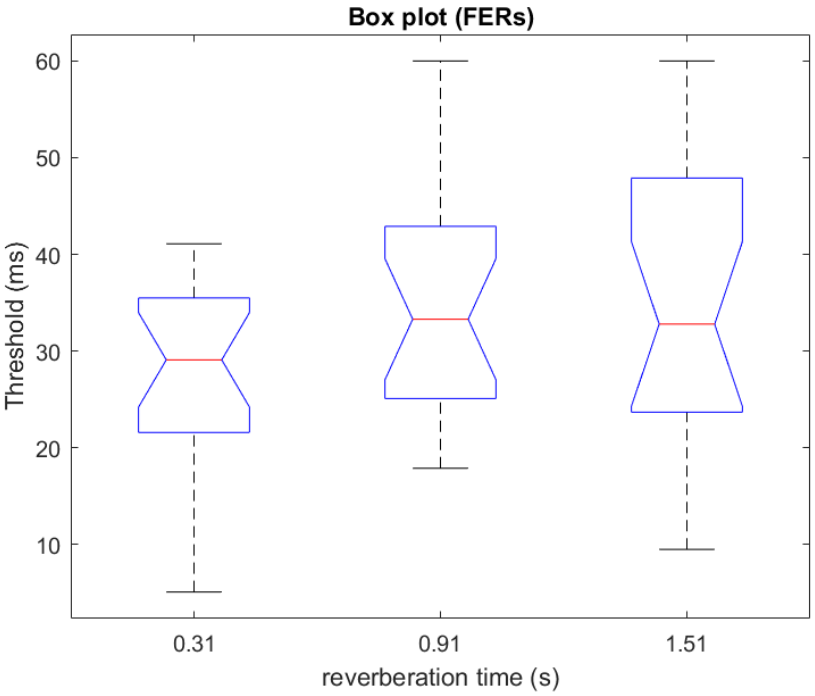
reverberation. Before this can be done, it is crucial to ensure the integrity of the data. Box plots are employed to visually represent the distribution of the data and to identify potential outliers. Outliers are defined as the values outside the upper and lower quartiles plus 1.5 times the interquartile range (IQR) [274]. Figure 3.5 (a) to Figure 3.5 (d) show the box plots of the thresholds for RER removal, FER removal, ITDG extension and LR removal, where each box plot contains three reverberation times of 0.31 s, 0.91 s and 1.51 s. The information provided by these box plots assists in evaluating whether outliers should be retained or removed from the dataset based on their potential impact on the analysis. This ensures that the calculated averages are representative and robust, minimizing the influence of outliers.

After checking, these outliers were not caused by input or measurement errors. The listening test was based on a subjective assessment and the outliers were caused by individual differences in auditory sensitivity, which refers to the ability of an individual to perceive sounds of varying intensities and frequencies. These outliers should therefore be retained as a natural part of the study and statistical analysis was conducted using non-parametric hypothesis tests to ensure robustness.

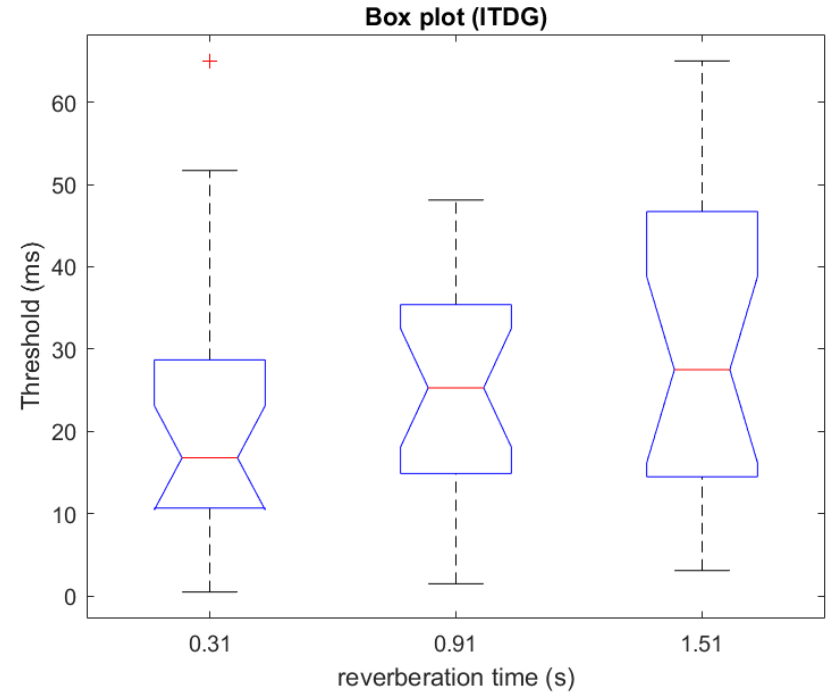
Then the data provided by these 20 experienced participants were averaged. Figure 3.6 (a) to Figure 3.6 (d) are the error bars presenting the mean thresholds with standard errors of RER removal, FER removal, ITDG extension and LR removal, where each plot contains three reverberation times of 0.31 s, 0.91 s and 1.51 s. As shown in Figure 3.6 (a), the average impact thresholds of RER removal are 17.59 ms, 21.17 ms and 21.71 ms corresponding to 0.31 s, 0.91 s and 1.51 s reverberation times, respectively. The slight increase in thresholds with increasing reverberation time suggests that participants find it slightly more difficult to dis-


Box plot (RERs)
Threshold (ms)
0
10
20
30
40
50
60
0.31
0.91
1.51
reverberation time (s)

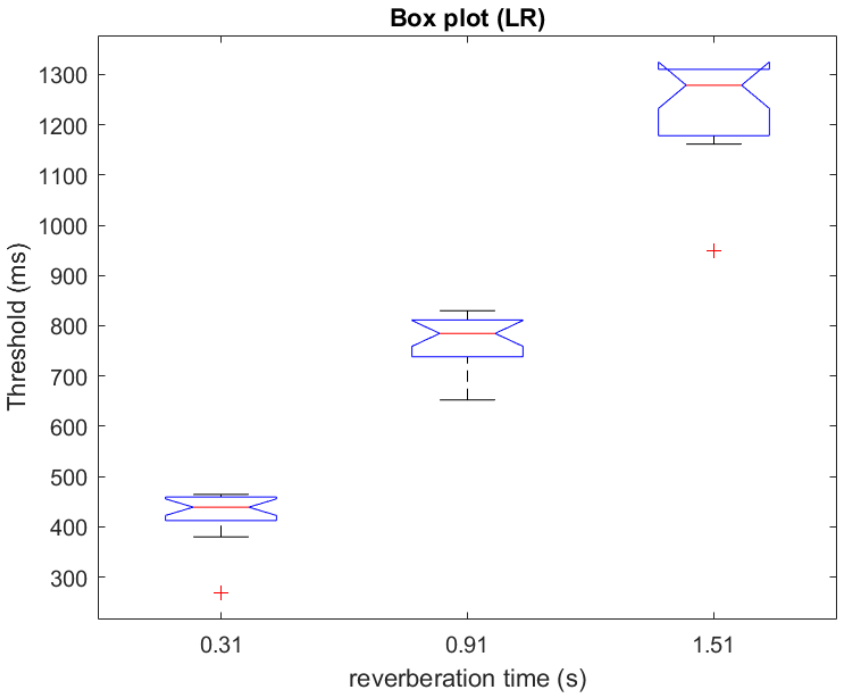
(a)


Box plot (FERs)
Threshold (ms)
10
20
30
40
50
60
0.31
0.91
1.51
reverberation time (s)

(b)



(c)



(d)

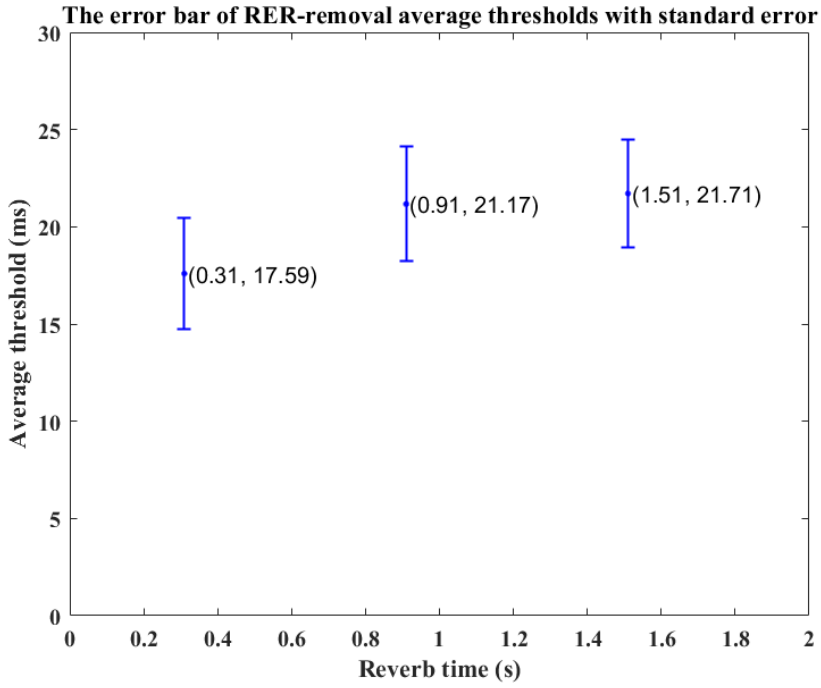
Figure 3.5: Box plots of the thresholds of RER removal, FER removal, ITDG extension and LR removal. These box plots show the interquartile range of the thresholds for each group, the first quartile at the bottom of the box, the third quartile at the top, and the median of the thresholds represented by the red horizontal line inside the box. The box whiskers extend from the first quartile to the minimum value and from the third quartile to the maximum value. Any point outside these ranges can be considered an outlier. (**a**) RER removal thresholds. (**b**) FER removal thresholds. (**c**) ITDG extension thresholds. (**d**) LR removal thresholds.

cern changes in RER as the reverberation time increased. However, this change was not substantial. For FER removal, as shown in Figure 3.6 (b), the corresponding average thresholds for these three reverberation times are 27.68 ms, 34.325 ms and 34.595 ms. There is a noticeable increase in the threshold from 0.31s to 0.91s and it stabilizes slightly above 34 ms for 0.91s and 1.51s. This suggests a higher sensitivity to FER changes at shorter reverberation times. However, its average thresholds are higher than the average thresholds of RER removal, so its effect on perceptual reverberation is less than RER removal. When ITDG is extended, as shown in Figure 3.6 (c), the average thresholds are 20.705 ms, 25.21 ms and 30.75 ms, respectively. The threshold for ITDG extension increases consistently with the increase in reverberation time. This indicates that it becomes increasingly difficult for listeners to perceive the extension of ITDG as the overall reverberation time increases. As shown in Figure 3.6 (d), the average impact thresholds of LR removal on perceptual reverberation are 427.155 ms, 771.16 ms and 1211.52 ms corresponding to 0.31 s, 0.91 s and 1.51 s reverberation times. The threshold values dramatically increase with longer reverberation times, which might suggest that listeners find it significantly more challenging to detect changes in late reverberation as the reverberation time extends. This parameter shows the most variability and the highest thresholds among the tested parameters.

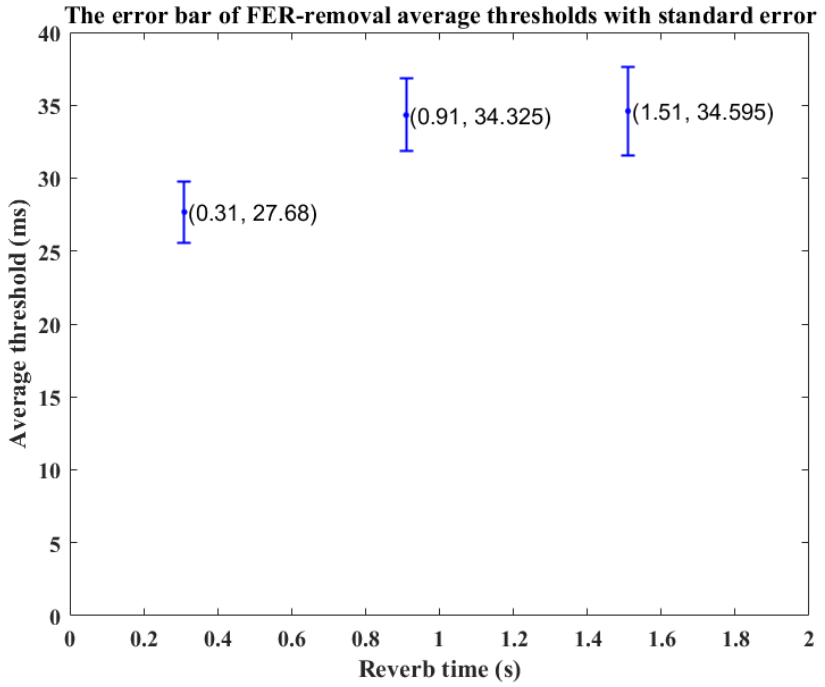
The analysis shows varying sensitivity across different reverberation times and parameters. Late reverberation changes are the hardest to perceive, especially as reverberation time increases, reflecting possibly a lower sensitivity of auditory perception to changes happening in the late reverberation in the sound signal. On the other hand, early reflections, including FERs and RERs, and ITDG also show increasing thresholds with longer reverberation times but to a lesser degree

compared to LR. The thresholds increase with the duration of reverberation, indicating that as sound environments become more complex (longer reverberation), the human auditory system's ability to discern changes diminishes or requires more substantial changes to notice differences.

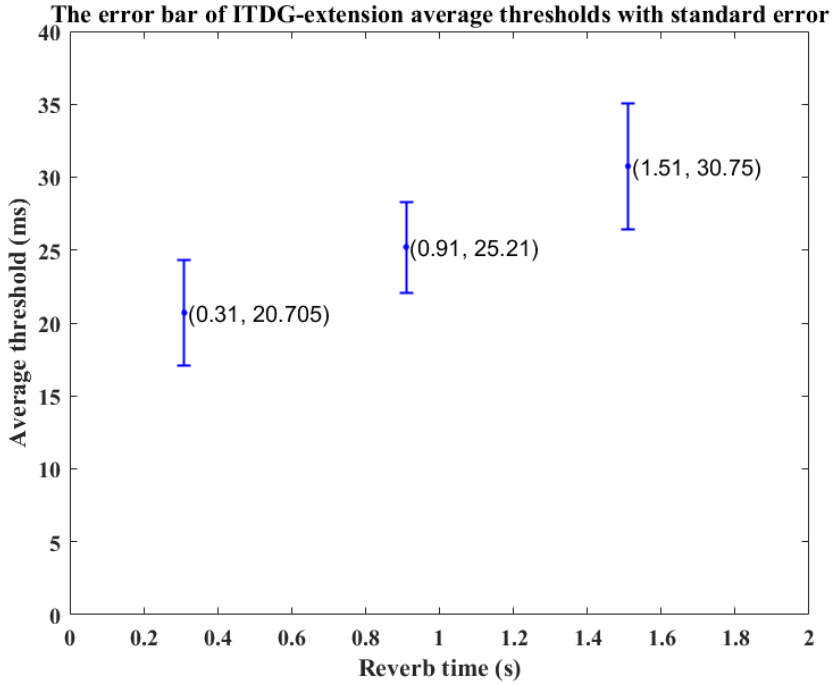
By calculating the data collected in Table 3.2 and 3.3, the standard deviation of RER removal thresholds are 12.87 ms, 13.19 ms and 12.43 ms corresponding to 0.31 s, 0.91 s and 1.51 s reverberation times. The standard deviation in thresholds for RER removal is fairly consistent across all reverberation times, suggesting a relatively uniform sensitivity among participants to changes in RERs. The standard deviation of FER removal thresholds are 9.35 ms, 11.14 ms and 13.72 ms, respectively. The increasing standard deviation with longer reverberation times indicates that participants' perceptual variability to changes in FERs increases as reverberation time increases. The standard deviation of ITDG extension thresholds are 16.11 ms, 13.89 ms and 19.39 ms, respectively. The variability in the standard deviation for the ITDG extension thresholds presents an interesting pattern where the middle reverberation time (0.91 s) has the lowest standard deviation compared to the shorter (0.31 s) and longer (1.51 s) times. It could be caused by that in very short reverberations, the temporal proximity of direct sound and early reflections may make it harder for participants to discern changes to the ITDG, increasing variability. Conversely, at the longest reverberation time, the dense overlap of reflections likely complicates the auditory scene, making consistent perception more difficult and increasing variability. The standard deviation of LR removal thresholds are 46.04 ms, 50.64 ms and 139.76 ms, respectively. The dramatically higher standard deviation at the longest reverberation time indicates significant variability in how participants perceive late reverberation. This suggests that late


The error bar of RER-removal average thresholds with standard error
Average threshold (ms)
Reverb time (s)
(0.31, 17.59)
(0.91, 21.17)
(1.51, 21.71)
0
5
10
15
20
25
30
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

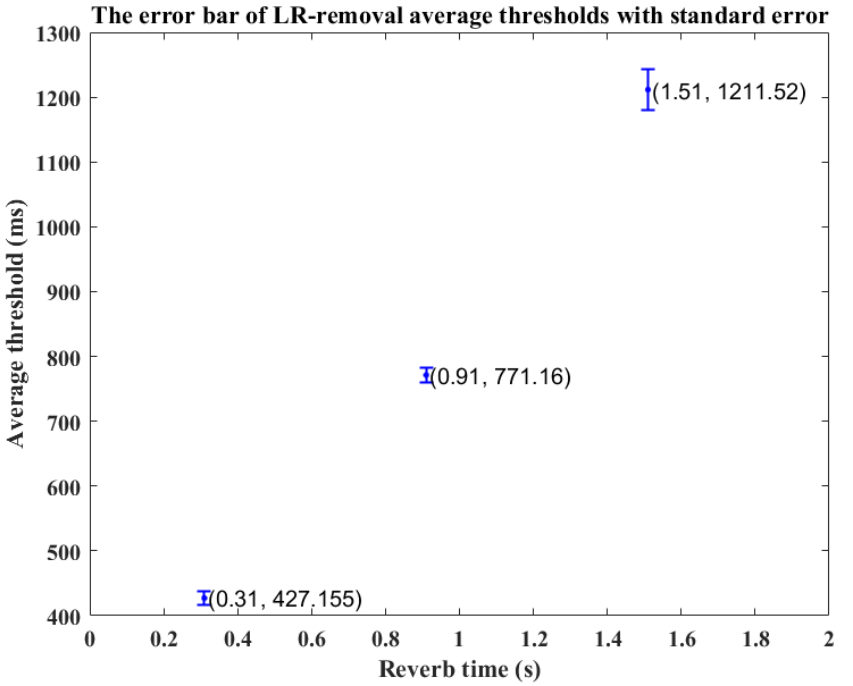
(a)


The error bar of FER-removal average thresholds with standard error
Average threshold (ms)
Reverb time (s)
(0.31, 27.68)
(0.91, 34.325)
(1.51, 34.595)
0
5
10
15
20
25
30
35
40
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

(b)



(c)



(d)

Figure 3.6: The error bars of the average thresholds distribution with standard errors of each parameter for different reverberation times. (**a**) The error bar of RER removal. (**b**) The error bar of FER removal. (**c**) The error bar of ITDG extension. (**d**) The error bar of LR removal.

reverberation are subject to high perceptual variability, with some participants being highly sensitive to changes in late reverberation, while others are less so.

The data reveals that LR removal presents the most significant challenge in perceptual consistency among listeners, especially as reverberation time increases. This might imply that the acoustic complexity introduced by late reverberation at longer durations makes it difficult for listeners to uniformly perceive changes.On the other hand, parameters like RER removal and FER removal show less perceptual variability, indicating more consistent auditory responses among participants. The smallest standard deviation at the medium reverberation time indicates a relatively uniform perception among participants when experiencing ITDG extensions.

3.2.2 Statistical Analysis

As can be seen in Figure 3.6, the longer the reverberation time, the greater the average impact threshold of the parameters. However, further significance test and post-hoc test are required to determine whether the reverberation time has a significant effect on the thresholds of perceptual reverberation.

Listening test data were checked for normality and variance homogeneity using the Lilliefors test [261] and Bartlett test [265]. Results show that not all data conform to normal distribution and homogeneity of variance. Also, because of the presence of outliers, to ensure robustness, a non-parametric Kruskal-Wallis one-way ANOVA test [266] with 95% confidence intervals [275] was run to determine whether there are significant differences between the different reverberation times for the average impact thresholds for each parameter.

Table 3.4 presents the Kruskal-Wallis one-way ANOVA test results of three re-

verberation times for the following four parameters: RER removal, FER removal, ITDG extension and LR removal. For RER removal, FER removal and ITDG extension, the p-values are bigger than the significance level 0.05, so there are no significant differences between three reverberation times. This means that reverberation time will not affect the final average threshold of RER removal, FER removal and ITDG extension. But for LR removal, the p-value is less than the significance level 0.05, so there are significant differences between three reverberation times. It means that reverberation time will affect the final average threshold of LR removal.

Table 3.4: The Kruskal-Wallis ANOVA test results of three reverberation times for RER removal, FER removal, ITDG extension and LR removal.

DF = 2 **Significance Level = 0.05**	***p* value**
RER Removal	0.390
FER Removal	0.185
ITDG Extension	0.177
LR Removal	$*< 0.001$

Note 1 Values marked with * indicate significant differences.

In statistical analysis, a post-hoc test is conducted after an initial test (like an ANOVA) to explore specific differences between groups when the test indicates significant differences overall. Therefore, after conducting a Kruskal-Wallis one-way ANOVA test on LR removal thresholds, further analysis was needed to identify specific group differences. This was achieved through pairwise comparisons, which assess significance between each pair of groups. The results, displayed in Table 3.5, list the p-values from these tests. A p-value below 0.05 indicates a statistically significant difference, affirming that the threshold levels for LR removal significantly different in reverberation times.

Table 3.5: The post-hoc test results of three reverberation times for LR removal.

LR removal	0.91s	1.51s
0.31s	$*< 0.001$	$*< 0.001$
0.91s		$*< 0.001$

Note 1 Values marked with * indicate significant differences.

3.2.3 Analysis of The Threshold Distribution for Each Parameter Type

The data from the histograms Figures 3.7 and 3.8 demonstrate the distribution of thresholds for RER and FER removal across different reverberation times, as reported by the 20 participants. For RER removal displayed in Figure 3.7, a significant majority of the thresholds fall within the 0-20 ms range. Specifically, 65% participants reported thresholds within this range for reverberation time of 0.31 s and 0.91 s, respectively, and 60% of thresholds fall within this range for reverberation time of 1.51 s. More precisely, 40% identified their thresholds within the 10-20 ms interval at 0.31 s reverberation time, while 55% of participants detected changes at this range for reverberation time of 0.91 s and 1.51 s, respectively, indicating a high sensitivity to small changes in RERs. Notably, for a 0.31 s reverberation time, 25% participants reported thresholds ranging from 0-10 ms. This suggests that participants can detect very subtle differences when RERs are slightly altered. This trend shows a general sensitivity to the removal of RERs, especially in settings with shorter reverberation times.

Figure 3.8 shows that the threshold distribution of FER removal is concentrated in 20-50 ms. Specifically, 45% of participants detected changes at 20-30 ms for reverberation time of 0.31 s, and 35% and 30% of thresholds fall within this range for reverberation time of 0.91 s and 1.51 s, respectively. Among all

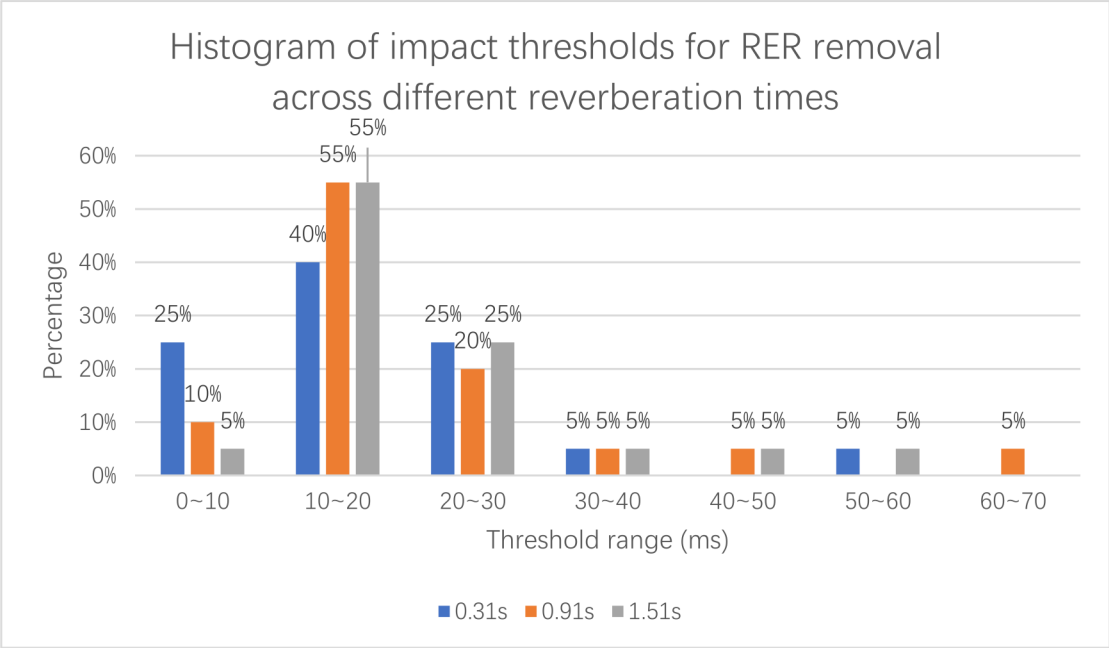


Figure 3.7: Histogram displaying the percentage distribution of threshold ranges for RER removal on perceptual reverberation. Data is categorised by three reverberation times: 0.31 s, 0.91 s, and 1.51 s. Each bar represents the proportion of participants whose thresholds fell within specified ranges, highlighting the sensitivity to changes in RER at different reverberation times.

reverberation times, the percentage of threshold distribution between 30-40 ms has slightly decreased, but remains relatively high. At 0.91 s and 1.51 s reverberation times, the threshold distribution between 40-50 ms is still relatively high (25% and 35%, respectively). The concentration of distribution in the 20-50 ms indicates that FER removal within this range is more perceptible to participants, showing moderate sensitivity to such changes. Very few participants (only no more than 15% across all reverberation times) can perceive changes when FER removal is less than 20 ms, indicating a low sensitivity to small changes in this parameter. Comparison of the distribution of thresholds shows that the effect of FER removal on perceptual reverberation is weaker than RER removal.

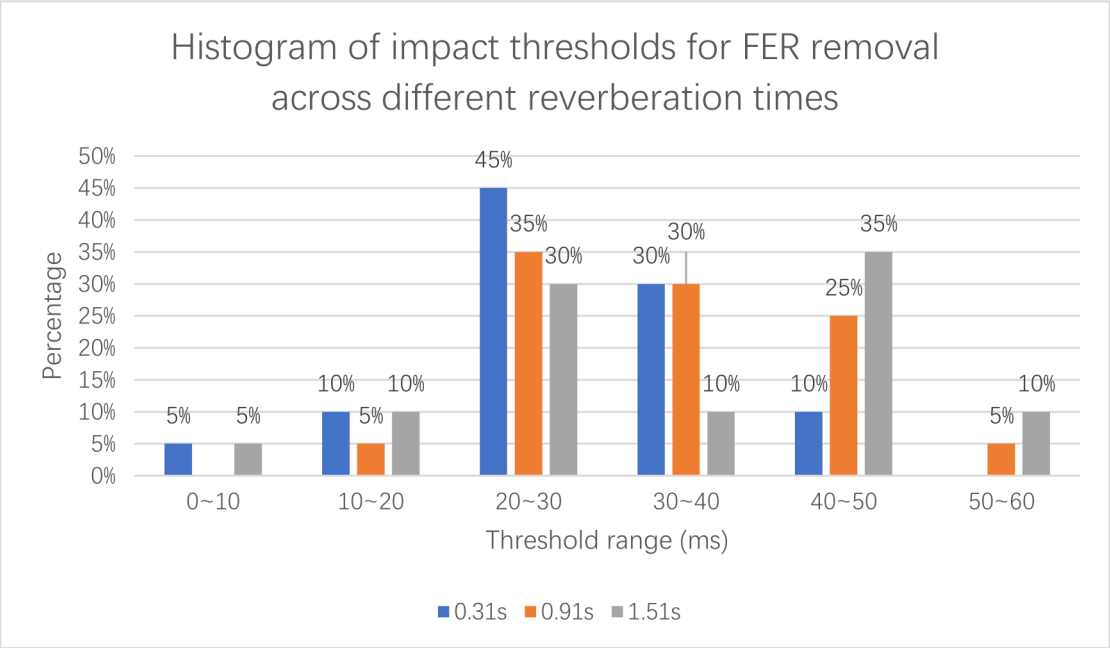


Figure 3.8: Histogram displaying the percentage distribution of threshold ranges for FER removal on perceptual reverberation. Data is categorised by three reverberation times: 0.31 s, 0.91 s, and 1.51 s. Each bar represents the proportion of participants whose thresholds fell within specified ranges, highlighting the sensitivity to changes in FER at different reverberation times.

The data depicted in Figure 3.9 illustrates the threshold distributions for ITDG extension at different reverberation times. At 0.31 s, the majority of thresholds are focused within the 0-20 ms range, indicating that participants are sensitive to even minor extensions in ITDG. For longer reverberation times of 0.91 s and 1.51 s, thresholds are more uniformly distributed across the 0-50 ms range. Despite this wider distribution, a slight extension of the ITDG is still perceptible to participants. This demonstrates a general ability among listeners to detect changes in ITDG. These results suggest that perceptual sensitivity to ITDG changes is more acute with shorter reverberation times and becomes more dispersed as the reverberation time lengthens. This might be due to the increasing complexity of the

auditory scene, where longer reverberation times potentially mask the perceptual cues that participants rely on to detect differences in ITDG.

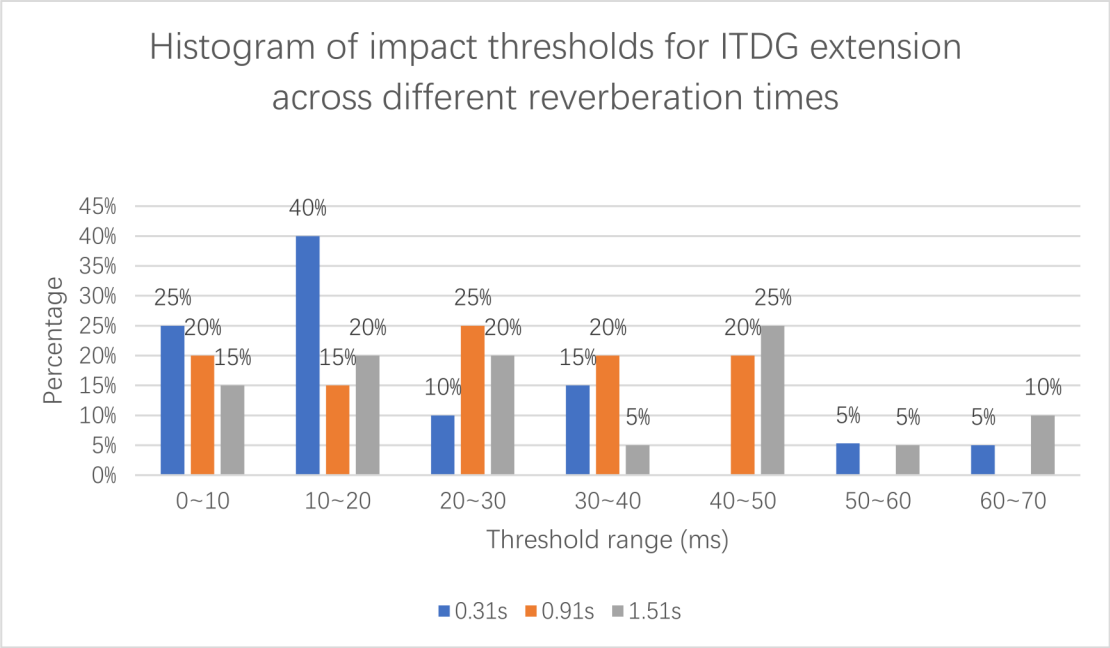


Figure 3.9: Histogram displaying the percentage distribution of threshold ranges for ITDG extension on perceptual reverberation. Data is categorised by three reverberation times: 0.31 s, 0.91 s, and 1.51 s. Each bar represents the proportion of participants whose thresholds fell within specified ranges, highlighting the sensitivity to changes in ITDG at different reverberation times.

Finally, reverberation time of BRIRs can influence the time of LR removal, so the ranges of parameter change are relatively large. As Figure 3.10 shows, for short reverberation time, the threshold distribution concentrates in 400-500 ms. At 0.91 s reverberation time, most impact thresholds distribute 750-850 ms, and most thresholds concentrate in 1250-1350 ms for 1.51 s reverberation time. As shown in Figures 3.11-3.13, the grey impulse responses represent the original impulse responses, while the red impulse responses illustrate the effect after LR has been removed. The difference between the two is the removed LR. The LR

time length values are calculated based on the late reverberation segment of the BRIRs through the following formula:

$$\text{LR\_time\_length} = (\text{Total\_BRIR\_length} - (\text{Direct\_sound\_length} + 0.08 \times \text{Fs})) \times \frac{1000}{\text{Fs}}, \tag{3.1}$$

where LR_time_length represents the time duration of the late reverberation, calculated by dividing the number of samples representing the late reverberation by the sampling rate (Fs) and multiplying by 1000 to convert it to milliseconds. $\text{Total}_{\text{B}}\text{RIR}_{\text{l}}\text{ength}$ refers to the total number of samples in the BRIR, and $\text{Direct}_{\text{s}}\text{ound}_{\text{l}}\text{ength}$ is the number of samples from the start of the impulse response to the end of the direct sound. A buffer of 0.08 (unit in seconds) accounts for time of early reflections.

Through calculation, the time length of the LR is 465 ms for the BRIR with 0.31 s reverberation time, 1329 ms for the BRIR with 0.91 s reverberation time and 1920 ms for the BRIR with 1.51 s reverberation time. A distinction is made here between the length of the LR and the reverberation time to avoid confusion as to why the length of the LR is greater than the reverberation time. By comparing the thresholds for LR removal with the total length of LR, it can be seen that perceptual reverberation is significantly affected only when more than 86% of the LR is removed for 0.31 s reverberation time, 56% for 0.91 s reverberation time, and 65% for 1.51 s reverberation time. This suggests that it is difficult to distinguish differences in perceptual reverberation caused by LR removal unless a substantial portion of the LR is removed. Consequently, slight LR removal does not seriously affect perceptual reverberation.

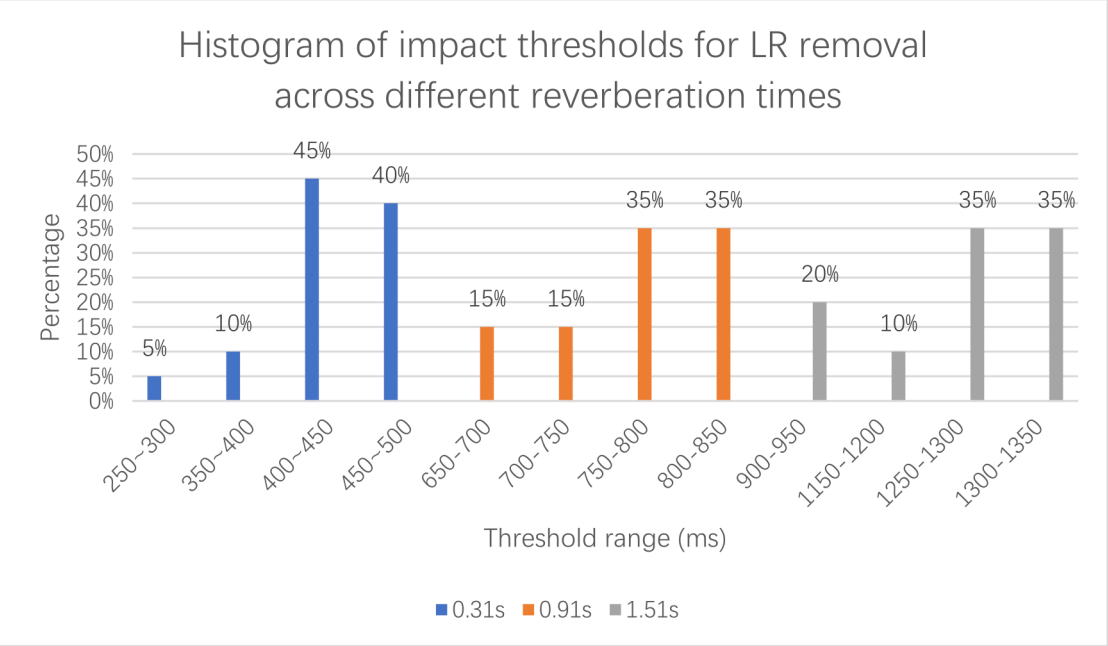


Figure 3.10: Histogram displaying the percentage distribution of threshold ranges for LR removal on perceptual reverberation. Data is categorised by three reverberation times: 0.31 s, 0.91 s, and 1.51 s. Each bar represents the proportion of participants whose thresholds fell within specified ranges, highlighting the sensitivity to changes in LR at different reverberation times.

3.3 Discussion

By reviewing Table 3.2 and 3.3, the maximum thresholds of LR removal at a long reverberation time appear most frequently. Further, some maximum thresholds of LR removal appear at a medium reverberation time. Although the maximum LR removal threshold does not occur at a short reverberation time, the average LR removal threshold for the short reverberation time is close to its LR length. These results suggest that people's sensitivity to LR removal is low, meaning that the perceptual impact of LR removal becomes significant only when a large portion of LR is removed, indicating that listeners are relatively insensitive

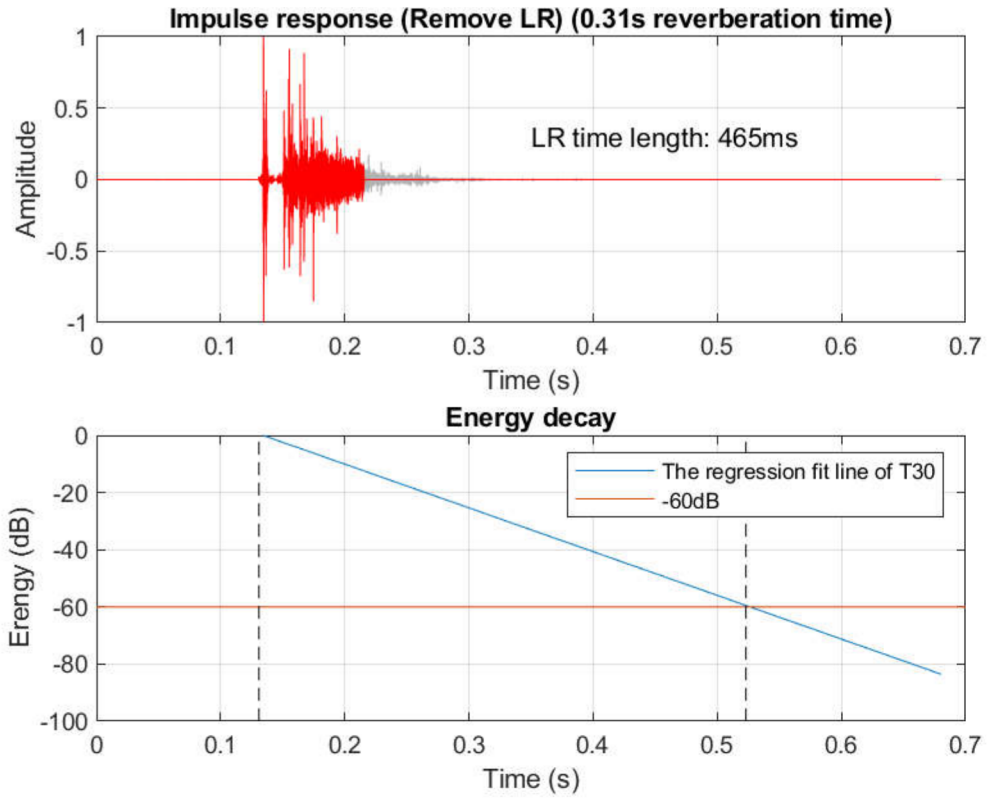


Figure 3.11: The time length of the LR for the BRIR with 0.31 s reverberation time.

to small changes in LR. However, some minimum thresholds of LR removal also appear in the long reverberation time, so this does not rule out the fact that for some people with extremely sensitive hearing, they can still distinguish LR removal clearly.

Through analysis, RER removal and ITDG extension have the most influence on perceptual reverberation of the parameters assessed. ERs improve speech intelligibility by increasing the loudness of the direct sound [276]. Table 3.6 shows a comparison between the energy corresponding to the reference BRIR versus the BRIR with early reflections removed at different reverberation times. The energy values were computed through integrated Loudness according to EBU R 128 stan-

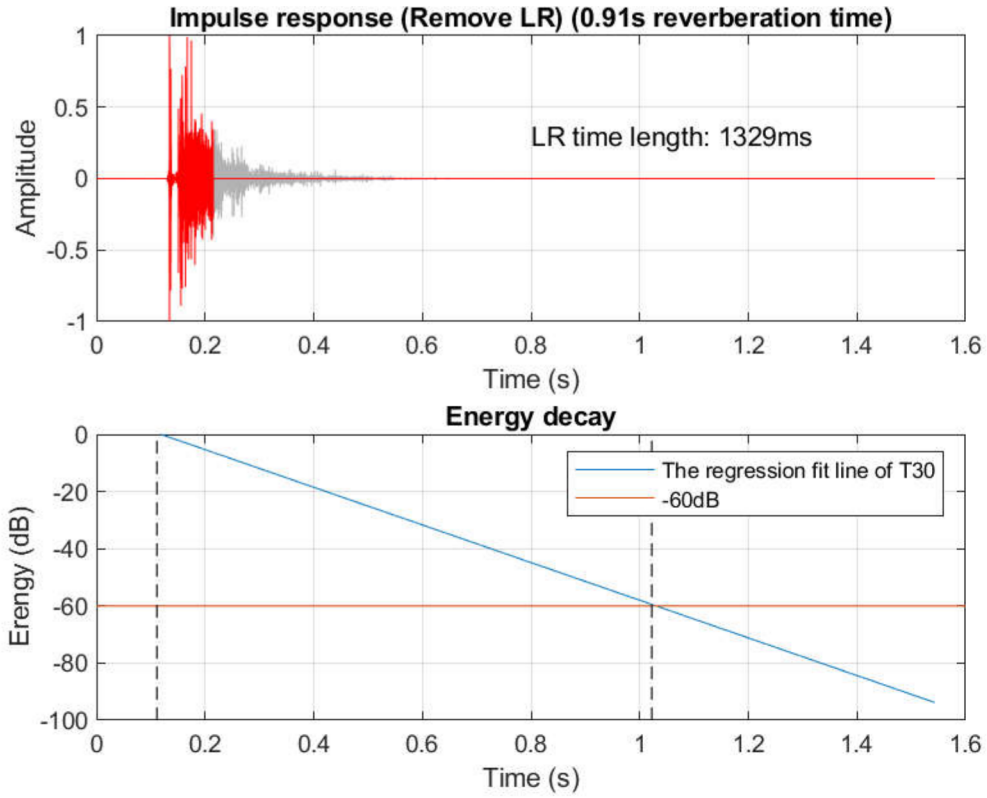


Figure 3.12: The time length of the LR for the BRIR with 0.91 s reverberation time.

dards [277] (measured in Loudness Units relative to Full Scale (LUFS)) and the ITU-R BS.1770-4 [278] (still refers to Loudness, K-weighted, relative to Full Scale (LKFS), but LKFS and LUFS are equivalent). ITU-R BS.1770-4 [278] also defines that the LKFS unit is equivalent to a decibel scale in that an increase in the level of a signal by 1 dB will cause the loudness reading to increase by 1 LKFS. It can be seen from Table 3.6 that for FER removal, the energy reduction is 2.0858 LUFS at 0.31 s, 3.0164 LUFS at 0.91 s, and 1.589 LUFS at 1.51 s reverberation times. For RER removal, the energy reduction is -1.344 LUFS at 0.31 s, indicating an increase in energy, and reductions of 0.7576 LUFS at 0.91 s, and 0.7352 LUFS at 1.51 s reverberation times. These findings indicate that the energy of the BRIR

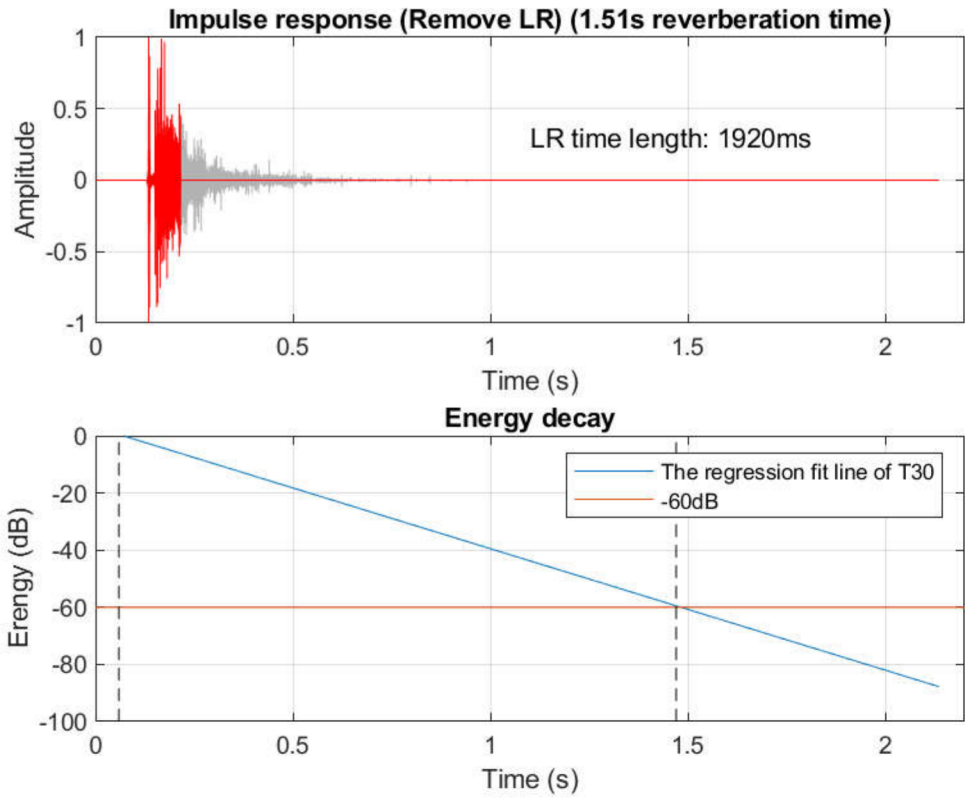


Figure 3.13: The time length of the LR for the BRIR with 1.51 s reverberation time.

is reduced by approximately 0.7 to 3 LUFS when the FER or RER is removed at the corresponding threshold. The exception noted is that the energy of the RER removed increases by 1.344 LUFS at 0.31 s reverberation time. It is pointed that listeners can distinguish a change in sound level of about 1 dB in their most sensitive sound level range (approximately 35 to 80 dB SPL) [279]. Therefore, the perceptible differences when early reflections are removed may be attributed to these SPL changes.

For a speech signal, the ERs are limited to around 50 ms. For RER removal under different reverberation times, their average resolution thresholds are from 17 ms to 22 ms. This means that when ERs are reversely removed 34% to 44%,

Table 3.6: Comparison between the energy corresponding to the reference BRIR versus the BRIRs with early reflections removed at different reverberation times (the amount of early reflection removal corresponds to the measured threshold in this thesis).

Reverb Time	**0.31s**	**0.91s**	**1.51s**
		Energy (LUFS)	
Reference BRIR	−22.4044	−19.436	−19.3806
BRIR with FER removal	−24.4902	−22.4524	−20.9696
Reference−FER removal	2.0858	3.0164	1.589
BRIR with RER removal	−21.0604	−20.1936	−20.1158
Reference−RER removal	−1.344	0.7576	0.7352

people can perceive the change of the reverberation. For FER removal under different reverberation times, their resolution thresholds are from 27 ms to 35 ms. When ERs are forward removed over 50% to about 70%, people can perceive the change of the reverberation. Compared with FER removal, RER removal should be a greater consideration when designing reverberation algorithms. When ITDG is extended by about 20-31 ms, the difference of reverberation can be perceived clearly. To contextualise, research shows that ITDG is typically in the range of 20 ms to 60 ms in a large concert hall and from about 8 ms to 27 ms in chamber music halls [177]. ITDG extension thresholds of 21-30 ms is enough to affect the perceptual reverberation, so ITDG should be another major consideration in designing artificial reverberation algorithms. For LR, the average thresholds measured at different reverberation times are 427 ms, 771 ms and 1211 ms, respectively. Before a difference was perceived, the LR was more than halved at all three reverberation times, and even over 90% removal for the short reverberation time, suggesting that people are less sensitive to the reverberation difference of male speech signals caused by LR changes. Therefore, to a certain extent, in the design of reverbera-

tion algorithms, it is not necessary to give priority to some small changes in the later reverberation, if these changes do not have major impact on the perception of reverberation overall. Theoretically, expert listeners and those experienced with acoustic experiments can more accurately distinguish the impact of BRIR parameters on perceptual reverberation. Therefore, with the test data provided by them, the calculated average threshold should be lower than the generic average threshold of the public, so their average threshold should represent the generic threshold.

Overall, comparison of the above four BRIR parameter types reveals that RER and ITDG require most attention in the design of artificial binaural reverberation algorithms, while slight changes in FER and LR are less critical.

However, because of the limitation of test time and hearing fatigue, this experiment just used male speech as the test signal. Actually, these parameters may have different effects on perceptual reverberation of different audio signals. Therefore, further experiments should test a variety of different noise and musical stimuli to further test the influence of BRIR parameters on perceptual reverberation. Also, this experiment is a static binaural reverberation parameter test rather than a dynamic one, so further experiments should also include head tracked conditions.

In this experiment, BRIRs modeled in ODEON were employed to investigate perceptual thresholds related to early reflections and reverberation. The geometric-based simulation capabilities of ODEON allowed for precise control and manipulation of some acoustic parameters, effectively meeting the experimental requirements. This level of control was essential, as it enabled the systematic variation of key parameters, such as reverberation time, to generate the necessary BRIRs. The controlled environment offered by ODEON ensured the isolation and examination of specific aspects of the acoustic experience, free from the variability

inherent in real-world measurements. However, it's important to acknowledge the limitations of using ODEON. As a geometric-based simulation tool, ODEON excels in rendering high-frequency sounds where sound can be approximated as rays, but it struggles with low-frequency sound waves due to its inherent limitations in modelling wave-based phenomena like diffraction. This means that ODEON might not accurately represent the full acoustic complexity of a space, particularly in the lower frequency ranges. To overcome these limitations, a hybrid approach or a fully wave-based approach might be necessary for simulations that require more accurate low-frequency rendering. Such methods would better capture the wave nature of sound and provide a more comprehensive simulation, particularly in scenarios where low-frequency accuracy is critical.

Furthermore, due to the COVID-19 pandemic environment, some experimental parameters were not controlled across subjects. This listening test was conducted online, so the listening environment and the volume of audio samples were dependent on the preference of the test subject. This should be avoided in the future experiments.

3.4 Summary

The purpose of this experiment is to find out the average perceptual thresholds of four BRIR parameters to represent the generic thresholds on perceptual reverberation, and determine the potential impact of these parameters on perceptual reverberation. By understanding which parameters have the greatest impact on perceptual reverberation, these parameters can be prioritised when designing artificial reverberation algorithms, potentially leading to the design of more realistic and efficient reverberation models. This research makes the following conclusion

through the measurement and the analysis for RER, ITDG, FER and LR:

- The average thresholds for RER removal are 17.59 ms, 21.17 ms, and 27.71 ms for reverberation times of 0.31 s, 0.91 s, and 1.51 s, respectively. Similarly, the average thresholds for ITDG extension are 20.705 ms, 25.21 ms, and 30.75 ms for the same reverberation times. These lower thresholds indicate high listener sensitivity, meaning small changes in these parameters are easily perceived by listeners. Therefore, these parameters should be given priority consideration in design of artificial reverberation algorithms to maintain realistic reverberation.

- The average thresholds of FER removal are 27.68 ms, 34.325 ms and 34.595 ms for reverberation times of 0.31 s, 0.91 s, and 1.51 s, respectively. Higher thresholds suggest lower sensitivity to changes. FER removal thresholds are higher compared to RER and ITDG, suggesting that listeners are less sensitive to changes in FER. FER can be less emphasized in algorithm design without significantly impacting perceptual reverberation.

- The average thresholds of LR removal are 427.155 ms, 771.16 ms and 1211.52 ms for reverberation times of 0.31 s, 0.91 s, and 1.51 s, respectively. Very high thresholds indicate that a substantial amount of LR needs to be removed before listeners notice a difference. This shows a low sensitivity to changes in LR. Small changes in LR are less critical, allowing more flexibility in algorithm design for computational efficiency without significantly affecting perceptual reverberation.

- The statistical test shows that reverberation time does not affect the thresholds of RER removal, ITDG extension and FER removal on perceptual re-

verberation, but it does impact the thresholds of LR removal. This suggests a consistent sensitivity for RER, ITDG, and FER across different reverberation times, while LR adjustments can be wider for longer reverberation times.

Early reflections are achieved by finite impulse response (FIR) delay lines and late reverberation by infinite impulse response (IIR) filters. Whilst focusing on hybrid FIR and IIR filters would certainly theoretically lead to more accuracy, exploiting perceptual sensitivity to reverberation could reduce computational resources for algorithm design. Based on the experimental findings, it appears that LR removal has a smaller effect on perceptual reverberation, and although FER does not affect perceptual reverberation as much as RER, overall ER can have an obvious effect on perceptual reverberation. ITDG can also have a significant effect on perceptual reverberation. In order to balance reverberation accuracy and algorithmic efficiency, perceptually motivated reverberation algorithms should focus on ERs and ITDG. When designing reverberation algorithms, early reflections are achieved by means of FIR delay lines. More FIRs require greater computational cost, so the number of FIRs can be minimised, depending on the measured threshold, without affecting reverberation perception. ITDG has a large impact on perceptual reverberation, so controlling the time period between direct sound and early reflections as accurately as possible when designing a reverberation algorithm can improve efficiency in the design. Future designs of artificial reverberation algorithms can take this conclusion into account to optimise the balance between perceptual accuracy and computational efficiency.

Chapter 4 proposes a new hybrid artificial reverberation algorithm based on the Schroeder and Moorer reverberation algorithms according to the conclusions

of this chapter, and evaluates the similarity of this algorithm and six other well-known reverberation algorithms to real measured reverberation perception.

4

New Reverberation Algorithm and Evaluation

This chapter proposes a binaural artificial reverberation algorithm based on the findings of Chapter 3. it evaluates the perceptual similarity of the proposed algorithm and other well-known reverberation algorithms compared to real reverberation and analyses their computational cost. Additionally, this chapter compares their key parameters such as reverberation time, direct-to-reverberant (D/R) energy ratio, interaural cross-correlation coefficient (IACC) and timbre, as these parameters are fundamental factors that influence the perception of reverberation in virtual environments.

The core objective of this research is the computationally efficient reproduction of plausible real-world perceptual reverberations in virtual acoustic rendering. As discussed in Section 2.3.4, traditional reverberation algorithms, based on geometric or wave-based models, have shown efficacy but also limitations, particularly in their ability to adapt dynamically to varied acoustic settings and maintain computational simplicity in real-time applications [135, 137]. Recursive reverberation algorithms based on delay network structures offer flexible parametric characteristics and better real-time performance compared to geometric acoustic methods, which cannot model wave properties, and computationally expensive wave-based methods [9, 135, 142]. Therefore, the artificial reverberation algorithms based on feedback delay networks are more suitable for achieving the research aims of this

thesis. Consequently, this thesis evaluates six artificial reverberation algorithms based on delay networks mentioned in Section 2.4.

These algorithms were chosen for several key reasons. As one of the pioneers in artificial reverberation, Schroeder's algorithm laid the groundwork for digital reverb techniques. His work is fundamental to understanding the evolution of reverb algorithms [178]. Moorer's contributions built upon Schroeder's work, introducing important improvements in digital reverb and offering a practical implementation that has been widely adopted in both academic research and practical applications [175]. Gardner's virtual acoustic algorithm added significant advancements in simulating complex acoustic environments [171]. Jot's algorithm provided crucial developments in digital reverb design, emphasising perceptual and spatial aspects, which are critical for creating immersive audio experiences [15]. Dattorro's work introduced a highly flexible and efficient framework for digital reverb, offering detailed control over reverb parameters, which is essential for fine-tuning reverb effects in various acoustic settings [179]. Alary's directional reverb algorithm represents a modern advancement, incorporating directionality in reverb processing, which is particularly relevant for applications requiring precise spatial audio rendering [19]. By selecting these six algorithms, the research aims to provide a comprehensive evaluation of traditional reverb techniques, covering a broad spectrum of historical and modern approaches in digital reverberation.

Although the six feedback network-based digital reverberation algorithms continue to evolve, they still face limitations in enhancing user perceptual experiences and ensuring computational efficiency. The author simulated the six reverberation algorithms mentioned above in the MATLAB environment to preliminarily evaluate their perceptual effect and computational efficiency. The results showed

that the Schroeder algorithm often produces metallic sounds and has insufficient echo density at longer reverberation times. While the Moorer algorithm addresses some metallic qualities of Schroeder's method by incorporating a low-pass filter to simulate air absorption, it still tends to produce metallic sounds with extended reverberation times. The Gardner algorithm offers enhanced user perceptual experiences through adjustable reverberation times and frequency damping. However, its performance is notably better for larger rooms, with less satisfactory results for smaller and medium-sized spaces. Despite theoretical control over attenuation characteristics and the ability to simulate air absorption, the FDN do not consistently deliver superior perceptual outcomes. Their complex feedback matrices also demand substantial computational resources, limiting their utility in real-time applications. The Dattorro algorithm achieves high echo density and can simulate diverse reverberation characteristics through its intricate tank structure. Nevertheless, its perceptual quality often falls short, and its intensive computational demands hinder its use in dynamic and mobile applications. The DFDN extends the FDN capabilities to produce directional reverberation times. While innovative, it still falls short in delivering satisfying reverberation effects, and the necessity to encode and process multi-channel signals significantly increases its computational complexity.

These existing algorithms are the foundation of digital reverberation technology and have made outstanding contributions to the development of reverberation, but often fail to meet the dual requirements of enhancing user perceptual experience and ensuring computational efficiency in dynamic virtual environments. They are primarily designed for static acoustic settings and do not adapt dynamically to user interactions or environmental changes. Furthermore, the computational intensity

of more complex models like the FDN and the Dattorro algorithms restricts their deployment in real-time applications on portable and low-power devices, which is a key aspect of modern virtual acoustic experience. These challenges highlight the need for new algorithms that can not only simulate natural reverberation more faithfully, but also run efficiently under the limitations of real-time processing.

In response, the author presents a new binaural reverberation algorithm based on the Schroeder and Moorer reverberation algorithms. Its linear structure is simpler than the topology of the Dattorro reverberation algorithm and the feedback matrix structure of the FDN and DFDN. Additionally, it is further enhanced by the adaptation of the traditional Schroeder and Moorer reverberation algorithms to more accurately simulate realistic reverberation. The motivation for developing this new algorithm is driven by the essential need to improve the immersive audio experience in virtual environments, by providing realistic and dynamically adaptable reverberation that enhances user perceptual experience while ensuring computational efficiency. Therefore, it is necessary to clarify that by evaluating the impact of room impulse response parameters on perceptual reverberation and the realisation of the manipulability of these parameters, can we create a perceptually plausible and efficient digital reverberator that sounds as good as real-world acoustic measurements Building on the foundational understanding of critical BRIR parameters and their perceptual thresholds from previous research, this development focuses on optimising these parameters to enhance realism efficiently.

This chapter will detail the algorithm's design and its evaluation methodology, contrasting its performance with both well-known reverberation algorithms and measured real-world reverberations to verify its effectiveness and practical applicability. The detailed structures of the first six reverberation algorithms were

described in Section 2.4, and their parameter configurations for the implementation of this study will be described in Section 4.1. The proposed new binaural reverberation algorithm by the author based on the Schroeder and Moorer reverberation algorithms will be described in Section 4.2. The similarities between their reverberation effects and real binaural reverberation were compared for audio stimuli for four different contexts: male speech, female singing, solo cello music and drum beat. A perceptual study was conducted to evaluate which of these reverberation algorithms most closely matched the real reverberation in each condition, which would suggest it most suitable for implementation in virtual rendering. In addition, the key parameters of the BRIRs generated by these algorithms were compared with those of the measured BRIRs at the same reverberation time settings to see how well they matched. The parameters involved include reverberation time, D/R energy ratio, IACC and timbre.

4.1 Implementation Details of Traditional Reverberation Algorithms

The detailed structure of the six traditional feedback-delay-network-based reverberation algorithms has been described in Section 2.4, and their detailed parameter configurations for the implementation of this study are fully described below.

The Schroeder reverberation algorithm is implemented through a user-oriented graphical user interface (GUI) in this thesis, where the reverberation time (RT_{60}) is set as a user-adjustable parameter (range 0.1 - 5 s). The user can input the desired reverberation time to produce the desired reverberation effect. All filter delays (T_d)

are set according to Schroeder's recommendations [141, 178], and filter gains (g) are calculated using Equation 2.30 [178]. The gains of the two all-pass filters are calculated from the parameters set in Table 4.1, and are exactly 0.7 as suggested by Schroeder [141, 178]. The gain of the four comb filters is calculated based on the user input reverberation time, which is equal to the overall reverberation time of the desired impulse response. The parameter values for each filter are listed in Table 4.1 [280].

Table 4.1: The parameter values of the Schroeder reverberation algorithm [280].

Filter	**Delay (s)**	**Reverb time (s)**
COMB1	0.0297	RT_{60}
COMB2	0.0371	RT_{60}
COMB3	0.0411	RT_{60}
COMB4	0.0437	RT_{60}
ALLPASS1	0.005	0.09683
ALLPASS2	0.0017	0.03292

Note 1 RT_{60} is used to calculated the gain of comb filters and is a user-entered parameter (range 0.1–5 s).

In this thesis, for Moorer reverberation algorithm, some adjustments have been made to the parameters of the original LPF comb filter. As in the Schroeder reverberation algorithm, the filter gains (g) are calculated from the delay time (T_d) and the reverberation time (RT_{60}) as in Equation 2.30. The parameter values of the FIR delay line are presented in Table 4.2 [175]. The low-pass-feedback gain values for all low-pass-feedback comb filters are 0.9, and the values of the late reverberation filters are shown in Table 4.3 [175].

In this study's implementation of the Gardner reverberation algorithm, each filter delay length is adjusted to create the correct reverberation time. The corresponding parameter values for each filter of Gardner reverberator are presented in Table 4.4 [171] for small, medium and large room reverberation structures.

Table 4.2: The parameter values of the early FIR of the Moorer reverberation algorithm [175].

Tap	Delay (s)	Gain
1	0.0043	0.841
2	0.0215	0.504
3	0.0225	0.491
4	0.0268	0.379
5	0.0270	0.380
6	0.0298	0.346
7	0.0458	0.289
8	0.0485	0.272
9	0.0572	0.192
10	0.0587	0.193
11	0.0595	0.217
12	0.0612	0.181
13	0.0707	0.180
14	0.0708	0.181
15	0.0726	0.176
16	0.0741	0.142
17	0.0753	0.167
18	0.0797	0.134

Table 4.3: The parameter values of the late reverberation filters of the Moorer reverberation algorithm [175].

Filter	Delay (s)	Reverb time (s)
LPF COMB1	0.04	RT_{60}
LPF COMB2	0.041	RT_{60}
LPF COMB3	0.043	RT_{60}
LPF COMB4	0.055	RT_{60}
LPF COMB5	0.059	RT_{60}
LPF COMB6	0.061	RT_{60}
ALLPASS1	0.007	0.09683
DELAY	0.0017	-

Note 1 RT_{60} is used to calculated the gain of low-pass-feedback comb filters and is a user-entered parameter (range 0.1–5 s).

Table 4.4: The parameter values of the the late reverberation filters of the Gardner reverberation algorithm [171].

Gardner reverberation	Filter	Delay (ms)	Delay Length	Gain
Small-size-room	Lowpass1	-	-	0.6
	Delay	24	24×8	-
	DNA_O	35	35×8	0.3
	DNA_I1	22	22×8	0.4
	DNA_I2	8.3	8.3×8	0.6
	SNA_O	66	66×8	0.1
	SNA_I	30	30×8	0.4
	Lowpass2	-	-	1/4200
	GAIN	-	-	0.5
Medium-size-room	DNA_O	35	35×15	0.3
	DNA_I1	8.3	8.3×15	0.7
	DNA_I2	22	22×15	0.5
	Delay1	5	5×15	-
	Allpass1	30	30×15	0.5
	Delay2	67	67×15	-
	Delay3	15	15×15	-
	SNA_O	39	39×15	0.3
	SNA_I	9.8	9.8×15	0.6
	Delay4	108	108×15	-
	Lowpass	-	-	1/2500
	GAIN	-	-	0.5
Large-size-room	Lowpass1	-	-	0.6
	Allpass1	8	8×15	0.3
	Allpass2	12	12×15	0.3
	Delay1	4	4×15	-
	Delay2	17	17×15	-
	SNA_O	87	87×15	0.5
	SNA_I	62	62×15	0.25
	Delay3	31	31×15	-
	Delay4	3	3×15	-
	DNA_O	120	120×15	0.5
	DNA_I1	76	76×15	0.25
	DNA_I2	30	30×15	0.25
	Lowpass2	-	-	1/2600
	GAIN	-	-	0.34

Note 1 The DNA presents the double nested all-pass filter, the SNA presents the signal nested all-pass filter, and O presents the outer and I presents the inner. GAIN present all gain filters in each structure.

In this thesis, the author adopts four-dimensional delay lines for FDN. In this structure, following Schroeder's advice [178], the delay length (m_i) of each delay line can be chosen as an integer power of a distinct prime number pri_i:

$$m_i = \text{pri}_i^{\text{ppwr}_i}, \tag{4.1}$$

where ppwr_i can be called the 'multiplicity' of the prime pri_i. Therefore, m_i can be derived from Equations 4.2 to 4.4.

The total minimum delay length M_{min} is defined as

$$M_{\text{min}} = \lceil 0.15 \times \text{Fs} \times \text{RT}_{60} \rceil, \tag{4.2}$$

where Fs is the sample rate and RT_{60} is the reverberation time. $\lceil\rceil$ is the ceil function, indicating rounding up to the nearest integer. This is to ensure sufficient mode density in the frequency domain according to Schroeder's suggestion, a mode density of 0.15 modes per Hz is adequate for a reverberation time of 1 s [15].

d_i is the preliminary estimate of the non-prime-number delay length for each delay line, and is expressed as

$$d_i = \lceil M_{\text{min}}/N + (i-1) \times M_{\text{min}}/(N \times 5) \rceil \tag{4.3}$$

The 'multiplicity' ppwr_i of the prime numbers pri_i can be derived from

$$\text{ppwr}_i = \left[\frac{\ln(d_i)}{\ln(\text{pri}_i)} \right], \tag{4.4}$$

where prime numbers $\text{pri}_i) \in \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, \dots\}$, [] is used to represent the rounding function, indicating rounding to the nearest integer.

The delay line length is then can be obtained from Equation 4.1.

The gain (g_i) for each delay line is calculated from a given reverberation time (RT_{60}) and sample rate (Fs) by Equations 4.5 to 4.7 according to Alary's [19] suggestion. g_{dB} is a per-sample attenuation gain, g_{lin} is a linear scale gain converted from a logarithmic scale gain, and g_i is the target gain of each delay line, given respectively as

$$g_{\mathrm{dB}} = \frac{-60}{\mathrm{RT}_{60} \times \mathrm{Fs}} \tag{4.5}$$

$$g_{\mathrm{lin}} = 10^{\frac{g_{\mathrm{dB}}}{20}} \tag{4.6}$$

$$g_{\mathrm{i}} = (g_{\mathrm{lin}})^{m_i} \tag{4.7}$$

According to Alary's study [19], to build the lossless prototype, the recirculating matrix of an FDN is constrained to be orthogonal. The orthogonality of Hadamard matrices is a crucial property for recirculating matrices in FDNs to ensure energy conservation and stable echo decay to help maintain the losslessness of the system, ensuring that the total signal energy is preserved over time. The feedback matrix (A) is implemented by a fourth-order Hadamard matrix:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{4.8}$$

A fourth-order Hadamard matrix is a specific type of Hadamard matrix that is 4x4 in size. The rows of the matrix are mutually orthogonal, and each element in the matrix is either +1 or -1 This means that the dot product of any pair of different rows (or columns, since Hadamard matrices are also symmetric) is zero.

The input gain (b_i), output gain (c_i), and direct sound gain (d) are all set as 1 in this structure, and can be adjusted as required.

Dattorro has proposed some default values for the parameters of the filtering structures, output delays and signs [141] in his reverberation algorithm. However, the decay parameter he proposed for controlling the reverberation time is a fixed value and the parameter is not directly related to the physical characteristics of the reverberation, so this parameter is recalculated in this thesis to ensure tunability of the reverberation time. The decay is derived from a given reverberation time (RT_{60}) and sample rate (Fs) by Equations 4.9 to 4.12:

$$m = 0.15 \times \mathrm{Fs} \times \mathrm{RT}_{60} \tag{4.9}$$

$$\mathrm{decay}_{\mathrm{dB}} = \frac{-60}{\mathrm{RT}_{60} \times \mathrm{Fs}} \tag{4.10}$$

$$\mathrm{decay}_{\mathrm{lin}} = 10^{\frac{\mathrm{decay}_{\mathrm{dB}}}{20}} \tag{4.11}$$

$$\mathrm{decay} = (\mathrm{decay}_{\mathrm{lin}})^m \tag{4.12}$$

The delay length is marked in Figure 2.32 [141], the value of the pre-delay is 0.001 s, and other parameters used in this structure (See Figure 2.32) are presented in Table 4.5 [141]. Also, the delays and signs used to generate the output of Dattoro's reverberator are given in Table 4.6.

Alary [19] adopted a four-delay-line DFDN using third-order Ambisonics in his DFDN reverberator, so his reverberator has four groups of sixteen delay lines, making a total of 64 delay lines. Therefore, the number of delay lines increases by a factor of 16 when compared to an equivalent conventional FDN. Considering

Table 4.5: The parameters of the Dattorro reverberation algorithm [141].

Filter	Gain name	Gain
LOWPASS 1	bandwidth	$18,000/\mathrm{Fs}$
ALLPASS 1	input_diffusion 1	0.75×0.8
ALLPASS 2	input_diffusion 1	0.75×0.8
ALLPASS 3	input_diffusion 2	0.625×0.8
ALLPASS 4	input_diffusion 2	0.625×0.8
ALLPASS 5	decay_diffusion 1	0.7×0.8
ALLPASS 5'	decay_diffusion 1	0.7×0.8
LOWPASS 2	damping	0.7
LOWPASS 2'	damping	0.7
ALLPASS 6	decay_diffusion 2	0.5×0.8
ALLPASS 6'	decay_diffusion 2	0.5×0.8

Table 4.6: Delays and signs used to generate the output of the Dattorro's reverberator [141].

Left Channel			Right Channel		
Output	**Delay**	**Sign**	**Output**	**Delay**	**Sign**
out_4	394	+	out_1	$[394 + 128 \times \mathrm{Delvar}]$	+
out_4	4401	+	out_1	$[4401 + 966 \times \mathrm{Delvar}]$	+
out_5	2831	-	out_2	$[2831 + (-1014) \times \mathrm{Delvar}]$	-
out_6	2954	+	out_3	$[2954 + 1002 \times \mathrm{Delvar}]$	+
out_1	2945	-	out_4	$[2945 + 179 \times \mathrm{Delvar}]$	-
out_2	277	-	out_5	$[277 + 219 \times \mathrm{Delvar}]$	-
out_3	1578	-	out_6	$[1578 + (-1399) \times \mathrm{Delvar}]$	-

Note 1 The [] is used to represent the rounding function, indicating rounding the calculated decimals to the nearest integer. The Delvar as a random constant, presents the delay difference between right channel and left channel. The minimum discrimination threshold for ITD is close to 10 µs for normal hearing listeners, as demonstrated by Thavam et al [281]. In this thesis, the output delay values suggested by Dattorro are accordingly strategically narrowed down, by setting the parameter Delvar to 0.1 to ensure that the delay is subtle enough not to cause disorienting echo effects but is sufficient to create a perceivable stereo width, but it can be adjusted according to the specific listening environment.

computational cost, the higher the order of Ambisonics, the longer the running time. When comparing the computational cost with a traditional FDN, the cost of the DFDN increases by a factor equal to the number of channels in the delay line group. Therefore, the DFDN used in this thesis adopts four delay lines using first-order Ambisonics in order to reduce computational costs. First-order Ambisonics includes four channels, meaning that the reverberator has four groups of four delay lines, making for a total of sixteen delay lines, which decreases the number of delay lines by a factor of four compared to an equivalent third-order Ambisonics DFDN. In this structure, the set of delay length (m_i) is the same as the FDN (see Equation4.9 to 4.12). The gain (g_i) for each delay line is also seen Equation 4.5 to 4.7.

For DFDN, the recirculating gain of a given channel should be limited to the corresponding channel within each delay line group to ensure the cohesion of the Ambisonic channels [19]. Therefore, the gain of the other channels in the recirculation matrix should be set to zero. To achieve this configuration, a new recirculation matrix $\boldsymbol{A}$ is realised by the Kronecker product ($\otimes$) of (see Equation 4.13) and an identity matrix I_Q of size corresponding to the number of channels Q, denoted as

$$\boldsymbol{A} = A \otimes I_Q = \begin{bmatrix} A_{11} \cdot I_Q & A_{12} \cdot I_Q & A_{13} \cdot I_Q & A_{14} \cdot I_Q \\ A_{21} \cdot I_Q & A_{22} \cdot I_Q & A_{23} \cdot I_Q & A_{24} \cdot I_Q \\ A_{31} \cdot I_Q & A_{32} \cdot I_Q & A_{33} \cdot I_Q & A_{34} \cdot I_Q \\ A_{41} \cdot I_Q & A_{42} \cdot I_Q & A_{43} \cdot I_Q & A_{44} \cdot I_Q \end{bmatrix} \tag{4.13}$$

where

$$I_Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.14}$$

The Kronecker product still retains orthogonality because these two matrices involved in the operation are orthogonal themselves [282, 283]. This recirculation matrix is therefore expended into a 16×16 matrix for a first-order Ambisonics with four delay lines.

The directional weighting matrix transform $\boldsymbol{T}$ can be used to modify the direction-dependent energy distribution of an Ambisonic signal, thereby altering the reverberation times. The transform is a matrix that operates on Ambisonic signals, so that a delay line group containing a set of Ambisonic channels is used as the input, and the transform is multiplied by the Ambisonic signal for each delay line group.

To construct the matrix $\boldsymbol{T}$, the direction-dependent gain $\mathrm{gT_i}(\phi, \theta)$ is first calculated from the direction-dependent decay time $T_{60}(\phi, \theta)$, as shown in Equation 4.15 to 4.17:

$$\mathrm{gT_{dB}}(\phi, \theta) = \frac{-60}{\mathrm{Fs}}\left(\frac{1}{T_{60}(\phi, \theta)} - \frac{1}{T_{60}^{\max}}\right) \tag{4.15}$$

$$\mathrm{gT_{lin}}(\phi, \theta) = 10^{\frac{\mathrm{gT_{dB}}(\phi, \theta)}{20}} \tag{4.16}$$

$$\mathrm{gT_i}(\phi, \theta) = (\mathrm{gT_{lin}}(\phi, \theta))^{\mathrm{mT}_i}, \tag{4.17}$$

where $T_{60}^{\max}$ is the maximum direction-dependent decay time. mT_i is calculated using $T_{60}^{\max}$ through Equation 4.9 to 4.12.

Then the full directional weighting matrix $\boldsymbol{T}_{\text{full}}$ is constructed by

$$\boldsymbol{T}_{\text{full}} = \frac{4\pi}{K}\boldsymbol{Y}_{\text{uni}}^{T}\text{diag}(\text{gT}_{\text{i}}(\phi,\theta))\boldsymbol{Y}_{Q}, \tag{4.18}$$

where $\text{diag}(\text{gT}_{\text{i}}(\phi,\theta))$ denotes the creation of a matrix containing the elements of $\text{gT}_{\text{i}}(\phi,\theta)$ on the main diagonal. Y_Q represents the SH matrix with the number of columns corresponding to the SH channels Q. The full directional weighting matrix generates more output channels than input channels, so, to maintain the original SH channels, the transform matrix need to be truncated to the number of SH channels Q.

The input gain vector ($\boldsymbol{b}_i$), output gain vector ($\boldsymbol{c}_i$) and direct gain vector ($\boldsymbol{d}$) are all set as $[1,1,1,1]^T$, and can be adjusted as required.

The direct sound is similarly encoded as an Ambisonic signal by the direct gain $\boldsymbol{d} = [1,1,1,1,1]^T$.

Finally, according to Figure 2.34, the output signal of DFDN is the sum of the direct signal $X(z)$ and reverberation signals, and the transfer function is

$$Y(z) = \sum_{i=1}^{N} g_i\boldsymbol{c}_i \odot \boldsymbol{Y}_i(z) + \boldsymbol{d}X(z), \tag{4.19}$$

where $\odot$ represents element-wise multiplication (Hadamard product), and the output signals $\boldsymbol{Y}_i(z)$ of each delay line are given by

$$\boldsymbol{Y}_i(z) = [A_{i,j}g_j\boldsymbol{T}_j\boldsymbol{Y}_j(z) + \boldsymbol{b}_iX(z)] \cdot z^{-m_i} \tag{4.20}$$

4.2 New Reverberation Algorithm

Driven by the need to create a plausible and efficient reverberation model, a new reverberation algorithm is proposed, capable of simulating the acoustical characteristics of rooms of different sizes. This algorithm builds upon the simplicity and computational efficiency of the Schroeder algorithm and the enhanced realism and flexibility of the Moorer algorithm. It is referred to as the Hybrid Moorer-Schroeder (HMS) reverberation algorithm in this thesis.

The design process began with selecting the Schroeder and Moorer reverberation algorithms as the simulation of late reverberation due to their foundational role in the development of artificial reverberation techniques and their complementary strengths. The Schroeder algorithm is known for its efficient and straightforward structure, which provides a solid basis for creating dense and smooth reverberation [178]. On the other hand, the Moorer algorithm introduces frequency-dependent decay and early reflections, providing a more natural and realistic reverberation effect [175].

Building on these foundations, the HMS algorithm was structured to separately handle direct sound and early reflections from late reverberation, as previous research in Chapter 3 indicated the significant impact of ERs and ITDG on perceptual reverberation. This separation allows for greater tunability of ERs and ITDG, enabling them to be simulated more accurately. The algorithm employs a FIR filter, as proposed in the Moorer algorithm, to model the early reflections, as it has been verified to have good perceptual effects [175]. Furthermore, FIR filters can be implemented using efficient digital signal processing techniques [284]. This choice balances perceptual accuracy with computational efficiency, critical for

real-time processing in virtual acoustic applications.

The basic mono structure design and binaural implementation of the HMS reverberation algorithm based on Schroeder and Moorer algorithms will be described in detail below.

4.2.1 The Basic Mono Structure of the Proposed Reverberation Algorithm

The HMS algorithm's basic mono structure includes key components such as FIR delay lines, comb filters, low-pass feedback comb filters, all-pass filters, delay filters, and first-order low-pass filters. The detailed structure is shown in Figure 4.1. The filter coefficients in the dashed boxes within this structure are dynamically adjusted based on the reverberation time ranges specified for different room sizes (illustrated in Table 4.7), ensuring that the algorithm can simulate small, medium, and large rooms effectively. This adaptability makes the HMS algorithm suitable for a wide range of acoustic environments. These time ranges are defined by the acoustic properties of the space in Section 2.4, referenced to Gardner's study.

Table 4.7: The corresponding reverberation time ranges for each HMS reverberator.

HMS reverberator structures	**Reverberation time ranges (s)**
Small-size-room reverberator structure	$0 < \mathrm{RT}_{60} \leq 0.57$
Medium-size-room reverberator structure	$0.58 \leq \mathrm{RT}_{60} \leq 1.29$
Large-size-room reverberator structure	$1.30 \leq \mathrm{RT}_{60} < \infty$

This algorithm uses an FIR delay line to simulate early reflections. The coffi-cients ($a1$ to aN) and delays (z^{-m_1} to z^{-m_N}) follow the Moorer algorithm's strategy, ensuring accuracy in initial sound reflections which are critical for perceptual realism.

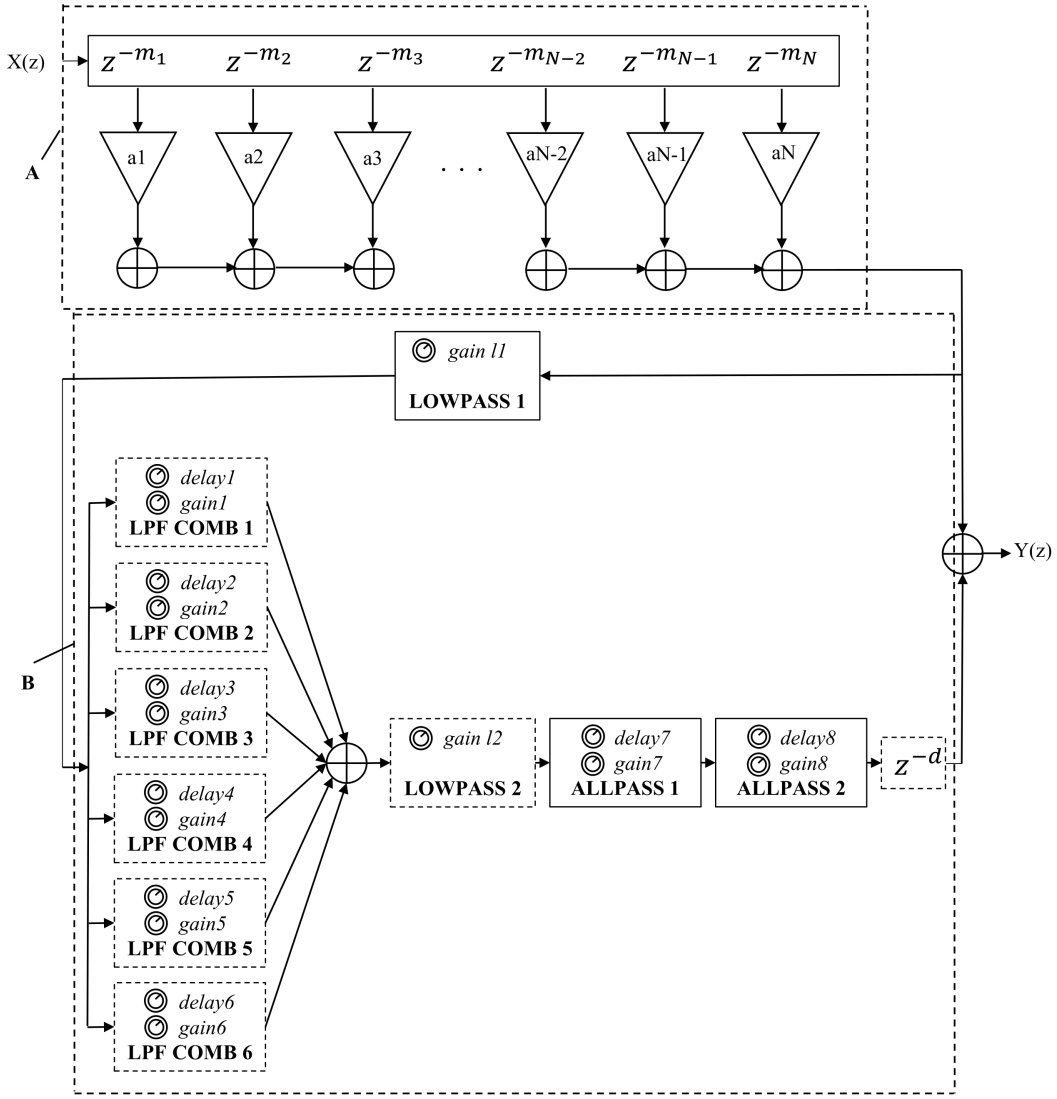


Figure 4.1: The structure of the HMS reverberation algorithm, where the part A is the FIR delay line for simulating early reflections and the part B that is composed of different filters is to simulate late reverberation.

For the large-room HMS reverberation algorithm, six low-pass feedback comb filters are employed to create the dense reverberation tail. These filters, characterised by different delay times ($delay1$ to $delay6$) and feedback gains ($gain1$ to $gain6$), generate the necessary diffusion [175]. Inserting a one-pole low-pass filter

between the FIR delay line and these six low-pass feedback comb filters helps to simulate the natural absorption of high-frequencies over time to smooth the audio [285]. Following the low-pass feedback comb filters, the signal passes through another low-pass filter to reduce high-frequency content, mimicking the energy decay in a real acoustic environment [285]. The output of the low-pass filter is further processed by two all-pass filters with their respective delays ($delay7$ and $delay8$) and gians ($gain7$ and $gain8$) to increase the number of echoes and ensure that their phase characteristics cause minimal interference [141, 286]. These all-pass filters contribute to the smoothness and natural decay of the reverberation tail, a technique borrowed from the Schroeder algorithm [178]. The final delay ensures that the overall late reverberation later than the early reflections. The parameter values of the FIR are the same as those for the Moorer reverberation algorithm shown in Table 4.2 [175]. The gain values of low-pass filters 1 and 2 are 0.6 and 0.3, respectively. This choice is made to ensure a gradual reduction in high-frequency content, mimicking the natural absorption of high frequencies in real rooms. This method helps to produce more realistic and natural reverberation effects. The low-pass feedback gain values for all low-pass feedback comb filters are 0.9. The high feedback gain results in a prolonged decay time, making the reverberation effect last longer to help in achieving a smooth and dense reverberation tail. it also means high frequencies are progressively attenuated more than low frequencies, mimicking real-world acoustic behavior where high frequencies tend to be absorbed faster by room surfaces. The parameter values of the other late reverberation filters are presented in Table 4.8. The delay times of these six low-pass feedback comb filters are mutually prime to ensure that the echoes generated by the filters do not align or reinforce each other at regular intervals.

When delay times are not mutually prime, certain frequencies can build up due to constructive interference, leading to resonance peaks. By choosing mutually prime delay times, the algorithm minimizes these resonance peaks, resulting in a more balanced frequency response. This contributes to the realism of the simulated reverberation. In addition, these delay times are longer than those proposed by Moorer [175]. Longer delay times contribute to higher echo density thus creating a more spacious and diffuse reverb tail, simulating larger acoustic spaces effectively. The reverberation time (RT_{60}) is set as a user-adjustable parameter (range 0.1 - 5 s). The user can input the desired reverberation time to produce the desired reverberation effect. The parameter settings of the two all pass filters follow Schroder's suggestion [178], with delay times of 0.005 s and 0.0017 s, and a gain of 0.7. The reverberation times of the all-pass filters in Table 4.8 are calculated using Equation 2.30. The final delay value of 0.02 s was chosen to fine-tune the alignment of the reverberation tail with the direct sound.

Table 4.8: The parameter values of the the late reverberation filters of HMS reverberation algorithm for larger size room.

Filter	**Delay (s)**	**RT_{60} (s)**
LPF COMB1	0.082	RT_{60}
LPF COMB2	0.091	RT_{60}
LPF COMB3	0.113	RT_{60}
LPF COMB4	0.123	RT_{60}
LPF COMB5	0.145	RT_{60}
LPF COMB6	0.203	RT_{60}
ALLPASS1	0.005	0.09683
ALLPASS1	0.0017	0.03292
DELAY	0.02/0.01	-

Note 1 Reverb time (RT_{60}) is a user-oriented and adjustable parameter (range 0.1–5 s).

Note 2 DELAY value 0.02 is for large size room and 0.01 is for medium size room.

The medium-room HMS reverberation algorithm is derived from the large-room HMS reverberation algorithm (illustrated in Figure 4.1) by setting the gain

(g) and delay coefficients (z^{-1}) of low-pass filter 2 (shown in Figure 2.17) to zero. This effectively removes low-pass filter 2 from the algorithm. A study by Harris et al. [287] indicated that the attenuation constant of the air is usually important only in large rooms or at high frequencies, so further simulation of high frequency noise attenuation and air absorption is not necessary in medium and small rooms. The gain value of the low-pass filter 1 is adjusted to 0.55 and the delay value of the delay filter is adjusted to 0.01 s. Slightly reducing the gain value of the low-pass filter 1 can ensure that the reverberation retains a certain amount of high-frequency content, making the reverberation sound more natural and realistic, suitable for medium-sized rooms with moderate high-frequency absorption. A shorter delay of 0.01 s is more appropriate for medium-sized rooms, where reflections arrive sooner than in large rooms. The parameter values of other filters are also presented in Table 4.8.

The small-room HMS reverberation algorithm is also derived from the large-room HMS reverberation algorithm (illustrated in Figure 4.1). Similar to the medium-room HMS algorithm, the coefficients of low-pass filter 2 are set to zero, as depicted in Figure 2.17. In smaller rooms, high-frequency attenuation due to air absorption and other factors is less pronounced compared to larger rooms [287]. Therefore, the second low-pass filter, which would simulate additional high-frequency damping, is unnecessary. Setting the coefficients to zero effectively removes it from the algorithm, simplifying the structure without compromising the reverberation quality for small rooms. Additionally, the gain (g and g_L) and delay coefficients (z^{-m} and z^{-1}) of low-pass-feedback comb filters 5 and 6 are set to zero, disabling these filters, as shown in Figure 2.23. Small rooms typically have shorter reverberation times and require fewer feedback loops to simulate the reverbera-

tion tail accurately. By setting the gains and delay coefficients of these filters to zero, they are disabled, which reduces the computational complexity and prevents an overly dense or prolonged reverberation tail that would be unrealistic for small spaces. The gain (g_L) and delay coefficients (z^{-1}) of the low-pass feedback sections of low-pass feedback comb filters 1 to 4 are also set to zero, converting them to regular comb filters 1 to 4, as illustrated in Figure 2.19. In smaller rooms, the effect of high-frequency damping is less significant. By setting the low-pass feedback coefficients to zero, these filters are converted to regular comb filters. This adjustment helps in maintaining the natural reverberation characteristics without excessive high-frequency damping, which aligns better with the acoustics of small rooms. The final delay filter coefficient is set to zero, rendering it non-functional. The final delay filter, which adds a specific amount of delay to align the reverberation tail with the direct sound, is not necessary in small rooms due to their naturally short reverberation times. These modifications are equivalent to incorporating an FIR delay line and a one-pole low-pass filter before the Schroeder reverberation algorithm [178]. The gain value of low-pass filter 1 is set to 0.3. The lower gain of the single-pole low-pass filter ensures that the high-frequency content is somewhat suppressed without overdoing it, resulting in a more realistic reverb effect. The parameter values for the other late reverberation filters are listed in Table 4.9, drawing from the parameters of the Schroeder reverberator [178]. The use of mutually prime comb filter delay times prevents constructive interference, ensuring a more diffuse and natural reverberation, which aligns with the characteristics of small rooms. The settings for the delay times and the reverberation times for the all-pass filters are also based on Schroeder's recommendations [178], which have been validated in his studies.

Table 4.9: The parameter values of the late reverberation filters of HMS reverberation algorithm for small size room.

Filter	**Delay (s)**	**RT_{60} (s)**
COMB1	0.0297	RT_{60}
COMB2	0.0371	RT_{60}
COMB3	0.0411	RT_{60}
COMB4	0.0437	RT_{60}
ALLPASS1	0.005	0.09683
ALLPASS1	0.0017	0.03292

The development of the HMS algorithm began with the construction of working prototype based on the Schroeder and Moorer algorithms. This prototype implemented the early and late reverberation processes using distinct approaches: FIR filters for early reflections to capture the initial sound-field characteristics accurately, and IIR filters for the dense reverberation tail to emulate the complex diffusion processes occurring in natural environments. Subsequently, parameter sets extracted from the Schroeder and Moorer algorithms were integrated to initially set parameters such as delay times and feedback gains in the HMS algorithm.

The development of the HMS reverberation algorithm involved a rigorous iterative testing and optimisation process to fine-tune its parameters, ensuring that it not only adhered to established practices but also achieved high perceptual realism in its reverberation effects. The author evaluated the similarity of the initial prototype of the HMS algorithm to real reverberation by conducting subjective listening tests. Based on the perceptual feedback, the filter parameters were adjusted, including the modification of delay lengths, feedback gains, and filter attenuation rates to better align with the natural perception of reverberation. After implementing these initial adjustments, the algorithm underwent another round of listening tests. The cycle of feedback and adjustment was repeated multiple times to iteratively refine the algorithm's settings. This process ensured that

each iteration moved closer to an optimal configuration. Furthermore, expert listeners were also brought in to evaluate the adjusted algorithm. Their valuable insights lead to further refinements. The expert listeners and the author participated in re-evaluations, critically analysing the algorithm's performance to ensure configure optimal parameters that would deliver a reverberation effect close to real environments.

Overall, established practices in the design of digital reverberation algorithms, such as those proposed by Schroeder, Moorer, and Gardner, provided a reliable basis for selecting parameter values. These algorithms have been widely adopted and validated in various applications, ensuring that the chosen parameters would yield high-quality reverberation effects. Iterative testing and optimization were also employed to fine-tune the parameter values, ensuring that the algorithm achieves the desired perceptual realism. This process involves repeatedly adjusting parameters and evaluating the resulting reverberation effects until the optimal configuration is found.

4.2.2 The Binaural Optimised Structure of Reverberation Algorithms

The author proposed a novel binaural optimised structure for all reverberation algorithms mentioned in Sections 4.1 and 4.2.1 to generate binaural reverberation, as shown in Figure 4.2. The process begins with a unit impulse as the input, and process it with one of the seven reverberation algorithms to generate the left channel reverb. To generate the right channel, a constant delay parameter (Δ) is applied to the delay coefficients of all filters of the reverberation algorithms, which is set to 0.1 in this chapter but can be adjusted. This delay introduces

a time difference between the left and right channel reverberations. The filter delay parameters for the right channel reverberation structure are calculated using Equation 4.21.

$$\delta_{\text{right}} = \delta_{\text{left}} \times (1 + \Delta), \tag{4.21}$$

where δ_{right} is the delay parameter for all filters in the right channel, δ_{left} is the delay parameter for all filters in the left channel, and Δ is the constant delay parameter mentioned above.

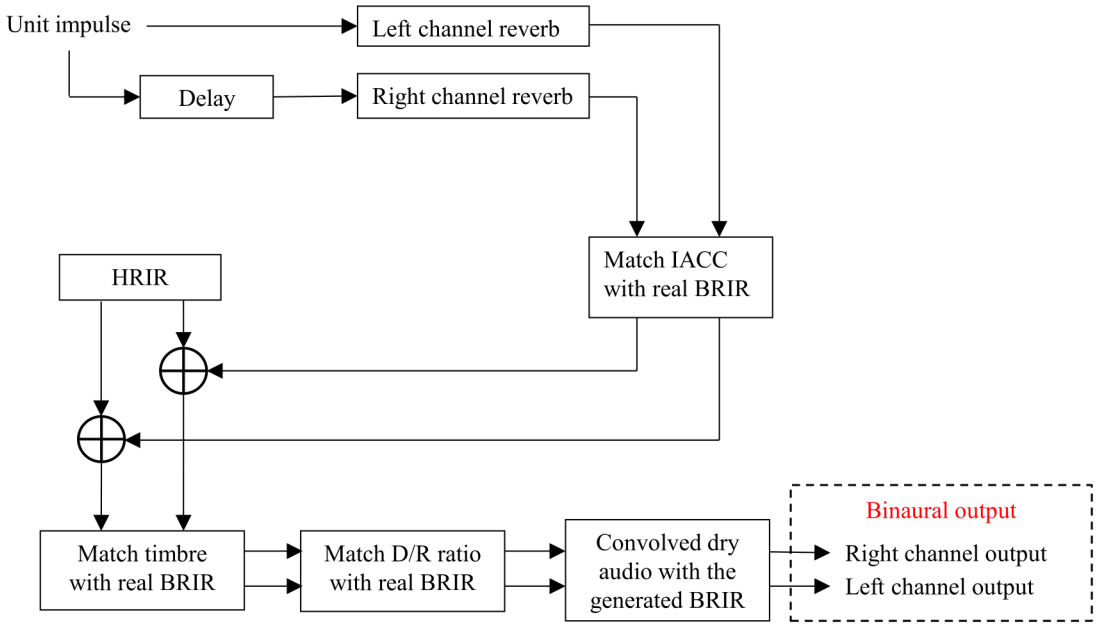


Figure 4.2: The structure of the binaural reverberation algorithm.

Next, the IACC of the left channel and the right channel are matched with a measured BRIR through a channel correlation parameter ($corr_{\text{scale}}$). The formula used to achieve this matching is

$$\text{Imp}_{\text{IACC}} = \text{Imp} + \text{Imp}_{\text{flipped}} \times \text{corr}_{\text{scale}}, \tag{4.22}$$

where Imp is the output binaural impulse response (a matrix with two columns, one for each ear) generated by the reverberation algorithms. $\text{Imp}_{\text{flipped}}$ is the output binaural impulse response flipped around the vertical axis in the left-right direction. This flipping process essentially swaps and reverses the left and right channels, creating a mirrored version of the impulse response. $\text{corr}_{\text{scale}}$ is the channel correlation parameter used to adjust the level of correlation between the left and right channels. Imp_{IACC} is the resulting binaural impulse response after the IACC has been matched.

By combining the original impulse response (Imp) with the flipped version ($\text{Imp}_{\text{flipped}}$) scaled by the channel correlation parameter ($\text{corr}_{\text{scale}}$), the IACC of the binaural signal is adjusted to match the measured BRIR. This adjustment ensures that the spatial characteristics, specifically the sense of envelopment and spaciousness, are preserved or enhanced in the virtual acoustic rendering. Large IACC values indicate greater correlation between the signals received by the two ears, corresponding to a higher degree of envelopment and a more enjoyable listening experience in auditoriums or other listening environments [156]. By carefully controlling the channel correlation parameter ($\text{corr}_{\text{scale}}$), the algorithm can fine-tune the perceived spatial attributes to closely resemble those of a real acoustic environment.

The minimum phase reconstructed version of the 0 degree azimuth and 0 degree elevation head-related impulse response (HRIR) from Subject D1 of the SADIE II database is added as the direct sound component [288]. Subject D1 is the KU100 dummy head, which is a widely recognised standard in the field of binaural audio [288]. It provides highly accurate and repeatable HRIR measurements, which are crucial for developing reliable and consistent reverberation algorithms. Us-

ing the KU100 dummy head ensures that the HRIR data is representative of a typical listener [288]. This representativeness is important for ensuring that the reverberation algorithm produces perceptually accurate results for a wide range of users. The minimum-phase version is taken to ensure the energy of the impulse response is moved to the start of the filter, implemented via the 'rceps' function in MATLAB. It is also ensured that the ITDG is determined by the first early reflection only. The adjustment for the ITDG can be achieved by delaying the early reflection part.

The timbre of the reverberation is then matched to the measured BRIR. Timbre matching is performed separately for the left and right channels, and is only performed on the reverberation, not on the direct sound. Timbre matching is achieved by aligning the root mean square values of the generated impulse response with those of the measured impulse response within the Equivalent Rectangular Bandwidth (ERB) scale. The ERB is a measure used in auditory perception and psychoacoustics to describe the bandwidth of auditory filters [289]. It gives an approximation to the bandwidths of the filters in human hearing, which is used to quantify the ear's ability to separate different frequencies. A 4096 tap ERB linear-phase critical band filter bank with 24 bands is used. This means that each channel is passed through 24 filters, each representing a different ERB band. According to Kearney's recommendation [162], the number of taps needs to be very high to get low frequency resolution, $2^{12} = 4096$ taps for a sample rate of 44.1 kHz. The root mean square (RMS) value of the impulse response of the reverberators is calculated within each ERB band by Equations 4.23 and 4.24.

$$\mathrm{Imp}_{\mathrm{ERB}} = \mathrm{fftfilt}(\mathrm{ERB}, \mathrm{Imp}_{\mathrm{RE}}, \mathrm{nfft}), \tag{4.23}$$

where fftfilt is a MATLAB function, representing fast Fourier transform (FFT)-based FIR filtering using overlap-add method. ERB is the linear-phase critical band filter bank with 24 bands. $\mathrm{Imp_{RE}}$ is the a single-channel reverb component of a BRIR. nfft represents the number of the taps, 4096 in this research. $\mathrm{Imp_{ERB}}$ is the filtered reverb component.

$$\mathrm{rms_{db}} = 20\log(\mathrm{rms}(\mathrm{Imp_{ERB}})), \tag{4.24}$$

where $\mathrm{rms_{db}}$ is a matrix, including the root mean square values of 24 bands (unit in dB).

These values are then compared to the RMS values of the corresponding band in the measured impulse response. The differences in RMS values between the simulated signal and the measured signal in each band are calculated by

$$\Delta\mathrm{rms} = \mathrm{rms}_R - \mathrm{rms}_S, \tag{4.25}$$

where rms_R represents the RMS values of the real impulse response, and rms_S is the RMS values of the simulated impulse response. Δrms is their differences.

The algorithmic reverb is scaled by these differences using Equation 4.26 to match the timbre of the measured BRIR.

$$\mathrm{Imp_{scale}} = \mathrm{Imp_{RE}} \times \mathrm{db2mag}(\Delta\mathrm{rms}), \tag{4.26}$$

where db2mag is a MATLAB function that converts decibels to magnitudes. $\mathrm{Imp_{scale}}$ indicates the reverb component of the scaled impulse response.

After scaling, the outputs from the ERB filter bank are summed. The direct sound component is then added back to the reverberated signal to create the final

output.

Subsequently, the D/R energy ratio of the impulse response generated by the reverberation algorithm is matched with the measured BRIR. The original D/R energy ratio of the algorithm-generated impulse response is first calculated by Equation 4.27, its reverberation portion is then scaled according to Equation 4.28 to match the D/R energy ratio of the desired measured impulse response.

$$\mathrm{DRR}_{\mathrm{orig}} = \mathrm{mean}(\mathrm{rms}(\mathrm{Direct}))/\mathrm{mean}(\mathrm{rms}(\mathrm{Reverb})), \tag{4.27}$$

where $\mathrm{DRR}_{\mathrm{orig}}$ is the original D/R energy ratio of the impulse response generated by the reverberation algorithm, mean represents the average of the left and right channels, rms represents the calculation of the root mean square, Direct represents the direct sound portion, and Reverb represents the reverberation portion.

$$\mathrm{Reverb}_{\mathrm{RA}} = \mathrm{Reverb} \times \mathrm{DRR}_{\mathrm{orig}}/\mathrm{DRR}_{M}, \tag{4.28}$$

where $\mathrm{Reverb}_{\mathrm{RA}}$ is the scaled reverberation portion, Reverb is the original reverberation portion, and DRR_{M} represents the D/R energy ratio of desired measured impulse response.

The final step in the reverberation process involves convolving the dry (original, unprocessed) audio signal with the generated BRIR to create the final binaural audio output. This convolution process integrates the spatial and temporal characteristics of the simulated room into the audio signal. Essentially, it blends the direct sound, early reflections, and late reverberation into the signal, mimicking how sound would interact with the environment in a real-world setting.

4.3 Objective Analysis

Prior to conducting perceptual listening tests, it is crucial to perform objective parametric analyses to compare the algorithmically generated impulse responses with the actual measured binaural impulse responses. This preliminary step ensures that the algorithms under evaluation are not only theoretically sound but also practically viable in simulating realistic acoustic environments. Objective analyses provide quantifiable metrics that can be used to assess the fidelity of the reverberation algorithms, serving as a foundational validation before subjective assessments are conducted.

Objective parametric analyses include evaluating key acoustic parameters. Due to the matching of IACC, timbre, and D/R energy ratio in the aforementioned binaural structure, and the use of reverberation time as a key parameter for user input, an objective analysis is conducted on these four parameters. These parameters are critical in determining how closely the algorithmic simulations match the characteristics of actual measured impulse responses. By conducting objective parametric analyses, we can identify discrepancies and refine the algorithms before they are subjected to subjective listening tests. This step ensures that the algorithms are in their optimal state, providing a solid basis for further perceptual evaluation. Moreover, objective analyses can highlight specific areas where algorithms excel or fall short, guiding the focus of subsequent subjective assessments.

For the analysis, BRIRs were chosen from small, medium, and large rooms with respective reverberation times of 0.266 s, 0.95 s, and 2.34 s. These specific environments, depicted in Figures 4.3 (a) to (c), were selected to cover a broad range of acoustic conditions, from tight, controlled rooms to expansive, echoic

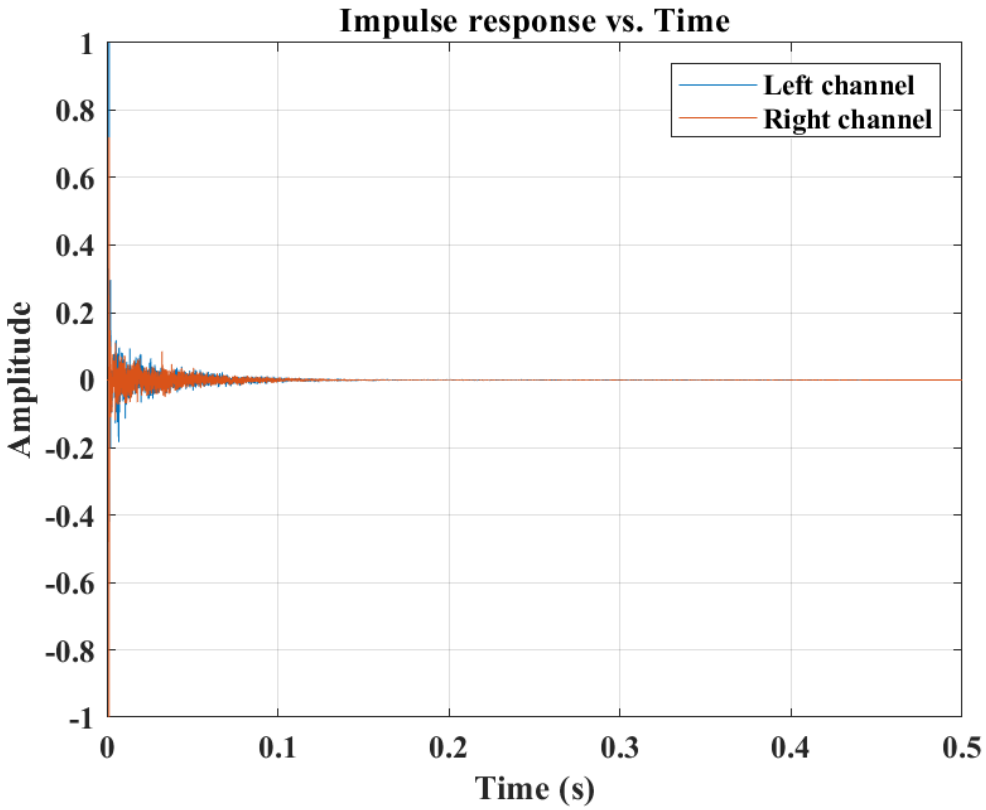
spaces.

The BRIR with a 0.266 s reverberation time, as shown in Figure 4.3 (a), was measured in Control Room 7 at WDR Broadcast Studios, Germany. This measurement is illustrated in Figure B.1, which is placed in Appendix B to provide additional context without disrupting the main narrative of the chapter. The reason of choosing this space is that it offers a reference for short reverberation times, helping the evaluation of how well the algorithms perform in such environments. The measurement was taken using a KU100 at 3 m from the source [146, 290].

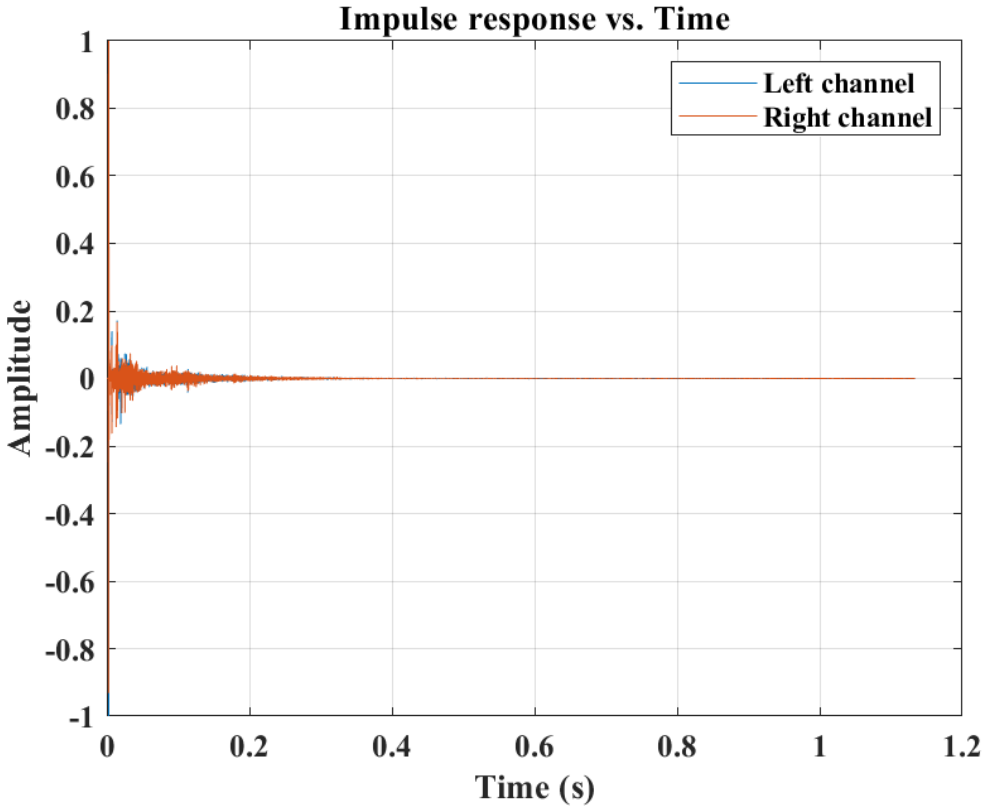
The BRIR with a 0.95 s reverberation time, as shown in Figure 4.3 (b), recorded in the Printing House Hall at Trinity College Dublin, Ireland (illustrated in Figure B.2 in Appendix B), was selected to represent a medium reverberation environment. This space offers a mid-range reverberation time that is crucial for testing how the algorithms handle more complex acoustic scenarios. The measurement was taken using a KU100 at 2 m from the source [291].

The BRIR with a 2.34 s reverberation time, as shown in Figure 4.3 (c), measured in the Lady Chapel at St Alban's Cathedral, United Kingdom (shown in Figure B.3 in Appendix B), was chosen to represent a long reverberation time typical of large, reverberant spaces such as cathedrals. This allows for the evaluation of the algorithms' performance in simulating spaces with extended reverberation tails, which is important for testing the realism and accuracy of the simulated reverberation. It was taken using a KU100 at 4.2 m from the source [292].

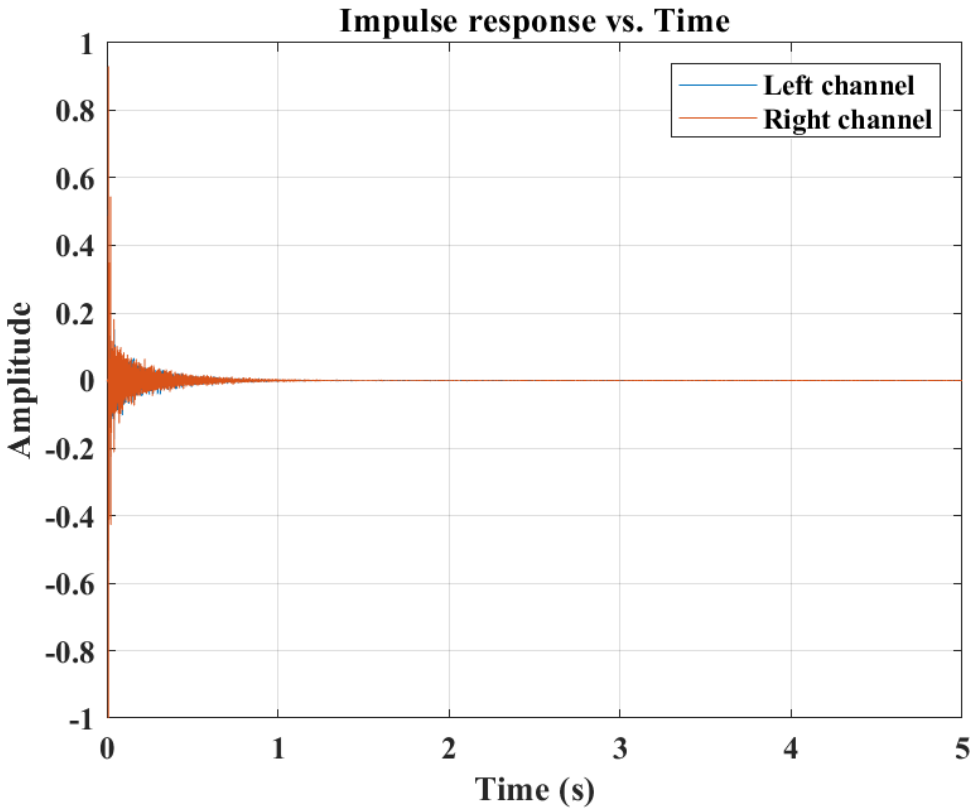
These BRIRs were selected not only for their varied acoustic characteristics but also because they were measured using a KU100 dummy head, ensuring high fidelity and realistic binaural recordings. This makes them particularly suitable for benchmarking the performance of the reverberation algorithms against real-world


Impulse response vs. Time
Left channel
Right channel
Amplitude
Time (s)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5

(a)


Impulse response vs. Time
Left channel
Right channel
Amplitude
Time (s)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.2
0.4
0.6
0.8
1
1.2

(b)



(c)

Figure 4.3: The BRIRs used in the listening test. (**a**) The BRIR with 0.266 s reverberation time. (**b**) The BRIR with 0.95 s reverberation time. (**c**) The BRIR with 2.34 s reverberation time.

acoustic conditions.

Reverberation time can vary across different frequency bands, but common practice simplifies this complexity by focusing on a representative value, typically T_{30}, at key frequencies like 500 Hz or 1000 Hz, which are standard for evaluating the acoustic suitability of a room [293]. In some cases, the mean value of T_{30} across these bands is taken to provide a more comprehensive evaluation of the room's acoustics [293]. In this chapter, it is primarily to consider T_{30} at 1000 Hz to evaluate reverberation time (RT_{60}), but the discussion is extended to include multiple frequency bands to ensure a detailed evaluation of the algorithms' performance in replicating varied acoustic environments.

D/R energy ratio is calculated from the direct sound and reverberation components. The direct sound component and the reverberation component of the BRIRs are derived from Equation 4.29.

$$\begin{aligned} \text{Direct} &= \text{Imp}(T_0 - C : T_0 + C) \\ \text{Reverb} &= \text{Imp}(T_0 + C + 1 : \text{end}), \end{aligned} \tag{4.29}$$

where Direct is the direct sound component, Reverb is the reverberation component, Imp is the matrix form of the impulse response, T_0 is the sample value corresponding to the time of the direct impulse, C is the sample value corresponding to a correction parameter of 2.5 ms, which is the defined duration of the direct path based on the approximate duration of the anechoic measured head-related impulse response [233,294], and end represents the last sample value of the impulse response matrix.

The D/R energy ratio is calculated from Equation 4.30.

$$\text{DRR} = \text{mean}(\text{rms}(\text{Direct}))/\text{mean}(\text{rms}(\text{Reverb})), \tag{4.30}$$

where DRR is the average D/R energy ratio value of the left and right channels of the binaural room impulse response, mean represents the average of the left and right channels, rms represents the calculation of the root mean square, Direct represents the direct sound component, and Reverb represents the reverberation component.

The IACC is calculated using MATLAB's library function xcorr [295] as shown in Equation 4.31.

$$\text{IACC} = \max(\text{abs}(\text{xcorr}(\text{Imp}_L, \text{Imp}_R, \text{maxlags}, \text{`normalized}'))), \tag{4.31}$$

where IACC is the interaural cross-correlation coefficient for the entire binaural room impulse response, max is used to obtain the maximum value of the cross-correlation function, abs takes all values of the cross-correlation function to be positive, xcorr calculates the values of the cross-correlation function for the two channels of the binaural impulse response, Imp_L and Imp_R represent the left and right channels of the impulse response respectively, maxlags specifies the maximum lag for correlation - 0.001 $\times$ sample rate, and normalized normalises the values of the cross-correlation.

The timbre is determined by calculating the logarithmic value of the root mean square of the matched impulse response within the ERB bands, using Equation 4.32.

$$\text{rms}_{\text{db}} = 20\log(\text{rms}(\text{Imp}_{\text{scale}})), \tag{4.32}$$

where $\text{Imp}_{\text{scale}}$ indicates the reverberation component of the impulse response that has been adjusted to match the timbre. rms_{db} is a matrix, including the root mean square values of 24 bands (unit in dB).

For the study, impulse responses with reverberation times of 0.266 s, 0.95 s, and 2.34 s were selected for small, medium, and large rooms respectively. Notably, the reverberation times calculated during this research showed minor discrepancies from those reported in cited studies. For instance, while the referenced study reported a reverberation time of 2.34 s, calculations in this study indicated a slightly different time of 2.37 s. Such variations could result from different approaches to windowing, smoothing, or addressing the noise floor during RT_{60} calculations. To maintain consistency and reliability, the original reverberation times were recalculated, resulting in revised values of 0.252 s, 0.95 s, and 2.37 s for small, medium, and large rooms in this research.

The parameter values of the measured and algorithm-generated BRIRs for small, medium, and large rooms are presented in Figures 4.4 to 4.9, respectively. These figures offer a detailed comparison of essential parameters such as reverberation time (RT_{60}) at the 1000 Hz octave band and IACC, as well as the reverberation time across all octave bands. This comparison helps evaluate how closely the algorithm-generated impulse responses mirror those measured in different room environments. Since the D/R energy ratio remains consistent with the measured values across all algorithms, it is not separately depicted in the visualisations. The algorithms accurately match the measured D/R energy ratios, 19.8523 in the small room, 31.4731 in the medium room, and 26.1826 in the large room. This fidelity is vital for the subsequent objective analysis aimed at assessing the precision and effectiveness of the reverberation algorithms in duplicating the acoustic

characteristics of actual spaces.

In Figure 4.4, the measured RT_{60} at 1000 Hz octave band in the small room is 0.252 s. According to ISO 3382-1, the JND for RT_{60} is approximately 5% [296]. In this analysis, only the differences for the Dattorro and FDN algorithms fall within this threshold. While other algorithms, except for Moorer and HMS, are relatively close to the measured value, they still exceed the JND threshold. The most significant deviations are observed with the Moorer (0.347 s) and HMS (0.333 s) algorithms, showing differences of about 38% and 32% from the measured value, respectively. These deviations are well beyond the JND threshold, indicating that they would likely be perceptible and could result in a noticeable difference in the perceived length of the reverberation tail.

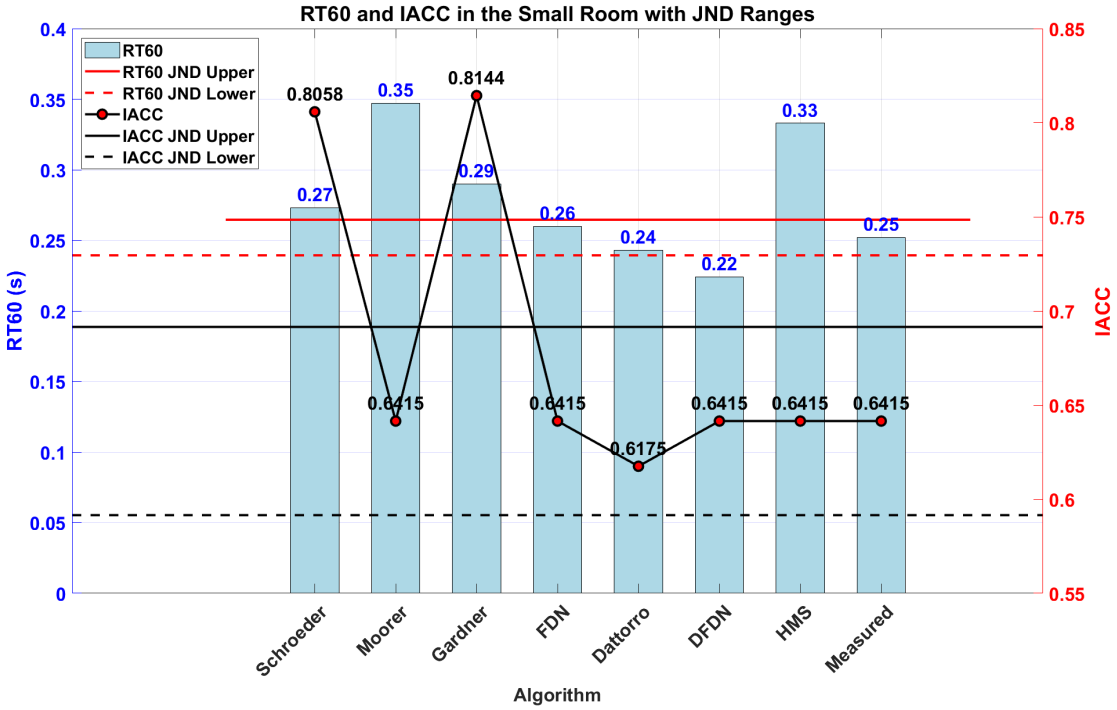


Figure 4.4: Comparison of matched RT_{60} at 1000 Hz octave band and IACC of the algorithmic-generated impulse response with those of the measured impulse response in a small room.

The measured IACC is 0.6415. The Gardner (0.8144) and Schroeder (0.8058) algorithms exhibit IACC values that deviate by approximately 0.1729 and 0.1643, respectively. According to ISO 3382-1, the JND for IACC is 0.05 [296]. These substantial deviations are likely to be perceptible, potentially altering the perceived spatial impression and envelopment, and making the sound image appear more focused or narrow compared to the actual room. The IACC value of the Dattorro algorithm (0.6175) is slightly lower than the measured value but remains relatively close. The other algorithms align perfectly with the measured IACC.

Figure 4.5 illustrates the comparison of reverberation times between each algorithm-generated impulse response and the measured values across the octave bands in the small room. In the 63 Hz octave band, all algorithms recorded reverberation times below the measured value, with Schroeder coming closest and DFDN displaying the largest deviation, suggesting that all algorithms struggle with accurately representing very low frequency reverberations. At 125 Hz, Gardner and Schroeder algorithms align more closely with the measured value. Moorer significantly overshoots the measured value, indicating an overestimation in this frequency band, while Dattorro and DFDN continue to show lower values, with DFDN again having the least reverberation. In the 250 Hz band, Dattorro closely matches the measured value. Moving into the mid-frequency range of 500 Hz to 2000 Hz, FDN shows a high fidelity with measured values, and at 1000 Hz, FDN and Dattorro align closely with the measured data. However, at the higher frequencies of 4000 Hz and 8000 Hz, the Schroeder and Gardner algorithms most closely approximate the measured values, especially at 8000 Hz.

Overall, the Schroeder algorithm demonstrates the greatest consistency with measured RT_{60} values in the low frequencies. FDN excels in the mid-frequency

range, while Schroeder and Gardner offer better performance in the high frequencies. On the other hand, DFDN and Dattorro tend to underestimate reverberation times across most bands, while Moorer and HMS typically overestimate reverberation times.

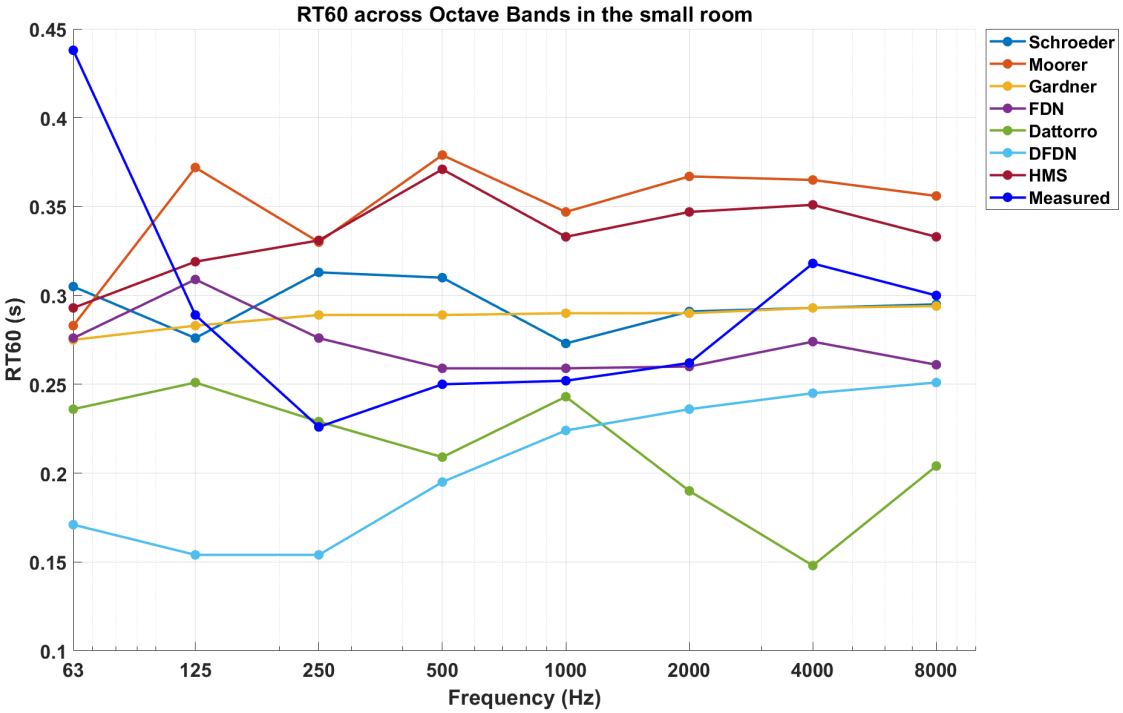


Figure 4.5: Comparison of matched RT_{60} at 63 - 8000 Hz octave bands of the algorithmic-generated impulse response with those of the measured impulse response in a small room.

In the medium room indicated in Figure 4.6, the measured RT_{60} is 0.95 s. The Schroeder (0.92 s) and Gardner (0.96 s) algorithms fall within the 5% JND threshold, suggesting these differences are likely imperceptible. However, the Moorer (0.43 s) and HMS (0.73 s) algorithms show significant deviations, with reductions of approximately 55% and 23% from the measured value, respectively. These differences exceed the JND and are likely to result in noticeable changes in the room's reverberant characteristics. The shorter RT_{60} values for the Moorer and HMS al-

gorithms could be due to differences in filter designs or feedback parameters, such as the use of low-pass feedback comb filters for high-frequency absorption. These factors can affect how quickly the reverberation decays, leading to a lower RT_{60}. The deviations in other algorithms slightly exceed the JND, indicating that the perceptual differences may not be as significant.

The measured IACC is 0.7625. The Dattorro algorithm, with an IACC of 0.6720, deviates by 0.0905, slightly exceeding the JND. This deviation might be perceptible depending on the listener and the specific acoustic environment. The other algorithms match the measured IACC exactly, indicating that the spatial impression they produce would be consistent with the actual room.

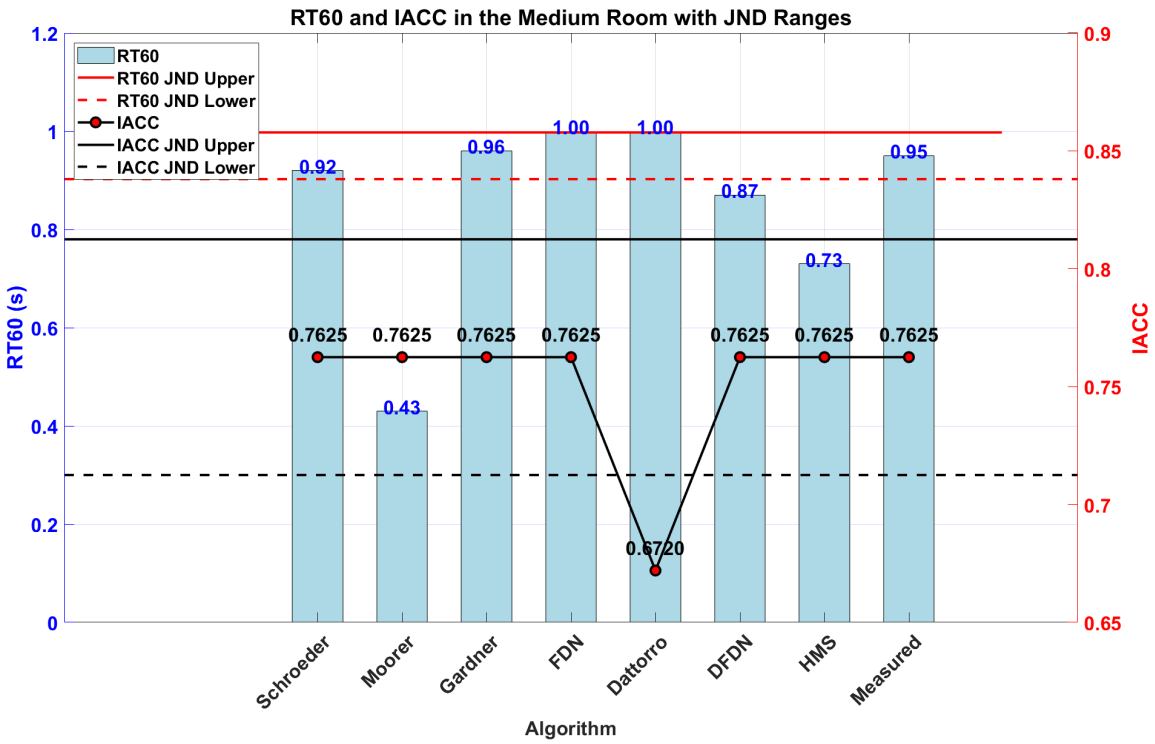


Figure 4.6: Comparison of matched RT_{60} at 1000 Hz octave band and IACC of the algorithmic-generated impulse response with those of the measured impulse response in a medium room.

Figure 4.7 examines the reverberation times in a medium room, comparing each algorithm's performance across various octave bands with the measured data. At the lowest frequency of 63 Hz, most algorithms closely approach the measured reverberation time, with the HMS algorithm showing the closest alignment. However, the Moorer, FDN, and Dattorro algorithms significantly underestimate at this frequency. At 125 Hz, all algorithms, except Dattorro, exceed the measured values, with FDN aligning most closely. The Moorer algorithm matches nearly with the measured value at 250 Hz. Between 500 Hz and 1000 Hz, most algorithms generally approximate the measured values well, with the Gardner algorithm notably close across this range, and the HMS algorithm precisely matching the measured value at 500 Hz. From 2000 Hz to 8000 Hz, although deviations increase, the HMS algorithm remains closest to the measured values.

Overall, while the HMS algorithm shows significant variances from the measured data at 125 Hz and 1000 Hz, it aligns well across multiple other bands. The Gardner algorithm demonstrates strong performance in the mid-frequency range, whereas the Moorer and Dattorro algorithms typically underestimate reverberation times across most frequencies.

As shown in Figure 4.8, the measured RT_{60} in the large room is 2.37 s. The Schroeder (2.35 s) and FDN (2.35 s) algorithms are within 1% of the measured value, which is well below the JND threshold, indicating that these differences would likely be imperceptible. However, the Moorer (0.61 s) and HMS (1.14 s) algorithms show significant deviations, with reductions of approximately 74% and 52%, respectively. These large discrepancies exceed the JND and would likely be perceptible, resulting in the simulated environment feeling much less reverberant than the actual room. The discrepancies in the Moorer and HMS algorithms' RT_{60}

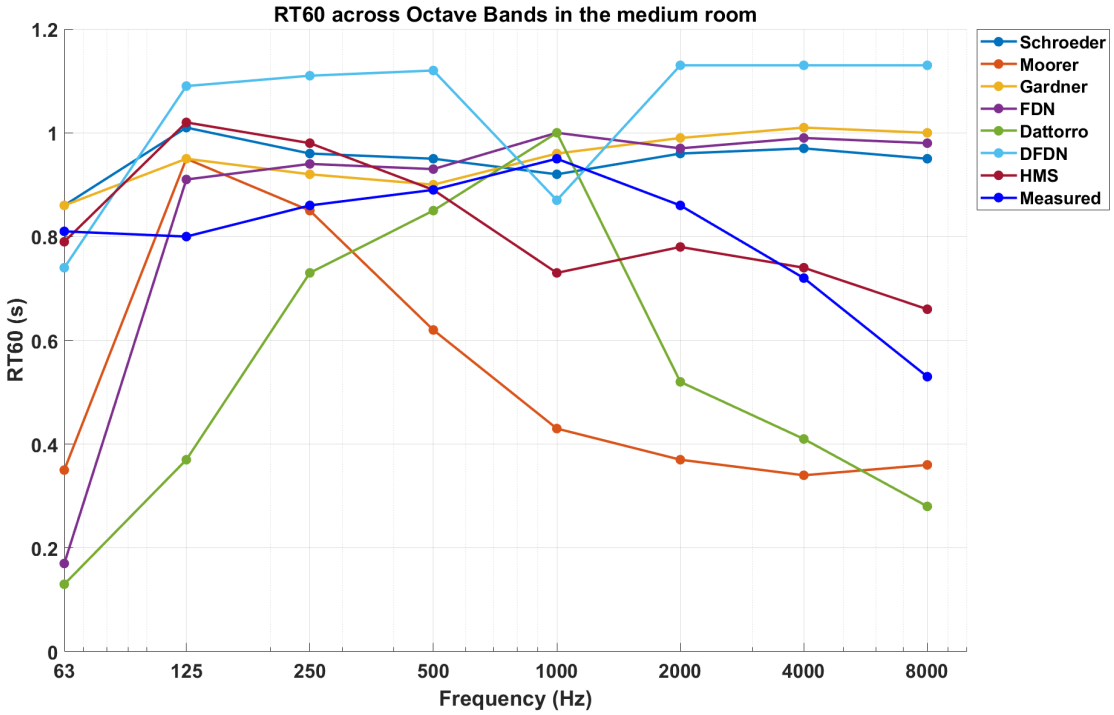


Figure 4.7: Comparison of matched RT_{60} at 63 - 8000 Hz octave bands of the algorithmic-generated impulse response with those of the measured impulse response in a medium room.

may stem from their inherent design, including the use of low-pass-feedback comb filters for high-frequency absorption, which might not be well-suited for accurately modelling large room reverberation. These algorithms may focus more on early reflections and shorter reverberation tails, leading to an underestimation of the reverberation time in larger spaces. The Gardner (2.03 s), Dattorro (2.16 s), and DFDN (2.63 s) algorithms are somewhat close but still exhibit deviations that slightly underestimate or overestimate the decay time, and all of these deviations exceed the JND, indicating that they could also be perceptible to listeners.

The measured IACC is 0.5339. Except for the Schroeder algorithm (0.5359), which is very close to the measured value, all other algorithms match exactly. These small deviations are within the JND threshold, meaning the perceived spa-

tial impression would likely be consistent with the actual room.

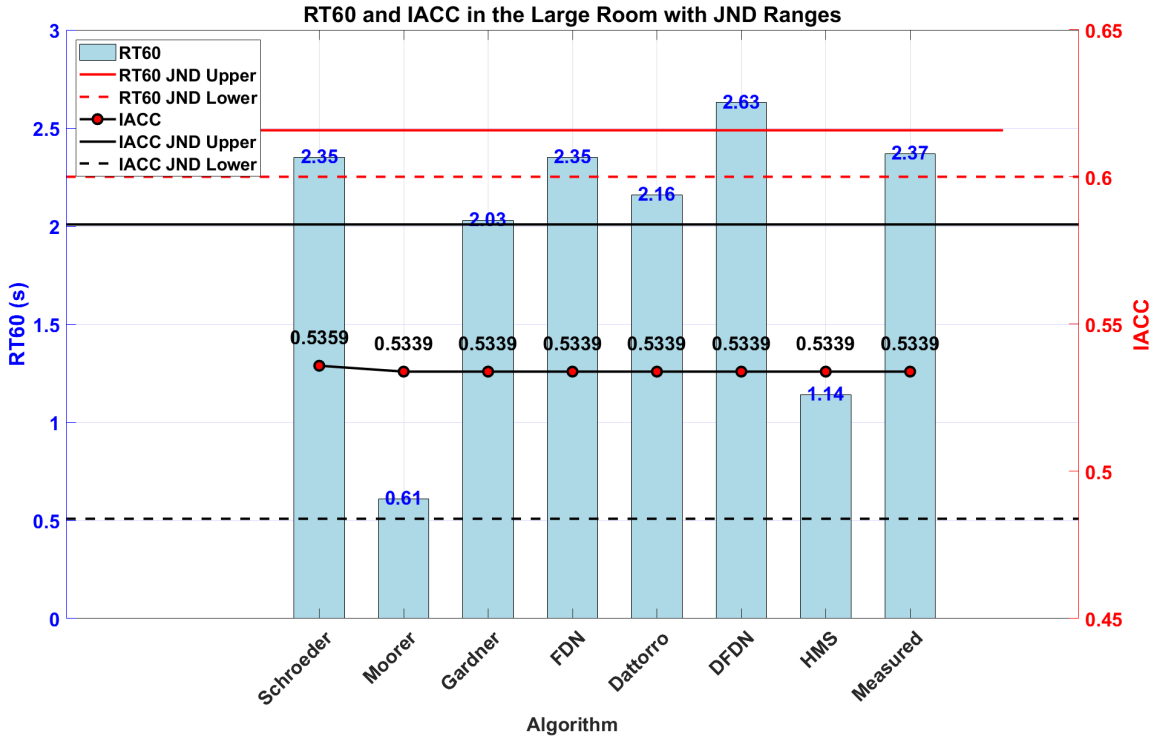


Figure 4.8: Comparison of matched RT_{60} at 1000 Hz octave band and IACC of the algorithmic-generated impulse response with those of the measured impulse response in a large room.

Figure 4.9 offers a detailed comparison of reverberation times across octave bands in a large room. At the lower frequencies of 63 to 250 Hz, all algorithms generally underestimate the measured values. As the frequency increases to the mid-range of 500 to 2000 Hz, the algorithms begin to approximate the measured values more closely. The DFDN algorithm shows consistent alignment with the measured values at both 500 and 2000 Hz, demonstrating its effectiveness in these frequency ranges. Both the Schroeder and FDN algorithms are notably accurate at 1000 Hz, capturing the mid-frequency reverberation characteristics well. At the higher frequencies of 4000 and 8000 Hz, discrepancies between the algorithms and

measured values become more evident. However, the Dattorro algorithm perform better compared with other algorithms, closely matching the measured data at these high frequencies.

Overall, while performance is lacking at lower frequencies, the DFDN, Schroeder, and FDN algorithms exhibit a balanced performance across the mid frequencies. The Dattorro algorithm excels in the high-frequency range. In contrast, both the Moorer and HMS algorithms struggle across the octave bands in modelling the reverberation characteristics of large spaces effectively, indicating challenges in accurately capturing extensive spatial acoustics.

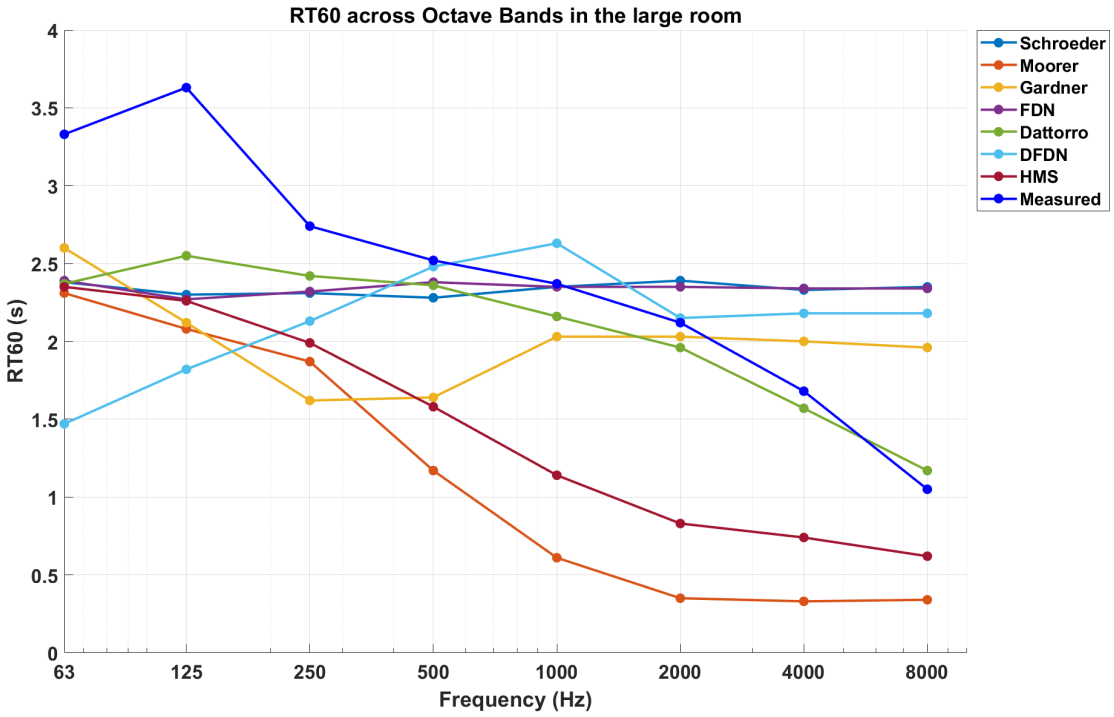


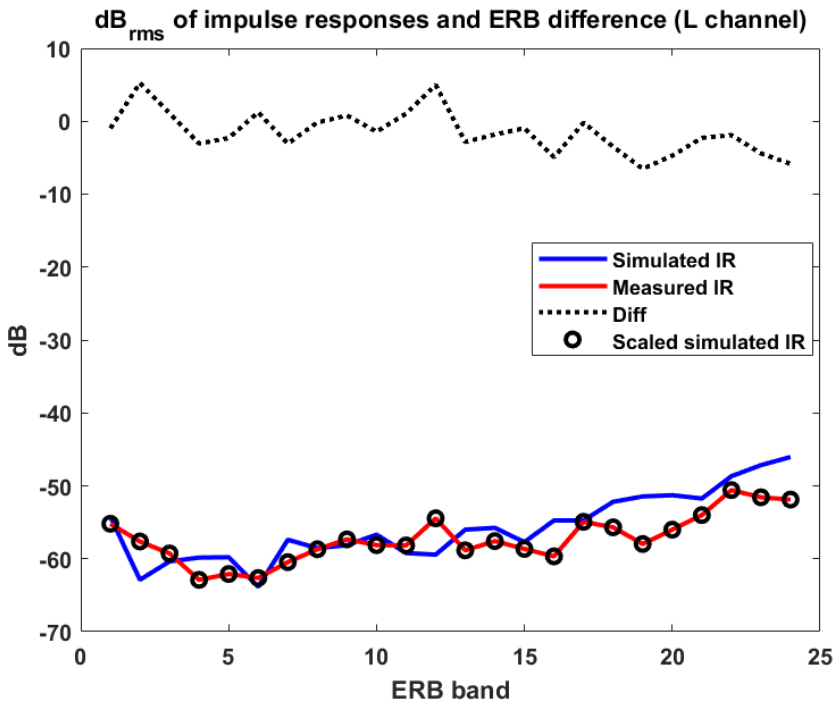
Figure 4.9: Comparison of matched RT_{60} at 63 - 8000 Hz octave bands of the algorithmic-generated impulse response with those of the measured impulse response in a large room.

Figure 4.10 illustrates the timbre matching process for the HMS-generated impulse responses to the measured impulse responses across three different rever-

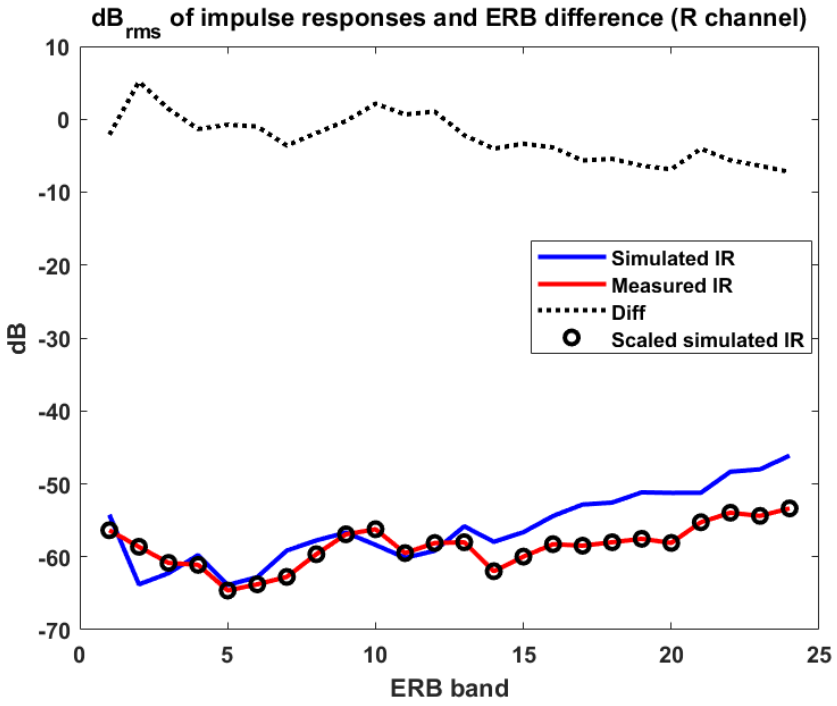
beration times (0.266 s, 0.95 s, and 2.34 s). For each reverberation time, the left and right channels are analysed separately. The matching is made by aligning the RMS values within ERB bands. The blue line represents the RMS values of the simulated impulse response, while the red line corresponds to the measured impulse response. The black circles indicate the RMS values of the scaled simulated impulse response after the timbre matching. The dotted black line shows the difference between the simulated and measured impulse responses.

Across all channels and reverberation times, the black circles align perfectly with the red line, indicating an excellent match between the scaled simulated impulse responses and the measured impulse responses across all ERB bands. This precise alignment suggests that the timbre of the simulated reverberation has been accurately adjusted to match the measured impulse response, even across different room contexts and reverberation times. The absence of any visible deviations means the effectiveness of the scaling applied to the simulated impulse responses, resulting in a nearly perfect timbre match throughout the frequency spectrum. This level of precision in matching is indicative of a well-tuned reverberation algorithm that successfully replicates the acoustic characteristics of the measured environment.

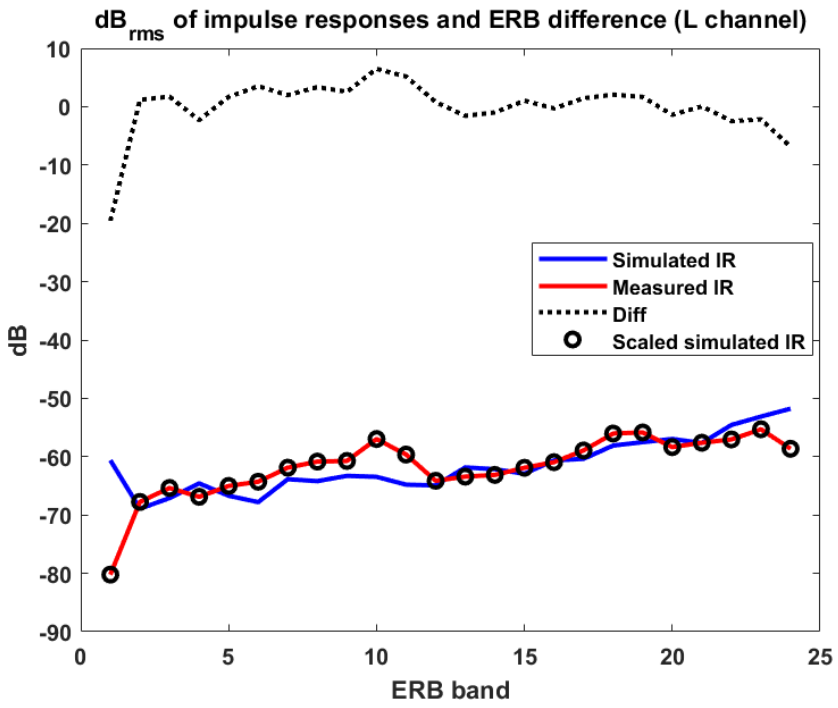
All the reverberation algorithms evaluated in this study, including Schroeder, Moorer, Gardner, FDN, Dattorro, DFDN, and HMS, have demonstrated a perfect match between the timbre of the simulated impulse response and the measured impulse response across all ERB bands. To illustrate this, the HMS algorithm is used as an example, with the Figures B.4 to B.9 for the remaining algorithms (Schroeder, Moorer, Gardner, FDN, Dattorro, and DFDN) provided in the Appendix B for reference.


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

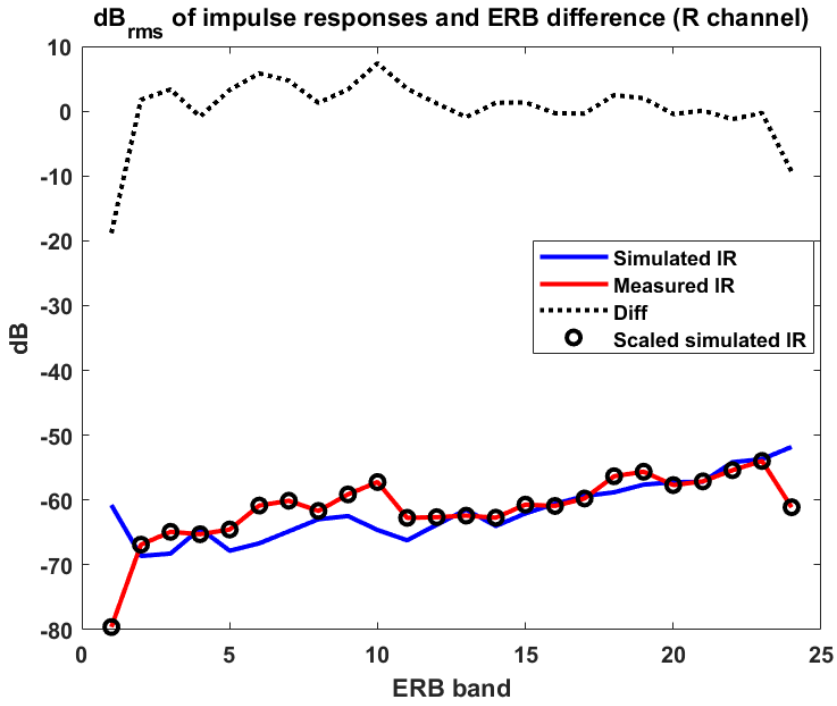
(a)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

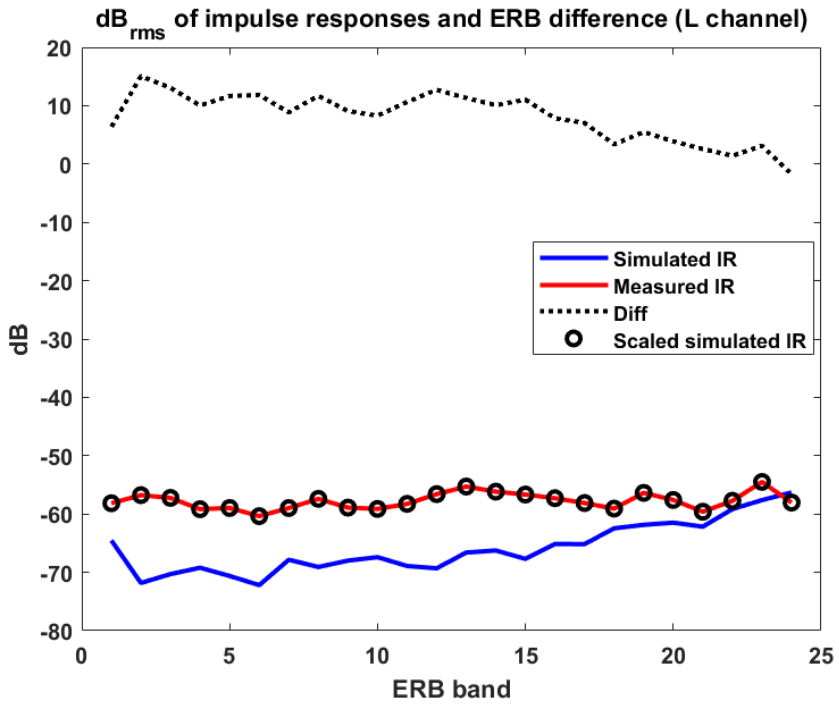
(b)


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

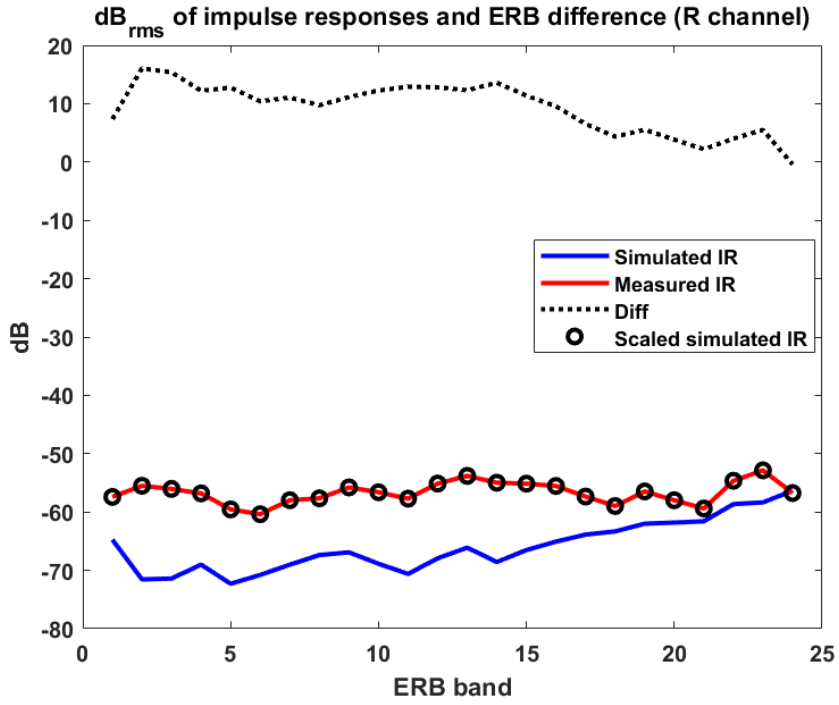
(c)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)



(e)



(f)

Figure 4.10: The timbre matching of the HMS-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.

These objective analyses reveal some deviations in the matching of reverberation time. The Schroeder algorithm is the most reliable algorithm for the small room, providing the most accurate reverberation times at both low and high frequencies. In the medium room, HMS performs best, providing the closest match across most octave bands. For the large room, no single algorithm is universally superior, instead, different algorithms excel in specific frequency ranges. However, algorithms like Moorer and HMS show significant deviations in the larger room, which would likely be perceptible to listeners. Despite these variations, most algorithms perform well in terms of energy distribution, spatial impression and timbre matching. Nevertheless, subjective listening tests are necessary to validate these findings and confirm their alignment with human auditory perception.

4.4 Listening Test

A listening test, following the MUSHRA paradigm [254], was designed to assess the similarity between artificial reverberation algorithms and measured reverberation. In this context, similarity refers to the closeness in timbre and perceived reverberation. A scale from zero to one hundred was used, where the reference condition was established as the standard. A test condition perceived as identical to the reference condition received a score of 100. The closer a test condition was to the reference, the higher its score.

4.4.1 Participants

A total of 26 participants completed the test, but only twenty of them were selected as expert listeners and experienced assessors following the post-screening

method outlined in the ITU-R BS.1543-3 recommendation [255]. This screening process excludes assessors who may not provide reliable evaluations by identifying those who assign very high scores to significantly impaired anchor signals or who frequently grade a hidden reference as if it were significantly impaired. Specifically, assessors were excluded if they rated the hidden reference condition lower than a score of 90 for more than 15% of the test items or if they rated the mid-range anchor higher than 90 for more than 15% of the items [255]. This rigorous selection ensured that only participants with consistent and accurate critical listening skills were included in the final evaluation group. Each participant was paid to take part in the test which lasted about 40 min. All participants were between 20 and 60 years old, ten identified as male and ten as female. All of these participants were members of York AudioLab or music related majors at the University of York or Beijing Film Academy.

4.4.2 Stimuli

The seven reverberation algorithms outlined in Sections 4.1 and 4.2.1 were implemented in MATLAB by the author specifically for this study. Each algorithm was custom-programmed to maintain a consistent structure, allowing precise parameter adjustments and ensuring comparability across all implementations. Following implementation, each algorithm was then binauralised according to the method detailed in Section 4.2.2. The test materials consist of commonly used male speech, female singing, a solo cello piece and drum beat [297], each rendered with the seven different reverberation algorithms with three different reverberation times (0.266 s, 0.95 s and 2.34 s). Jouni Paulus et al. used these same materials as part of their experimental material when studying the perceived level of late

reverberation in speech and music [101]. The four audio samples were chosen for the following reasons:

- Male speech is a common and recognised sound source with familiar timbre.
- Female singing is a familiar musical source with less energy at low frequencies.
- A solo cello piece is used as a low frequency source.
- Drum beat is used as an example of more transient sounds.

These four test materials are all one-channel anechoic audio of 10 s length with 44.1 kHz sample rate and 24 bits bit depth.

The reference conditions are established using the four audio samples processed with measured BRIRs, each corresponding to one of the three reverberation times, 0.266 s, 0.95 s, and 2.34 s, as illustrated in Figure 4.3.

In each trial, there are nine stimuli: seven stimuli convolved with different BRIRs generated by seven different reverberation algorithms, one hidden reference, and one low anchor (dual-mono 3.5 kHz low pass filtered audio), for a total of 108 stimuli and 12 trials in the whole listening test.

4.4.3 Procedure

The listening test was built on the MUSHRA online platform [254]. The participants were informed of the purpose of the experiment and the protocol of the experiment prior to conducting their trial on the online platform. After participants reviewed the information sheet and agreed the consent form, they could access the whole listening test via a MUSHRA URL. At the end of the listening test, demographic information such as email, type of headphone used, age,

and gender were collected. This study received approval from the University of York Physical Sciences Ethics Committee under the approval code Mi070821. For further details, please refer to Appendix C.

Before commencing the main listening test, participants were required to complete a brief training session that lasted approximately 3 min. During this session, participants adjusted their headphone volume to a safe and comfortable listening level. The session also provided them with an opportunity to carefully read through the instructions and familiarise themselves with the user interface, which is depicted in Figure 4.11. The primary task during the training session involved playing back an audio sample, with participants using the interface to control playback, set audio loop points, and rate the audio using vertical sliders. The purpose of this task was to ensure participants were comfortable with the controls and understood the importance of rating standards, which is critical for the accuracy of the listening test results.

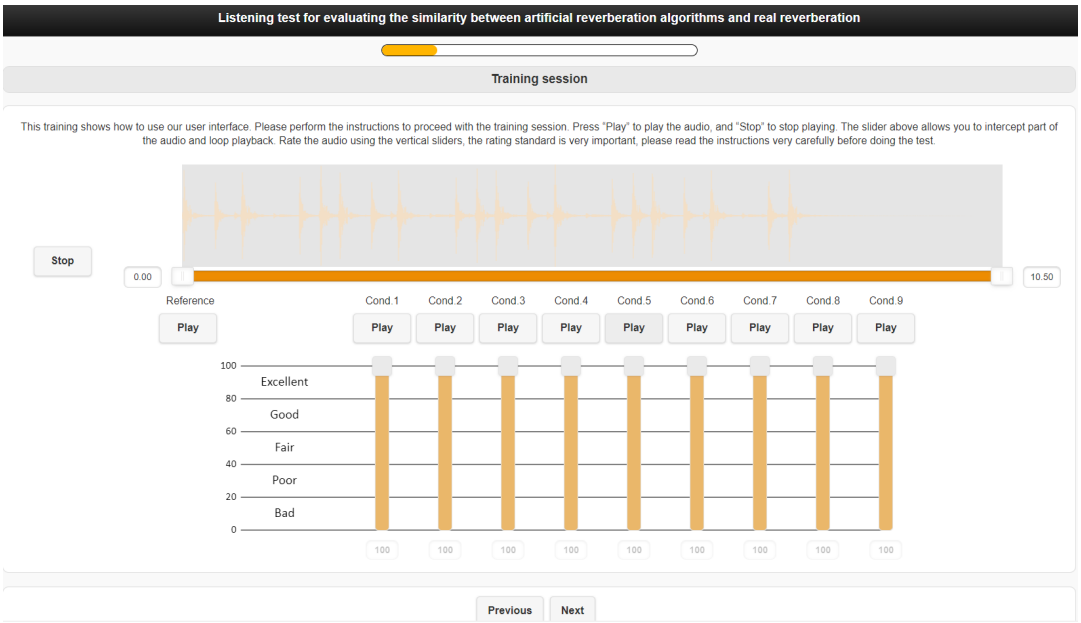


Figure 4.11: The interface of the training session.

Participants were asked to listen to the reference audio and each stimulus carefully and to judge the similarity between each stimulus and the reference audio sample on a scale of zero to a hundred. The stimulus which was the same as the reference audio sample was expected to score a hundred, and the more similar the other stimuli were to the reference audio sample, the higher the expected score.

4.4.4 Apparatus

Due to the COVID-19 pandemic, the listening test was conducted remotely online, utilising the participants' personal computers or laptops and their own headphones, and participants were instructed to perform the test in a quiet listening environment. Therefore, the listening environment, type of headphones (although Beyerdynamic DT990 was advised), and volume of audio samples were dependent on the preference of the test subjects.

The headphone types used by participants in this experiment included seven Beyerdynamic DT990, two BeyerDynamic DT770, two Beyerdynamic DT240, two ATH-M50x, two AKG K701, one QDC Anole V6, one Audeze LCD-X, one Harman AKG N60NC (plugged in), one NEUMANN NDH20, and one Sony MDR7506. Different headphones may cause the potential impact on the listening experience. For example, variations in headphone models can introduce differences in frequency response, soundstage, distortion levels, and comfort.

Different headphones have unique frequency response curves, meaning that some may emphasise certain frequencies over others. For instance, one model might boost bass frequencies while another could highlight the midrange or treble. These discrepancies can alter the perception of reverberation, especially in low or high-frequency regions, potentially leading to variations in the perceptual

similarity between the artificial and measured reverberations. The ability of headphones to accurately reproduce spatial cues (such as the direction and distance of sounds) varies from model to model. Some headphones have a wider soundstage and better imaging capabilities, which could make the reverberation appear more or less spatially accurate. This is particularly relevant in tests comparing binaural reverberation algorithms where spatial accuracy is a critical factor. Different headphones may introduce varying levels of distortion, particularly at higher volumes. Distortion can mask subtle differences in timbre or the decay of reverberation, potentially leading to less accurate ratings in the listening test. The comfort and fit of headphones can also impact the listening experience. Uncomfortable headphones may lead to listener fatigue, which could affect the consistency of ratings over time. Additionally, the seal provided by over-ear or in-ear headphones can influence bass response and overall sound isolation, further impacting perception.

To address these potential issues, a significance test was conducted to compare the results obtained with the Beyerdynamic DT990 against those obtained with other headphone models. The results of this test, presented in Table B.1 in Appendix B, indicated no significant differences between the Beyerdynamic DT990 and the other headphones used. This finding suggests that, despite the variability in headphone models, the differences in the perceived audio quality were not statistically significant, thereby justifying the inclusion of all participants in the final analysis.

4.5 Results

The rating of stimuli in MUSHRA test of each participant are shown in Table B.2 to B.13 in Appendix B. In this experiment, the average score of the reverber-

ation algorithms was used as a criterion for judgement. The higher the score, the better the simulation effect. After all the results were collected, box plots were employed first to visually represent the distribution of the data and to identify potential outliers from the raw data. The outliers were then considered to determine whether they were removed from the analysis. The data were then subjected to mean analysis to determine which algorithm scored the highest and most similar to the real reverberation. This was followed by a significance test and a post-hoc test to check whether there were significant differences between the algorithms. In addition, the processing times of each reverberation algorithm were analysed to compare their computational costs.

4.5.1 Mean Value Analysis

Box plots were employed to visually represent the distribution of the data and to identify potential outliers from the raw data. In this section, the box plots are used to illustrate the distribution of scores across various reverberation algorithms when applied to different audio sources (female singing, male speech, cello piece, and drum beat) under three different reverberation times (0.266 s, 0.95 s, and 2.34 s). As an example, the box plots for female singing are presented here, while those for the other three sound sources are included in Figure B.10 to B.12 in the Appendix B for reference. The plots include seven different algorithms, a reference (Ref), and an anchor. Outliers are visually represented as points outside the whiskers of each box plot, which correspond to values outside the upper and lower quartiles plus 1.5 times the interquartile range (IQR) [274].

These box plots play a critical role in evaluating the performance and consistency of each reverberation algorithm. For each audio source and reverberation

time, the spread of scores indicates how consistently each algorithm matched the reference. The position and size of the interquartile range and the presence of outliers provide valuable insights into the robustness of each algorithm under varying conditions, helping to identify which algorithms maintain stable performance and which are more susceptible to variations in auditory perception.

For the female singing with a 0.266 s reverberation time, Figure 4.12 (a) shows that the Gardner, Moorer, HMS, and Schroeder algorithms received generally high scores from the 20 participants. The medians for these algorithms are close to the upper quartile, with narrow IQRs, indicating that the majority of participants consistently rated these algorithms highly. The Dattorro, DFDN and FDN algorithms exhibit lower scores and more variability, as shown by the lower medians with wider IQRs. This suggests that while some participants rated these algorithms well, others had differing perceptions, possibly due to individual auditory sensitivity.

At 0.95 s reverberation time (see Figure 4.12 (b)), the distribution of scores becomes slightly more varied across participants, particularly for the Gardner and Moorer algorithms. While these algorithms still maintain relatively high median scores, the wider IQRs suggest that some participants rated them lower than others. HMS and Schroeder algorithms continue to perform well with a consistent range of scores, as indicated by the narrow IQRs. The DFDN and Dattorro algorithms again show lower medians and wider spreads, indicating their weaker performance in handling the longer reverberation time.

Under the 2.34 s reverberation time shown in Figure 4.12 (c), there is a noticeable increase in score variability for most algorithms, which is reflected in the broader IQRs. The Gardner, HMS, and Schroeder algorithms still receive generally

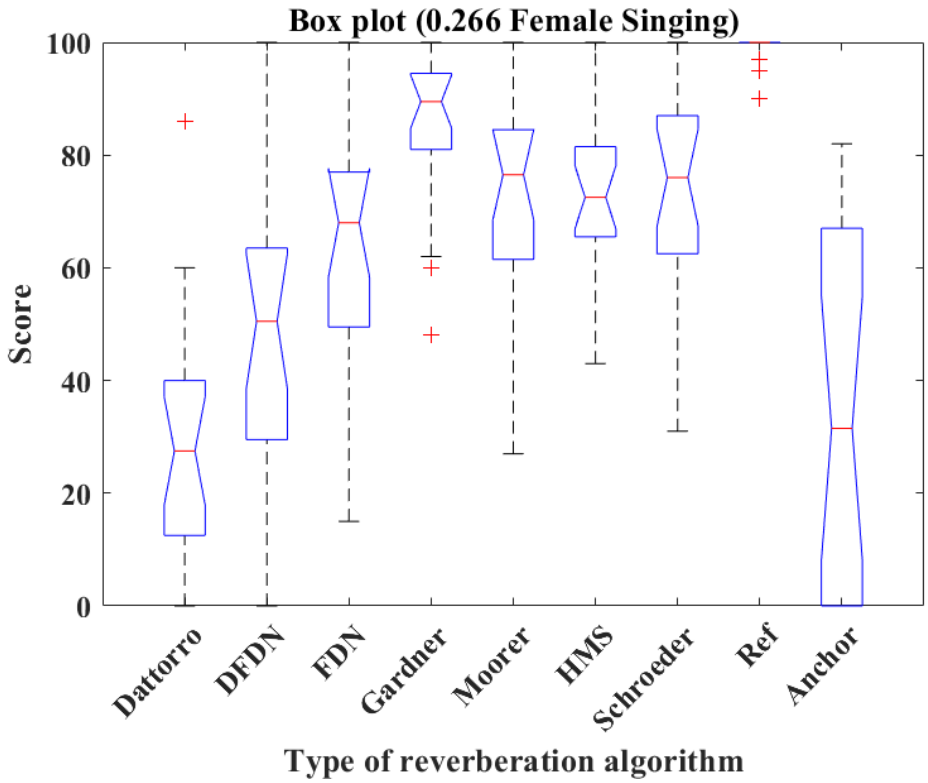
high median scores, and the FDN achieved higher score, but with more participants rating them lower, as shown by the wider spreads. The DFDN and Dattorro algorithms continue to struggle, with low median scores and significant variability, suggesting that participants found these algorithms less effective at simulating the longer reverberation time for female singing.

For male speech with a 0.266 s reverberation time, Figure B.10 (a) reveals that the HMS algorithm once again achieve higher median scores with relatively narrow IQR. This suggests that the majority of the 20 participants found that this algorithm performed well in this context. However, the Dattorro algorithm shows a lower median and relatively narrow IQR, indicating participant generally dissatisfied with its effectiveness.

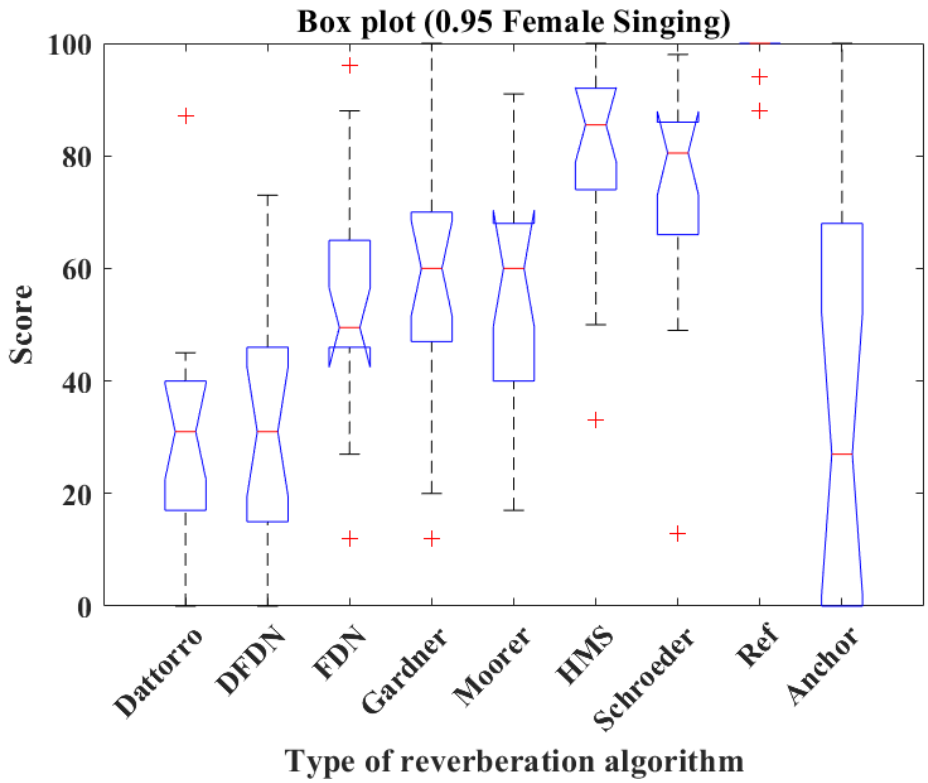
At 0.95 s reverberation time illustrated in Figure B.10 (b), HMS maintains a relatively high median with a narrow IQR, suggesting consistent performance among participants. DFDN and Dattorro algorithms display lower median scores and broader spreads, reflecting mixed opinions among participants regarding their ability to handle male speech at this reverberation time.

With a 2.34 s reverberation time as shown in Figure B.10 (c), the Gardner algorithm receives higher median score, though with a wide IQR, indicating that while many participants rated them well, some found them less effective under these conditions. The DFDN and Dattorro algorithms still show low median scores.

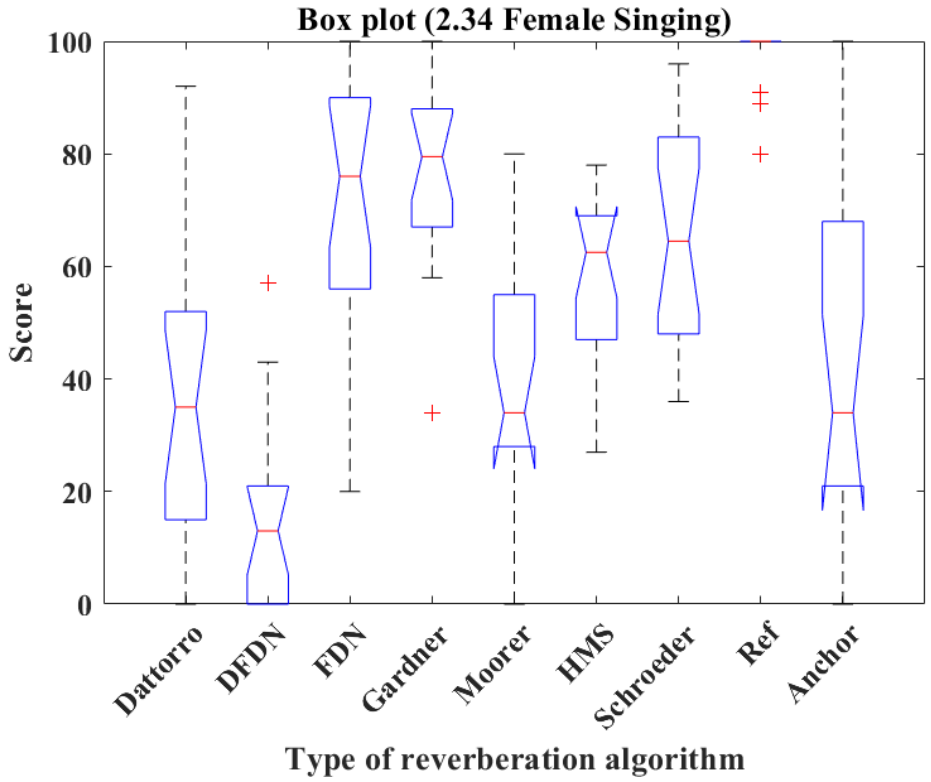
For the cello with a 0.266 s reverberation time illustrated in Figure B.11 (a), the FDN, Gardner, and HMS algorithms receive higher median scores from participants, with narrow IQRs, indicating consistent positive ratings. The Moorer algorithm also performs well, though its slightly wider IQR suggests more variability in the opinions of the participants. However, the Dattorro and Schroeder


Box plot (0.266 Female Singing)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(a)


Box plot (0.95 Female Singing)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(b)



(c)

Figure 4.12: Box plots of the scores of seven reverberation algorithms, reference and anchor simulating female singing with 0.266 s, 0.95 s, and 2.34 s reverberation time. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s ('+' in figures presents outliers).

algorithms show lower medians, indicating that some participants rated it poorly.

At a 0.95 s reverberation time, Figure B.11 (b) reveals that the HMS and Schroeder algorithms receive high ratings, with the narrow IQRs, indicating their good performance. The DFDN and Dattorro algorithms continue to show lower medians and wider spreads.

For a 2.34 s reverberation time (see Figure B.11 (c)), the performance of most algorithms declines as shown by the broader IQRs and lower median scores across the board. Despite this, the Gardner algorithm still achieves higher median with a narrow spread. DFDN algorithm performs poorly, with a low median score and a wide IQR, indicating less favourable evaluations by participants.

For drum beat with a 0.266 s reverberation time, Figure B.12 (a) shows that the Moorer algorithm stands out with a high median and a narrow IQR, indicating that it was rated well by participants. The HMS algorithm also performs well with the highest median score, though with a slightly broad spread. The DFDN and Dattorro show lower median scores, indicating that participants had less satisfaction on their effectiveness.

At a 0.95 s reverberation time as shown in Figure B.12 (b), the scores for drum beat show greater variability across algorithms, with HMS maintaining a high median. The DFDN and Dattorro again show lower performance, indicating that participants found these algorithms less effective at handling drum beat at this reverberation time.

Under the 2.34 s reverberation time, illustrated in Figure B.12 (c), most algorithms show a decline in performance, as evidenced by broader IQRs and lower median scores. The Gardner algorithm still maintains relatively higher rating than others, though with increased variability among participants, as shown by

the wider IQR. The FDN, DFDN, and Dattorro perform poorly, with low median scores, indicating that participants found these algorithms less effective for drum beat at the longer reverberation time.

These box plots demonstrate that the reference conditions consistently received high scores with minimal variability, reinforcing their effectiveness as objective high-quality standards. This confirms that participants were able to correctly identify the reference stimuli, a fundamental requirement for valid MUSHRA testing. However, the anchor did not exhibit the expected consistency. Its IQR was notably broad, indicating substantial variability in how participants rated it. Instead of consistently receiving the lowest scores, the anchor's ratings spanned a wide range, suggesting that some participants did not perceive it as the absolute worst condition.

However, certain reverberation algorithms, particularly Dattorro and DFDN, consistently received the lowest scores across different conditions. This suggests that these algorithms were perceived as significantly degraded, potentially influencing how participants evaluated the anchor. If these algorithms exhibited unnatural decay characteristics or reverberation artifacts, some participants may have perceived them as even worse than the anchor itself. Consequently, instead of using the anchor as a fixed low-quality reference, participants may have dynamically adjusted their rating scale. In this scenario, the worst-performing algorithm was placed at the lowest end of the scale, while the anchor was rated slightly higher in comparison. This shifted perception could explain the higher-than-expected scores for the anchor and its wide variability across participants.

If the worst-performing algorithm was perceived as more degraded than the anchor, participants might overestimate the anchor's quality, compressing the rating

scale. This may reduce the contrast between mid-range and high-quality conditions, making it harder to distinguish performance differences among the tested algorithms. However, given that Dattorro and DFDN consistently ranked as the worst-performing algorithms, they can function as secondary anchors to reinforce the reliability of the results. Since these algorithms were consistently rated poorly, their presence confirms that participants were able to distinguish lower-quality reverberation from higher-quality conditions, ensuring that the relative ranking of test algorithms remains valid. Besides, the ranking of test conditions remains meaningful because the reference was consistently rated highly, demonstrating that participants correctly identified the high-quality stimuli. These factors indicate that the test still meets MUSHRA standards, though refinements in anchor selection and participant training will enhance future research.

To further strengthen the reliability of the MUSHRA test in the future, it will be considered to use a stronger anchor with more severe perceptual degradation, ensuring that the anchor is unambiguously perceived as the lowest-quality condition to prevent overlap with poorly performing algorithms. Also, provide explicit training on expected rankings to help participants understand the role of the anchor and its intended function as the lowest-quality reference.

The outliers observed in these box plots represent individual scores that deviate significantly from the majority of the data points. These outliers are likely due to the subjective nature of auditory perception, which can vary from person to person. Each participant may have different levels of sensitivity to certain audio characteristics, leading to unique responses that fall outside the expected range. Since the listening test is based on subjective assessments, these outliers reflect the natural variability in human auditory perception. They do not indicate

errors in data collection or processing but rather highlight the individual differences among the participants. Retaining these outliers in the analysis is important because they contribute to a more comprehensive understanding of how different algorithms perform across a diverse group of listeners. Additionally, the non-parametric statistical methods used in this study help ensure that the presence of these outliers does not unduly influence the overall results, allowing for a more robust analysis.

Overall, these box plots illustrate the effectiveness of different reverberation algorithms as perceived by the 20 participants across various audio sources and reverberation times. The Gardner, Moorer, HMS, and Schroeder algorithms generally perform well, particularly at shorter reverberation times, while DFDN and Dattorro struggle, especially as the reverberation time increases.

The data provided by these 20 experienced participants were then averaged to determine which algorithm was perceived as having reverberation most similar to the real reverberation. The detailed visual representations of the mean scores with standard errors for female singing, male speech, cello piece, and drum beat simulated by the reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times are provided in Figures B.13 to Figure B.16 in Appendix B.

The mean scores are summarised in Table 4.10, with the highest scores highlighted in bold.

For female singing, the Gardner and HMS algorithms consistently achieve the highest scores across all reverberation times, with Gardner leading in long reverberation time (2.34 s) and short reverberation time (0.266 s). HMS excels particularly in the medium reverberation time (0.95 s). The performance of other algorithms

such as Dattorro, DFDN and FDN varies significantly, with generally lower scores indicating less effective simulation of the reverberation effect in comparison to Gardner and HMS.

In the case of male speech, the HMS algorithm scores the highest at the short and medium reverberation times. Gardner performs particularly well at the long reverberation time. The other algorithms, especially Dattorro and DFDN, show much lower scores, indicating less effectiveness in simulating the desired reverberation effect for male speech.

The cello piece results also highlight the strong performance of the HMS and Gardner algorithms. At short and medium reverberation times, HMS outperforms all others, while Gardner takes the lead at the long reverberation time. The other algorithms generally lag behind, with some showing moderate effectiveness but none matching the consistent performance of HMS and Gardner.

For the drum beat, a percussive and transient sound source, HMS again demonstrates superior performance at the short and medium reverberation times. Gardner also performs well at the long reverberation time. The lower scores of the other algorithms, such as Dattorro and DFDN, indicate their limited effectiveness in handling the reverberation of percussive sounds.

Overall, the data from Table 4.10 underscores the dominance of the Gardner and HMS algorithms across all audio stimuli and reverberation times. Gardner excels particularly in longer reverberation times, making it more suited for simulating reverberation in environments where sound decay is gradual and prolonged. On the other hand, HMS proves to be more versatile, performing exceptionally well in both short and medium reverberation times and across different types of sound stimuli. The variability in the performance of other algorithms, particularly

Dattorro and DFDN, suggests that they may not be robust or effective across a wide range of conditions.

Table 4.10: The mean scores of the above reverberation algorithms simulating female singing, male speech, cello piece and drum beat at short (0.266 s), medium (0.95 s) and long reverberation times (2.34 s) respectively.

Stimuli	Reverberation Time / Reverberation Algorithms	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder
Female singing	**0.266 s**	29.10	50.65	64.10	**85.05**	***72.65***	***73.05***	***73.90***
	0.95 s	30.35	31.70	53.45	57.25	58.00	**81.25**	***74.90***
	2.34 s	36.45	15.65	***69.50***	**77.10**	37.60	56.10	***65.80***
Male speech	**0.266 s**	17.45	51.00	***62.25***	***57.15***	51.00	**69.90**	***57.90***
	0.95 s	24.15	19.30	38.55	40.50	56.10	**72.15**	***68.35***
	2.34 s	29.15	11.35	28.10	**64.10**	37.85	***49.00***	43.25
Cello piece	**0.266 s**	49.95	65.10	***75.70***	***77.05***	***77.05***	**80.40**	***65.90***
	0.95 s	45.20	35.45	59.45	57.40	65.60	**82.85**	***70.90***
	2.34 s	54.60	22.20	62.70	**78.40**	45.85	***66.20***	48.55
Drum beat	**0.266 s**	17.70	35.50	***52.90***	***57.45***	***59.70***	**60.15**	***53.60***
	0.95 s	13.50	11.85	34.15	42.00	***63.55***	**68.55**	***56.15***
	2.34 s	24.00	11.55	22.15	**48.85**	***46.05***	***42.15***	***42.35***

Note 1 The highest scores are highlighted in bold, while scores in bold italics indicate values that are not significantly different from the highest score.

4.5.2 Statistical Analysis

Listening test data were checked for normality and variance homogeneity using the Lilliefors test [261] and Bartlett test [265]. The results show that not all data conform to normal distribution and homogeneity of variance. In addition, because of the presence of outliers and in order to ensure robustness, a non-parametric Kruskal-Wallis one-way ANOVA test [266] with 95% confidence intervals [275] was run to determine whether there were significant differences between the seven reverberation algorithms. The results showed that for all conditions (short, medium, and long reverberation times, and across all types of stimuli), the p-values were uniformly less than 0.001, far below the confidence level of 0.05, indicating significant differences between the reverberation algorithms in all cases.

After the Kruskal-Wallis one-way ANOVA test, post-hoc tests were performed to confirm where differences between groups occurred. Pairwise Comparison in Multiple Comparison was performed to determine which two groups were significantly different from each other. The post-hoc test results reveal whether the highest-scoring reverberation algorithm is significantly different from the other reverberation algorithms.

Tables 4.11 to 4.13 summarise the post-hoc test results for the pairwise comparisons of reverberation algorithms across different stimuli and reverberation times. The first column lists the different stimuli (Female Singing, Male Speech, Cello piece, and Drum beat). The first row and the second column list the reverberation algorithms being compared. The values in these tables represent p-values, which indicate whether differences between algorithms are statistically significant. A p-value below 0.05 (marked with *) suggests a statistically significant difference, meaning that one algorithm performed notably better or worse than another. Conversely, it suggests that the two algorithms performed similarly.

An example Interpretation from Table 4.11 is that in the female singing category, the Dattorro vs. Gardner comparison has a p-value < 0.001, meaning Gardner performed significantly differently from Dattorro. The Gardner vs. Moorer comparison has a p-value of 0.068, which is greater than 0.05, meaning the difference between these two algorithms is not statistically significant.

As described in Section 4.5.1, the algorithms with the highest score are shown in bold in Table 4.10. Reverberation algorithms that are not significantly different from the highest scoring reverberation algorithm are marked in bold italics in Table 4.10.

It can be seen that at short reverberation times, the HMS achieved the highest

Table 4.11: The post-hoc test results for short reverberation time (0.266 s).

Reverberation time 0.266s		DFDN	FDN	Gardner	Moorer	HMS	Schroeder
Female singing	Dattorro	0.052	*< 0.001	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	DFDN		0.141	*< 0.001	*0.006	*0.009	*0.004
	FDN			*0.002	0.197	0.249	0.151
	Gardner				0.068	0.050	0.094
	Moorer					0.891	0.884
	HMS						0.777
Male speech	Dattorro	*< 0.001	*< 0.001	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	DFDN		0.132	0.451	0.956	*0.015	0.451
	FDN			0.453	0.119	0.354	0.453
	Gardner				0.418	0.094	1.000
	Moorer					*0.013	0.418
	HMS						0.094
Cello	Dattorro	0.110	*< 0.001	*< 0.001	*< 0.001	*< 0.001	*0.042
	DFDN		0.086	0.072	*0.038	*0.017	0.662
	FDN			0.936	0.720	0.509	0.200
	Gardner				0.780	0.561	0.173
	Moorer					0.762	0.101
	HMS						0.052
Drum	Dattorro	0.050	*< 0.001	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	DFDN		*0.022	*0.011	*0.001	*0.001	*0.023
	FDN			0.806	0.372	0.328	0.983
	Gardner				0.517	0.464	0.789
	Moorer					0.932	0.361
	HMS						0.317

Note 1 Values marked with * indicate significant differences.

Table 4.12: The post-hoc test results for medium reverberation time (0.95 s).

Reverberation time 0.95 s		DFDN	FDN	Gardner	Moorer	HMS	Schroeder
Female singing	Dattorro	0.791	*0.006	*0.001	*< 0.001	*< 0.001	*< 0.001
	DFDN		*0.013	*0.003	*0.002	*< 0.001	*< 0.001
	FDN			0.622	0.596	*< 0.001	*0.006
	Gardner				0.970	*0.003	*0.025
	Moorer					*0.003	*0.028
	HMS						0.458
Male speech	Dattorro	0.599	0.086	0.064	*< 0.001	*< 0.001	*< 0.001
	DFDN		*0.025	*0.017	*< 0.001	*< 0.001	*< 0.001
	FDN			0.893	*0.042	*< 0.001	*< 0.001
	Gardner				0.058	*< 0.001	*< 0.001
	Moorer					*0.045	0.109
	HMS						0.684
Cello	Dattorro	0.247	0.118	0.218	*0.023	*< 0.001	*0.002
	DFDN		*0.007	*0.017	*< 0.001	*< 0.001	*< 0.001
	FDN			0.742	0.472	*0.002	0.131
	Gardner				0.294	*< 0.001	0.066
	Moorer					*0.018	0.429
	HMS						0.114
Drum	Dattorro	0.836	*0.011	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	DFDN		*0.006	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	FDN			0.355	*< 0.001	*< 0.001	*0.008
	Gardner				*0.010	*0.002	0.087
	Moorer					0.631	0.393
	HMS						0.182

Note 1 Values marked with * indicate significant differences.

Table 4.13: The post-hoc test results for long reverberation time (2.34 s).

Reverberation time 2.34 s		DFDN	FDN	Gardner	Moorer	HMS	Schroeder
Female singing	Dattorro	*0.031	*< 0.001	*< 0.001	0.888	*0.024	*< 0.001
	DFDN		*< 0.001	*< 0.001	*0.022	*< 0.001	*< 0.001
	FDN			0.340	*< 0.001	0.113	0.697
	Gardner				*< 0.001	*0.011	0.179
	Moorer					*0.035	*< 0.001
	HMS						0.232
Male speech	Dattorro	*0.014	0.826	*< 0.001	0.225	*0.007	0.068
	DFDN		*0.025	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	FDN			*< 0.001	0.152	*0.004	*0.041
	Gardner				*0.004	0.149	*0.021
	Moorer					0.141	0.542
	HMS						0.389
Cello	Dattorro	*< 0.001	0.256	*< 0.001	0.272	0.105	0.434
	DFDN		*< 0.001	*< 0.001	*0.014	*< 0.001	*0.006
	FDN			*0.027	*0.025	0.629	0.055
	Gardner				*< 0.001	0.084	*< 0.001
	Moorer					*0.007	0.752
	HMS						*0.016
Drum	Dattorro	0.067	0.740	*< 0.001	*0.002	*0.010	*0.009
	DFDN		0.133	*< 0.001	*< 0.001	*< 0.001	*< 0.001
	FDN			*< 0.001	*< 0.001	*0.004	*0.003
	Gardner				0.797	0.467	0.503
	Moorer					0.638	0.681
	HMS						0.953

Note 1 Values marked with * indicate significant differences.

scores for three out of four stimuli, except for female singing. However, there were no significant differences among the HMS, Gardner, and Schroeder reverberation algorithms, indicating that all three algorithms performed comparably across the four stimuli. At the medium reverberation time, the HMS consistently scored the highest in simulating real reverberation. However, there was no significant difference between the HMS and Schroeder algorithms, this indicates these two algorithms' effectiveness in capturing medium reverb characteristics. At the long reverberation time, the Gardner algorithm demonstrated the best overall performance, but in most cases (except female singing), the HMS showed no significant difference from the Gardner. This suggests that the HMS could be a viable alternative to the Gardner for handling long reverb times, making it a reliable choice for most stimuli, except when dealing with female singing, where Gardner is clearly superior.

4.5.3 Computational Cost Analysis

The processing overhead for each reverberation algorithm was extracted through the 'Profiler' in MATLAB, with the results listed in Table 4.14. The comparison shows that DFDN is far less computationally efficient than the other algorithms. At low reverberation times, Schroeder, Gardner, and FDN have the lowest computational costs. At medium reverberation times, the Schroder and Gardner algorithms are more computationally efficient than the others. At long reverberation times, the Schroeder and Moorer reverberation algorithms are far more computationally efficient than the others, followed by the Gardner algorithm, which is slightly more computationally efficient than the rest. The computational cost of our proposed HMS reverberation algorithm is slightly higher than the Gardner,

Moorer, and Schroeder algorithms.

Table 4.14: Computational cost of seven reverberation algorithms at three different reverberation times.

Reverberation Time \ Algorithms	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder
0.266 s	1–1.5 s	2–2.5 s	0.5–1 s	0.5–1 s	1–1.5 s	1–1.5 s	0.5–1 s
0.95 s	4–5 s	24–25 s	3–4 s	1–2 s	1.5–2.5 s	3.5–4.5 s	1–2 s
2.34 s	13–14 s	129–130 s	12–13 s	8.5–9.5 s	3–4 s	10–11 s	2–3 s

4.6 Discussion

A perceptual score exceeding 80 out of 100 in listening tests is often considered a benchmark for high-quality audio production, indicating that the reverberation closely matches real-world acoustics, while a reverberation algorithm is deemed effective when it achieves a score of 60 or higher, as this suggests an acceptable level of perceptual quality [255]. While the HMS algorithm did not achieve a score of 80 or above across all stimuli, it consistently achieved high perceptual scores, particularly for short and medium reverberation times, where it exceeded 60 points, demonstrating its effectiveness in accurately simulating reverberation across various stimuli. The HMS achieved the highest scores in seven out of twelve tested stimuli, highlighting its strong overall performance. Additionally, the HMS performed comparably to the highest-scoring algorithm in eleven out of twelve cases, with the only significant exception being female singing under long reverberation conditions, where the Gardner demonstrated clear superiority. The Schroeder algorithm ranked second overall, as it showed no significant differences from the highest-scoring algorithm in ten out of twelve cases, except for male speech and cello piece with the long reverberation time. Although Schroeder did not achieve

any highest perceptual scores, it maintained strong performance across most stimuli while offering significant computational advantages.

The experimental results indicate that the Schroeder algorithm emerges as the optimal choice when computational efficiency is a primary concern, particularly in scenarios where slight compromises in perceptual accuracy are acceptable. For instance, in real-time applications such as live sound reinforcement, gaming, or interactive simulations, where low latency and minimal computational load are crucial, the Schroeder algorithm's ability to deliver reasonable reverberation with minimal processing time makes it highly suitable. Although specific thresholds for processing time are not definitively established, it is generally understood that faster algorithms are more suitable for real-time processing. The exact requirements for processing time would depend on the specific application.

On the other hand, if the goal is to achieve high perceptual accuracy, especially in contexts where the realism and naturalness of the reverberation are paramount, such as in professional audio production, high-fidelity virtual reality experiences, or detailed acoustic simulations, the HMS reverberation algorithm is preferable. Although it demands more computational resources, the HMS algorithm offers a closer match to the desired acoustic properties, as demonstrated by its higher perceptual scores in the listening test. This makes it suitable for use in environments where computational cost is less of a concern, or where processing can be done offline.

Alternatively, a hybrid approach could be employed, leveraging the strengths of different algorithms depending on the reverberation time. For short and medium reverberation times, the Schroeder algorithm or the HMS algorithm can be utilized to strike a balance between speed and quality. The Schroeder algorithm,

being computationally efficient, is ideal for situations where quick processing is necessary, whereas the HMS algorithm is suitable when higher perceptual accuracy is required, even at a slightly higher computational cost.

For long reverberation times, the Gardner reverberation algorithm shows particular strengths in perceptual accuracy, making it an excellent choice when the highest fidelity is needed. Although slightly more computationally costs than Schroeder, Gardner's performance remains manageable and significantly better in terms of perceived audio quality. In applications where a blend of efficiency and perceptual quality is necessary, such as in mixed-use environments like virtual concert halls or simulation systems, combining Schroeder or HMS for shorter reverberation times with Gardner for longer reverberations could offer an optimal solution.

In summary, the decision between these algorithms depends on the specific requirements of the application. If perceptual quality can be compromised for the sake of efficiency, Schroeder is the preferred choice. Conversely, if fidelity is paramount and computational resources are available, the HMS algorithm should be prioritised. When both perceptual quality and efficiency are needed, a hybrid approach utilising Schroeder or HMS for shorter reverb times and Gardner for longer ones may provide the best overall results. This approach ensures that the reverberation effects are tailored to the needs of the environment, balancing computational load with the desired level of acoustic realism.

The observation that male speech and drum beat samples received lower scores compared to female singing and cello pieces suggests that the artificial reverberation algorithms tested in this experiment are less effective in simulating speech signals and percussive sounds to make their perceptual effects similar to real re-

verberation. Notably, the drum beat scores were particularly low, indicating a challenge in accurately reproducing the perceptual effects of reverberation for percussive sounds.

The lower scores for male speech and drum beat samples in the reverberation tests can be attributed to the distinct acoustic characteristics of these types of sounds and how they interact with reverberation algorithms. Listeners are highly sensitive to changes in speech intelligibility and clarity. Even slight distortions or mismatches in the reverberation of speech can be perceived negatively, as they can interfere with the understanding of the spoken content. This heightened sensitivity can lead to lower scores for reverberation algorithms when simulating male speech, as any deviation from the expected natural reverberation is more readily detected. Drum beat, being percussive, are characterised by sharp, transient attacks and short durations, which make them particularly challenging for reverberation algorithms to simulate accurately. The artificial reverberation may struggle to replicate the precise decay and spatial characteristics of such transient sounds, leading to a perceived mismatch between the simulated and real reverberation.

These factors contribute to the generally lower scores for male speech and drum beat compared to female singing and cello pieces. This observation aligns with findings by Frissen et al. [298], who showed that while stimulus type does not affect a person's ability to distinguish different reverberation effects, it does influence the absolute amount of perceived reverberation. Their study found that speech stimuli led to higher estimates of reverberation than singing or drums, particularly for longer reverberation times (i.e., greater than 1.8 s). Moreover, the absolute amount ofperceived reverberation for drums was significantly different from that for vocal stimuli (including speech and singing). This suggests that

the type of sound stimulus does have a notable effect on the perceived success of reverberation algorithms.

The COVID-19 pandemic created challenges in recruiting a large pool of participants, which limited this experiment to a sample of approximately 20 participants. According to the ITU-R BS.1543-3 recommendation [255], when listening test conditions are tightly controlled, both technically and performance, data from as few as 20 evaluators may be sufficient to reach reliable conclusions. However, it is important to acknowledge that this specific test was conducted remotely, which inherently limits the level of control over the listening environment, headphone quality, and other variables that can affect the results.

The use of 20 participants aligns with ITU-R BS.1543-3 recommendation [255] guidelines, but the lack of tightly controlled conditions, such as ensuring a consistent listening environment for all participants, may introduce variability in the data that wouldn't be present in a more controlled setting. Therefore, while the sample size is sufficient to detect general trends and significant differences between conditions, the lack of tight control over the testing environment suggests that the findings might have a higher margin of error. Thus, while the conclusions drawn from this study are likely valid within the context of the experiment, they should be applied cautiously to broader settings. Due to the fact that remote experiments conducted under COVID-19 restrictions could affect the listener's perception of the stimuli, thus, it is important to confirm the findings with in-person testing in further studies.

It should be noted that this experiment is a static binaural reverberation test rather than a dynamic one. However, the static case provides a valuable foundation before attempting fully head-tracked dynamic scenario conditions.

4.7 Summary

In this chapter, a novel reverberation algorithm with a binaural structure was proposed, representing an advancement in the field of audio processing. The introduction of this binaural structure not only enhances the spatial accuracy and realism of the reverberation effects but also provides a new idea for how binaural processing can be integrated into reverberation algorithms. By applying this binaural structure to other popular reverberation algorithms, the research demonstrates its versatility and potential to improve existing technologies. This work contributes to the field by offering a new approach to binaural audio that can be leveraged in various applications, such as immersive audio experiences, virtual reality, and audio plug-in design, ultimately pushing the boundaries of how spatial audio is perceived and produced.

This experiment evaluated the similarity between different artificial reverberation algorithms and measured reverberation, including the assessment of a novel reverberation algorithm, the HMS. The motivation for the study was to consider the suitability of reverberation algorithms to be applied to virtual rendering in the future to create more plausible reverberation effects.

The Schroeder algorithm demonstrated superior computational efficiency, making it an ideal choice for real-time applications where processing speed is critical. However, its perceptual performance, while reasonable, is not as high as other algorithms (like HMS and Gardner), particularly in scenarios where accurate simulation of reverberation is paramount. The HMS algorithm, on the other hand, provided the best perceptual performance, particularly for medium and short reverberation times, although at a slightly higher computational cost. This makes

it suitable for applications where sound quality is prioritised over computational efficiency. The Gardner algorithm consistently performed well in simulating long reverberation times, making it a valuable tool for environments where longer decay times are needed, such as in large spaces or for certain musical genres. A combination of the Schroeder and Gardner algorithms or the HMS and Gardner algorithms could be the way forward for virtual rendering in future research.

The artificial reverberation algorithms tested in this experiment are less effective in simulating speech signals and percussive sounds. This finding has informed subsequent research and development by identifying specific challenges that current reverberation algorithms face, particularly in handling transient sounds and speech signals. Future work will focus on refining algorithms to improve their performance in these areas, potentially by developing specialised techniques for transient sounds and speech signals reproduction.

This work contributes to the field of audio signal processing by providing a comprehensive evaluation of popular reverberation algorithms, highlighting their strengths and weaknesses in different contexts. The findings offer valuable insights into how these algorithms can be employed for specific applications, whether in real-time environments or high-fidelity audio production. The detailed analysis of computational costs and perceptual performance helps inform future developments in reverberation algorithms, guiding the design of more efficient and accurate reverberation tools. Furthermore, the insights gained regarding the performance of algorithms with different types of stimuli will influence future algorithmic adjustments to better handle percussive sounds and speech signals. By addressing the identified weaknesses, future algorithms can achieve more accurate and versatile reverberation effects, enhancing their applicability in diverse audio contexts,

including music production, gaming, and virtual reality environments.

Chapter 5 investigates the combination of the Gardner, Schroeder and HMS reverberation algorithms into a set of algorithms that simulate room models with different reverberation times and making them into a dynamic audio plug-in that can be applied in a virtual rendering environment in order to facilitate the implantation and application of this reverberation plug-in in future research.

5 Dynamic Realisation

The development and application of dynamic reverberation algorithms have become increasingly essential in the context of virtual acoustic rendering environments. These environments demand not only high-fidelity audio but also dynamic adaptability to the interactions and movements within a virtual space. This chapter focuses on the dynamic realisation of reverberation algorithms, detailing the creation and implementation of a dynamic audio plug-in, the Hybrid Moorer-Schroeder-Gardner (HMSG) plug-in, which enables real-time processing of reverberation effects within virtual environments. By leveraging the findings from Chapters 3 and 4, particularly the perceptual evaluation of various reverberation algorithms, this chapter outlines the process of integrating these algorithms into a versatile tool that can be applied across a range of virtual scenarios and presents an objective evaluation of the plug-in's performance.

5.1 Motivation and Relevant Published Key Work

The motivation behind the dynamic realisation of the HMSG plug-in stems from the need to address the limitations of static audio effects in delivering immersive and interactive experiences within virtual environments. Traditional reverberation algorithms, while effective in static scenarios, often struggle to deliver

the dynamic, responsive audio experiences required in immersive applications. The HMSG plug-in was developed to bridge this gap by simulating real-world acoustics in real-time, dynamically adapting to user interactions and environmental changes to create a more believable and engaging virtual experience.

Previous work in this field has explored various approaches to dynamic reverberation, but many of these methods involve trade-offs between computational efficiency and perceptual accuracy, particularly in real-time applications. The HMSG plug-in is designed to overcome these limitations by integrating the perceptually significant parameters identified in Chapter 3. This plug-in leverages the strengths of existing algorithms while addressing their shortcomings, offering a more efficient and perceptually accurate solution for dynamic reverberation in virtual environments.

5.1.1 Research Motivation

To enhance the immersive experience within virtual environments, spatial audio, an essential foundation for virtual acoustic rendering, is continually evolving. Achieving a deeper sense of immersion in virtual acoustics requires not only accurate acoustic simulation characteristics but also real-time, interactive, and dynamic features. These elements are crucial for successfully implementing immersive virtual acoustics in virtual rendering environments. Among the key characteristics of spatial audio, reverberation plays a significant role in enhancing the spatial quality and fullness of the audio experience. In Chapter 4, the perceptual similarity between various artificial reverberation algorithms and real-world reverberation has been assessed, and it is also proposed that future work will explore the combination of the Gardner, Schroeder and HMS reverberation algorithms to create a suite of

algorithms capable of simulating room models with varying reverberation times. To support the implementation and application of this combined reverberation algorithm in future research, this chapter details its implementation as a dynamic audio plug-in that can be integrated into virtual rendering environments.

5.1.2 Relevant Published Key Work

For artificial reverberation algorithms to be used flexibly for various sound processing applications, their reverberation effects need to closely replicate the perceptual characteristics of real-world environments. Additionally, they must support real-time processing and user-oriented interactivity. Audio plug-ins are a good way of satisfying the real-time and interactive nature of audio processing. In general, an audio plug-in is a tool that helps users create, enhance, analyse and process sound in a DAW. Audio plug-ins are divided into three main categories: tool plug-ins, processing plug-ins, and analyser plug-ins [299]. Tool plug-ins are usually used to create sounds through virtual synthesisers or samplers. Audio processing plug-ins are intended to change or reshape the effect of the input audio signal, such as the usual filters, compression limiters, reverbs, delays, etc. The main purpose of analyser plug-ins is to analyse audio preferences for characteristics such as loudness, stereo width or frequency components. A DAW is a software application used to record, edit and produce sound files. It is also utilised in the production of radio, television, film, podcasts, games and any other context where complex audio signals processing is required [300].

From a technical point of view, a plug-in is code that processes an input audio signal and returns an output signal, and can generate a user interface (UI) containing UI widgets that can be plugged into the DAW to enhance its functionality.

To process streaming audio data, the DAW calls the plug-in, passing in a frame of input audio data and receiving back a frame of processed output audio data. MATLAB's Audio Toolbox supports code generation for the most common plug-in format, Steinberg's Virtual Studio Technology (VST) [300].

At present, many artificial reverberation algorithms have realised interactive real-time plug-in processing. A real-time audio plug-in implementation of the FDN algorithm was proposed by Prawda et al. [142] in 2020. Their plug-in provides precise control of the reverberation time in ten octave frequency bands, and also several possibilities to control the elements of the FDN structure, such as the feedback matrix and delay lines. Its graphical user interface also shows the attenuation filter's magnitude response, the corresponding reverberation time curves and the resulting RIR. A complete view of the decay characteristics and quality of the synthesised reverb is also provided.

Borß [301] presented a newly developed VST reverberation effect plug-in named HybridReverb, which is based on synthetic room impulse responses. This plug-in delivers a natural-sounding reverb by utilising physical principles, allowing users to select appropriate RIR synthesis parameters as presets for convolution-based reverberation effects. It is designed for stereo and surround sound plug-ins, offering low latency and a consistent processing load for signal processing.

A VST 2 audio plug-in that performs convolution reverb using synthetic RIR generated via a Genetic Algorithm was proposed by Ly and Villegas [302]. The plug-in generates the initial room impulse response through a custom Gaussian noise method and then evolves it through truncation selection, randomly weighted average crossover and mutation via Gaussian multiplication to generate RIRs similar to those recorded in the real world. The plug-in uses parameters such as

reverberation time, early decay time and clarity to determine the fitness value of the potential RIR so that the user can control the shape of the final RIR. In addition, it is possible to generate a stereo RIR by assigning two different RIRs to the left and right channels.

Ducceschi et al. [303] proposed a plug-in that does not depend on the sampled impulse responses or the delay algorithms. It uses a purely physical model of a rectangular plate as the core of a real-time effects plug-in and uses the dimensions of the plate, tension and decay rate as user adjustable parameters. The plug-in can dynamically change input and output positions during runtime, and the model is optimised via central processing unit (CPU) vector intrinsics to allow the plug-in to compute large-scale plate reverbs in real time on consumer hardware.

A real-time audio plug-in was developed by MathWorks, Inc. [304] based on the Dattorro plate-class reverberation topology [179,305] to add reverb to input audio. It is a great example of an artificial reverb plug-in with the flexibility to adjust the output reverb with parameters such as PreDelay, Diffusion, DecayFactor and WetDryMix et al.

5.2 Implementation of the Audio Plug-in

Based on the evaluation results in Chapter 4, the HMS reverberation algorithm worked best overall in simulating measured reverberation, the HMS and Schroeder reverberation algorithms were closer to measured reverberation for short and medium reverberation times, while the Gardner reverberation algorithm was perceived to be closer to measured perceptual reverberation for long reverberation times. However, when converting the HMS algorithm's medium- and large-size-room structures into an audio plug-in, a significant number of underrun samples

occurred. 'Underrun' refers to a situation where the audio processing cannot keep up with the real-time playback, causing the output buffer to run out of data, which can result in audio dropouts or glitches [306]. This issue arose because the six parallel low-pass feedback comb filters in its reverberation structure required a substantial amount of computational time. To achieve an optimal audio plug-in that balances reverberation effect and computational efficiency, this chapter focuses on a combined approach. Specifically, for a real-time, user-oriented interactive plug-in, the HMS reverberation algorithm is used to simulate small rooms, the Schroeder algorithm is employed for medium rooms, and the Gardner algorithm is applied for large rooms, based on reverberation time ranges as indicated in Table 4.7. By integrating these structures, the plug-in can deliver high-quality reverberation while maintaining efficient performance. The presented plug-in in this section is called the HMSG reverb plug-in. The combined algorithms were implemented using MATLAB's audio plug-in. In the MATLAB environment, an audio plug-in refers to a class that is derived from the 'audioPlugin' base class or the 'audioPluginSource' base class [307].

5.2.1 Adjustable Parameters of the Audio Plug-in

In this implementation of the reverb plug-in, five parameters are used to adjust the output reverb audio: the reverb time (RT_{60}), the left and right channel correlation coefficient ($Corre_{coe}$), the ITDG extension ($ITDG_{extension}$), the direct sound scale ($Direct_{scale}$) and the reverb scale ($Reverb_{scale}$). The selection of these five parameters is based on the impulse response parameters covered in Chapters 3 and 4. Reverberation time is a primary input parameter of the reverberation algorithm, and it is the main parameter controlling the reverberation characteris-

tics. The left and right channel correlation coefficient controls the IACC, affecting spatial width and envelopment. the ITDG is crucial for perceptual reverberation, helping to define the perceived size of the space. Adjustments to the direct sound scale and the reverb scale control the D/R energy ratio, which adjusts the balance between the direct and reverberant sounds. Existing reverb plug-ins typically offer a wide range of parameters, which can vary depending on the complexity and intended use of the plug-in. Some of the most common parameters include: reverb time, pre-delay, room size,damping, diffusion, early reflections level, and reverb mix (dry/wet). This plug-in focuses on perceptually relevant parameters like IACC, ITDG, and D/R energy ratio, which directly shape perceived reverberation characteristics. This streamlined approach allows for precise spatial manipulation while avoiding excessive parameter complexity, making it more intuitive yet still flexible for professional use. These five parameters are controlled in the graphical user interface (GUI) by five horizontal sliders, as shown in Figure 5.1. The value of each parameter can be changed by dragging the corresponding slider or by entering the corresponding value in the value box. Reverb time ranges from 0.1 to 5 s, ITDG extension ranges from 0 to 0.1 s and the other three parameters range from 0 to 1.

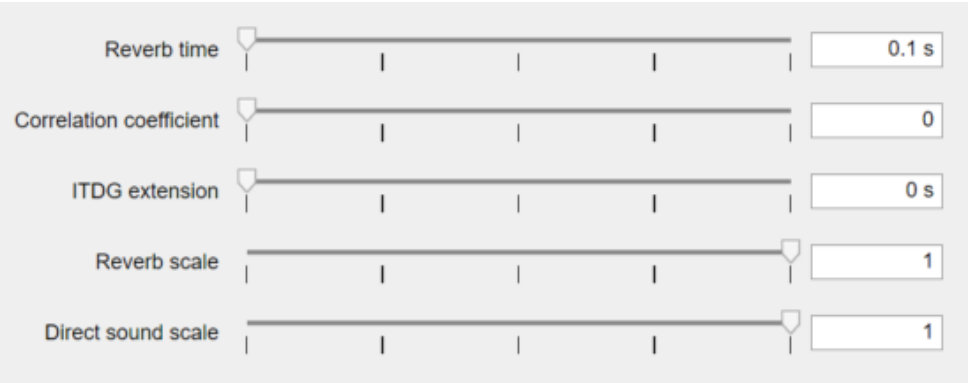


Figure 5.1: The GUI of the reverberation audio plug-in.

The reverberation component of the plug-in is generated by three artificial reverberation algorithms. When the reverberation time parameter ranges from 0.1 to 0.58 s, the plug-in uses the HMS small room reverberation algorithm internally. When the reverberation time parameter range is 0.58 to 1.3 s, the plug-in uses the Schroeder reverberation algorithm. When the reverberation time parameter range is greater than 1.3 s, the plug-in uses Gardner large room reverberation algorithm.

The direct sound component is generated by convolving the dry sound input with the normalised minimum phase reconstructed version of the 0 degree azimuth and 0 degree elevation HRIR of subject D1 from the SADIE II database [288]. The minimum-phase version is taken to ensure the energy of the impulse response is moved to the start of the filter and that the impulse response representing the direct sound is located before the reflected sound to ensure that the adjustment of the ITDG is achieved by delaying the reflected sound. The minimum phase reconstruction version is implemented via the 'rceps' function [308] in MATLAB.

The reverberation component is delayed by a delay function to realise the ITDG extension. When the parameter $\mathrm{ITDG}_{\mathrm{extension}}$ is 0, the reverberation component is not delayed on the original ITDG. When the $\mathrm{ITDG}_{\mathrm{extension}}$ is not 0, the reverberation component is delayed by $\mathrm{ITDG}_{\mathrm{extension}} \times$ sample rate samples on the original ITDG. For example, if ITDG is extended by 0.1 s and the audio sample rate is 44.1 kHz, the reverberation component is delayed by 4410 (0.1×44100) samples on the original ITDG.

The $\mathrm{Direct}_{\mathrm{scale}}$ and the $\mathrm{Reverb}_{\mathrm{scale}}$ control the level of the direct sound component and the reverberation component, allowing the D/R energy ratio to be adjusted. Both the direct sound and reverb scales default to 1. The user can adjust the scale of the direct sound and reverberation components separately to

experience the acoustic effects at different D/R energy ratios. The $\text{Direct}_{\text{scale}}$ and the $\text{Reverb}_{\text{scale}}$ are implemented by

$$\text{output}_{\text{MD}} = \text{Direct}_{\text{scale}} \times \text{output}_{\text{Dir}} + \text{Reverb}_{\text{scale}} \times \text{output}_{\text{Reverb}}, \tag{5.1}$$

where $\text{output}_{\text{MD}}$ is the mixed output audio, $\text{output}_{\text{Dir}}$ is the direct sound component and $\text{output}_{\text{Reverb}}$ is the reverberation component.

The left and right channel correlation coefficient ($\text{Corre}_{\text{coe}}$) is used to estimate the IACC. It reflects, to some extent, the spatial impression of the reproduced sound [156]. Since the reverberation component of the impulse response of the right channel is generated based on the reverberation component of the impulse response of the left channel, and since the direct sound component used is itself somewhat correlated, there is some correlation between the two channels, i.e. IACC is not 0. Therefore, even if $\text{Corre}_{\text{coe}}$ is set to 0, the IACC can only reach a minimum value. As the $\text{Corre}_{\text{coe}}$ between the left and right channels increases, the IACC also increases. A $\text{Corre}_{\text{coe}}$ of 1 for the left and right channels means that the left channel is the same as the right channel, i.e. the IACC is also 1. The correlation between the left and right channels is implemented by Equation 5.2.

$$\text{output} = \text{output}_{\text{MD}} + \text{output}_{\text{MDflipped}} \times \text{Corre}_{\text{coe}}, \tag{5.2}$$

where output is the final audio output, $\text{output}_{\text{MDflipped}}$ is the flipped column of the mixed output audio around the vertical axis in the left-right direction, and $\text{Corre}_{\text{coe}}$ is the left and right channel correlation coefficient mentioned above.

5.2.2 Implementation of the Reverberation Component

Four different methods were utilised for the implementation of the plug-in. Three of them did not satisfy the computational efficiency or created an unrepresentative reverberation effect. Therefore, the adopted implemented was the one with the most efficient computation and the best reverberation effect. All four methods are described below.

1. **Filter the streamed input audio through the 'filter' function**

   The first method attempted to apply reverberation by directly filtering the streamed input audio using MATLAB's built-in 'filter' function [309]. The expectation was that this function would apply the necessary filtering operations to introduce reverberation effects in real-time. However, this method resulted in an output with an amplitude of zero, meaning no reverberation effect was produced.

   Reverberation effects are commonly implemented using convolution with an impulse response or filtering techniques. The 'filter' function in MATLAB applies a set of polynomial coefficients to the input signal, making it a simple and computationally efficient choice for real-time processing. It was expected that applying filter directly to the input audio stream would allow for efficient and continuous reverberation.

   Unlike traditional batch processing, real-time audio plug-ins process sound in small chunks (frames) rather than as a continuous signal. In this case, the input audio was divided into frames of varying sizes, typically following a power-of-two structure (from 2^1 to 2^{13}, $2^2 - 1$ to $2^{13} - 1$, $2^1 + 1$ to $2^{13} + 1$). Each frame was processed individually before moving to the next, ensuring

low-latency real-time performance.

The 'filter' function expects a continuous input signal. It applies polynomial coefficients across the entire waveform rather than small independent frames. It requires past samples to compute the current output. Since the audio plug-in processes each frame separately, the 'filter' function lacks access to previous frames, making it unable to accumulate the reverberation effect over time. Each time a new frame was processed, the system reset, meaning any previously processed data was lost. Since reverberation builds up over time, the 'filter' function failed to maintain a continuous decay and sustain effect. This causes it to produce an output with all-zero values, resulting in an output of silence.

This method is unsuitable for real-time audio streaming because it does not preserve internal states between frames. While the 'filter' function is useful for processing entire audio signals in an offline setting, it fails to apply reverberation effectively in a frame-based real-time environment. This limitation required exploring alternative approaches that could maintain state persistence and allow reverberation to accumulate over multiple frames.

2. **Filter the discrete time unit impulse through the 'filter' function**
   The second method explored for implementing the audio plug-in involved using the 'filter' function [309] in MATLAB to filter the discrete time unit pulse in order to generate a BRIR. This BRIR would then be convolved with the input streaming audio to produce the reverberation effect. However, this approach proved to be computationally inefficient and resulted in significant issues, such as a large number of underrun samples and partial silences in

the output signal [306].

The problem arises from the constraints of real-time audio processing. When processing streamed audio, each frame of audio data must be processed within a specific time budget, which is determined by the formula from the document in MATLAB [310]:

$$\text{Time Budget} = \frac{\text{Frame Size}}{\text{Sampling Rate}} \tag{5.3}$$

In this method, the BRIR is generated by filtering an unit impulse signal whose size is determined by the product of the reverberation time and the sampling rate:

$$\text{Impulse Size} = \text{Reverberation Time} \times \text{Sampling Rate} \tag{5.4}$$

The issue is that the filters used in the reverberation algorithm take significantly more time to process this long impulse signal than the available time budget per frame allows. As a result, the processing cannot keep up with the real-time requirements, leading to underrun samples. Underrun samples occur when the audio processing cannot produce output data quickly enough, resulting in gaps or silences in the output signal.

Due to this inefficiency and the inability to meet the real-time processing demands, this method also proves unsuitable for implementing the audio plug-in. The need to perform computations within the strict time limits of real-time audio processing means that alternative methods must be explored to achieve the desired reverberation effects without compromising the output quality.

3. **Filter the streamed input audio through the 'dsp.IIRfilter' system object**

   The third method explored for implementing the audio plug-in involved directly filtering the streamed input audio using MATLAB's 'dsp.IIRFilter' system object [311]. This system object is optimised for speed and is designed to improve computational performance, making it a potentially viable option for real-time audio processing.

   The 'dsp.IIRFilter' system object works effectively when dealing with filters that have small polynomial coefficients, typically meaning filters with a low number of terms and coefficients that are relatively low in magnitude. Such coefficients are easy and quick to process, resulting in efficient real-time performance. For instance, filters with polynomial coefficients of a few terms (e.g., 2nd or 3rd order polynomials) and small values (e.g., coefficients in the range of single to low double digits) are considered small and are handled efficiently by the 'dsp.IIRFilter'.

   However, the reverberation algorithms being implemented involve a large number of comb and all-pass filters, which, by nature, require polynomial coefficients composed of large, sparse vectors. A large sparse vector might have hundreds or thousands of coefficients, and with many zeros interspersed among them, and only a few non-zero values spread throughout. These large sparse vectors significantly increase the computational complexity of the filtering process because the 'dsp.IIRFilter' must still process all positions in the vector, including the zeros, leading to inefficiencies.

   As a result, even though 'dsp.IIRFilter' is optimised for speed, it struggles

to efficiently process these large sparse vectors. This inefficiency leads to an excessive amount of time being required to process each frame of audio data. Therefore, while the 'dsp.IIRFilter' system object offers improved performance in scenarios with small polynomial coefficients, it is not suitable for this particular application due to the computational demands imposed by the large polynomial coefficients required by the reverberation algorithms. The resulting underrun samples [306] and partial silences in the output signal mean that this method does not meet the requirements for implementing the audio plug-in effectively.

4. **Construct recursive equations for algorithm filters**

   The fourth approach explored for implementing the audio plug-in involved constructing recursive equations for the filters used in the reverberation algorithms. This method involves processing the streamed input audio on a sample-by-sample basis using a pointer-indexed method. Essentially, instead of processing entire frames of audio data at once, each individual sample is processed sequentially, with the filters applied recursively.

   This method draws inspiration from the reverberator plug-in based on the Dattorro reverberation algorithm developed by MathWorks, Inc [304]. By constructing recursive equations for those filters used in the Dattorro algorithm, the processing is broken down into smaller, more manageable operations, reducing the overall computational load.

   In this method, the key filters in the reverberation algorithm, such as comb filters and all-pass filters, are implemented using pointer-based indexing. This allows for efficient memory access and processing, as each sample is

processed using previously computed values stored in memory. The recursive nature of the equations ensures that the filter's response is continuously updated as new audio samples are processed, maintaining the correct reverberation effect.

To further enhance computational efficiency, these recursive filter equations are packaged into functions that can be compiled into MATLAB executables, commonly known as mex files. Mex files are highly optimised for performance because they allow MATLAB code to run at speeds comparable to C or C++ programs. By compiling the filter functions into mex files, the processing speed is significantly improved, enabling the real-time processing required for the audio plug-in.

The first three methods encountered significant challenges, primarily due to computational inefficiencies and the inability to process audio within the strict time constraints required for real-time applications, leading to either the absence of a reverberation effect or poor-quality audio with gaps and silences.

In contrast, the fourth method, which involved constructing recursive equations for the filters and optimising them with mex files, effectively addressed these issues. This approach offered several key advantages, including increased computational efficiency by processing each audio sample individually and using optimised mex files, ensuring that the processing time per frame remained within the available budget, which is crucial for minimising latency and maintaining continuous, gap-free audio output. Additionally, the recursive processing accurately replicated the behavior of the filters, producing the correct reverberation effect. The method's flexibility and scalability, due to its sample-by-sample processing approach, allowed

for easy adaptation to different reverberation algorithms and filter configurations, making it a versatile solution for various audio plug-ins. Given these benefits, the fourth method was chosen as the optimal approach for generating the reverberation audio plug-in, ensuring both high computational efficiency and accurate audio output in real-time applications.

5.3 Objective Analysis

In order to explain in detail the effect of these dynamic parameters on the artificial reverberation, specifically, how the parameters of the impulse response generated by the HMSG plug-in vary as the reverberation measured at different positions in space, the acoustic parameters of measured impulse responses were compared and analysed with those of the artificial impulse response generated by this dynamic reverberation plug-in. In this chapter, the multi-point binaural room impulse responses measured in the meeting room and lecture room from the Aachen Impulse Response Database [312] and the grid impulse responses measured in the classroom from the Database of Omindirectional and B-format Room Impulse Responses [313] were used to represent the impulse response demonstrations of small, medium and large rooms, respectively. These selected rooms provide high-quality, multi-point and grid impulse responses, which are crucial for analysing reverberation across various acoustic conditions.

According to the original materials [312, 313], all impulse responses in the Aachen Impulse Response Database are 48 kHz and 24 bit WAV audio files, while those in the Database of Omindirectional and B-format Room Impulse Responses are 96 kHz and 32 bit WAV audio files. However, upon verification, it was found that the classroom impulse responses from the the Database of Omnidirectional

and B-format Room Impulse Responses were actually 96 kHz and 24-bit, rather than the reported 96 kHz and 32-bit. To control for variables in subsequent analyses, all impulses used were converted to 48 kHz and 16 bit WAV audio files.

The impulse responses from the Database of Omnidirectional and B-format Room Impulse Responses were in B-format. However, the impulse responses used in this research required conversion to a binaural format to fit the context of binaural audio applications, so they were decoded into BRIRs [314] to represent binaural rendering.

This decoding process was performed using MATLAB's ambisonic-to-binaural decoding function [314], which involves filtering the B-format signals with HRTFs to simulate how sound would reach the ears of a listener in a 3D space. The HRTF dataset used in this decoding process is derived from the ARI NH2 subject in the HRTFs database of the Acoustics Research Institute of the Austrian Academy of Sciences [315], because it is provided and supported by MATLAB's documentation, ensuring compatibility and ease of integration within the MATLAB environment. The decoding format employed channel ordering in FuMa format and normalisation in maxN format [316] to maintain compatibility with common ambisonic practices and to ensure that the resulting binaural signals accurately capture the spatial characteristics of the original B-format recordings.

The impulse response measurement layout of the meeting room, representing the small room (with a short reverberation time) demonstration, is shown in Figure 5.2. This room is 8 m long, 5 m wide and 3.1 m high. Since there are only five impulse response measurement points at different locations in the small room, all five impulse responses were selected for the comparison of the grid impulse response parameters for the small room. The measurement is conducted by keeping the

dummy head measurement system stationary and capturing the impulse responses from loudspeakers placed at different positions on the opposite side of the head. The distance of the loudspeaker at the five different positions from the dummy head measurement system is, in order, 1.45 m, 1.7 m, 1.9 m, 2.25 m and 2.8 m. The five positions are referred to here as positions 1 to 5, in order from closest to farthest from the dummy head. It's important to note that the large table in the center of the room may introduce additional reflections and diffraction effects, which could influence the measured impulse responses by adding complexity to the sound field.

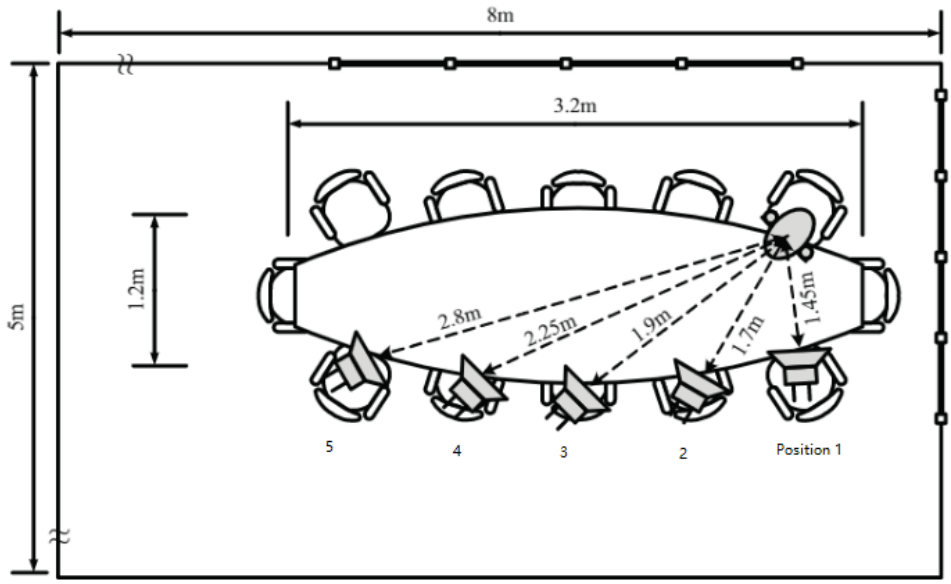


Figure 5.2: The impulse response measurement layout of the meeting room (small room demonstration) [312].

The impulse response measurement layout of the lecture room, representing the medium room (with a medium reverberation time) demonstration, is shown in Figure 5.3. This room is 10.9 m long, 10.8 m wide and 3.15 m high. Similar to the small room, the impulse responses measured at these five measurement points

were all selected for comparison of the grid impulse response parameters. In this case, the loudspeaker is fixed at the lectern, and the microphone system is moved to different rows with increasing distances from the loudspeaker. The distances between the loudspeaker and the microphone at these positions are 4 m, 5.56 m, 7.1 m, 8.68 m and 10.2 m, referred to as positions 1 through 5 from nearest to farthest. The presence of benches within the room could also impact the impulse responses by introducing additional reflections, diffraction, and possible obstructions, which may alter the acoustic characteristics captured by the measurements. However, since the parameters were derived from real-world measurements and the HMSG parameters were adjusted after parameter extraction, the influence of the table and benches is inherently accounted for in the measured data. As a result, these objects do not significantly impact the ability to match reverberation parameters in terms of overall room characteristics.

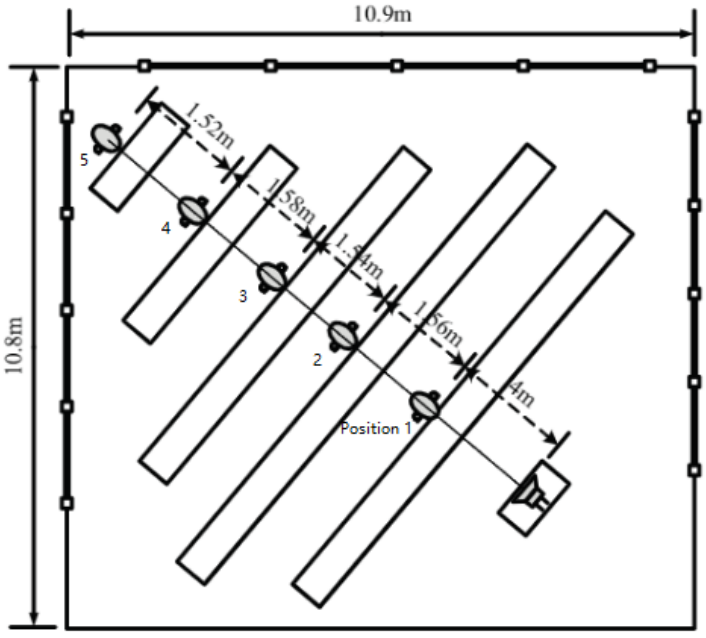


Figure 5.3: The impulse response measurement layout of the lecture room (medium room demonstration) [312].

The impulse response measurement layout of the classroom, representing the large room (with a long reverberation time) demonstration, is shown in Figure 5.4. This room has approximate dimensions of 9 m by 7.5 m by 3.5 m. Unlike the small and medium rooms, this layout features a much denser grid of measurement points, totaling 130 different locations. These points are arranged in a grid with 10 rows and 13 columns, spaced 50 cm apart, with the 8th column aligned directly with the axis of the loudspeaker. For the purposes of parametric analysis, only 9 specific locations, highlighted in red in Figure 5.4, are selected. These positions, labeled from '00x00y' to '60x45y', are referred to as positions 1 through 9, arranged from left to right and bottom to top. The extensive number of measurement positions in this large room setup allows for a far more detailed and comprehensive analysis of the room's acoustic properties compared to the smaller and medium rooms, capturing a finer spatial resolution of the sound field.

Since the impact thresholds of the parameters ITDG, early reflections and late reverberation of the room impulse response have been evaluated in Chapter 3, this chapter focuses on the objective analysis of the parameters directly related to these three components. This chapter lists the RT_{60}, ITDG, D/R energy ratio and IACC of the impulse responses measured at different locations in the same space and checks whether the HMSG reverberation plug-in can match, or closely approximate, the measured impulse response parameters by allowing the user to dynamically adjust the corresponding acoustic settings.

Reverberation time can vary across different frequency bands, and different frequencies can have different decay rates within the same room. However, to simplify analysis and ensure consistent comparison, it is common practice to use a single representative value for reverberation time, such as T_{30} at 500 Hz or

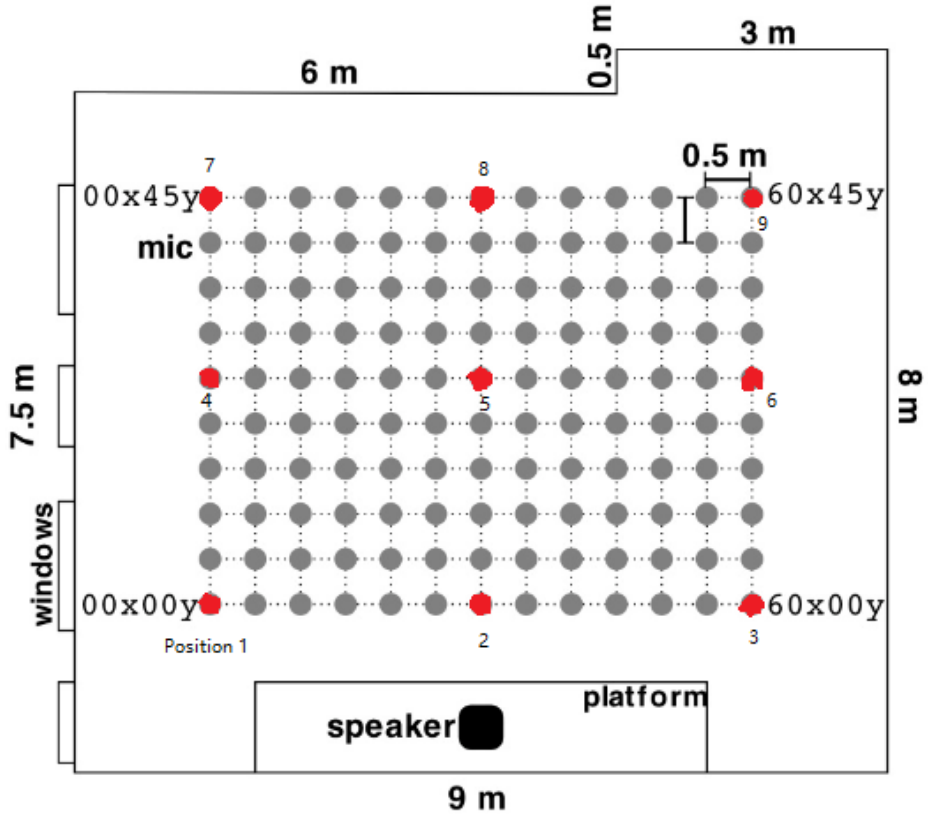


Figure 5.4: The impulse response measurement layout of the classroom (large room demonstration) [313].

1000 Hz [293]. In some cases, the mean value of T_{30} across these bands is taken to provide a more comprehensive evaluation of the room's acoustics [293]. In this chapter, T_{30} in the 1000 Hz octave band is selected as the representative value for evaluating the reverberation time, but the discussion is extended to include multiple frequency bands to ensure a detailed evaluation of the plug-ins' performance in replicating varied acoustic environments. Both ITDG and D/R energy ratio are calculated from the direct sound and reverberation components. The direct sound component and the reverberation component of the measured BRIRs are derived from Equation 4.29. The direct and reverberation components

of the BRIRs generated by the HMSG plug-in are originally two separate parts, so there is no need to calculate them via Equation 4.29.

The ITDG is calculated from Equation 5.5

$$\text{ITDG} = (\text{mean}(T_1) - \text{mean}(T_0))/\text{Fs}, \tag{5.5}$$

where ITDG is the average ITDG value of the left and right channels of the binaural room impulse response, mean represents the average of the left and right channels, T_1 is the sample value corresponding to the time of the first reflection impulse, T_0 is the sample value corresponding to the time of the direct impulse, and Fs is the sample rate.

The D/R energy ratio is calculated from Equation 4.30. The IACC is calculated using MATLAB's library function *xcorr* [295] as shown in Equation 4.31.

The HMSG reverb plug-in proposed in this chapter matches the reverberation time of the measured binaural impulse responses by adjusting the parameter RT_{60}, the delay between the direct sound and the first reflection of the measured binaural impulse responses by adjusting the parameter $\text{ITDG}_{\text{extension}}$, the D/R energy ratio of the measured binaural impulse responses by adjusting the parameters $\text{Direct}_{\text{scale}}$ and $\text{Reverb}_{\text{scale}}$, and by adjusting the parameter $\text{Corre}_{\text{coe}}$ to match the interaural cross-correlation coefficient of the measured binaural impulse responses.

The parameter values of the BRIRs measured at different locations in the small, medium, and large rooms, as shown in Figures 5.5, 5.6 and 5.7, are compared with those generated by the HMSG plug-in. This comparison is made to determine how well the HMSG plug-in replicates the variations in acoustic parameters as the listener moves to different locations within the space. It should be noted that the values of the D/R energy ratio in Figures 5.5, 5.6, and 5.7 are converted from

proportional values to logarithmic values (unit in dB) by Equation 5.6, in order to ensure consistency with the unit of the JND.

$$\mathrm{DRR_{dB}} = 10\log(\mathrm{DRR}) \tag{5.6}$$

Taking the binaural room impulse response at position 1 in the small room as an example, the HMSG plug-in was first adjusted to match the measured reverberation time. The $\mathrm{RT_{60}}$ parameter in the plug-in was set to 0.34 s, which resulted in a calculated $\mathrm{RT_{60}}$ of 0.37 s, an ITDG of 0.0042 s, a D/R energy ratio of 6.1853, and an IACC of 0.3349.

To better align with the measured ITDG of 0.0059 s, the $\mathrm{ITDG_{extension}}$ was adjusted to 0.017 s (the difference between the desired and current ITDG). However, since the minimum ITDG value in the generated impulse response is fixed, if this minimum value is greater than the ITDG of the measured impulse response, further reduction to match the measured value is not possible. After this adjustment, the D/R energy ratio increased to 8.3905 and the IACC to 0.4781.

The D/R energy ratio of the generated impulse response was 0.281 times that of the measured impulse response. To match the measured D/R energy ratio of 29.8688, the $\mathrm{Reverb_{scale}}$ was adjusted to 0.281, which brought the D/R energy ratio close to the target value of 29.8688, resulting in a final calculated value of 29.8596. If the generated D/R energy ratio is higher than the measured value, the $\mathrm{Direct_{scale}}$ should be reduced to match it.

At this point, the IACC reached 0.8420, but since the $\mathrm{Corre_{coe}}$ had already been reduced to zero, it could not be adjusted further downward. If the generated IACC is lower than the measured value, it is possible to increase it by adjusting the $\mathrm{Corre_{coe}}$ upwards. However, since the exact relationship between the $\mathrm{Corre_{coe}}$ and

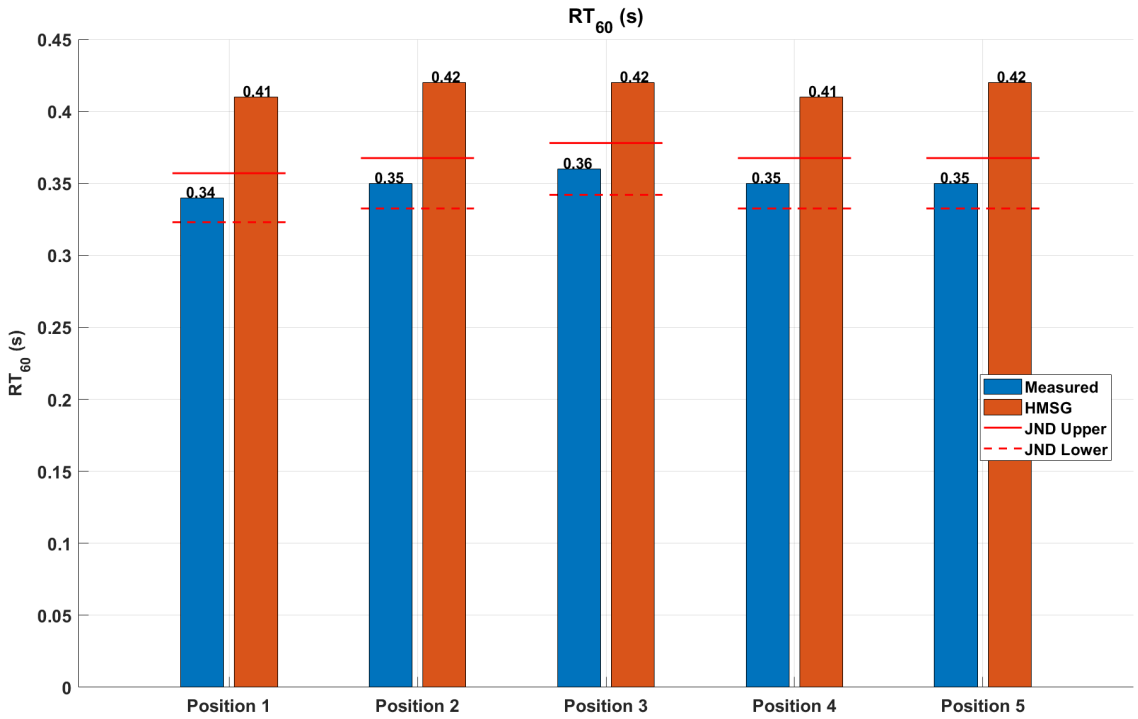
IACC is not easily predictable, fine adjustments would rely on subjective auditory perception.

Finally, because adjustments to other parameters can influence the reverberation time, the resulting impulse response parameters were: RT_{60} of 0.41 s, ITDG of 0.0059 s, D/R energy ratio of 29.8596, and IACC of 0.8420.

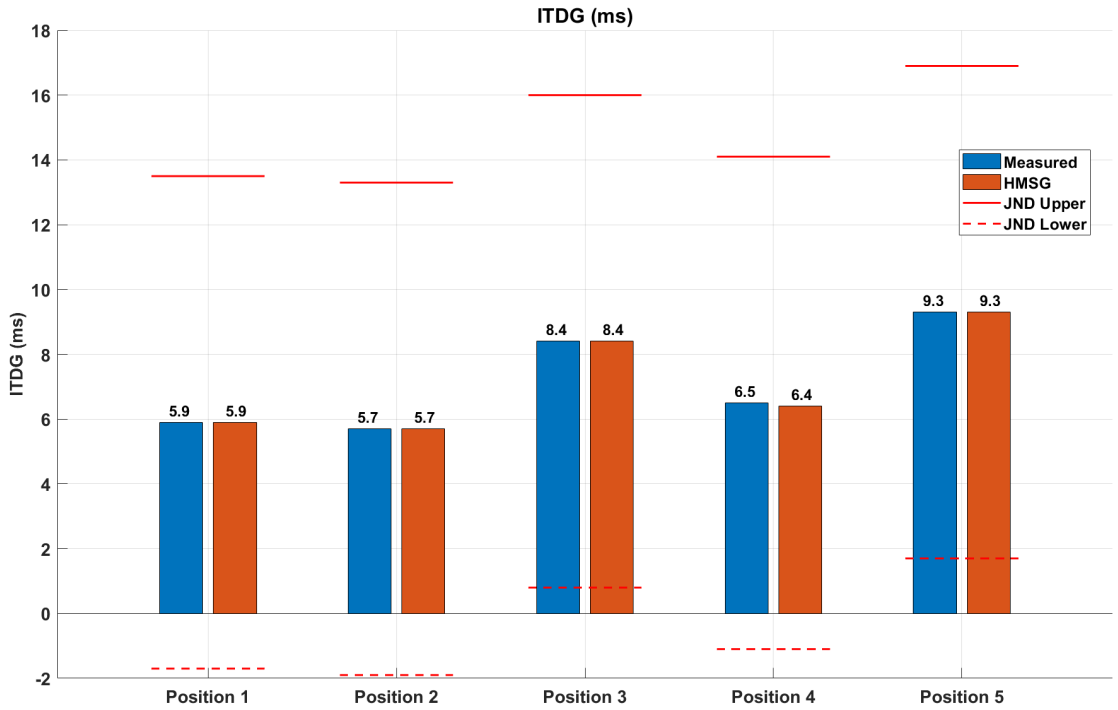
Following the outlined steps, the parameters of the impulse response measured at other locations within the small room were matched, with the results displayed in Figure 5.5. The data in Figure 5.5 (a) indicates that the HMSG plug-in generally produces RT_{60} values that are slightly longer than the measured values, with differences ranging from 0.06 to 0.07 seconds. When compared to the JND for RT_{60}, which ISO 3382-1 approximates as 5% [296], these differences (0.06 to 0.07 s) are significantly larger than the calculated 5% threshold of 0.017 to 0.018 s. Given that these differences exceed the JND, they could potentially influence the perceived acoustics, particularly in environments where precise control over reverberation is crucial. While the HMSG plug-in's performance is generally within acceptable bounds, these larger deviations could be noticeable in some situations and may impact the listener's experience.

Reichardt et al. [317] established a JND of approximately 7 ms $\pm$ 0.6 ms for the ITDG for lateral reflections. The ITDG data from the Figure 5.5 (b) indicates that the HMSG plug-in closely matches the measured values, with deviations well within this JND range. This suggests that any differences in ITDG between the measured and HMSG data are unlikely to be perceptible, and thus would have minimal impact on the perception of spatial clarity or the initial perception of sound sources.

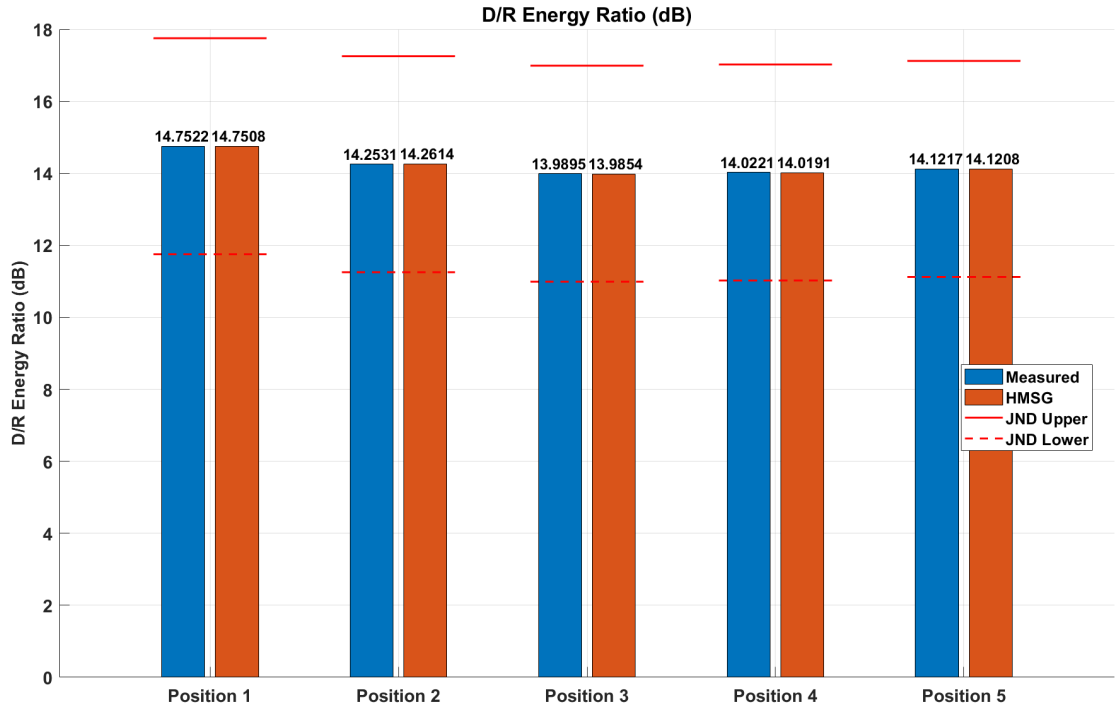
In terms of the D/R energy ratio, the smallest JNDs reported in full cue con-


RT60 (s)
RT60 (s)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.34
0.41
0.35
0.42
0.36
0.42
0.35
0.41
0.35
0.42
Position 1
Position 2
Position 3
Position 4
Position 5
Measured
HMSG
JND Upper
JND Lower

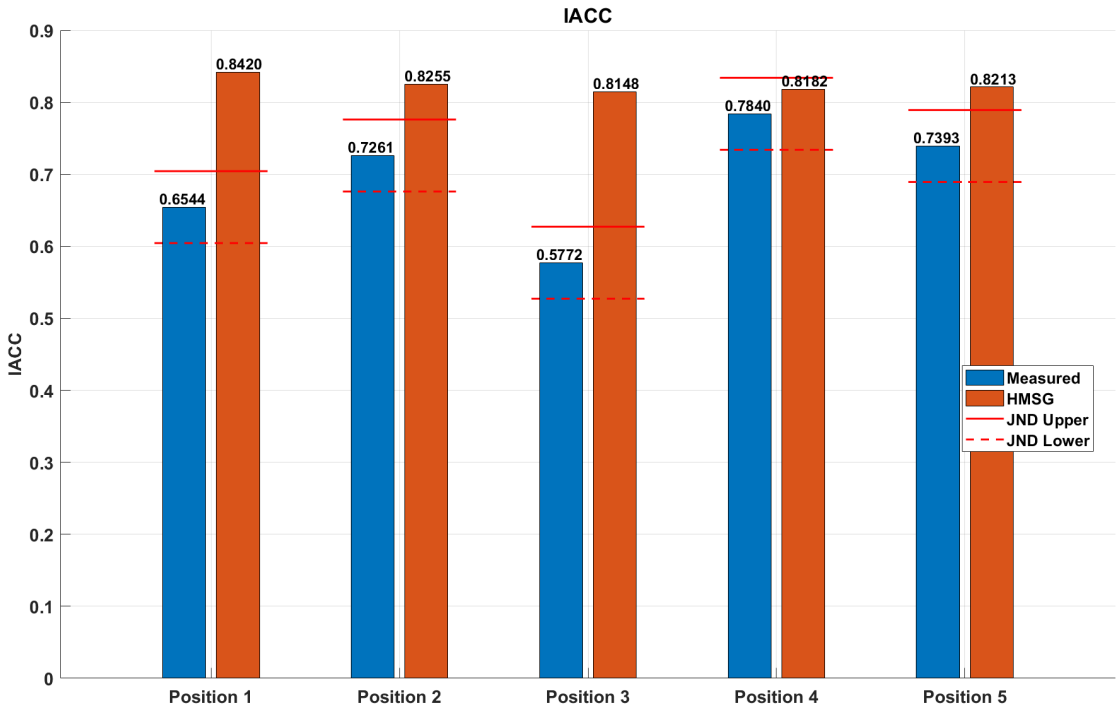
(a)


ITDG (ms)
ITDG (ms)
18
16
14
12
10
8
6
4
2
0
-2
5.9
5.9
5.7
5.7
8.4
8.4
6.5
6.4
9.3
9.3
Position 1
Position 2
Position 3
Position 4
Position 5
Measured
HMSG
JND Upper
JND Lower

(b)



(c)



(d)

Figure 5.5: Parametric comparison between measured BRIRs and HMSG-generated BRIRs at different locations in a small room. (**a**) RT_{60}. (**b**) ITDG. (**c**) D/R energy ratio. (**d**) IACC.

ditions are approximately 2-3 dB [127]. Figure 5.5 (c) shows that the differences between the measured and HMSG values for the D/R energy ratio are very small, indicating that these differences fall below the perceptual threshold. This suggests that the HMSG plug-in accurately replicates the balance between direct and reverberant sound, which is critical for maintaining overall sound clarity.

Lastly, the IACC values are critical for spatial impression, with ISO 3382-1 quoting a JNDs of 0.05 [296]. The data in Figure 5.5 (d) shows that the discrepancies of the IACC values between the measured and HMSG in some positions exceed this threshold. This indicates potential perceptual differences in spatial envelopment, which could be significant in environments where spatial qualities are a primary concern.

Figure 5.6 illustrates the reverberation times (RT_{60}) measured at different positions within the small room and compares them with the results obtained using the HMSG plug-in across a range of octave bands from 63 Hz to 8000 Hz. At 63 Hz, there is a noticeable disparity between the measured RT_{60} and those generated by the HMSG plug-in across all positions, with the HMSG values being significantly lower than the measured values. This could indicate the plug-in's limitations in accurately simulating the room's low-frequency response or an inherent challenge in the room's acoustical treatment at these frequencies. At 125 Hz, the HMSG's values start to approximate the measured values more closely, and begin to exceed the measured values at 250 Hz. This suggests that the plug-in may be better tuned for slightly higher frequencies but still struggles with the lowest bass tones. Across 500 to 4000 Hz, the HMSG-generated values close the measured values, although they are higher than them. This indicates relatively effective modelling by HMSG in these frequency ranges, where human hearing is most sensitive and where most

room interactions occur. At 8000 Hz, the generated and measured values tend to diverge again. This could be due to the plug-in not fully achieving the absorption properties of high frequencies.

Overall, while the HMSG plug-in performs relatively well in the critical mid-frequency range, adjustments are necessary to improve its accuracy, particularly at the frequency extremes.

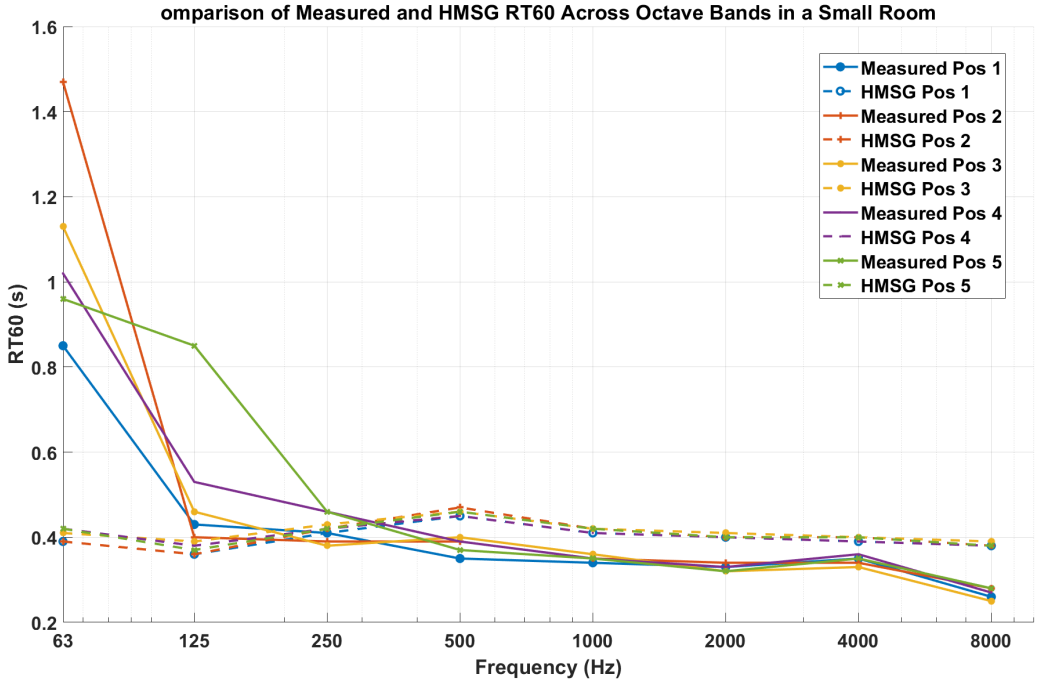
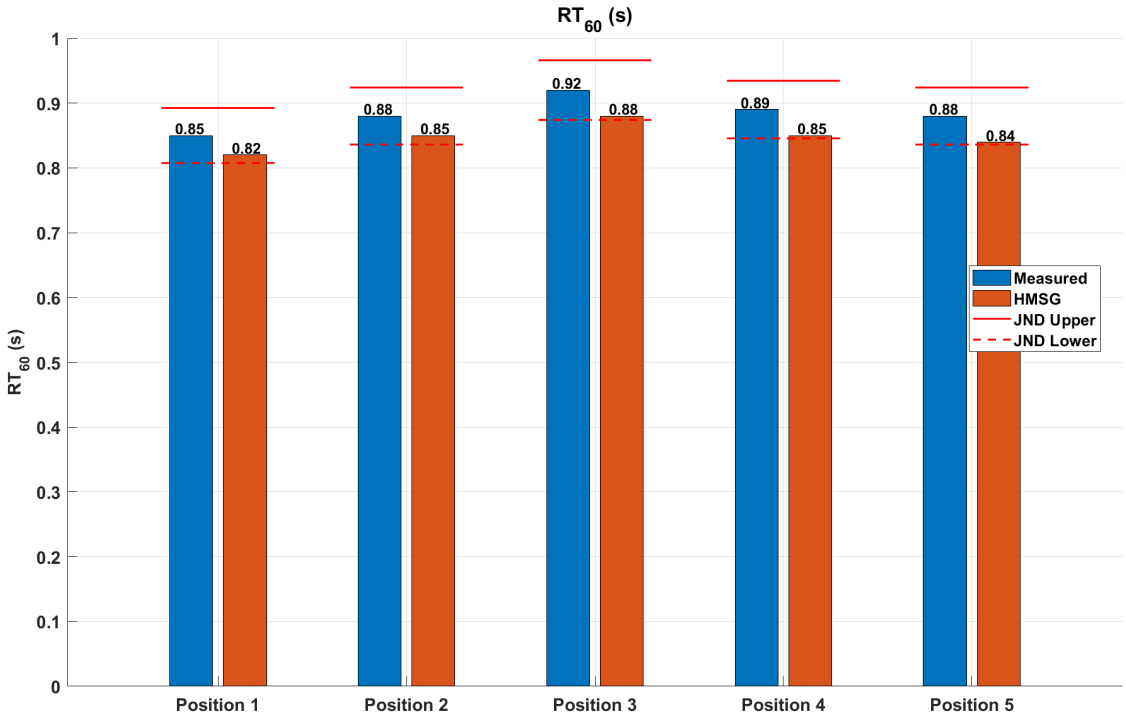


Figure 5.6: Comparison of RT_{60} across octave bands from 63 to 8000 Hz between measured and HMSG-generated impulse responses in a small room.

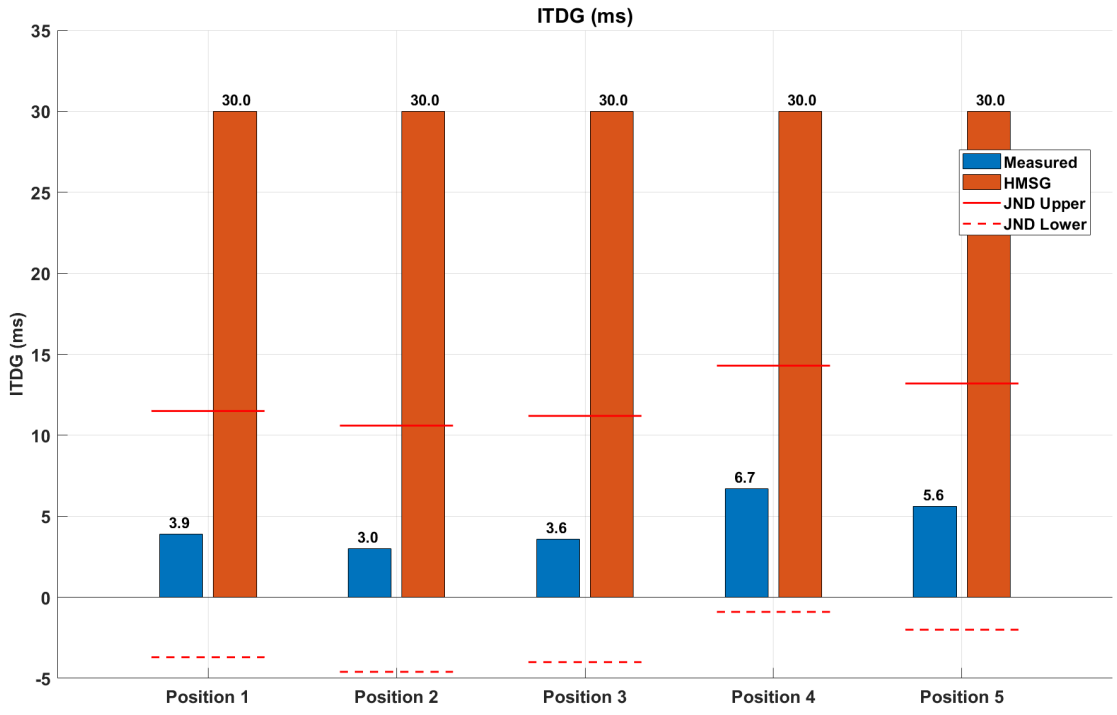
Following the same procedure as applied in the small room, the impulse response parameters measured in the medium room were matched and compared, with the results displayed in Figure 5.7. The data from Figure 5.7 (a) indicate that the reverberation time of the BRIRs generated by the HMSG plug-in is consistently 0.03 to 0.04 s shorter than the measured BRIRs. Given that ISO 3382-1 establishes a JND for RT_{60} at approximately 5% [296], the observed differences

fall slightly below this perceptual threshold (0.0425 to 0.046 s). When observed differences fall below the JND, it typically indicates that these differences are likely imperceptible to most listeners. In other words, the deviation from the measured value is so small that it does not reach the threshold where it would be noticeable by the average human ear. This suggests that the slight underestimation by the HMSG plug-in would not have a noticeable impact on the perceived acoustics of the room. Therefore, the HMSG plug-in's performance is validated as sufficiently accurate for practical purposes, as the simulated acoustics are perceptually indistinguishable from the actual measured response. For most applications, especially those that do not require extreme precision, these minor differences confirm that the simulation tool is reliable and its results are accurate enough for use. Overall, the small differences in reverberation time are unlikely to detract from the overall auditory experience, thereby validating the plug-in as a reliable tool for practical acoustic applications in similar environments.

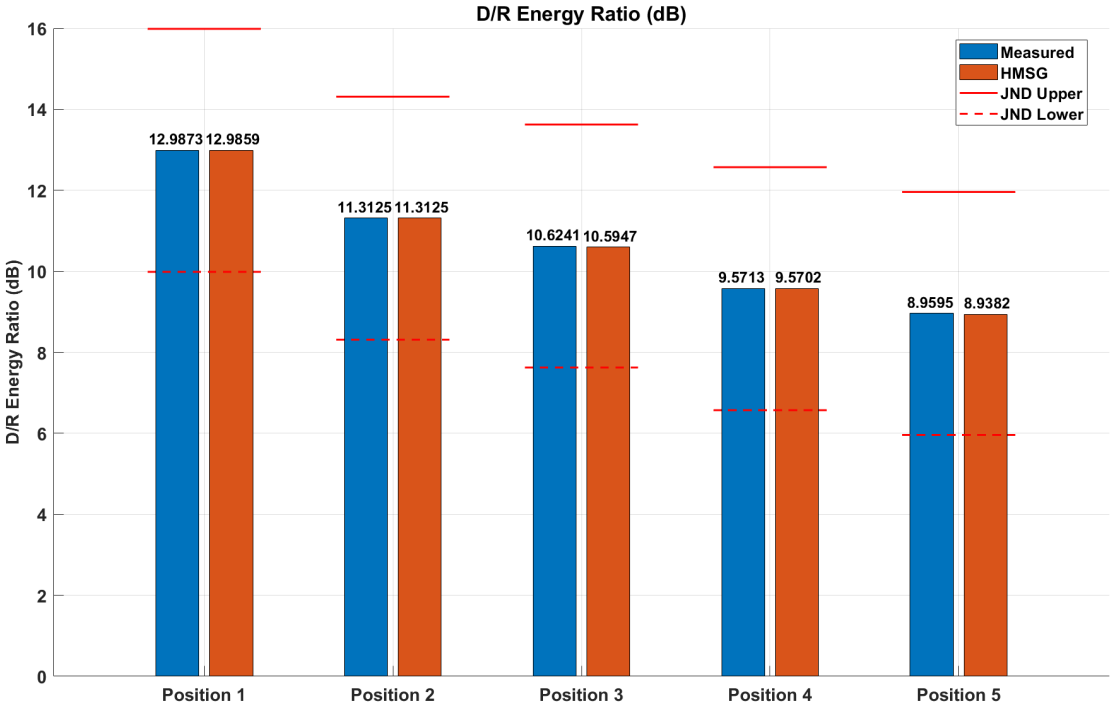
Regarding the ITDG values in Figure 5.7 (b) for the HMSG-generated impulse response, the minimum ITDG was fixed at 0.03 s, which is significantly larger than the measured ITDG values (ranging from 0.003 s to 0.0067 s across different positions). This discrepancy not only exceeds the JND threshold of approximately 7 ms $\pm$ 0.6 ms for ITDG [317] but also contradicts the discussion in Chapter 3, which emphasises the critical role of ITDG in perceptual realism. Specifically, the ITDG extension threshold for medium reverberation times was found to be approximately 25 ms, suggesting that even small deviations in ITDG can significantly affect spatial perception. The primary cause of this discrepancy lies in the use of the Schroeder reverberation algorithm to simulate medium room reverberation in the HMSG plug-in. The inherent structure of the Schroeder algorithm


RT60 (s)
RT60 (s)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.85
0.82
0.88
0.85
0.92
0.88
0.89
0.85
0.88
0.84
Measured
HMSG
JND Upper
JND Lower
Position 1
Position 2
Position 3
Position 4
Position 5

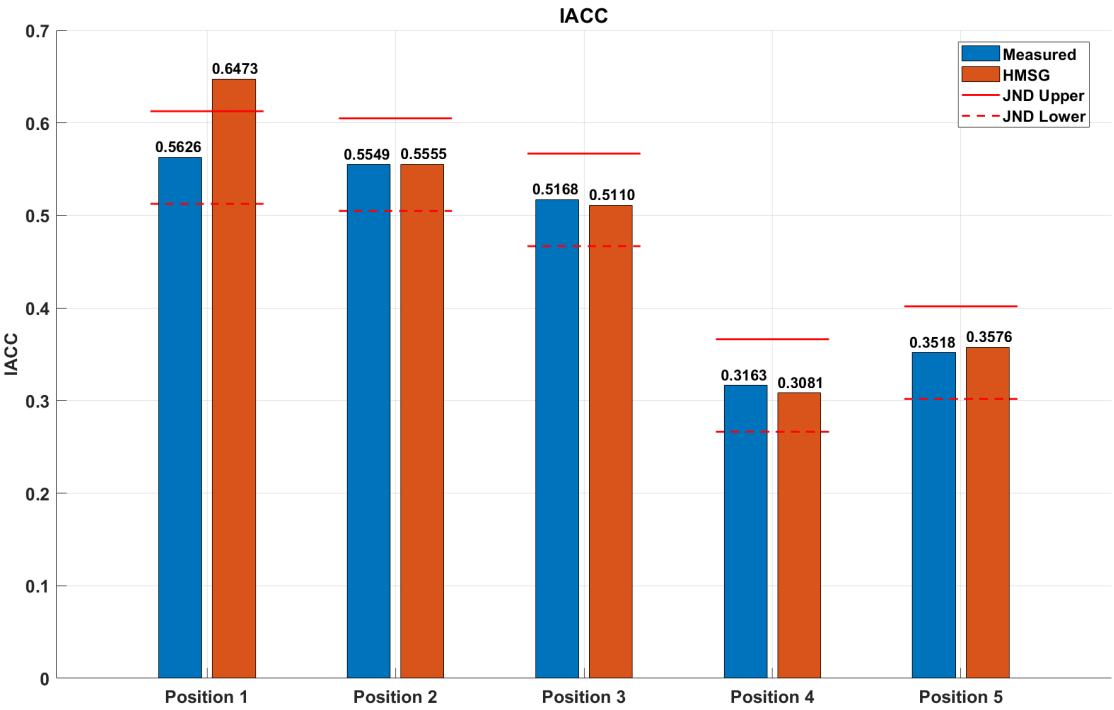
(a)


ITDG (ms)
ITDG (ms)
35
30
25
20
15
10
5
0
-5
30.0
30.0
30.0
30.0
30.0
3.9
3.0
3.6
6.7
5.6
Measured
HMSG
JND Upper
JND Lower
Position 1
Position 2
Position 3
Position 4
Position 5

(b)



(c)



(d)

Figure 5.7: Parametric comparison between measured BRIRs and HMSG-generated BRIRs at different locations in a medium room. (**a**) RT_{60}. (**b**) ITDG. (**c**) D/R energy ratio. (**d**) IACC.

introduces a delay between the direct sound and the reverberant component, leading to an artificially large initial ITDG value. This fundamental characteristic of the algorithm limits its ability to accurately replicate early reflections, which play a crucial role in spatial clarity and source localisation. This mismatch in ITDG highlights a significant challenge in simulating medium-sized rooms, where precise timing of early reflections is essential for achieving a natural and immersive auditory experience. The inability to match the measured ITDG values in HMSG may result in perceptual differences, particularly affecting the sharpness of spatial cues and the listener's ability to localise sound sources accurately. Future refinements should focus on improving ITDG flexibility to bring the simulated ITDG values closer to real-world measurements.

On the other hand, the D/R energy ratio displayed in Figure 5.7 (c) shows close alignment between the measured and simulated data. The differences in the D/R energy ratio across positions are generally within 0.03 dB, which falls well below the JND for this parameter (2-3 dB in full cue conditions [127]), suggesting that the HMSG plug-in accurately simulates the balance between direct and reverberant energy. This is crucial for maintaining speech intelligibility and overall sound clarity in medium-sized rooms.

Similarly, the IACC values shown in Figure 5.7 (d) are closely matched, with differences generally within the JND threshold of 0.05 as cited by ISO 3382 [296]. This indicates that the HMSG plug-in effectively preserves the spatial impression and envelopment, which are critical for creating a realistic and immersive auditory environment.

Figure 5.8 displays the comparison of reverberation times (RT_{60}) measured and generated by the HMSG plug-in across five different positions in a medium

room, over a range of octave bands from 63 Hz to 8000 Hz. At the lowest frequency band (63 Hz), there is a pronounced divergence between the measured and HMSG-generated RT_{60} values, with the HMSG values generally lower than the measured ones across all positions. This could indicate that the HMSG algorithm may struggle to accurately simulate lower frequency reverberations in this particular environment. As the frequency increases, particularly around the 250 Hz to 4000 Hz bands, the HMSG-generated values tend to converge more closely with the measured values. This suggests better performance in these mid-frequency ranges, where the HMSG seems to replicate the room acoustics with greater accuracy. In the highest band (8000 Hz), the RT_{60} values generated by the HMSG plug-in again deviate from the measured values.

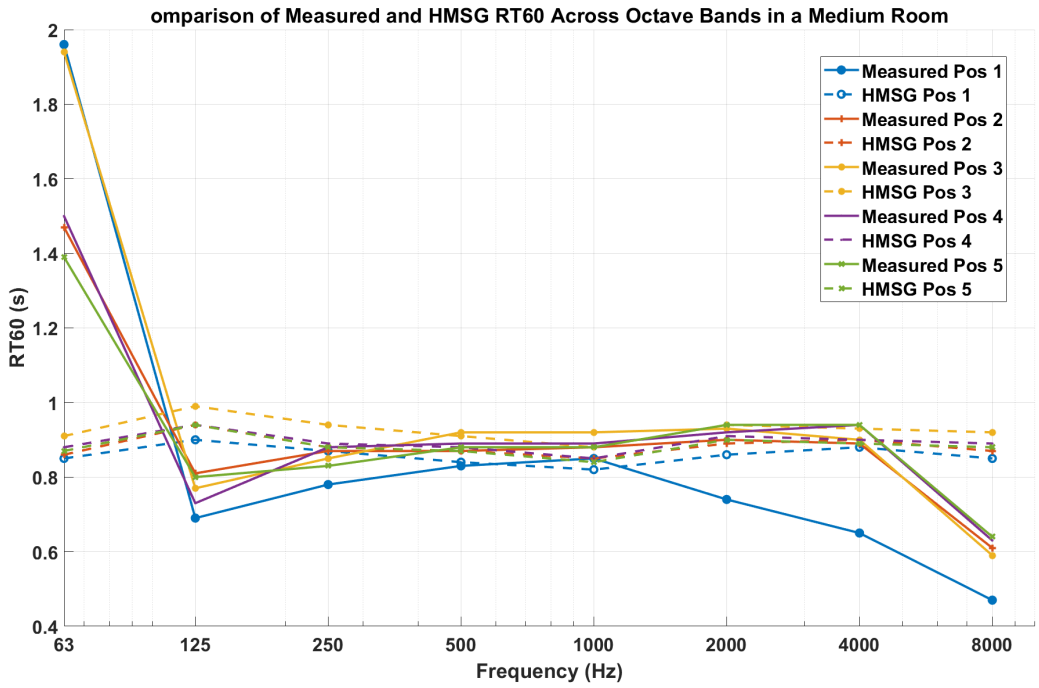


Figure 5.8: Comparison of RT_{60} across octave bands from 63 to 8000 Hz between measured and HMSG-generated impulse responses in a medium room.

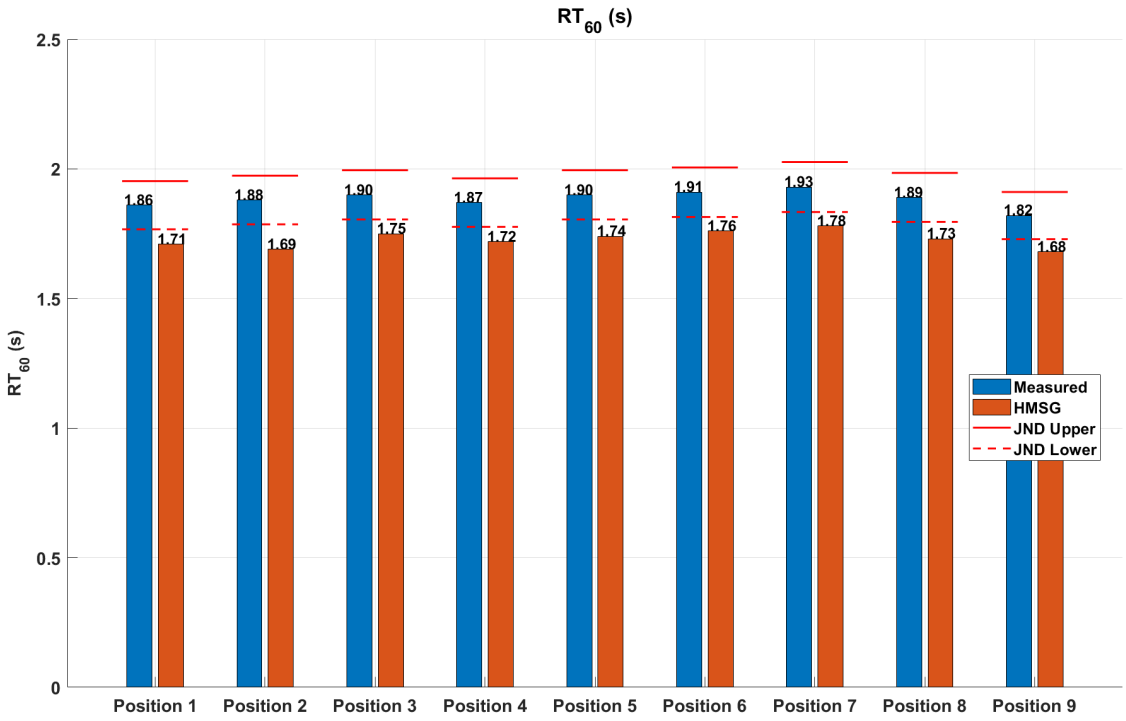
Overall, the HMSG plug-in needs further tuning and optimisation when dealing

with acoustic properties at very high and very low frequencies to improve the accuracy and reliability of the predictions.

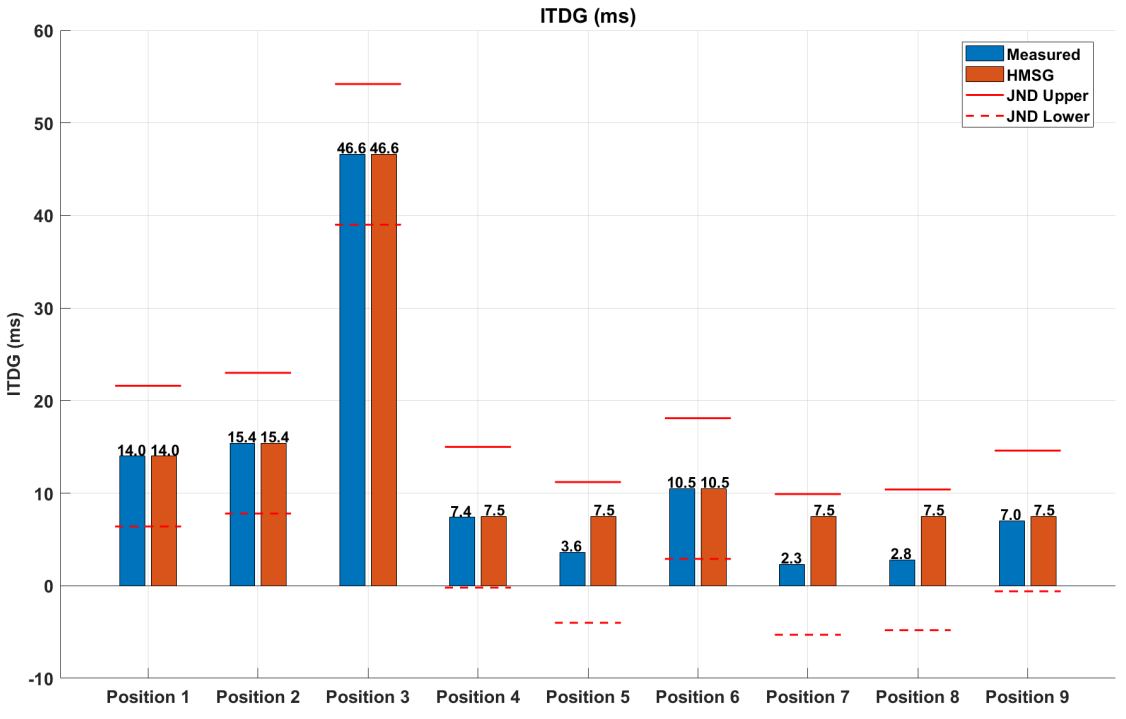
Following the same procedure as applied in the small and medium rooms, the impulse response parameters measured in the large room were matched and compared, with the results displayed in Figure 5.9. The data from Figure 5.9 (a) indicate that the reverberation time of the BRIRs generated by the HMSG plug-in is consistently 0.14 to 0.19 s shorter than the measured BRIRs. Given that ISO 3382-1 establishes a JND for RT_{60} at approximately 5% [296], for RT_{60} values ranging from 1.82 to 1.93 s, the JND would be approximately 0.091 to 0.0965 s. The observed differences exceed this perceptual threshold, suggesting that the deviations could be perceptible to listeners. This underestimation by the HMSG plug-in may lead to a noticeable impact on the perceived acoustics of the large room, particularly in environments where precise reverberation is critical. Therefore, while the HMSG plug-in generally performs well, the larger deviations in this case indicate a potential area for improvement in simulating RT_{60} for larger spaces.

Regarding the ITDG in Figure 5.9 (b), the generated impulse response had values that closely matched the measured ITDG values, with discrepancies being minimal (mostly within 0.0005 s), and all deviations remain below the JND. This close alignment suggests that the HMSG plug-in effectively replicates the ITDG, maintaining accurate temporal reflections. The precise matching of ITDG is crucial for preserving spatial clarity and the accurate localisation of sound sources, which is especially important in larger rooms where reflections can significantly influence auditory perception.

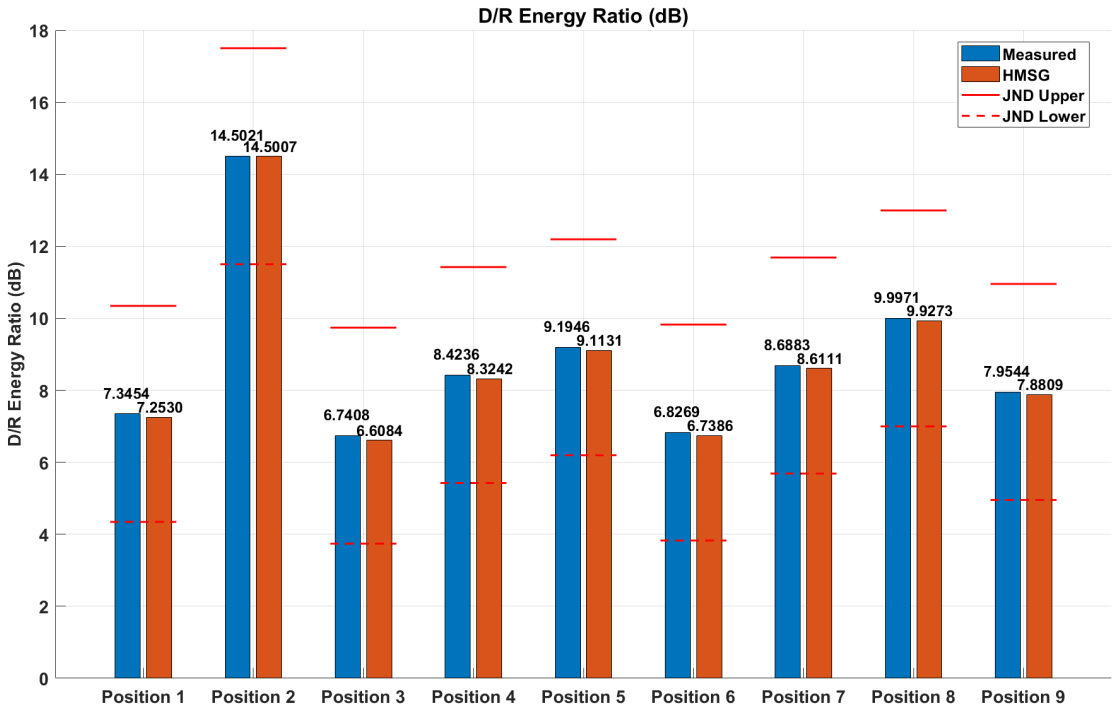
The D/R energy ratio displayed in Figure 5.9 (c) shows close alignment between


RT60 (s)
RT60 (s)
2.5
2
1.5
1
0.5
0
1.86
1.71
1.88
1.69
1.90
1.75
1.87
1.72
1.90
1.74
1.91
1.76
1.93
1.78
1.89
1.73
1.82
1.68
Measured
HMSG
JND Upper
JND Lower
Position 1
Position 2
Position 3
Position 4
Position 5
Position 6
Position 7
Position 8
Position 9

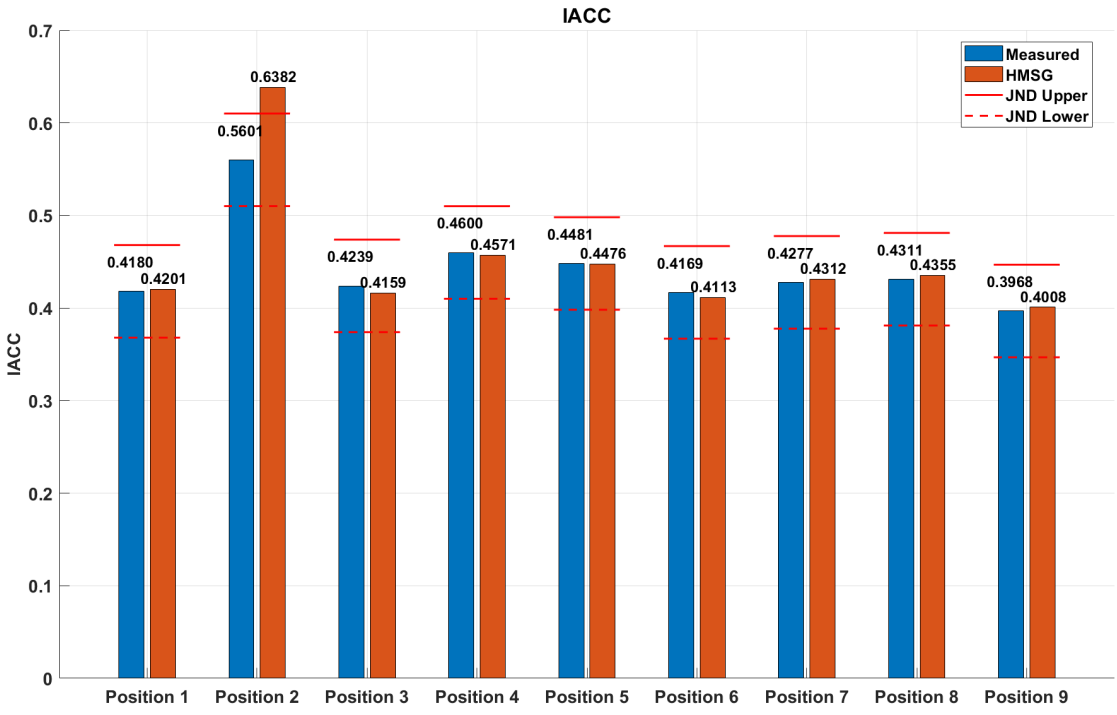
(a)


ITDG (ms)
ITDG (ms)
60
50
40
30
20
10
0
-10
14.0 14.0
15.4 15.4
46.6 46.6
7.4 7.5
3.6 7.5
10.5 10.5
2.3 7.5
2.8 7.5
7.0 7.5
Measured
HMSG
JND Upper
JND Lower
Position 1
Position 2
Position 3
Position 4
Position 5
Position 6
Position 7
Position 8
Position 9

(b)



(c)



(d)

Figure 5.9: Parametric comparison between measured BRIRs and HMSG-generated BRIRs at different locations in a large room. (**a**) RT_{60}. (**b**) ITDG. (**c**) D/R energy ratio. (**d**) IACC.

the measured and simulated data. The differences across positions are generally within 0.1 dB, which falls well below the JND for this parameter. This suggests that the HMSG plug-in accurately simulates the balance between direct and reverberant energy, ensuring that speech intelligibility and overall sound clarity are maintained in the large room. The accurate simulation of the D/R energy ratio is particularly important in large spaces, where the balance of direct and reflected sound can greatly influence the listener's experience.

Similarly, the IACC values shown in Figure 5.9 (d) are closely matched, with differences generally within the JND threshold of 0.05 as cited by ISO 3382-1 [296]. This indicates that the HMSG plug-in effectively preserves the spatial impression and envelopment, which are critical for creating a realistic and immersive auditory environment in a large room. The accurate preservation of IACC ensures that listeners experience a natural and enveloping sound field, which is essential for both speech and music in large spaces.

Figure 5.10 presents a comparison of measured and HMSG-generated RT_{60} values across different octave bands in a large room. The solid lines represent measured values, while the dashed lines represent HMSG-generated values. Each color corresponds to a different position in the room, labeled Position 1 to Position 9. At low frequencies (below 250 Hz), there is noticeable deviation between measured and HMSG-generated RT_{60}. The largest difference is observed at 125 Hz, where measured values peak significantly higher than HMSG. This suggests HMSG may underestimate low-frequency reverberation effects, potentially due to insufficient modal density or absorption mismodeling in the algorithm. At mid-to-high frequencies (250 Hz – 8000 Hz), HMSG begins to closely follow measured RT_{60} across all positions, particularly at 1000 -4000 Hz, minimal deviation is ob-

served between the two sets of values. This indicates HMSG plug-in effectively models mid and high-frequency reverberation characteristics.

This analysis suggests that while HMSG plug-in is a strong performer in mid-to-high frequency ranges, improvements at low frequencies could further enhance its accuracy in room simulation.

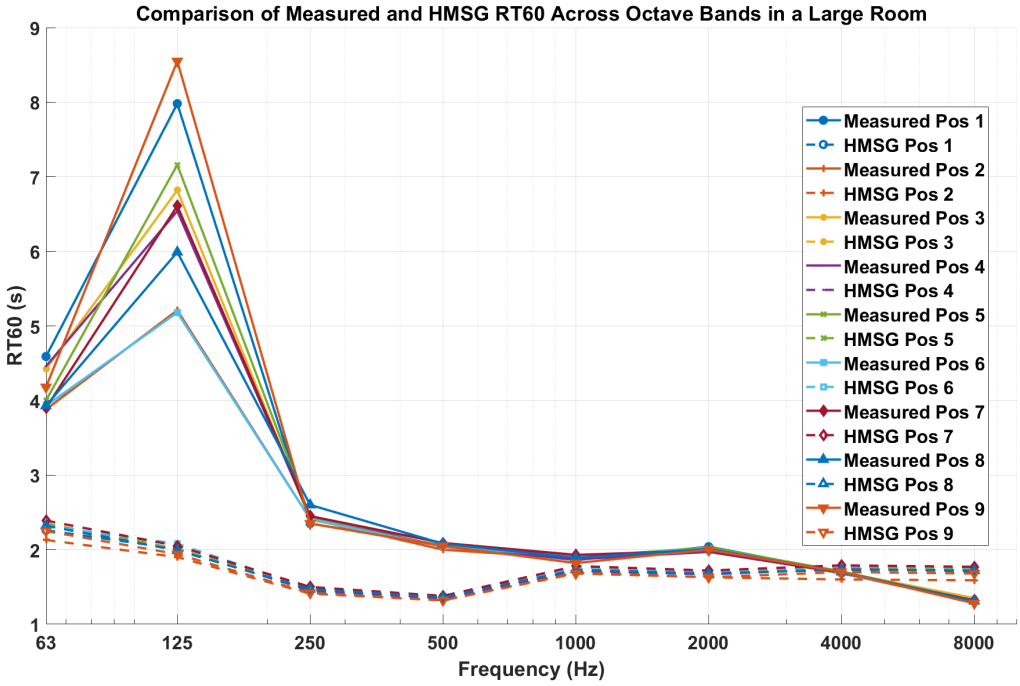


Figure 5.10: Comparison of RT_{60} across octave bands from 63 to 8000 Hz between measured and HMSG-generated impulse responses in a large room.

5.4 Discussion

The evaluation of HMSG-generated impulse responses against the measured BRIRs across small, medium, and large rooms has demonstrated varying levels of accuracy across different acoustic parameters. The HMSG plug-in is able to replicate key reverberation characteristics with reasonable accuracy in most scenarios.

However, some discrepancies arise in specific scenarios, which can be summarized as follows:

Mid-frequency reverberation is well-modeled, but deviations at low and high frequencies suggest an inaccurate representation of absorption and reflection. At the low frequency bands (below 125 Hz), HMSG underestimates reverberation time, likely due to insufficient modelling of low-frequency room modes. At the high frequency band (8000 Hz), HMSG overestimates RT_{60}, which may be caused by inaccuracies in absorption modelling. It could be good to Implementing frequency-dependent decay compensation to ensure more accurate modelling across different frequency bands, and introduce adaptive reverberation tail scaling to better replicate long decay times in large spaces.

The HMSG plug-in exhibits discrepancies in the simulated RT_{60} values for small and large rooms, where the deviations exceed the JND. RT_{60} tends to be too long in the small room and too short in the large room. These discrepancies suggest that the plug-in may not fully capture the nuances of reverberation time in environments with very short or long reverberation characteristics. The overestimation or underestimation of RT_{60} could impact the perceived realism of the simulated acoustic environment, particularly in critical listening applications where precise reverberation modelling is crucial.

In small and large rooms, HMSG plug-in closely matches ITDG values, but in the medium room, the minimum ITDG in HMSG is fixed at 0.03 s, whereas measured values are below 0.03 s. This limitation affects the tunability and accuracy of early reflections, which are vital for conveying a sense of spatial presence and the perceived size of the acoustic space. The inability to closely match the ITDG may lead to a less convincing simulation of the room's acoustic properties,

especially in scenarios where accurate spatial localisation is important.

Across all rooms and measurement positions, HMSG closely matches the D/R ratio, with differences below 0.03 dB. This is well within the 2-3 dB JND, confirming that HMSG accurately balances direct and reverberant sound.

Most measured positions have IACC differences within the 0.05 JND, but some locations exceed this threshold, indicating that HMSG may struggle to fully reproduce spatial expansion characteristics in certain cases.

To further enhance the accuracy of the HMSG plug-in, several optimis ations can be considered. Improving low-frequency modelling would better simulate room modes and standing waves, reducing errors at low frequency bands, while refining high-frequency absorption characteristics could more accurately capture the natural decay behavior above 8000 Hz. Enhancing RT_{60} adaptation by reducing excess late reverberation in small rooms and extending decay tails in large rooms would help achieve a more realistic match to real environments. Additionally, providing finer ITDG adjustments would allow for smaller values in medium-sized rooms, better simulating early reflections. Finally, optimising binaural spatialisation processing would improve IACC matching, ensuring consistent spatial accuracy across different measurement positions.

5.5 Summary

Based on the results of Chapter 4, the Schroeder algorithm offers good perceptual accuracy for short and medium reverberation times while excelling in computational efficiency, making it ideal for real-time applications where processing speed is critical. HMS provides the highest perceptual accuracy for short and medium reverberation times, though at a slightly higher computational cost. Gardner per-

forms best for long reverberation tails, making it particularly well-suited for large spaces. The HMSG reverberation plug-in was developed by dynamically integrating these three algorithms, HMS for small rooms, Schroeder for medium rooms, and Gardner for large rooms. The plug-in is now ready to be embedded in the DAW and has been implemented in REAPER, where it can be controlled in real time, tracking data, and realise the auralisation.

Unlike traditional static models that rely on a single algorithm with fixed parameters, the HMSG plug-in dynamically selects the optimal algorithm based on room size. This ensures a more accurate simulation of reverberation characteristics, improving adaptability for diverse applications in VR, AR, and immersive audio environments. This dynamic approach also allows the plug-in to adjust key acoustic parameters in real-time, closely matching the measured BRIRs across various room environments. By leveraging the strengths of each algorithm, HMSG achieves a balance between perceptual accuracy and computational efficiency, making it a robust solution for real-time reverberation modelling.

The development process involved testing four different methods to tackle the challenges of real-time audio processing. Initial approaches using standard MATLAB filters and digital signal processor (DSP) functions struggled with efficiency and stability, particularly when handling large sparse vectors in real-time. The final approach, employing recursive equations for filters and optimised mex file integration, proved to be the most effective and computationally efficient. This method effectively addressed the challenges encountered with other approaches, such as processing inefficiencies and audio dropouts. By processing each audio sample individually and leveraging mex file optimisations, the plug-in achieves the necessary real-time performance without compromising the quality of the rever-

beration effect. The recursive, sample-by-sample processing approach not only improves computational efficiency but also enhances the flexibility and scalability of the plug-in. This approach allows for easy adaptation to different reverberation algorithms and filter configurations, making the plug-in a versatile tool for various audio processing applications. Documenting these challenges and solutions provides valuable insights for future researchers and developers working in real-time audio processing.

The objective analysis indicates that HMSG plug-in effectively simulates reverberation across different room sizes and listening positions, achieving perceptually accurate results in most cases. Its strongest performance is in the mid-frequency range, where it successfully captures reverberation decay and spatial characteristics. However, refinements are needed in frequency-dependent decay modelling, particularly in low and high-frequency bands, and in adjusting ITDG flexibility for medium-sized rooms. Despite minor parameter mismatches in some cases, HMSG demonstrates strong overall performance, making it a viable tool for real-time, user-adjustable reverberation processing. With further refinements, particularly in low-frequency modal behaviour and binaural spatialisation, HMSG plug-in could achieve even greater accuracy and realism in artificial reverberation applications.

Although the plug-in has been objectively validated through comparisons with measured BRIRs, comprehensive subjective listening tests have not yet been conducted to assess the perceptual impact of the identified discrepancies. Without these tests, it remains uncertain whether the deviations in RT_{60} and ITDG, and other parameters are perceptible to listeners and to what extent they might influence the overall user experience in practical applications. Future studies should include controlled perceptual evaluations to determine whether these variations

affect spatial perception, speech intelligibility, or overall immersion. Additionally, listener preferences could be examined to assess whether certain deviations are acceptable or even preferable in specific use cases, such as music production, gaming, or virtual reality applications.

The HMSG reverberation plug-in delivers dynamic, real-time reverberation modelling by seamlessly integrating multiple algorithms. This approach enhances reverberation accuracy across different room sizes while ensuring computational efficiency, making it a powerful tool for immersive sound environments. Future improvements in low/high-frequency modelling, spatialisation, and ITDG flexibility, along with perceptual validation through listening tests, will further refine its performance. This work sets the foundation for the next generation of real-time audio processing technologies, with broad applications in VR, AR, gaming, and professional audio production.

6 Conclusions and Future Work

6.1 Conclusions

The work in this thesis is summarised as follows. Chapter 1 introduces the context and motivation for the research on virtual reverberation rendering, emphasising the importance of reverberation in creating realistic auditory experiences. It presents the aim of the research, to implement a real-time dynamic binaural reverberation algorithm that can reproduce plausible real-world perceptual reverberation in virtual acoustic rendering, and primary focus on enhancing user perceptual experience and ensuring computational efficiency. An overview of the research question, how can a plausible real-world reverberation effect be simulated in real-time in virtual acoustic rendering, using a computationally efficient reverberation algorithm? The key contributions of the research are highlighted. It also detailed the ethical approval for the experiments conducted in this thesis and presented the overall structure of the thesis.

Chapter 2 provides a detailed review of the fundamental concepts related to sound and reverberation, essential for understanding virtual reverberation rendering. It starts with an explanation of sound propagation and sound fields, followed by an overview of spatial audio implementation. The chapter then delves into the principles of reverberation and RIRs, discussing how sound interacts with

environments to create reverberation effects. It also explores different types of reverberation algorithms and their potential application scenarios in virtual acoustic rendering, setting the stage for the development of the new reverberation algorithm proposed later in the thesis.The chapter also introduces methods for conducting and analysing listening tests, which are crucial for assessing the perceptual effectiveness of reverberation algorithms. It outlines different testing methodologies, including paired comparison, the staircase method, category judgment, ranking, and semantic differential techniques. Furthermore, it discusses statistical analysis methods used to interpret listening test results, ensuring a rigorous evaluation of reverberation algorithms.

Chapter 3 presents an evaluation of the impact of different parameters of binaural room impulse responses on perceptual reverberation, including ITDG extension, forward ER removal, reverse ER removal and LR removal. Through a listening test, the chapter established perceptual thresholds for these parameters and analyses their influence on perceptual reverberation. The results of this evaluation provide crucial insights into which parameters are most important for achieving perceptually plausible reverberation in virtual acoustic environments.

Chapter 4 introduces and evaluates a new binaural reverberation algorithm, referred to as the HMS algorithm. This algorithm is developed by integrating and refining the Schroeder and Moorer reverberation algorithms, with particular emphasis separately optimising the ITDG and early reflections based on the experimental results presented in Chapter 3. The algorithm is evaluated against measured BRIRs and other popular reverberation algorithms, demonstrating its superior performance in terms of perceptual plausibility, especially in simulating small and medium size rooms. However, it has to be acknowledged that its compu-

tational efficiency did not meet the expected levels, indicating a trade-off between accuracy and processing demands.

Chapter 5 details the implementation of a dynamic reverberation plug-in, called the HMSG reverb plug-in. This plug-in integrates the HMS, Schroeder, and Gardner algorithms based on the experimental results in Chapter 4, allowing for real-time adjustment of parameters such as reverberation time, ITDG, direct sound scale, reverberation scale and binaural correlation coefficient. The chapter discusses the challenges encountered during implementation, particularly regarding computational efficiency, and explains how these challenges are addressed to create a functional and effective tool for virtual acoustic rendering. An objective analysis is also conducted to evaluate the performance of the plug-in, comparing its output parameters with those of measured impulse responses to verify its effectiveness in simulating reverberation in different spaces or at different locations in the same space.

6.2 Restatement of Research Question

The research question originally presented in Section 1.1, which provides the basis for the work in this thesis, is restated as follows:

How can a plausible real-world reverberation effect be simulated in real time in virtual acoustic rendering, using a computationally efficient reverberation algorithm?

This question was further explored through three sub-questions:

1. Evaluating the most perceptually relevant reverberation parameters for manipulation in digital reverberation algorithms (Chapter 3).

2. Identifying whether a perceptually plausible and computationally efficient digital reverberator can be designed based on these parameters (Chapter 4).

3. Assessing the effectiveness of the proposed reverberation algorithm in a real-time dynamic rendering scenario and comparing it to real-world measurements (Chapter 5).

Each of these questions was addressed in different chapters of the thesis, leading to a series of novel contributions.

A key contribution of this thesis is the threshold evaluation of BRIR parameters, which establishes a quantitative understanding of perceptual sensitivity to reverberation characteristics. Through a staircase method, perceptual thresholds for ITDG, forward and reverse early reflection strengths, and late reverberation strength were determined for different room sizes. These findings provide a foundation for optimising digital reverberation algorithms, ensuring that manipulated parameters accurately to enhance realistic virtual acoustics.

Another significant contribution is the development of a novel binaural reverberation algorithm, HMS, which was designed by combining and refining features of the Schroeder and Moorer algorithms and structured to separately handle direct sound and early reflections from late reverberation, as previous research in Chapter 3 indicated the significant impact of ERs and ITDG on perceptual reverberation. This algorithm effectively simulates reverberation for small, medium, and large rooms, incorporating key improvements such as high-frequency noise filtering, air absorption simulation, and binaural channel separation using delay constants. The algorithm also achieves accurate matching of the D/R energy ratio, IACC, and timbre, closely resembling real-world reverberation. Comparative evaluations demonstrated that HMS outperforms existing reverberation algorithms in

terms of perceptual plausibility to measured BRIRs.

Building upon these findings, this thesis also developed and verified the HMSG real-time dynamic reverberation plug-in, which dynamically integrates multiple reverberation algorithms to adapt to different reverberation conditions in real-time. The HMSG plug-in utilises the HMS algorithm for small rooms, Schroeder for medium rooms, and Gardner for large rooms, ensuring an optimised balance between perceptual accuracy and computational efficiency. Unlike static reverberation models that rely on a single algorithm with fixed parameters, the HMSG plug-in dynamically selects the optimal reverberation algorithm based on room size, enhancing adaptability for virtual immersive applications. The computational efficiency and real-time performance of the plug-in were validated through successful integration into a DAW (REAPER), where it demonstrated low-latency audio processing while maintaining high acoustic fidelity.

The perceptual validity of the HMSG plug-in was confirmed through the listening test for seperate algorithm in Chapter 4, which showed that the generated reverberation effects were perceptually optimal from real-world acoustic environments. Additionally, an objective evaluation compared the plug-in's output to measured BRIRs, demonstrating a high correlation in most of the parameters (RT_{60}, ITDG, D/R energy ratio, and IACC), further affirming the accuracy and plausibility of the simulated reverberation.

Despite these promising results, some limitations were identified. The RT_{60} values in small and large rooms, and low and high frequencies exhibited deviations exceeding JND thresholds, and the ITDG was fixed at 0.03s for medium rooms, whereas measured values ranged from 0.003 s to 0.0067s. These discrepancies indicate areas for further refinement, particularly in frequency-dependent decay

modeling and adaptive ITDG adjustments to improve reverberation time and early reflection accuracy. Future perceptual listening tests should be conducted to assess whether these discrepancies significantly impact user experience, guiding further improvements in the plug-in.

The findings of this research confirm that HMSG effectively simulates real-world reverberation in real-time virtual acoustic rendering, meeting the objectives outlined in the research question. The novel contributions of this work lie in the development of a hybrid, dynamically reverberation model, integrating multiple algorithms to optimise both perceptual accuracy and computational efficiency. Additionally, the quantitative perceptual threshold study establishes a scientific basis for BRIR parameter manipulation, and the successful implementation of a DAW-compatible real-time reverb plug-in demonstrates that low-latency processing can be achieved while preserving acoustic realism.

Therefore, the results of using the HMSG plug-in to simulate plausible real-world reverberation effects in virtual acoustic rendering are positive. The combination of perceptual tests, objective analysis, and real-time performance evaluation collectively substantiates the rationality of the plug-in's design and its effectiveness in achieving reverberation plausibility in virtual acoustic environments.

6.3 Future Work

Throughout the thesis, a number of recommendations are made for future research from the work undertaken. These are as follows:

6.3.1 Threshold Measurements for Other Parameters of the Binaural Impulse Response and Different Stimuli

In Chapter 3, the impact thresholds of the four parameters ITDG extension, forward ER removal, reverse ER removal and LR removal on the perceptual reverberation were evaluated. However, the objective parametric analysis in Chapter 5 revealed differences in the reverberation time, ITDG, D/R energy ratio and IACC of the binaural impulse response measured at different locations in the same space. This provides a potential avenue for future work in which the impact thresholds of parameters such as reverberation time, D/R energy ratio and IACC on perceptual reverberation can be evaluated to determine a reasonable range of parameter errors in the binaural room impulse responses generated by the artificial reverberation algorithm.

If these thresholds are identified, they could lead to the development of more robust and perceptually accurate reverberation algorithms that maintain their plausibility even when slight deviations occur in critical acoustic parameters. This refinement could significantly impact the field by enabling the creation of virtual acoustic environments that are more realistic and convincing, even under varying spatial conditions.

Moreover, the parameter thresholds evaluated in Chapter 3 were determined using only male speech as the test signal. In practical applications, different types of audio signals, such as noise, music, or female singing, can influence the perception of these parameters differently. By conducting experiments with a wider range of stimuli, it could be anticipated that the results might reveal how certain audio types are more sensitive to specific parameters. This could lead to more tailored reverberation algorithms that can adapt their processing based on the type

of audio signal, further enhancing the realism of virtual environments.

6.3.2 Optimising the Computational Efficiency of the Proposed HMS Reverberation Algorithm

In Chapter 4, a new reverberation algorithm called the HMS reverberation algorithm was presented. Its overall performance for real-world perceptual reverberation simulations is optimal. but its computational efficiency is not as effective as desired when simulating medium and large room reverberation due to the long coefficients of filters. When implemented as a reverberation plug-in using MATLAB, the medium and large room-size structure of this reverberation algorithm required substantial computational time, resulting in a large number of underrun samples during real-time audio processing. This issue prevented the successful implementation of a real-time audio plug-in.

To address this limitation, future research could focus on several specific optimisation strategies. One approach would be to reduce the length of the filter coefficients or explore alternative filter designs, such as employing more efficient FIR or IIR filters, which could deliver comparable acoustic results with lower computational demands. Another potential optimisation could involve refining the algorithm's structure to better balance the trade-off between accuracy and processing speed, possibly through techniques like filter coefficient quantisation, multi-rate filtering, or implementing more efficient mathematical operations.

Additionally, achieving computational efficiency gains through hardware acceleration could be explored. This might involve leveraging a dedicated digital signal processor (DSP) chip or field-programmable gate array (FPGA) to perform multiple parallel operations. These hardware solutions could significantly reduce

processing times by offloading computationally intensive tasks from the general-purpose CPU to specialised hardware optimised for high-speed, low-latency audio processing.

By pursuing these optimisations, future research could enhance the real-time performance of the HMS algorithm, ensuring that it remains both perceptually accurate and computationally efficient, even in complex acoustic scenarios involving medium and large rooms. This would make the algorithm more suitable for real-time applications, such as live audio processing in virtual and augmented reality environments.

6.3.3 Optimisation of the Adjustable Parameters of the HMSG Reverb Plug-in to Increase the Match with the Measured Impulse Response

In Chapter 5, the HMSG reverberation plug-in was developed by integrating the HMS, Schroeder, and Gardner reverberation algorithms to achieve real-time dynamic reverberation rendering. While the plug-in is capable of closely matching the parameters of the measured BRIRs by adjusting the corresponding parameters in real-time, certain deviations were observed in specific scenarios.

In the simulation of small and large rooms, noticeable deviations in reverberation time were noted, as well as in the mismatching of reverberation time at low and high frequencies in different environments. Additionally, minor mismatches in IACC were identified in some cases, indicating potential room for improvement in spatial accuracy. In medium room simulations, the initial ITDG value was found to be too large, leading to limited tunability and reduced flexibility in early reflection control.

To address reverberation time deviations, future work could focus on optimising the filter design to better model the acoustic characteristics of different room sizes. Specifically, this could involve adjusting the filter structures or fine-tuning the filter coefficients to more accurately model the acoustic characteristics of different room sizes. Approaches such as reconfiguring the feedback and feedforward paths in the filter, employing multi-stage filtering, or dynamically adjusting the coefficients based on real-time analysis of the room's acoustic response to improve frequency-dependent reverberation characteristics could help achieve more accurate and consistent reverberation times across various room simulations.

The large initial ITDG value in medium room simulations results in limited tunability. To improve this, future research could explore two specific approaches: reducing the initial ITDG value or extending ITDG in reverse. The first approach could involve modifying the algorithm to initialise with a smaller ITDG value, providing a greater range for adjustment. Techniques such as fine-tuning the delay line coefficients could be investigated to achieve a more precise ITDG adjustment capability. Another approach could be to explore reverse extension of the ITDG, where the initial ITDG value is artificially reduced by applying a negative delay or offset. This could be achieved through designing custom delay lines that allow for negative time delays. Such methods could provide more flexibility in tuning the ITDG, enabling finer control over the reverberation characteristics in medium-sized rooms.

Additionally, increasing BRIR channel differentiation would better optimise binaural spatialisation processing to improve IACC matching, ensuring consistent spatial accuracy across different measurement positions.

6.3.4 Implement Perceptual Evaluation for Developed HMSG Plug-ins in VR and AR Environments

Through the work in Chapters 3 and 4, it was demonstrated that it is feasible to use reverberation algorithms to dynamically simulate plausible real-world reverberation effects in virtual acoustic rendering. However, the plug-in has not currently been implemented to evaluate the perception of reverberation in VR and AR environments because it is difficult to design an effective perceptual test for this situation.

Given that the common listening test paradigms such as MUSHRA or ABX cannot work effectively because of the need to compare perceptual reverberation with real-world references, future work could focus on developing custom perceptual test paradigms tailored to VR and AR environments. For instance, a dynamic AB comparison test could be designed where participants can switch between real-world and virtual reverberation in real-time while staying in the same physical location, thus maintaining consistency in their spatial position.

Implementing a listening test for AR means that the evaluation must take into account the appropriate pass-through mode on headphones, which could alter the timbre and localisation cues. Therefore, future work could involve developing calibration and compensation algorithms. These algorithms would adjust the audio signal to account for the coloration introduced by the pass-through mode, ensuring that the perceived reverberation in AR environments is as close as possible to the intended effect.

In a six degrees of freedom evaluation environment, participants can move freely, which can affect the consistency of perceptual tests. To mitigate this, future work could focus on developing algorithms that track the participant's position in

real-time and stabilise the test conditions. This could involve using a combination of head tracking and room mapping to ensure that participants remain in the same relative position when switching between different audio samples.

To ensure that algorithm parameters do not change drastically between different positions, future research could explore techniques like parameter interpolation and smoothing. This would create a more seamless transition between different reverberation settings as the participant moves, minimising abrupt changes that could unfairly influence the test results.

Future work should integrate advanced head tracking systems and evaluate the plug-in's performance under real head-tracked conditions, particularly ensuring low latency and maintaining computational efficiency. By minimising latency, the plug-in can provide a seamless and responsive audio experience, enhancing the overall realism and immersion of the virtual environment. This integration would provide a more realistic and immersive experience in VR and AR environments.

Given the complexity of evaluating perceptual reverberation in VR and AR environments, future research could focus on developing a comprehensive evaluation framework that combines subjective testing with objective measures. This framework would take into account the unique challenges of VR and AR, such as varying positions, head movements, and pass-through effects, providing a more holistic assessment of the plug-in's performance.

Given the scope and complexity of these challenges, addressing them fully could constitute the topic for a dedicated PhD project. This research could delve into one or more of the following areas: developing new perceptual test paradigms for VR and AR, designing calibration algorithms for pass-through modes, or creating advanced head tracking systems tailored to dynamic virtual environments.

By addressing these detailed aspects in future work, the evaluation of the HMSG reverberation plug-in in VR and AR environments could be significantly improved, leading to more accurate and reliable assessments of its performance in these advanced settings.

6.4 Final Thoughts

In this thesis, a dynamic reverberation algorithm for use in virtual acoustics has been proposed. This development was based on an in-depth evaluation of the thresholds of various parameters of the BRIRs. Through the evaluation of the plausibility and computational efficiency of this algorithm compared to other existing algorithms, the optimal algorithms were integrated into a plug-in that can be used in real-time virtual environments. The plug-in allows to dynamically adjust the user interface parameters to closely align with those of real-world BRIRs, thereby enabling real-time simulation of perceptual reverberation across different spaces or varying locations within the same space. While the plug-in does not perfectly match all parameters of real impulse responses nor simulate real-world perceptual reverberation with absolute fidelity, it significantly enhances the realism and similarity of simulated reverberation effects. The comparative analysis with measured reverberation and its successful application within a DAW demonstrate that this plug-in is a viable tool for simulating plausible real-world reverberation effects in virtual acoustic rendering.

The research presented in this thesis lays the groundwork for future exploration and development in the field of virtual acoustic rendering. Specifically, it opens up avenues for further refinement of dynamic reverberation algorithms to better simulate real-world perceptual reverberation, taking into account the complexities

of different acoustic environments. Future research could build upon these findings to optimise the algorithm for greater accuracy, explore the integration of head tracking and personalised HRTFs, and extend its application to more diverse and complex virtual environments.

Moreover, this work highlights the potential for these technologies to be applied in VR and AR environments, where accurate and responsive reverberation is critical for creating a truly immersive experience. The challenges associated with perceptual evaluation in such dynamic contexts, particularly in relation to real-time adjustments and spatial variations, suggest that there is significant scope for further investigation. This area of study, encompassing the simulation of real-world perceptual reverberation in virtual environments, could form the basis for future PhD research, focusing on refining the algorithm, enhancing computational efficiency, and developing robust methods for subjective and objective evaluation in VR and AR settings.

Appendices

A

Supplementary Plots and Tables for Chapter 3

This appendix presents supplementary plots and tables for Chapter 3. Table A.1 presents room surface materials used in ODEON [148] modelling to generate impulse responses of different reverberation durations.

Figure A.1 shows the room model used to generate impulse responses in ODEON [148].

Table A.1: The materials used in the room surface for different reverb times (The material description for each material number can be found in ODEON [148]).

Surface Number	Surface Name	Material Number (0.31s Reverb Time)	Material Number (0.91s Reverb Time)	Material Number (1.51s Reverb Time)	Area (m^2)
1001	Podium floor	70	20	20	78.00
1002	Main audience floor	70	40	20	259.26
2001	End wall behind podium	1004	1004	1004	75.00
−2002	Podium side wall, South + North	1004	1004	1004	58.70
2002	Podium side wall, South + North	1004	1004	1004	58.70
−2003	Side wall, audience area South + North	11,009	11,009	11,009	139.52
−2003	Side wall, audience area South + North	11,009	11,009	11,009	139.52
2004	Rear wall behind audience	1	11,009	11,009	119.04
3001	Podium ceiling	3023	3023	3023	84.50
3002	Ceiling over audience	3023	3023	3023	256.00

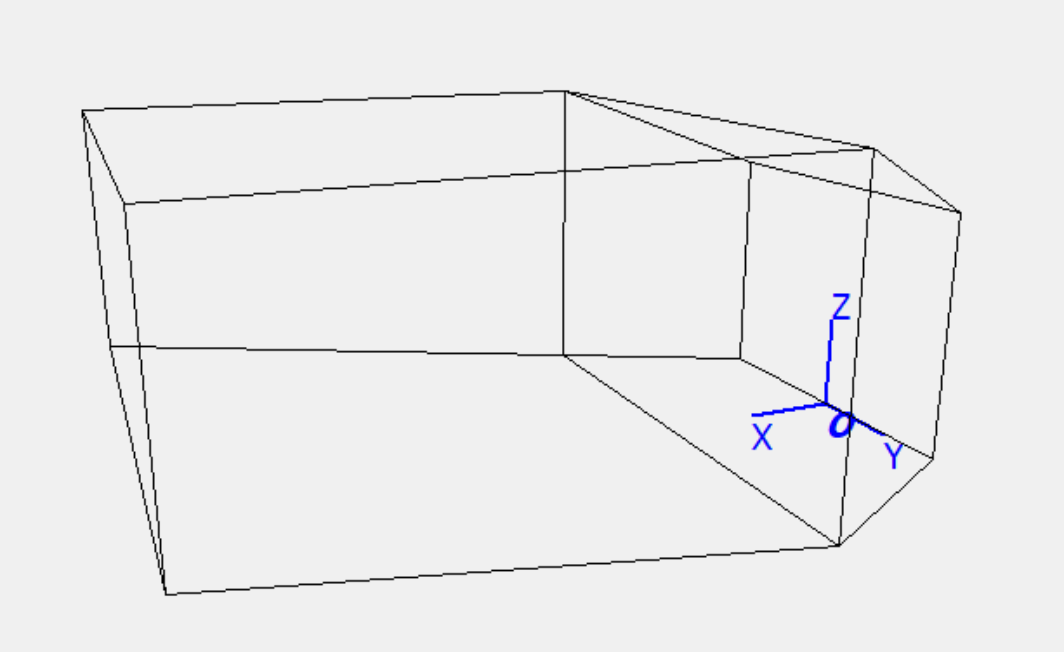


Figure A.1: The room model used to generate impulse responses, where $X_{\max} - X_{\min}$ is 22 m, $Y_{\max} - Y_{\min}$ is 16 m, and $Z_{\max} - Z_{\min}$ is 10 m [148].

B

Supplementary Plots and Tables for Chapter 4

This appendix presents supplementary plots and tables for Chapter 4. Table B.1 presents the ANOVA test results between Beyerdynamic DT990 and other headphones. Table B.2 to B.13 are the rating of stimuli in MUSHRA test of each participant.

Figure B.1 shows the Control Room 7 at WDR Broadcast Studios, Germany, where the BRIR with 0.266 s reverb time was measured [146]. Figure B.2 shows the Printing House Hall at Trinity College Dublin, Ireland, where the BRIR with 0.95 s reverb time was measured [291]. Figure B.3 shows the Lady Chapel at St Alban's Cathedral, United Kingdom, where the BRIR with 2.34 s reverb time was measured [292]. Figures B.4 to B.9 illustrate the timbre matching results for Schroeder, Moorer, Gardner, FDN, Dattorro, and DFDN reverberation algorithm, respectively. Figures B.10 to B.12 represent the box plots for male speech, cello piece, and drum beat simulated by the reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times, respectively. Figures B.13 to B.16 represent the mean scores with standard errors for female singing, male speech, cello piece, and drum beat simulated by the reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times, respectively.

Table B.1: The ANOVA test results between Beyerdynamic DT990 and other headphones.

DF = 1 Significance level = 0.05	p **value**
0.266s female singing	0.875
0.266s male speech	0.057
0.266s cello piece	0.614
0.266s drumbeat	0.358
0.95s female singing	0.125
0.95s male speech	0.664
0.95s cello piece	0.064
0.95s drumbeat	0.206
2.34s female singing	0.130
2.34s male speech	0.093
2.34s cello piece	0.474
2.34s drumbeat	0.545

Table B.2: The rating of stimuli in MUSHRA test of each participant (0.266s Female Singing).

Stimuli / Participants	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	40	62	77	89	71	68	83	100	60
2	40	50	84	94	70	88	86	100	79
3	1	25	75	100	100	100	75	95	73
4	21	100	86	100	84	67	100	100	82
5	26	38	50	62	82	80	91	100	0
6	0	0	49	48	60	70	31	100	0
7	0	65	42	94	83	76	74	100	72
8	40	40	80	60	80	90	80	100	40
9	52	60	77	81	73	76	70	100	5
10	49	47	100	100	96	75	74	100	23
11	27	26	43	96	44	82	96	100	27
12	28	57	70	77	40	51	62	100	0
13	86	91	69	95	98	67	100	97	72
14	11	51	55	85	27	62	31	100	62
15	14	31	49	81	63	81	61	100	0
16	0	28	15	82	46	67	80	100	0
17	24	86	65	94	100	64	88	90	38
18	28	59	67	82	71	43	77	100	11
19	60	71	60	91	85	90	60	100	36
20	35	26	69	90	80	64	63	100	0
Average	29.10	50.65	64.10	85.05	72.65	73.05	73.90	99.10	34.00
Standard Error	5.00	5.59	4.31	3.18	4.65	3.06	4.23	0.56	7.02

Table B.3: The rating of stimuli in MUSHRA test of each participant (0.95s Female Singing).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	22	15	34	53	57	76	76	100	50
2	31	31	41	70	60	89	81	100	81
3	0	28	83	60	31	92	49	100	3
4	44	73	88	60	61	100	92	100	73
5	44	28	48	54	69	90	86	100	0
6	0	0	12	12	30	50	13	100	5
7	26	0	48	47	60	74	83	94	28
8	40	0	50	20	40	70	70	100	100
9	45	38	55	80	49	88	85	100	15
10	38	47	46	100	65	86	89	88	33
11	40	46	52	69	91	98	80	100	26
12	31	24	53	41	68	78	66	100	0
13	87	64	75	88	68	96	98	100	59
14	35	48	65	55	17	33	83	100	69
15	12	31	49	69	48	85	79	100	0
16	12	10	27	43	24	80	62	100	0
17	26	44	95	71	84	100	92	100	68
18	17	37	48	68	81	64	61	100	0
19	26	49	60	34	77	87	63	93	36
20	31	21	39	51	80	89	90	100	0
Average	30.35	31.70	53.45	52.75	58.00	81.25	74.90	98.75	32.30
Standard Error	4.26	4.57	4.55	4.78	4.70	3.79	4.21	0.72	7.34

Table B.4: The rating of stimuli in MUSHRA test of each participant (2.34s Female Singing).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	33	43	50	72	66	63	82	100	53
2	40	21	72	79	30	51	83	100	100
3	13	0	100	86	15	47	59	191	21
4	39	20	80	89	63	68	44	100	39
5	52	15	90	88	35	70	62	100	0
6	0	0	30	34	0	30	40	100	0
7	60	3	86	65	44	70	87	100	74
8	10	0	20	90	40	70	50	80	100
9	15	14	81	58	27	65	58	100	21
10	28	25	69	82	31	69	96	100	36
11	23	24	90	94	33	62	79	100	68
12	37	12	56	80	33	44	36	100	0
13	92	57	93	67	60	78	83	100	64
14	48	7	100	63	28	36	48	89	82
15	30	12	33	75	55	54	80	100	0
16	14	0	56	74	0	27	39	100	27
17	59	16	56	100	80	65	95	100	32
18	54	0	86	81	54	49	67	100	29
19	43	28	63	78	29	54	37	100	50
20	39	16	79	87	29	50	91	100	0
Average	36.45	15.65	69.50	77.10	37.60	56.10	65.80	98.00	39.80
Standard Error	4.80	3.37	5.24	3.31	4.65	3.20	4.68	1.17	7.41

Table B.5: The rating of stimuli in MUSHRA test of each participant (0.266s Male Speech).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	19	61	54	46	40	78	57	100	16
2	31	51	85	77	45	71	51	100	19
3	11	79	91	78	100	97	174	100	5
4	38	89	85	79	64	84	100	100	22
5	10	38	79	89	24	68	45	100	0
6	0	12	30	42	12	50	28	100	1
7	0	77	29	81	65	60	76	100	46
8	2	60	80	71	70	80	50	100	30
9	17	27	58	46	53	70	52	100	10
10	38	55	82	84	75	76	76	100	18
11	10	66	86	26	37	76	41	100	5
12	18	38	60	49	21	58	47	100	0
13	64	60	79	92	73	92	94	100	63
14	12	17	52	70	29	62	47	100	79
15	14	59	50	38	65	80	80	100	0
16	0	47	57	29	22	36	20	100	0
17	20	72	46	50	62	67	80	100	11
18	0	57	40	12	56	50	35	100	27
19	26	21	62	48	37	52	36	100	40
20	19	34	40	36	70	91	69	100	0
Average	17.45	51.00	62.25	57.15	51.00	69.90	57.90	100.00	19.60
Standard Error	3.60	4.80	4.43	5.25	5.11	3.55	4.94	0.00	5.02

Table B.6: The rating of stimuli in MUSHRA test of each participant (0.95s Male Speech).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	42	22	32	28	52	59	84	100	40
2	15	30	71	59	65	81	69	100	31
3	0	5	81	26	46	79	60	100	0
4	57	43	64	49	78	77	88	100	18
5	18	37	35	28	57	62	87	100	0
6	0	0	0	30	30	26	34	100	4
7	61	0	42	22	38	56	47	100	74
8	22	0	0	60	79	76	60	100	50
9	20	14	20	67	70	82	79	100	3
10	10	15	29	47	52	87	87	100	12
11	30	58	38	52	89	69	86	100	20
12	16	7	34	34	29	78	46	100	0
13	56	55	72	55	67	87	88	100	84
14	50	31	32	41	56	62	44	96	90
15	15	11	17	53	50	75	75	100	0
16	3	0	12	44	27	83	64	100	0
17	0	8	43	8	27	73	37	100	42
18	17	0	42	19	58	75	81	100	44
19	31	33	61	49	77	74	68	100	39
20	20	17	46	39	75	82	83	100	0
Average	24.15	19.30	38.55	40.50	56.10	72.15	68.35	99.80	27.55
Standard Error	4.38	4.15	5.11	3.50	4.26	3.14	4.11	0.20	6.61

Table B.7: The rating of stimuli in MUSHRA test of each participant (2.34s Male Speech).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	22	9	22	61	42	51	40	100	19
2	28	5	20	35	29	34	41	100	41
3	0	1	52	85	1	7	83	100	0
4	42	18	20	80	41	42	21	100	62
5	48	15	61	90	35	56	62	100	0
6	0	0	0	12	0	30	35	100	0
7	21	42	61	64	0	49	68	100	31
8	40	0	20	90	40	60	20	100	50
9	22	22	14	58	37	50	47	100	8
10	36	28	39	90	62	56	56	100	42
11	28	18	45	96	60	62	56	100	22
12	28	9	33	55	42	46	35	100	0
13	83	12	50	67	72	72	52	100	75
14	29	0	25	33	21	26	23	100	42
15	19	12	16	65	65	30	41	100	0
16	0	0	0	41	0	67	25	100	58
17	24	8	26	76	67	76	0	100	80
18	36	0	27	61	71	51	26	100	40
19	34	15	11	50	12	40	51	100	19
20	43	13	20	73	60	75	83	100	0
Average	29.15	11.35	28.10	64.10	37.85	49.00	43.25	100.00	29.45
Standard Error	4.20	2.44	4.03	4.96	5.66	3.98	4.79	0.00	5.94

Table B.8: The rating of stimuli in MUSHRA test of each participant (0.266s Cello).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	37	71	84	61	100	51	80	88	22
2	61	70	89	89	70	90	81	100	85
3	100	88	97	95	97	100	97	100	4
4	80	100	100	100	100	100	100	100	61
5	38	68	79	68	78	89	50	100	0
6	4	22	11	49	57	42	21	100	0
7	73	74	88	85	59	78	52	100	0
8	70	50	80	90	100	100	80	100	40
9	46	62	82	88	90	86	40	100	11
10	55	72	84	84	92	86	84	100	16
11	57	92	95	93	83	81	99	94	54
12	25	46	65	34	40	56	31	100	0
13	81	85	85	97	93	97	90	100	49
14	18	71	73	53	49	72	49	100	84
15	30	54	65	81	75	71	57	100	0
16	23	27	39	75	15	60	57	100	0
17	58	40	84	80	97	97	58	100	87
18	63	66	75	74	100	84	27	100	0
19	41	74	78	74	82	79	74	100	33
20	39	70	61	71	64	89	91	100	0
Average	49.95	65.10	75.70	77.05	77.05	80.40	65.90	99.10	27.30
Standard Error	5.42	4.55	4.62	3.87	5.27	3.82	5.58	0.66	7.17

Table B.9: The rating of stimuli in MUSHRA test of each participant (0.95s Cello).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	53	32	44	70	62	84	72	100	26
2	30	19	59	50	51	70	70	100	49
3	94	47	93	41	94	100	100	100	0
4	84	80	87	87	86	91	91	100	42
5	59	36	47	62	64	92	81	100	0
6	2	0	1	17	34	69	32	100	0
7	43	38	57	82	72	100	82	78	0
8	0	39	30	85	50	90	90	100	40
9	66	54	77	70	58	86	80	100	6
10	51	46	52	86	76	86	71	100	26
11	36	12	73	42	69	89	87	100	2
12	58	17	56	33	63	48	37	90	0
13	72	71	90	96	87	95	91	100	72
14	24	61	50	31	0	66	45	100	77
15	51	14	54	71	66	85	80	100	0
16	0	0	21	27	54	73	16	100	5
17	65	28	87	64	90	96	80	100	39
18	64	63	74	59	81	68	83	100	0
19	35	24	60	24	76	78	51	100	42
20	17	28	77	51	79	91	79	100	0
Average	45.20	35.45	59.45	57.40	65.60	82.85	70.90	98.40	21.30
Standard Error	6.05	5.04	5.36	5.30	4.85	3.01	5.05	1.18	5.74

Table B.10: The rating of stimuli in MUSHRA test of each participant (2.34s Cello).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	60	40	81	86	35	50	77	100	26
2	61	10	87	55	69	49	30	100	100
3	15	6	79	87	25	89	53	100	0
4	80	72	86	94	78	87	87	100	64
5	74	11	47	91	27	90	32	100	0
6	30	0	50	50	30	50	50	100	30
7	40	52	47	68	0	41	54	100	52
8	50	0	80	80	51	70	37	100	40
9	47	25	44	77	30	61	74	100	15
10	74	38	69	78	74	74	66	100	32
11	68	24	89	90	44	55	45	100	0
12	71	22	54	100	31	29	16	48	0
13	66	24	80	72	54	96	77	100	62
14	45	59	58	80	48	61	42	100	80
15	45	11	45	71	60	58	49	100	0
16	36	0	45	78	29	83	29	100	0
17	81	16	63	93	64	72	13	100	81
18	52	0	68	75	49	51	36	100	71
19	58	12	67	84	48	67	61	100	36
20	39	22	15	59	71	91	43	100	0
Average	54.60	22.20	62.70	78.40	45.85	66.20	48.55	97.40	34.45
Standard Error	3.97	4.61	4.33	2.96	4.49	4.21	4.58	2.60	7.37

Table B.11: The rating of stimuli in MUSHRA test of each participant (0.266s Drum).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	12	39	52	34	40	44	20	90	19
2	11	71	55	60	55	68	50	100	29
3	0	0	25	67	47	49	56	100	0
4	40	59	69	94	68	63	88	100	17
5	20	40	60	42	50	80	36	100	0
6	0	0	12	30	0	29	0	100	0
7	29	0	51	35	65	44	49	100	8
8	5	40	70	75	80	70	90	100	30
9	13	36	47	48	51	24	46	100	6
10	32	28	57	100	65	78	58	100	26
11	29	33	66	89	88	30	70	100	4
12	24	47	76	40	64	66	40	100	0
13	77	69	82	82	93	75	96	100	44
14	8	27	33	51	61	72	34	100	64
15	11	20	54	41	51	63	34	100	0
16	0	11	59	38	54	72	21	100	10
17	4	41	42	100	64	63	75	100	38
18	2	38	34	23	61	67	56	100	5
19	17	48	70	59	69	56	73	100	24
20	20	63	44	41	68	90	80	100	0
Average	17.70	35.50	52.90	57.45	59.70	60.15	53.60	99.50	16.20
Standard Error	4.07	4.82	3.96	5.51	4.31	4.05	5.75	0.50	4.01

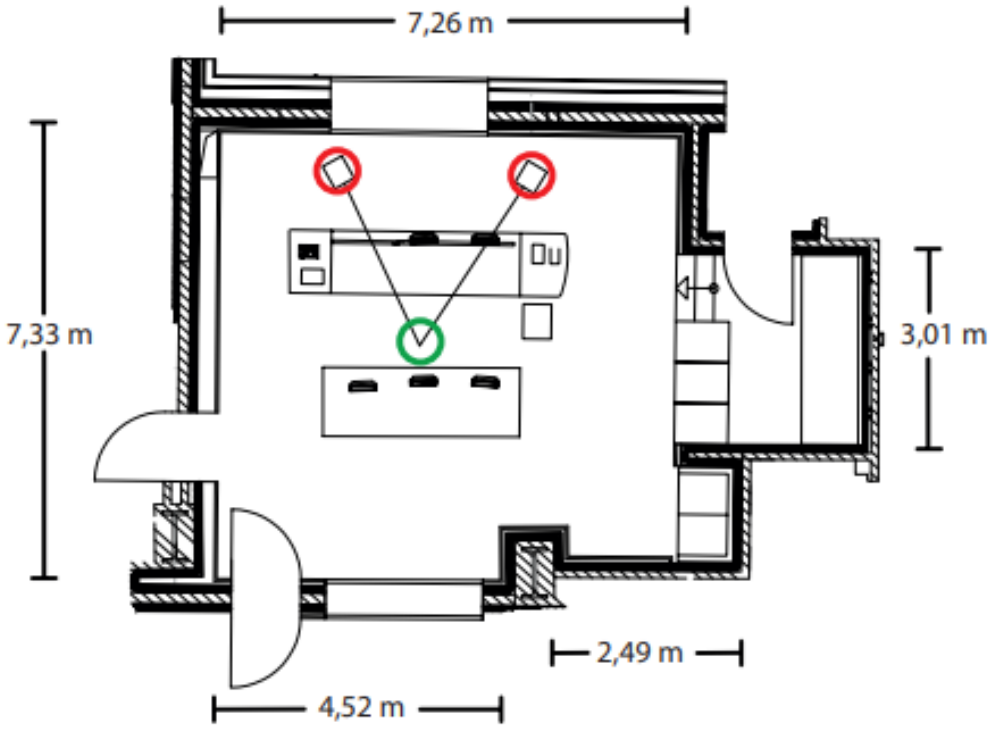
Table B.12: The rating of stimuli in MUSHRA test of each participant (0.95s Drum).

Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	13	9	20	47	57	64	82	100	4
2	10	5	60	45	51	72	59	100	21
3	16	6	0	39	79	27	26	98	0
4	21	24	64	56	56	83	54	100	22
5	21	12	25	69	91	56	67	100	0
6	5	0	0	29	49	49	30	100	0
7	0	16	38	38	43	58	63	100	34
8	0	1	36	19	60	80	40	100	50
9	10	14	30	17	70	59	50	100	0
10	24	8	34	76	67	74	67	100	0
11	7	11	69	38	80	90	39	98	20
12	14	13	31	22	73	37	22	100	0
13	11	16	50	56	78	93	88	100	29
14	0	30	15	27	50	81	50	100	71
15	17	13	19	45	59	73	65	100	0
16	0	0	0	37	50	86	62	100	22
17	32	4	73	36	40	77	76	100	60
18	27	0	22	32	64	47	56	100	33
19	31	39	45	63	75	76	68	100	30
20	11	16	52	49	79	89	59	100	0
Average	13.50	11.85	34.15	42.00	63.55	68.55	56.15	99.80	19.80
Standard Error	2.27	2.28	5.01	3.61	3.18	4.11	3.98	0.14	4.90

Table B.13: The rating of stimuli in MUSHRA test of each participant (2.34s Drum).

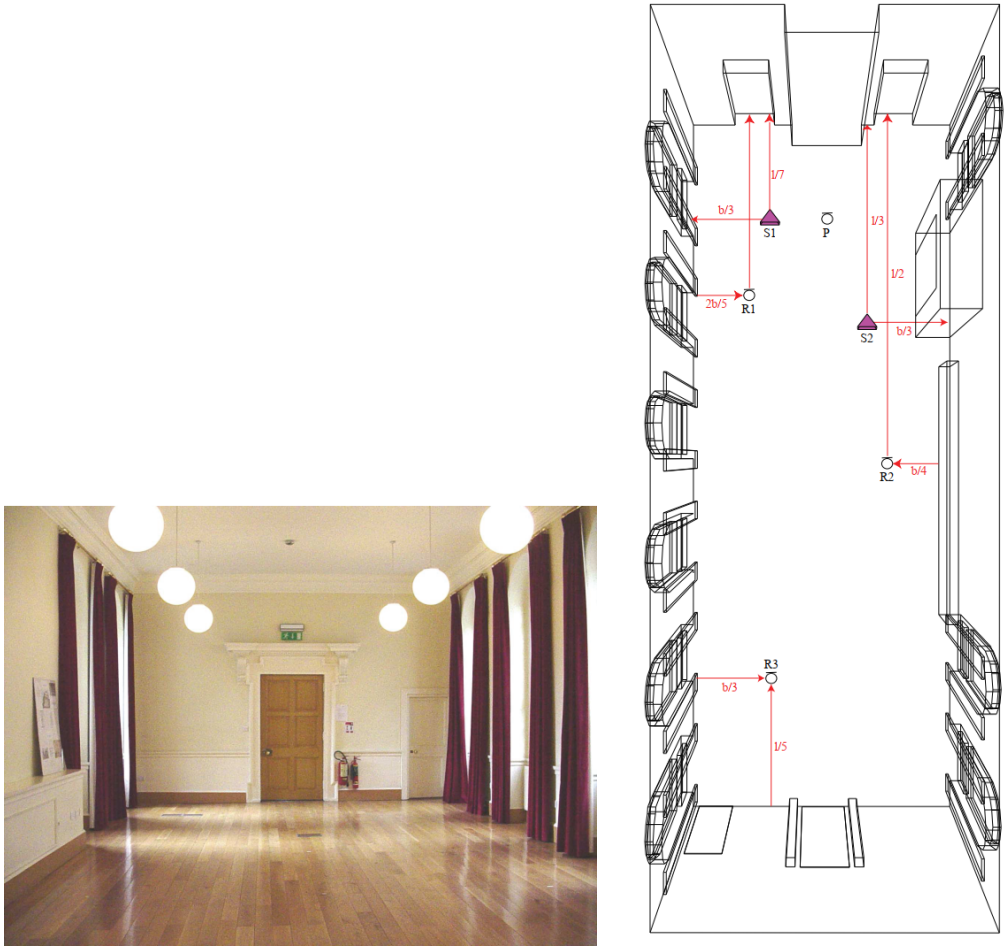
Participants \ Stimuli	Dattorro	DFDN	FDN	Gardner	Moorer	HMS	Schroeder	Ref	Anchor
1	24	14	20	17	48	62	68	100	38
2	10	5	40	59	29	30	20	100	50
3	3	4	5	0	13	0	20	100	0
4	15	23	52	40	70	38	38	100	59
5	29	10	18	85	42	78	71	100	0
6	0	0	0	0	10	10	10	100	0
7	27	0	16	74	54	57	32	100	64
8	38	0	70	50	40	38	70	100	60
9	22	11	16	30	36	24	50	100	11
10	35	32	27	60	52	38	62	100	12
11	6	32	32	36	70	35	52	100	5
12	32	16	22	57	61	50	48	100	0
13	44	0	0	58	90	73	60	100	37
14	26	0	31	57	34	54	12	100	67
15	10	12	12	45	56	48	40	100	0
16	0	10	9	38	21	14	22	100	0
17	20	1	7	57	58	60	33	100	80
18	47	13	0	79	26	63	44	100	58
19	53	27	31	45	57	45	42	100	29
20	39	21	35	90	54	26	53	100	0
Average	24.00	11.55	22.15	48.85	46.05	42.15	42.35	100.00	28.50
Standard Error	3.56	2.41	4.08	5.52	4.54	4.67	4.24	0.00	6.41

(a)



(b)

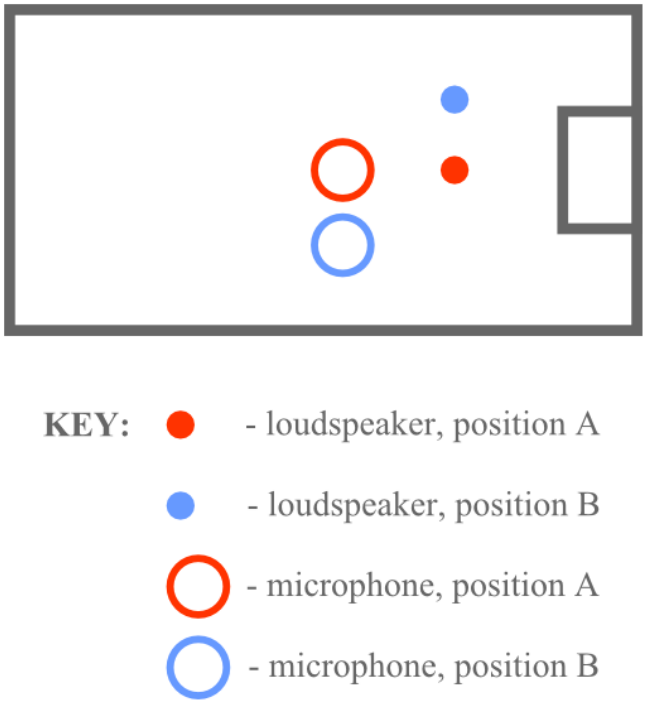
Figure B.1: The Control Room 7 at WDR Broadcast Studios, Germany, where the BRIR with 0.266s reverb time was measured [146]. (**a**) The realistic scene. (**b**) The plane layout.



(a) (b)

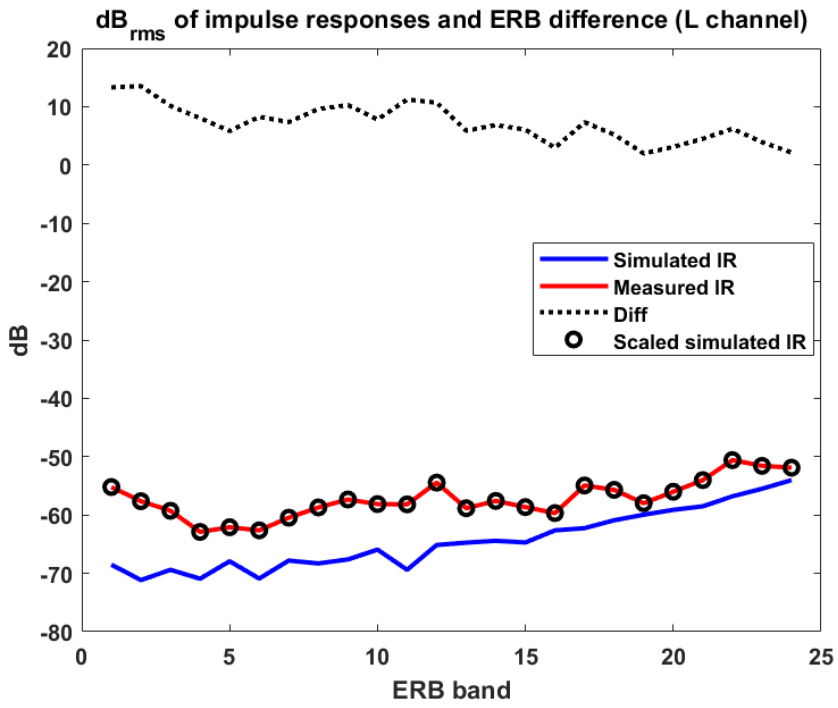
Figure B.2: The Printing House Hall at Trinity College Dublin, Ireland, where the BRIR with 0.95s reverb time was measured [291]. (**a**) Side perspective of staggered loudspeakers. (**b**) Subject perspective.

(a)

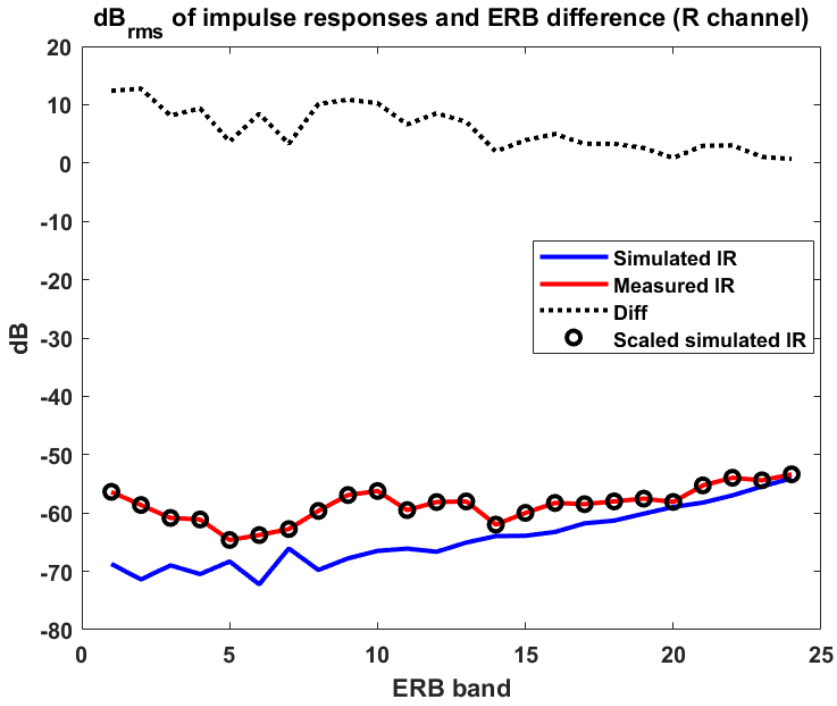


(b)

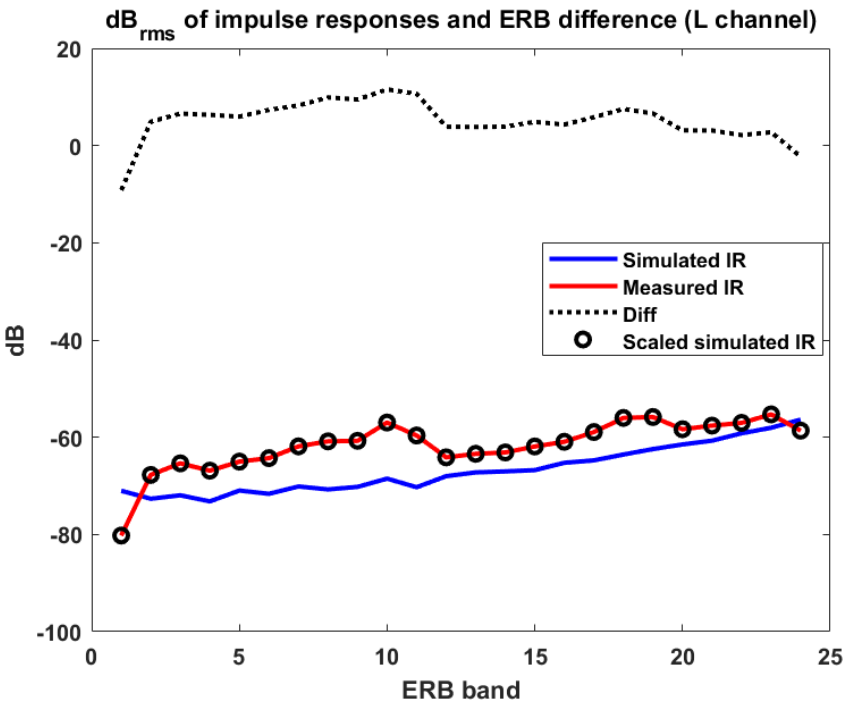
Figure B.3: The Lady Chapel at St Alban's Cathedral, United Kingdom, where the BRIR with 2.34s reverb time was measured [292]. (**a**) The realistic scene. (**b**) The plane layout.


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

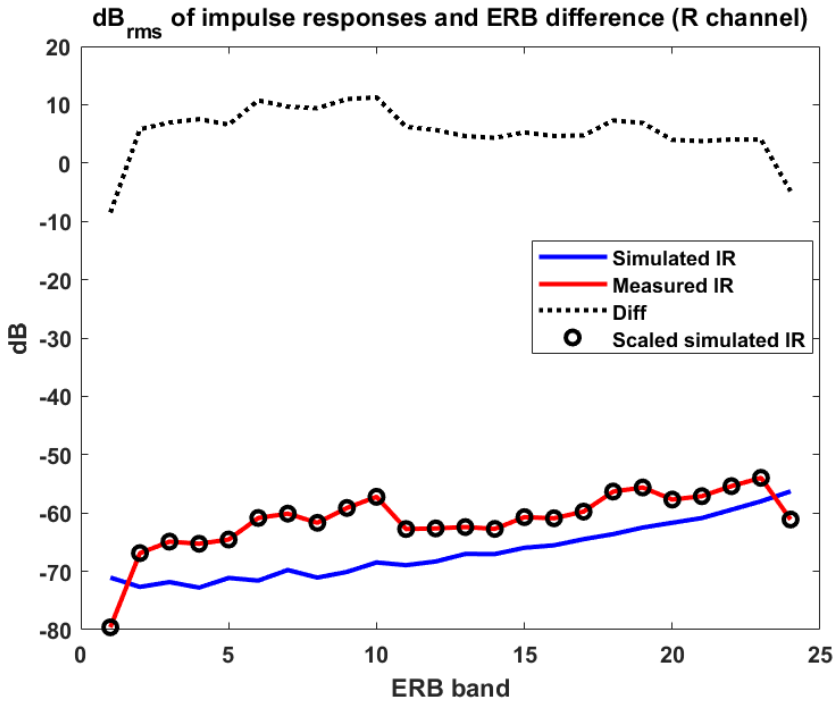
(a)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

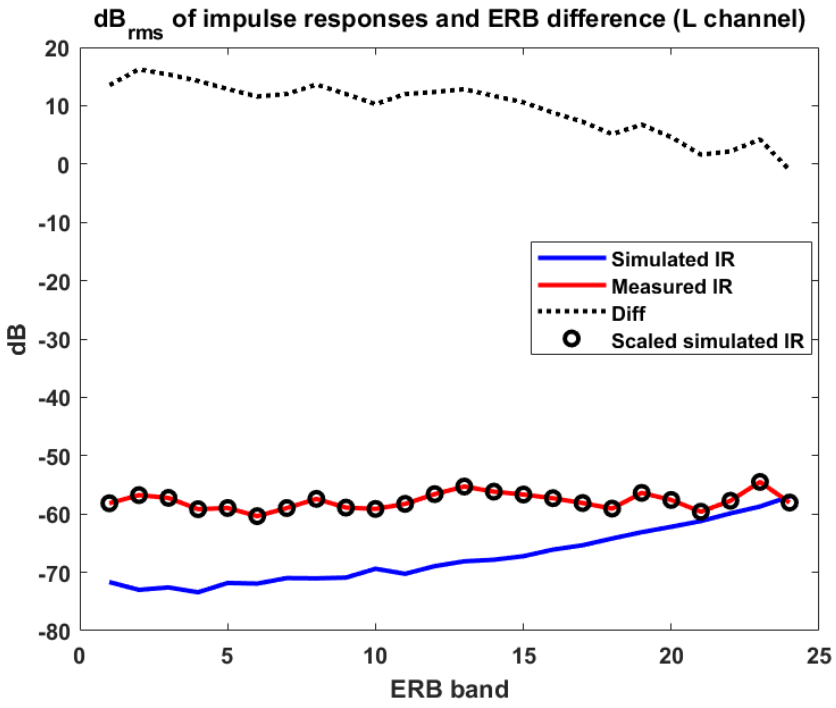
(b)


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

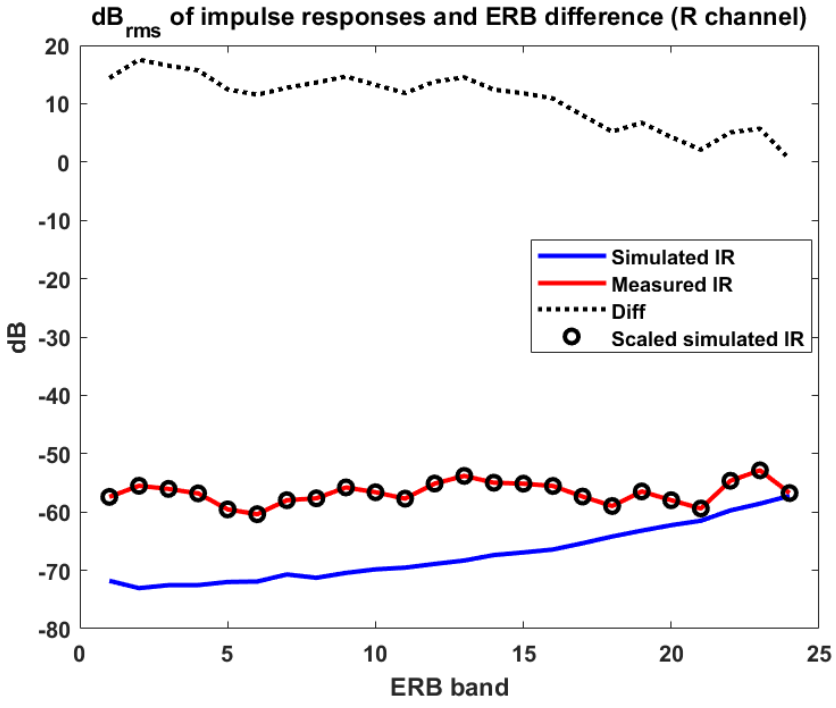
(c)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)

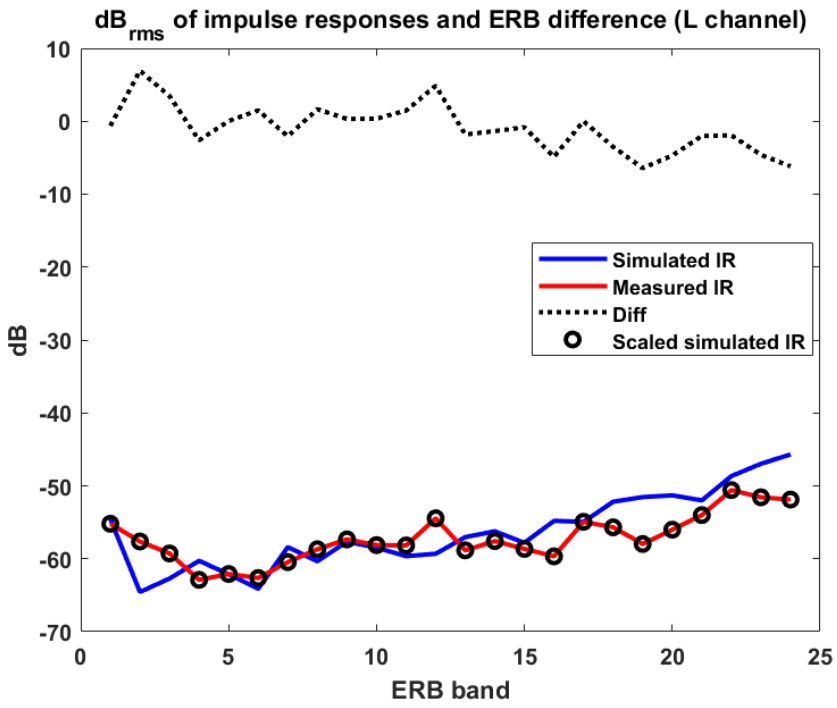


(e)

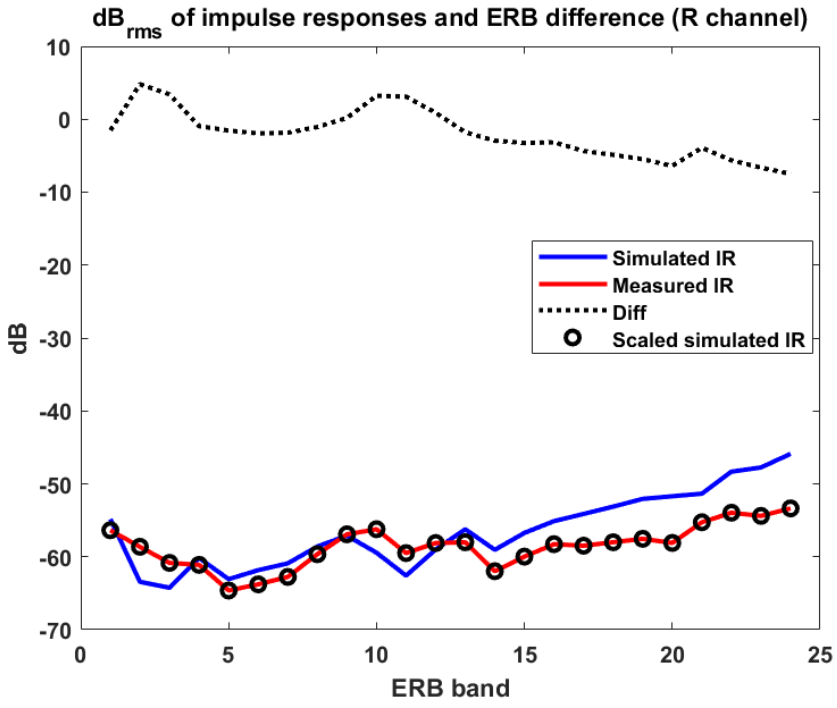


(f)

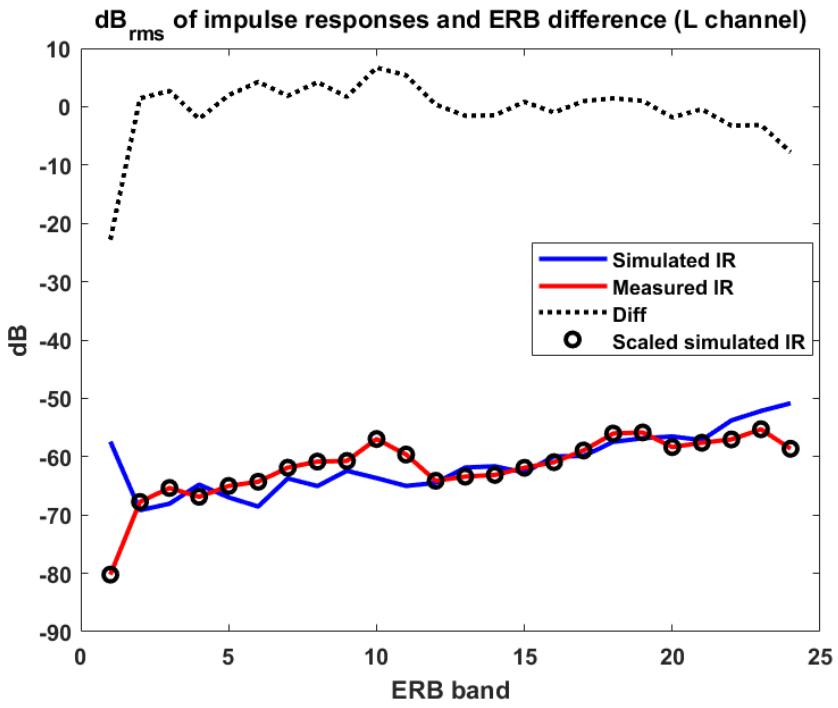
Figure B.4: The timbre matching of the Schroeder-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.


dB rms of impulse responses and ERB difference (L channel)
10
0
-10
-20
-30
-40
-50
-60
-70
dB
0
5
10
15
20
25
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

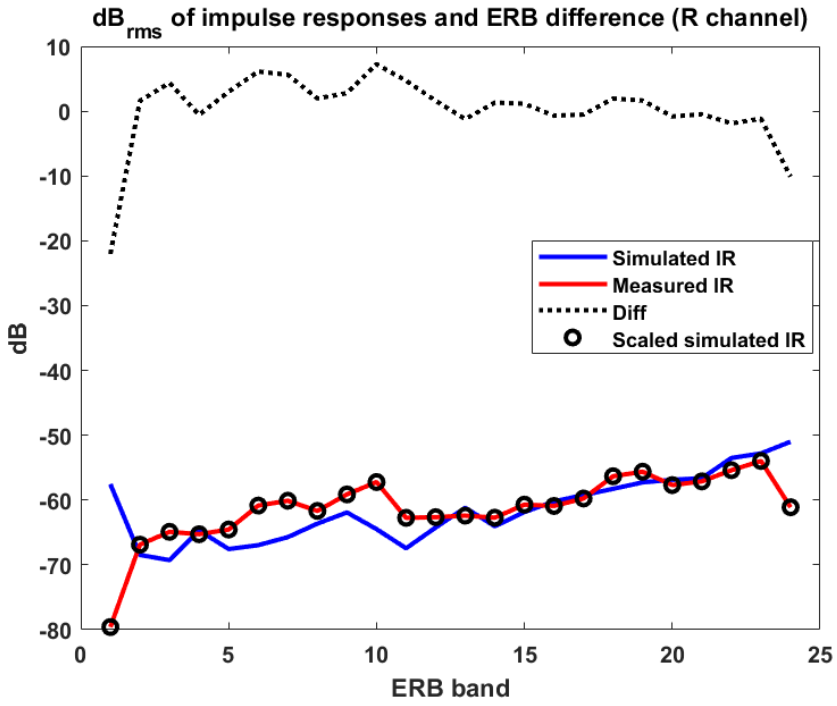
(a)


dB rms of impulse responses and ERB difference (R channel)
10
0
-10
-20
-30
-40
-50
-60
-70
dB
0
5
10
15
20
25
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

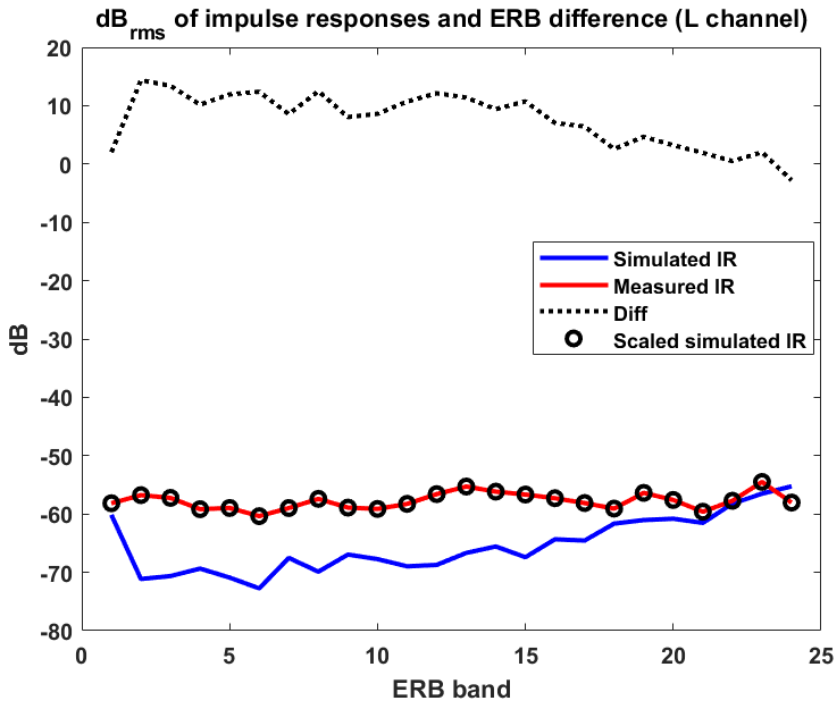
(b)


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

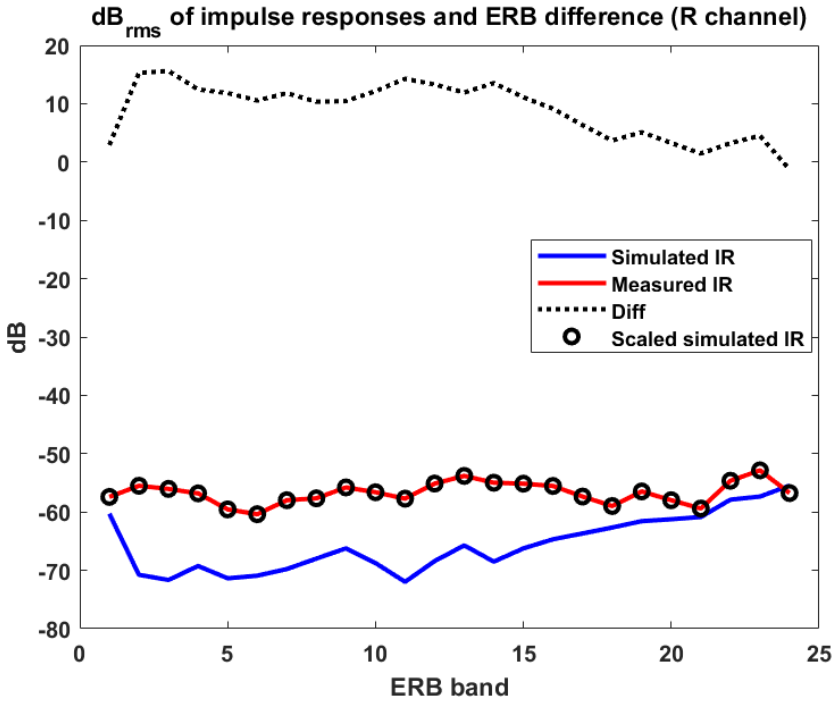
(c)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)

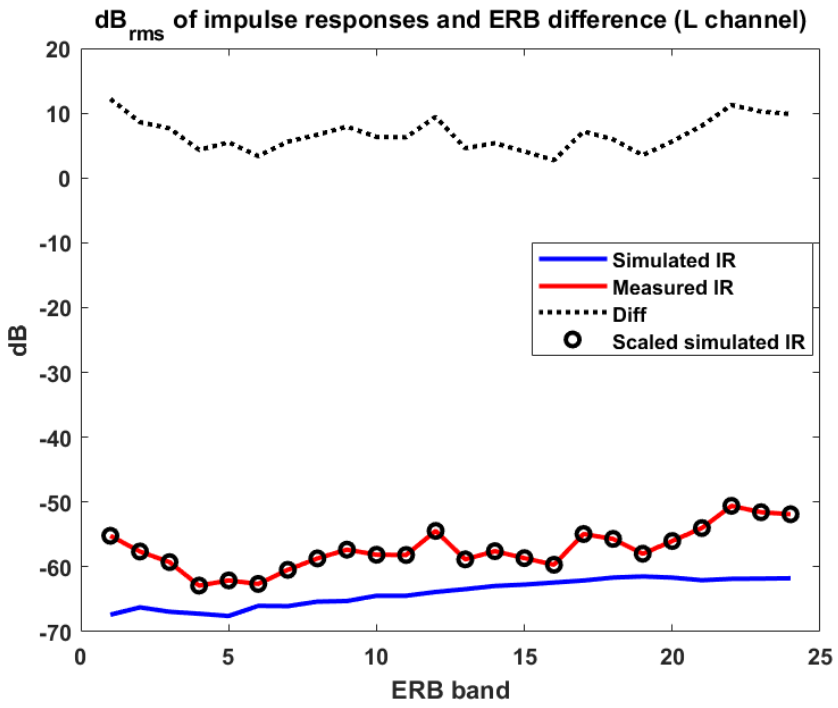


(e)

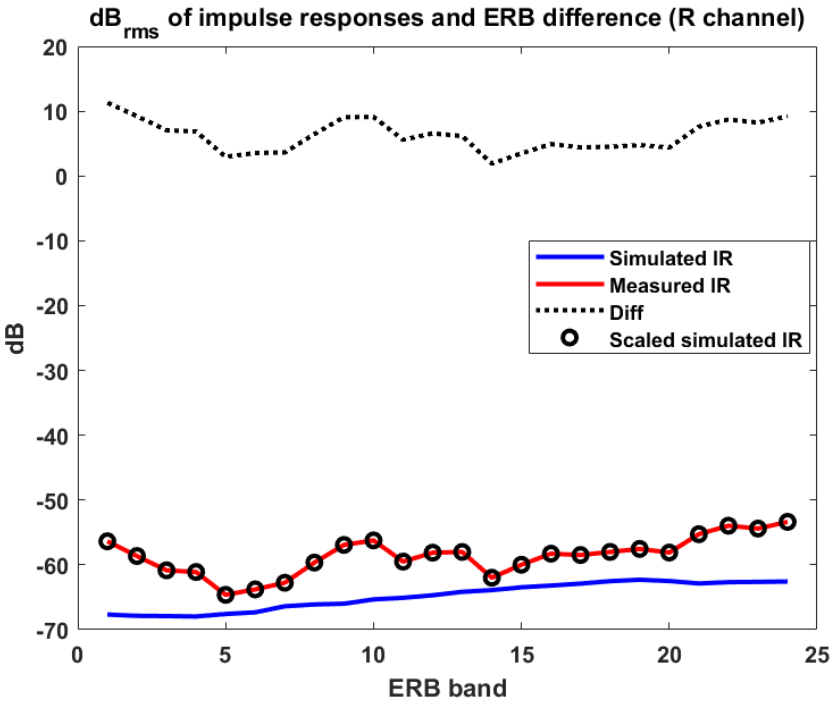


(f)

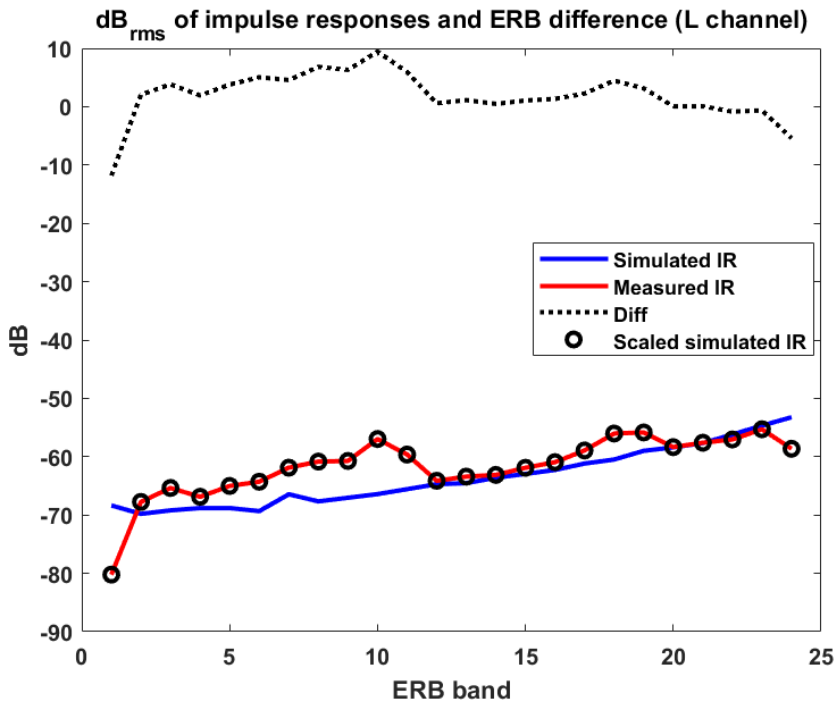
Figure B.5: The timbre matching of the Moorer-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

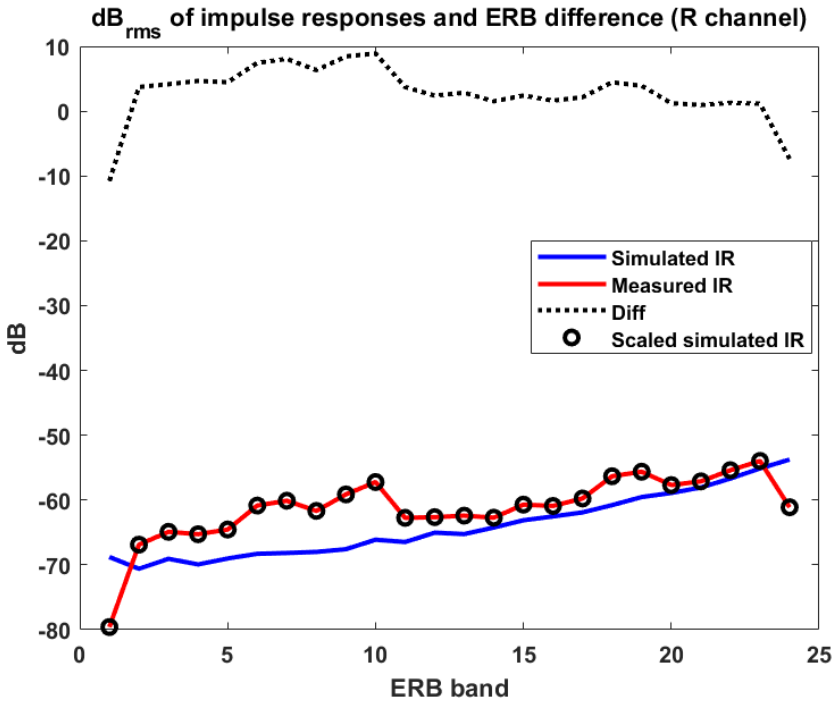
(a)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

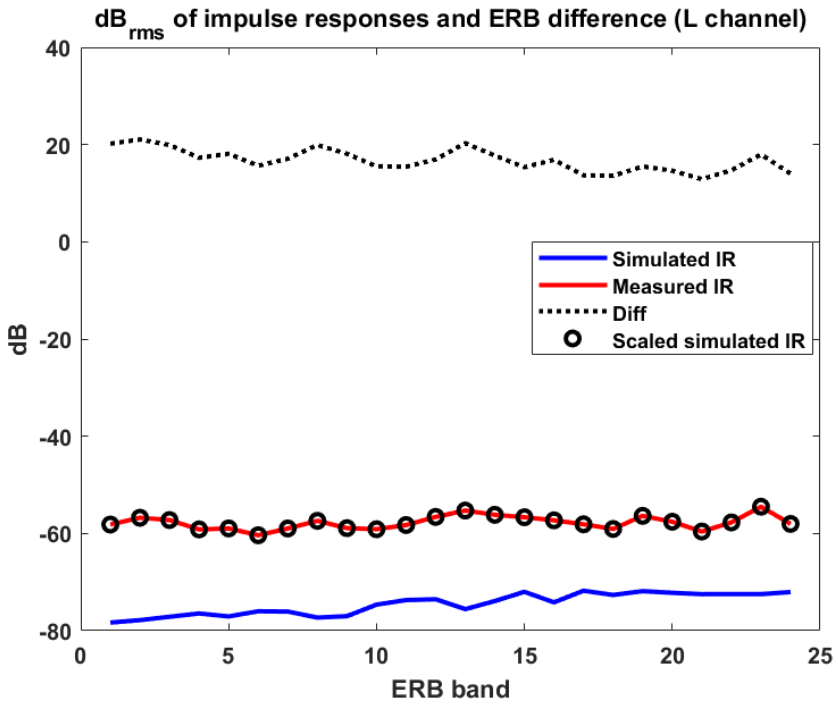
(b)


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

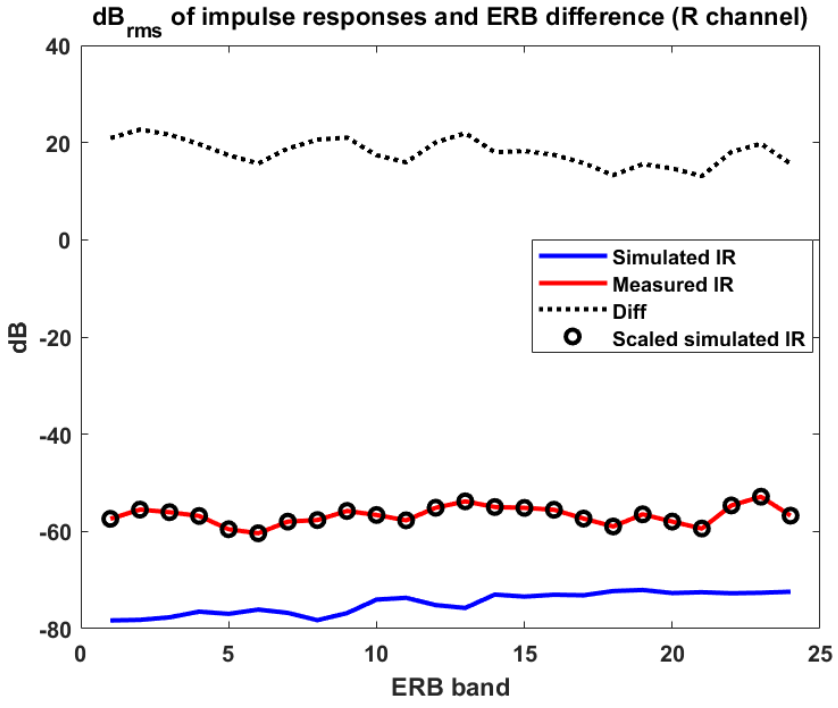
(c)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)

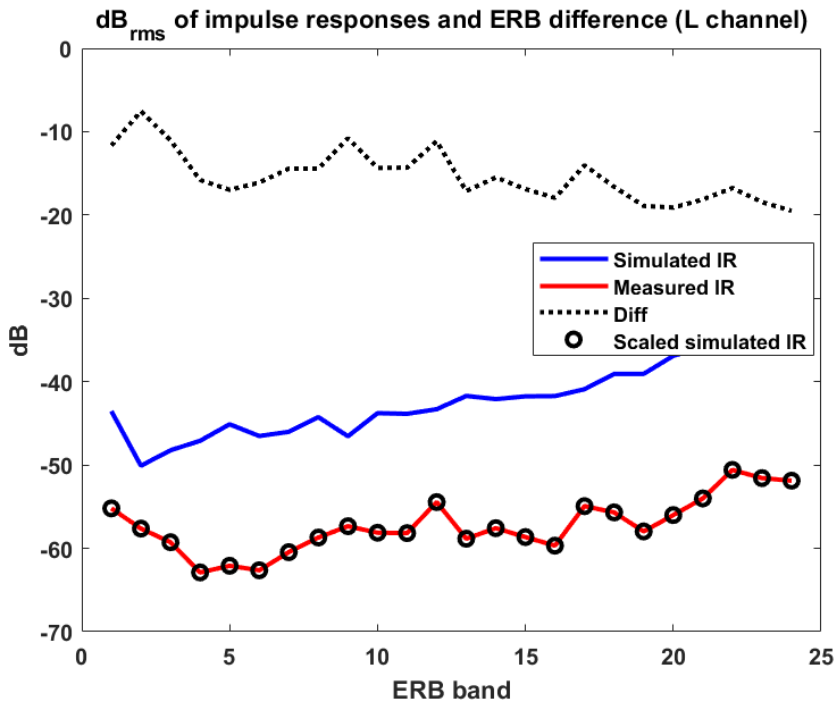


(e)

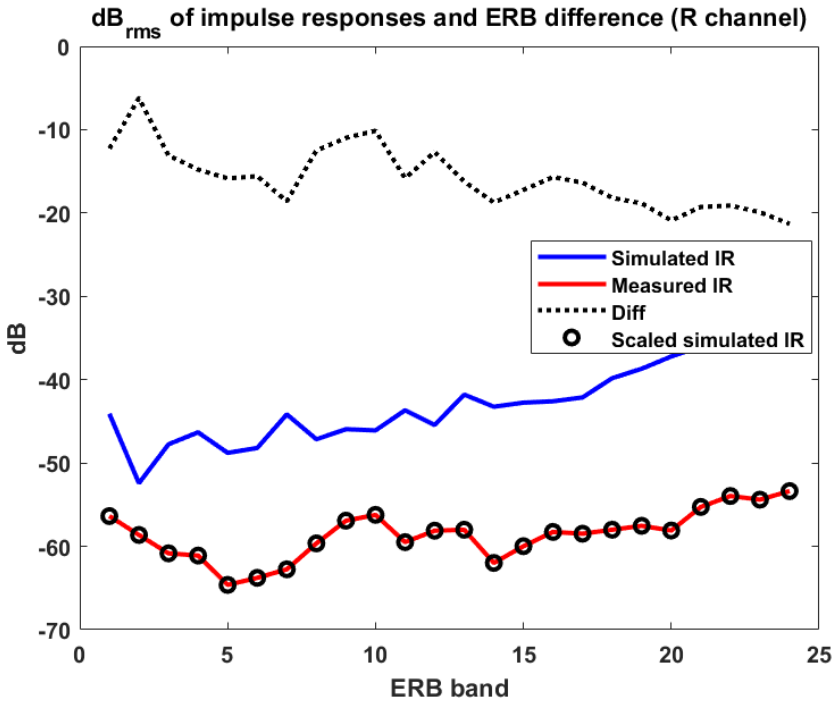


(f)

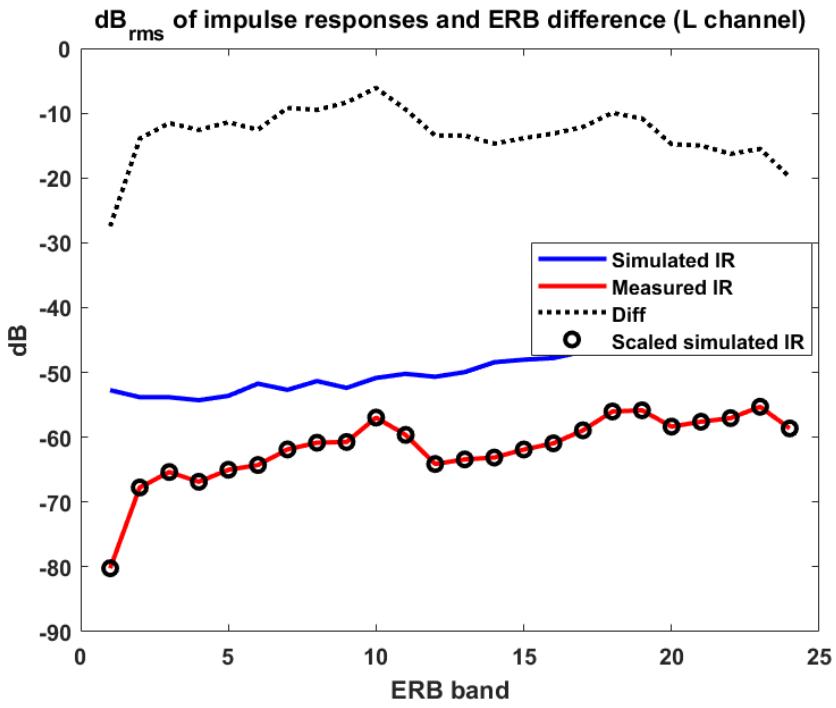
Figure B.6: The timbre matching of the Gardner-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR
0
-10
-20
-30
-40
-50
-60
-70
0
5
10
15
20
25

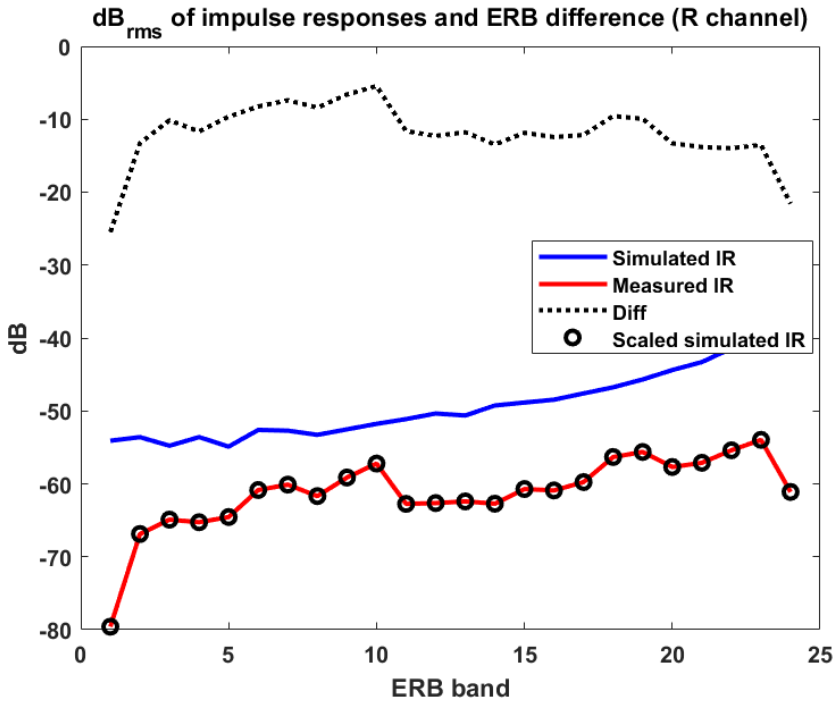
(a)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR
0
-10
-20
-30
-40
-50
-60
-70
0
5
10
15
20
25

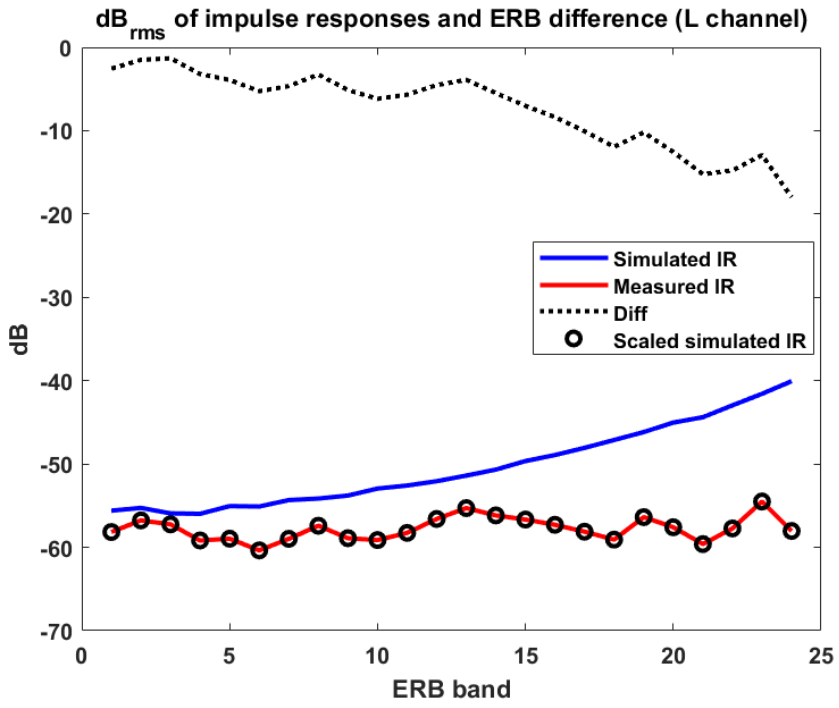
(b)


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

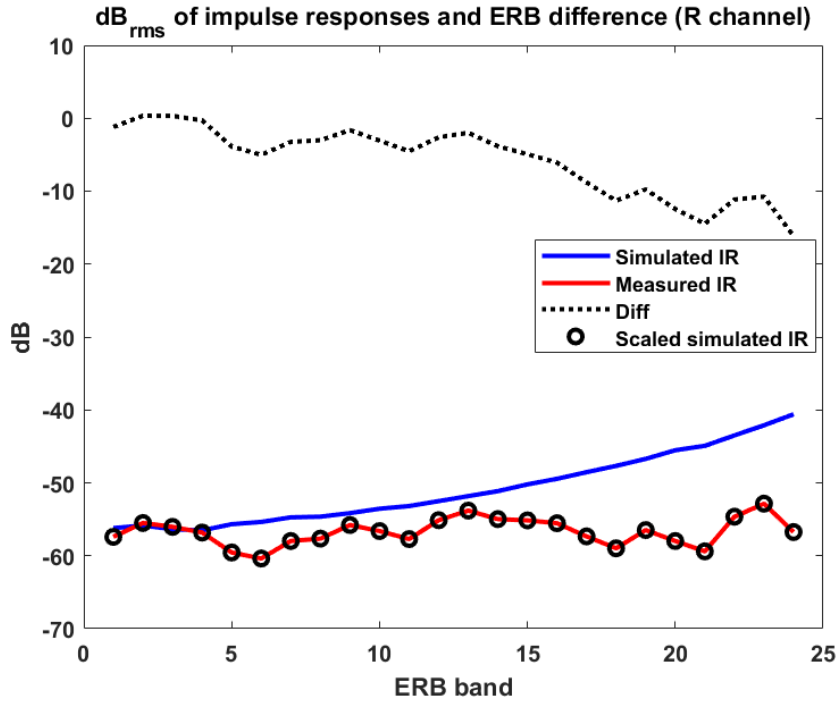
(c)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)

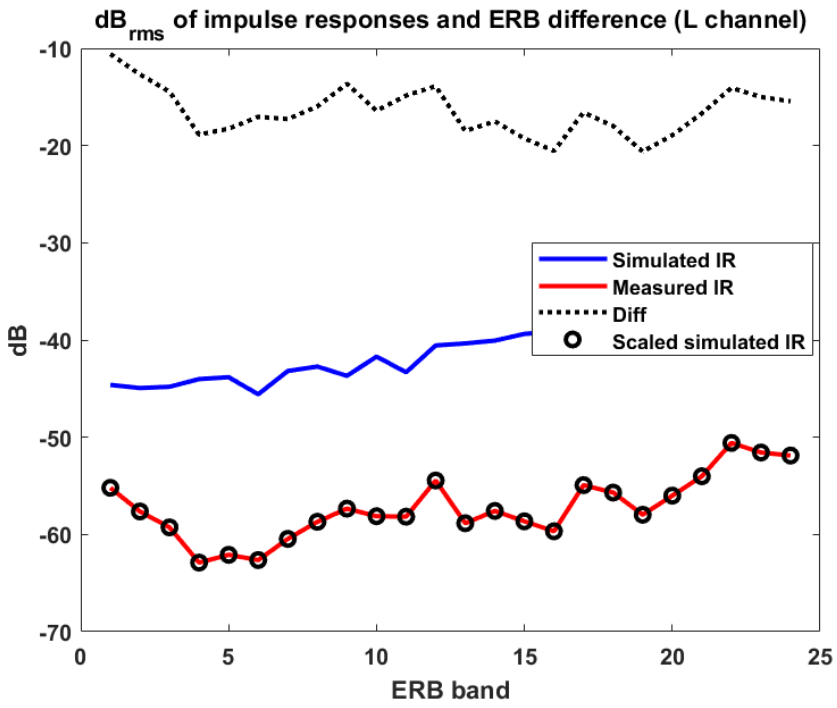


(e)

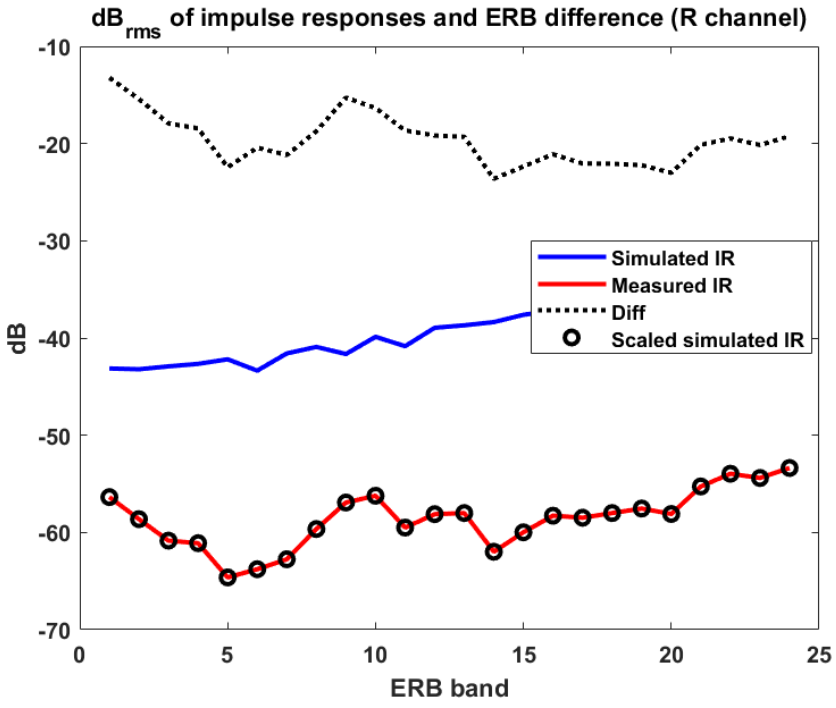


(f)

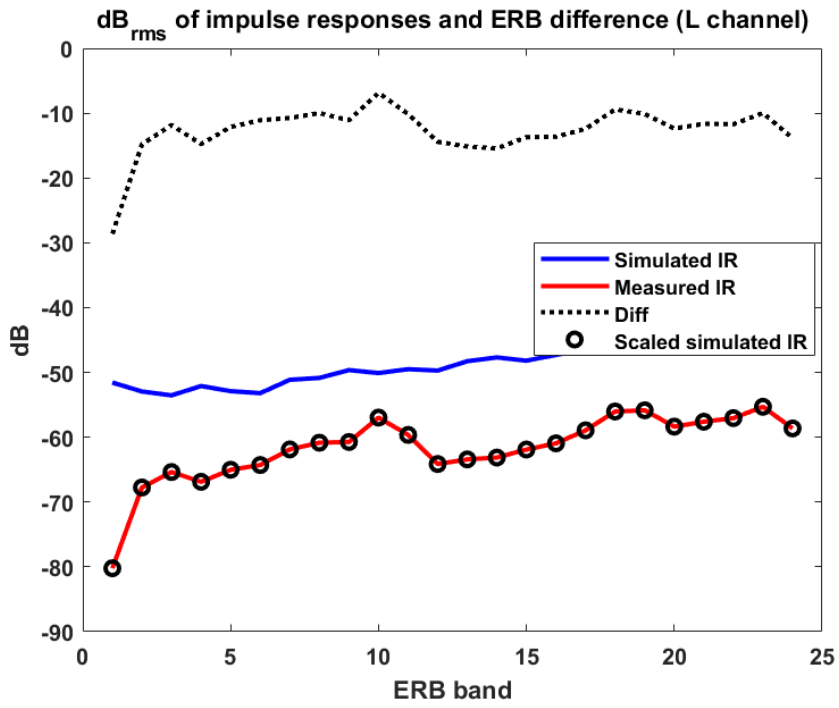
Figure B.7: The timbre matching of the FDN-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.


dB rms of impulse responses and ERB difference (L channel)
-10
-20
-30
-40
-50
-60
-70
dB
0
5
10
15
20
25
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

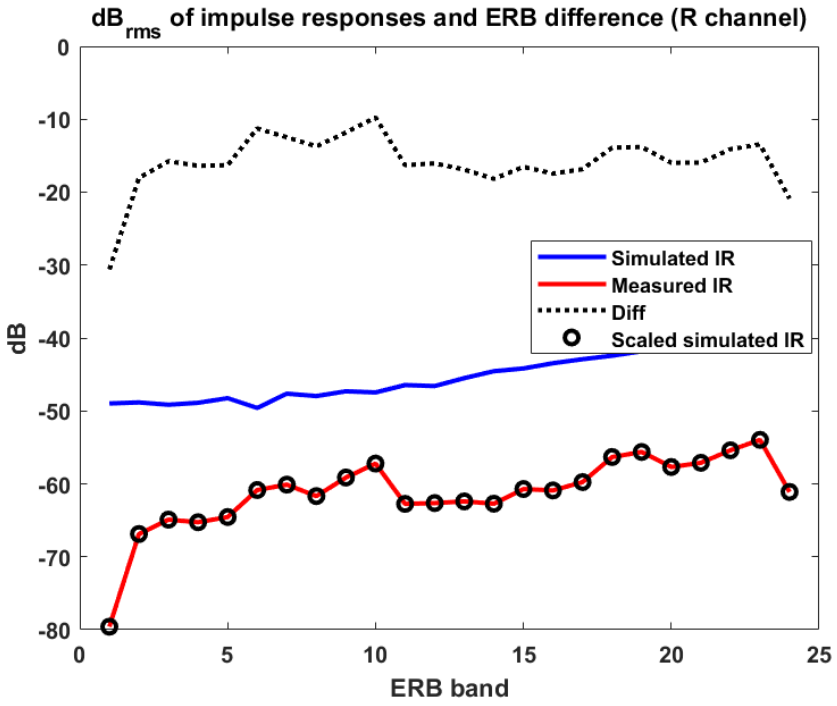
(a)


dB rms of impulse responses and ERB difference (R channel)
-10
-20
-30
-40
-50
-60
-70
dB
0
5
10
15
20
25
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

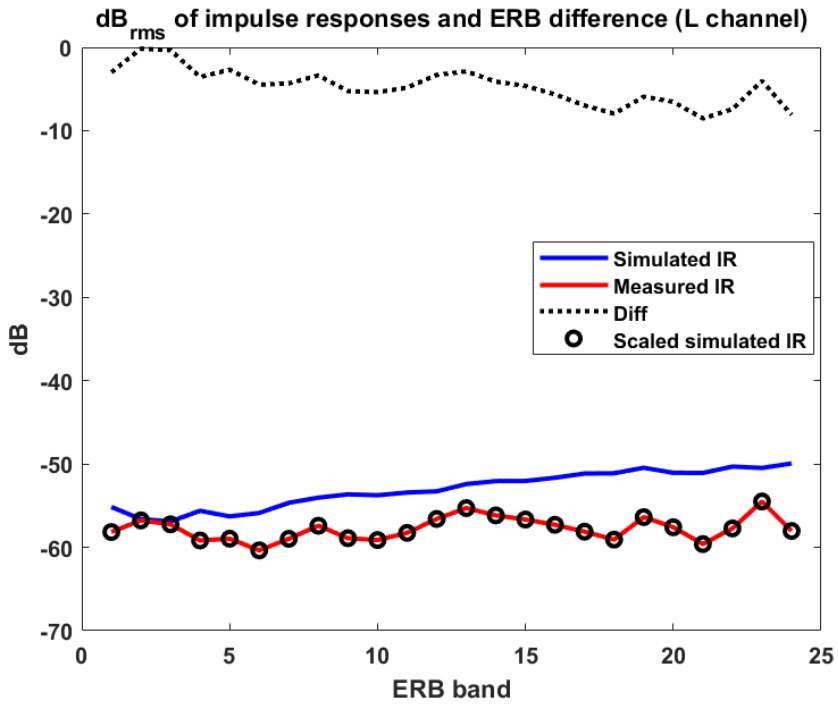
(b)


dB rms of impulse responses and ERB difference (L channel)
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
dB
0
5
10
15
20
25
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

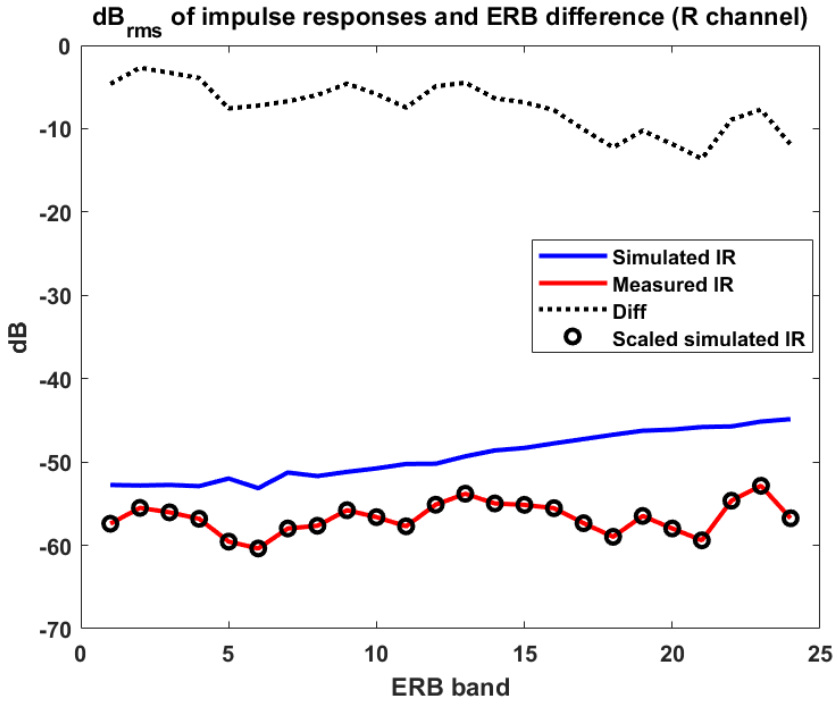
(c)


dB rms of impulse responses and ERB difference (R channel)
0
-10
-20
-30
-40
-50
-60
-70
-80
dB
0
5
10
15
20
25
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)

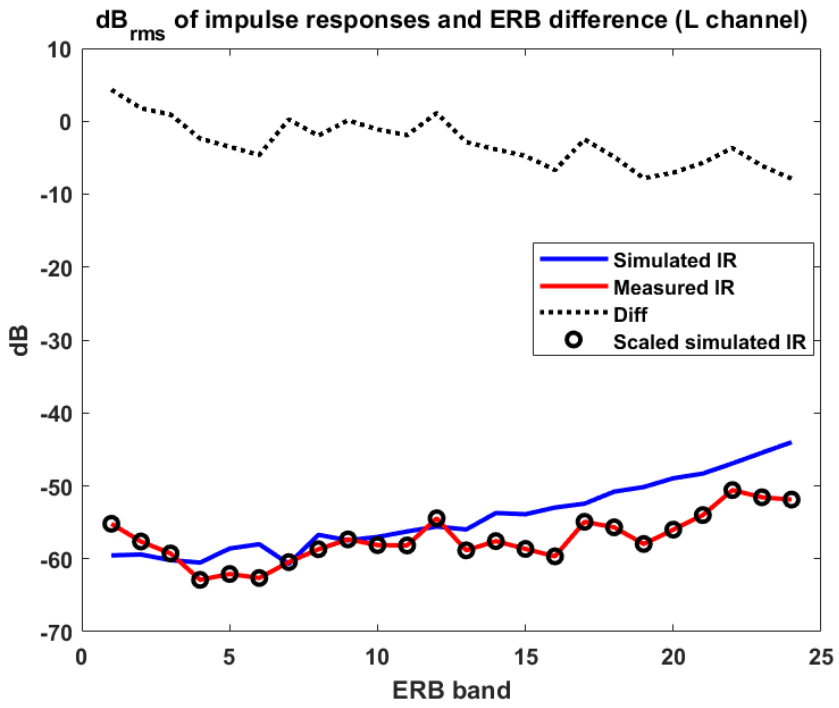


(e)

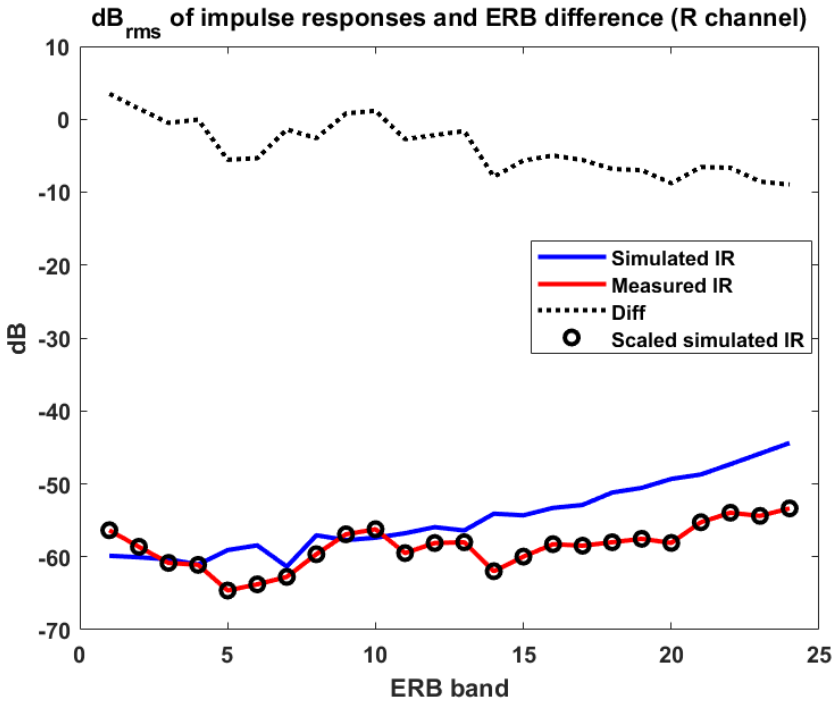


(f)

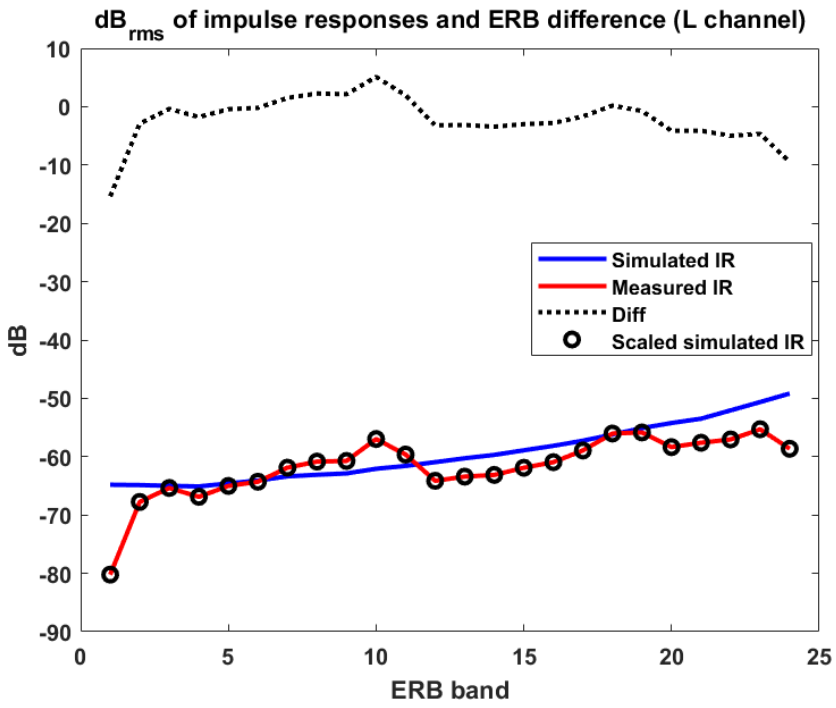
Figure B.8: The timbre matching of the Dattorro-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

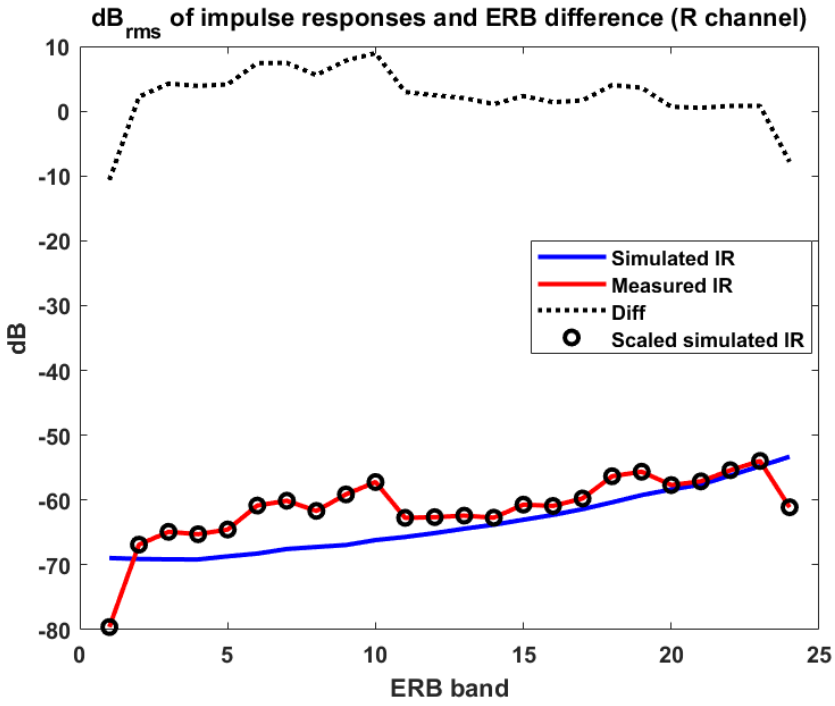
(a)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

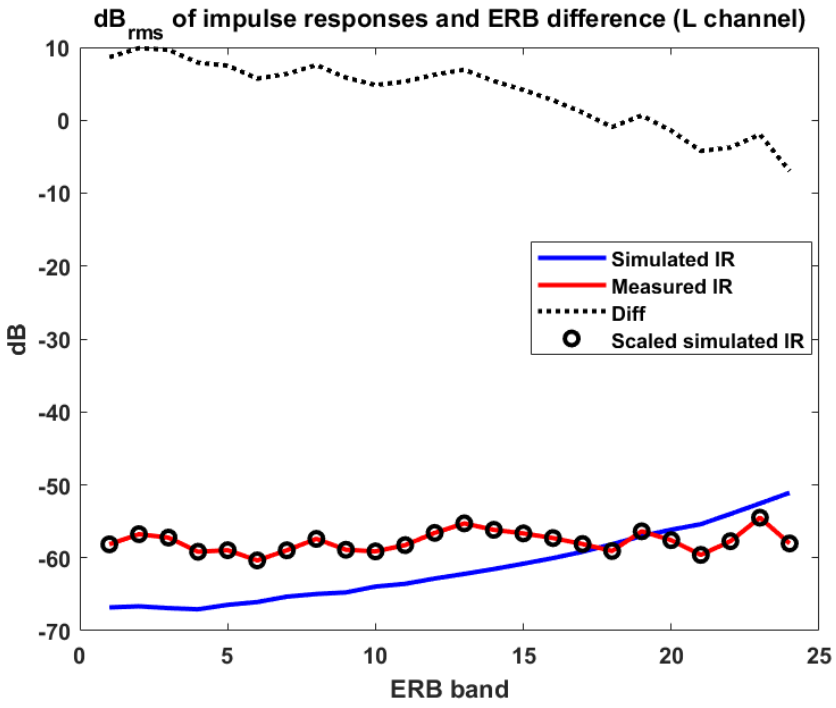
(b)


dB rms of impulse responses and ERB difference (L channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

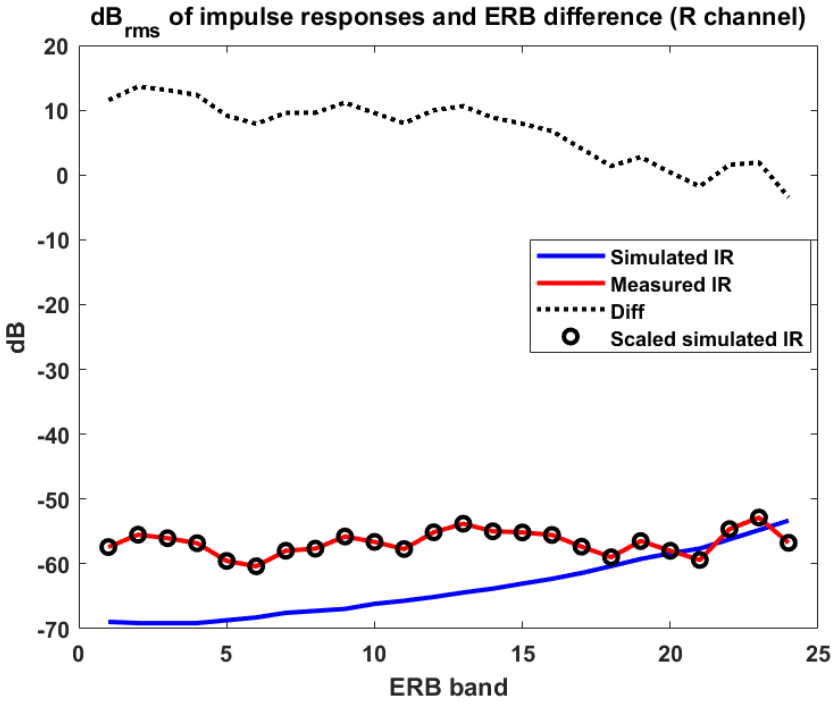
(c)


dB rms of impulse responses and ERB difference (R channel)
dB
ERB band
Simulated IR
Measured IR
Diff
Scaled simulated IR

(d)

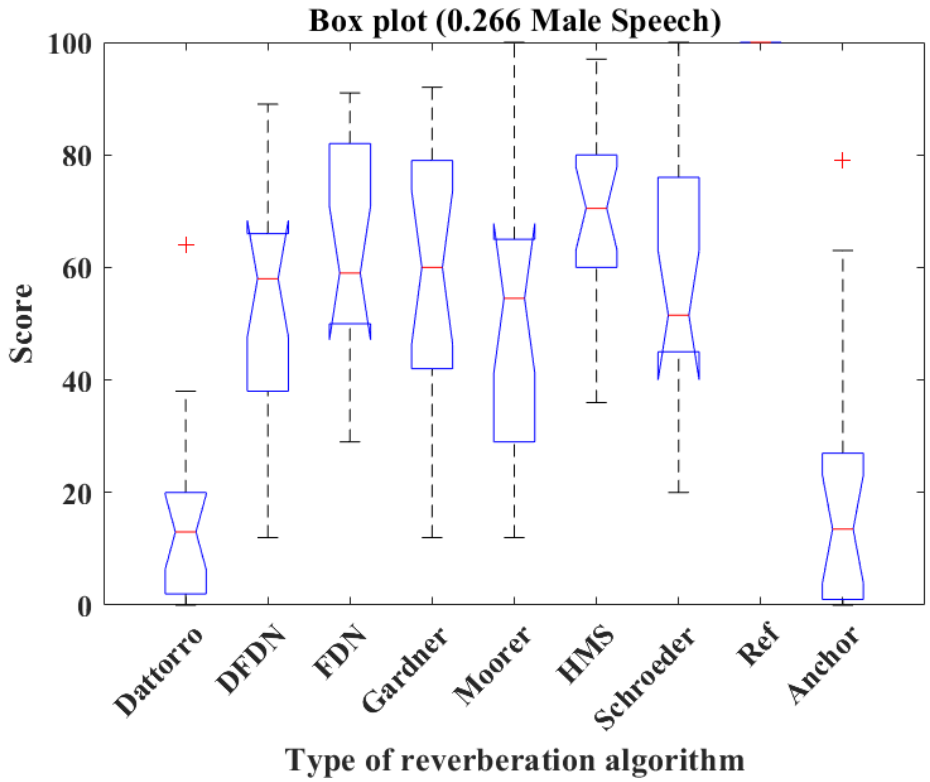


(e)

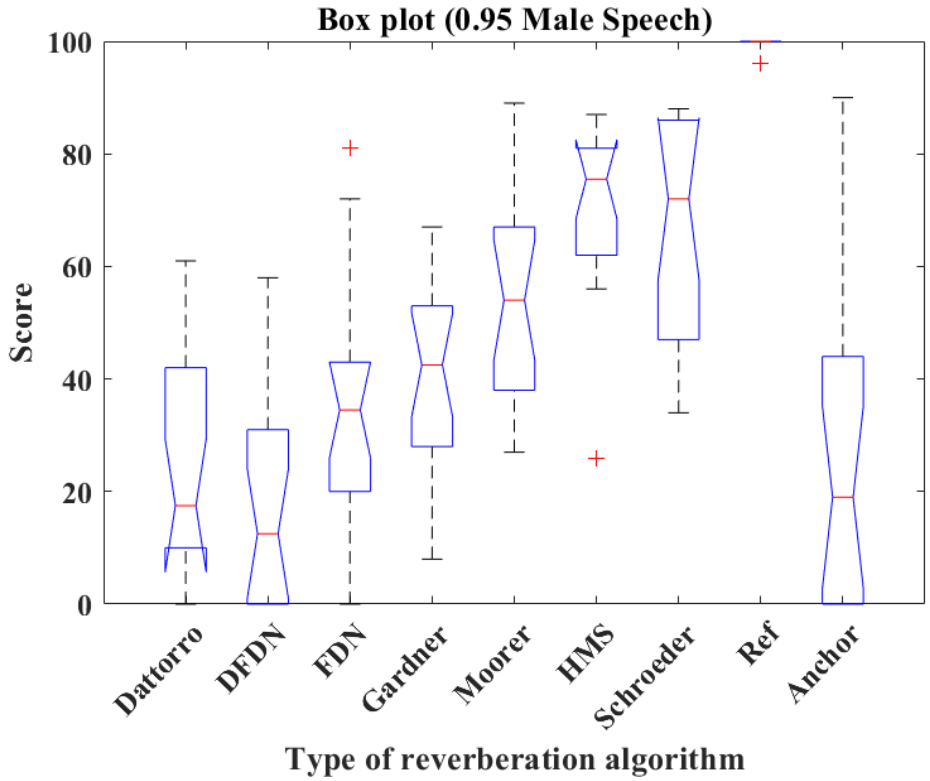


(f)

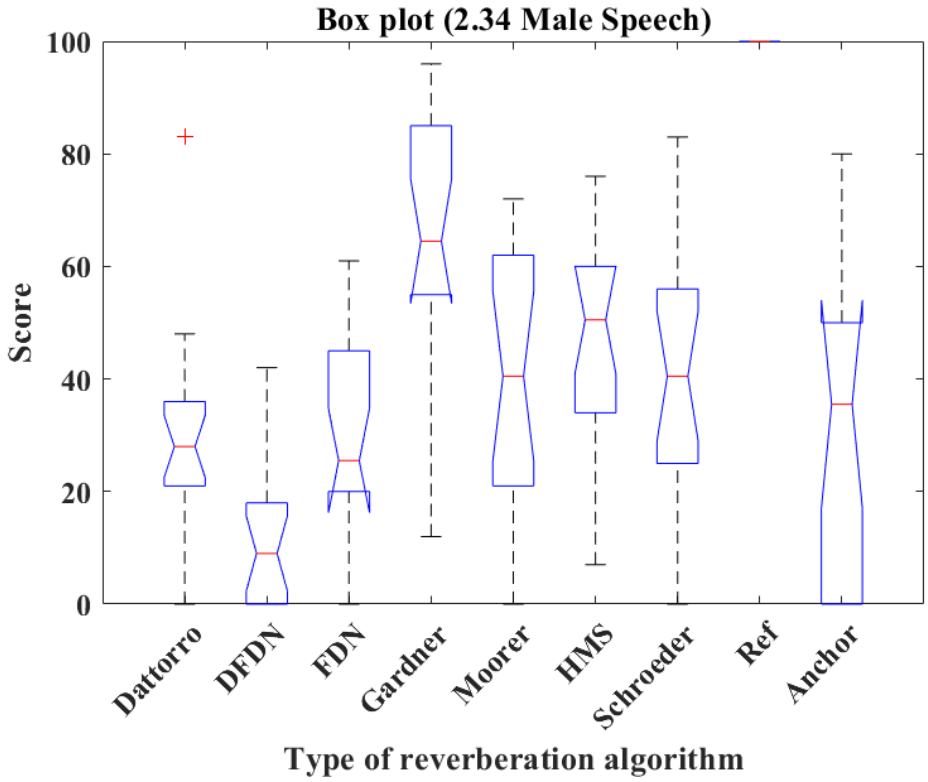
Figure B.9: The timbre matching of the DFDN-algorithm-generated impulse response to the measured impulse response (accomplished by aligning the RMS values within ERB bands). (**a**) The left channel of the BRIR with 0.266 s reverberation time. (**b**) The right channel of the BRIR with 0.266 s reverberation time. (**c**) The left channel of the BRIR with 0.95 s reverberation time. (**d**) The right channel of the BRIR with 0.95 s reverberation time. (**e**) The left channel of the BRIR with 2.34 s reverberation time. (**f**) The right channel of the BRIR with 2.34 s reverberation time.


Box plot (0.266 Male Speech)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(a)

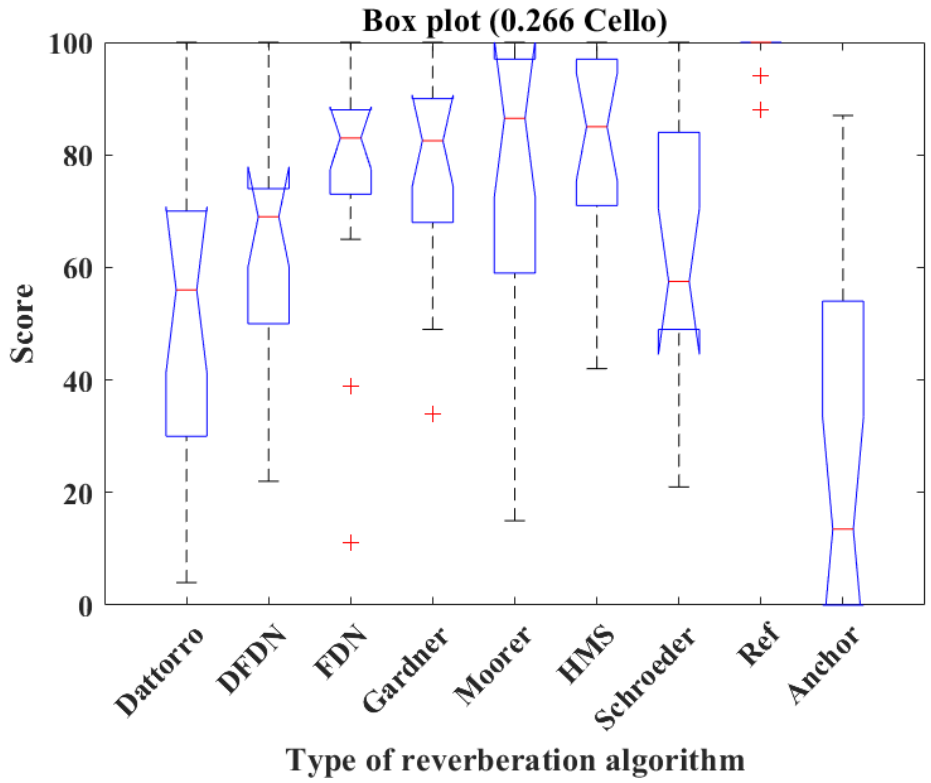

Box plot (0.95 Male Speech)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(b)

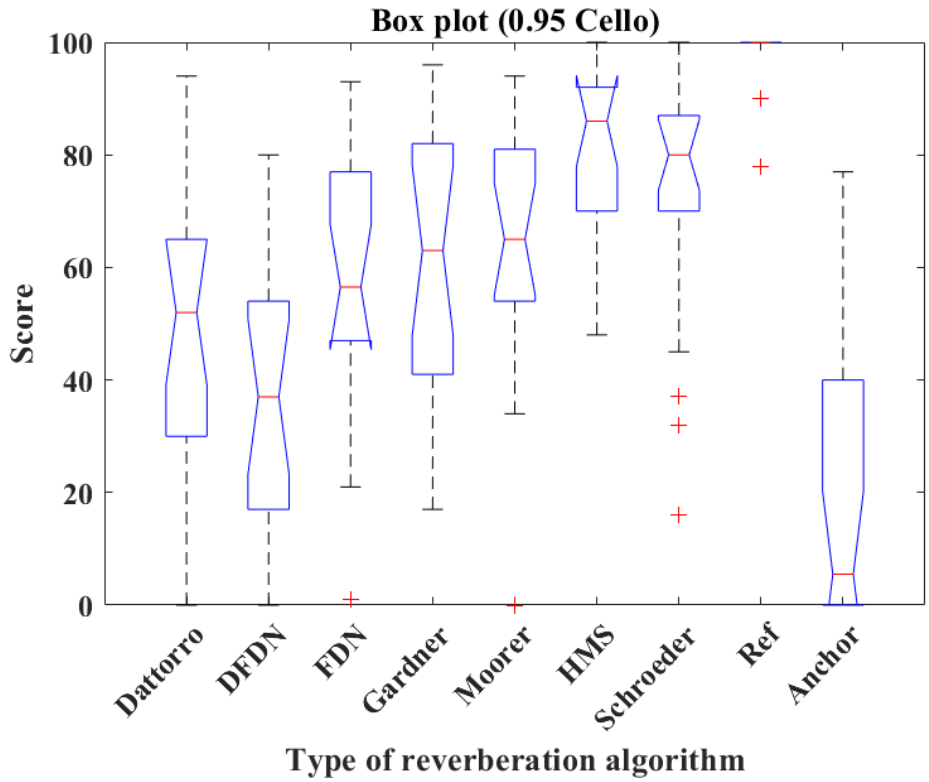


(c)

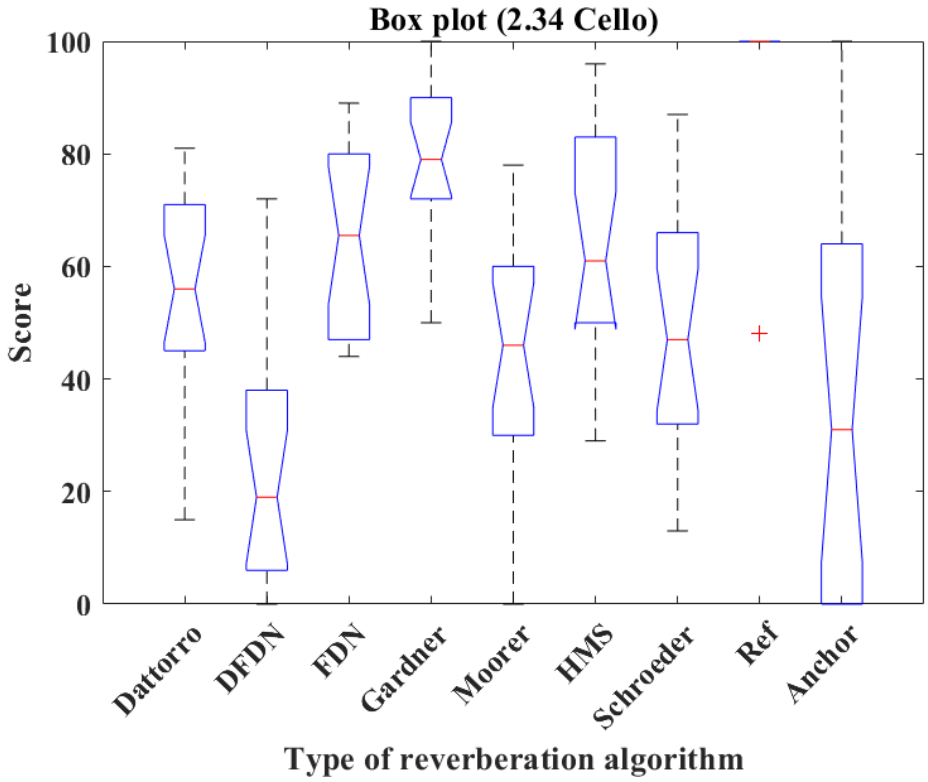
Figure B.10: Box plots of the scores of seven reverberation algorithms, reference and anchor simulating male speech with 0.266 s, 0.95 s, and 2.34 s reverberation time. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s ('+' in figures presents outliers).


Box plot (0.266 Cello)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(a)

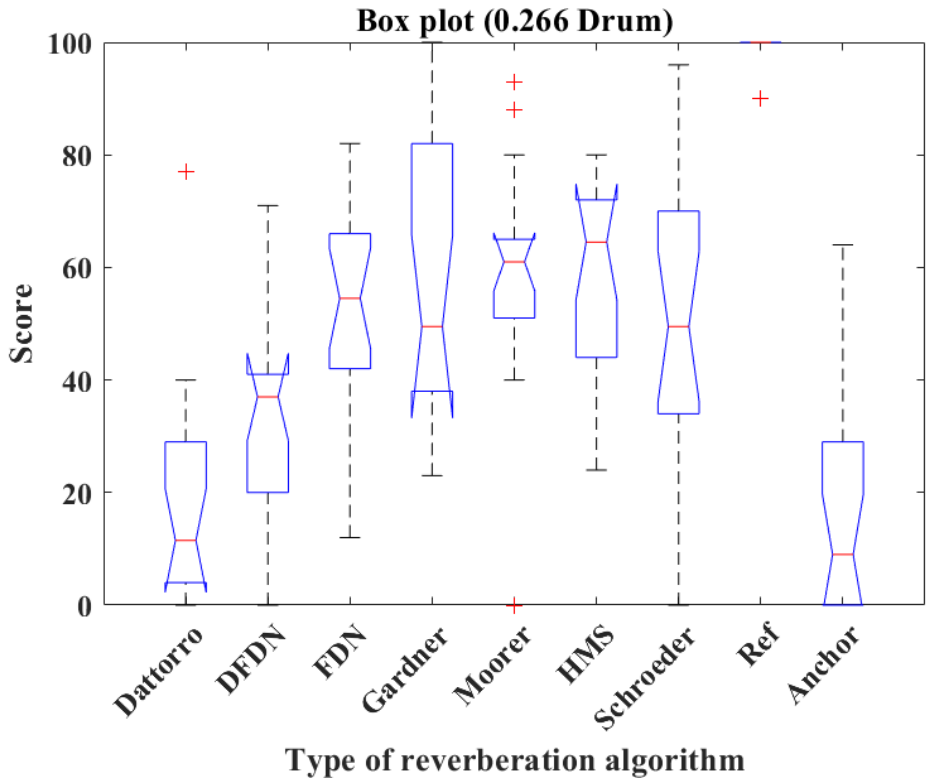

Box plot (0.95 Cello)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(b)

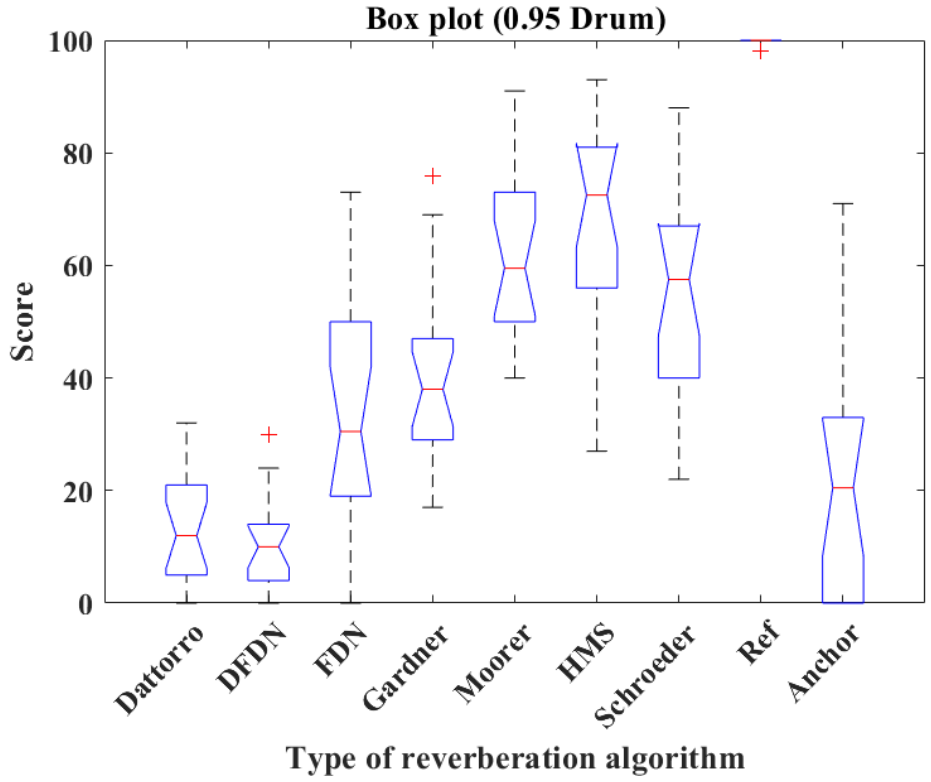


(c)

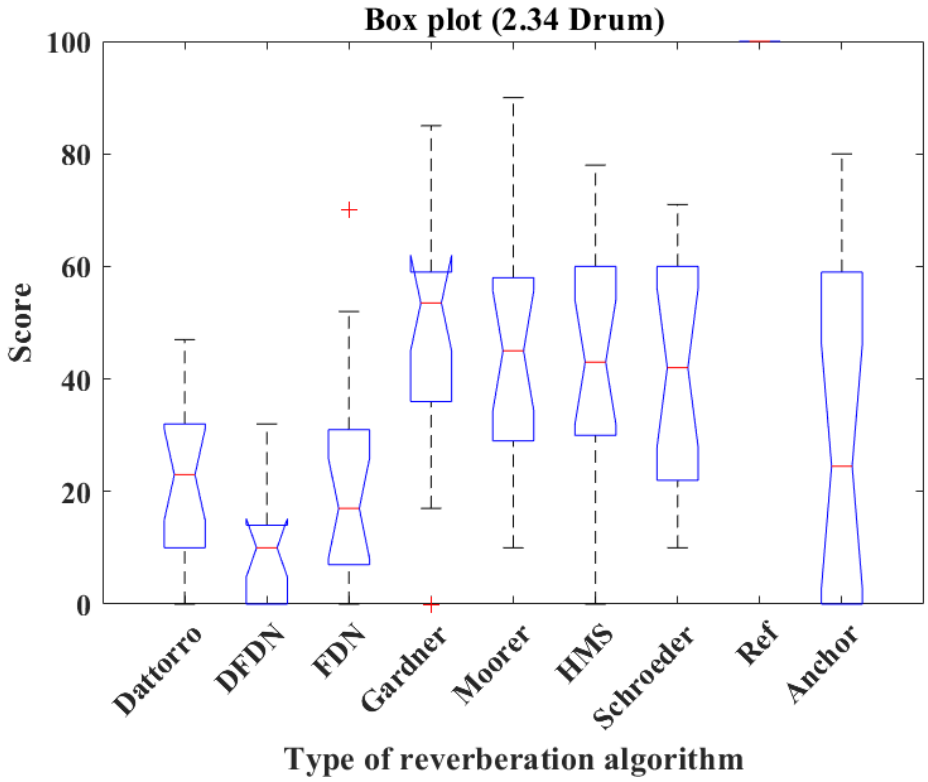
Figure B.11: Box plots of the scores of seven reverberation algorithms, reference and anchor simulating cello piece with 0.266 s, 0.95 s, and 2.34 s reverberation time. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s ('+' in figures presents outliers).


Box plot (0.266 Drum)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(a)

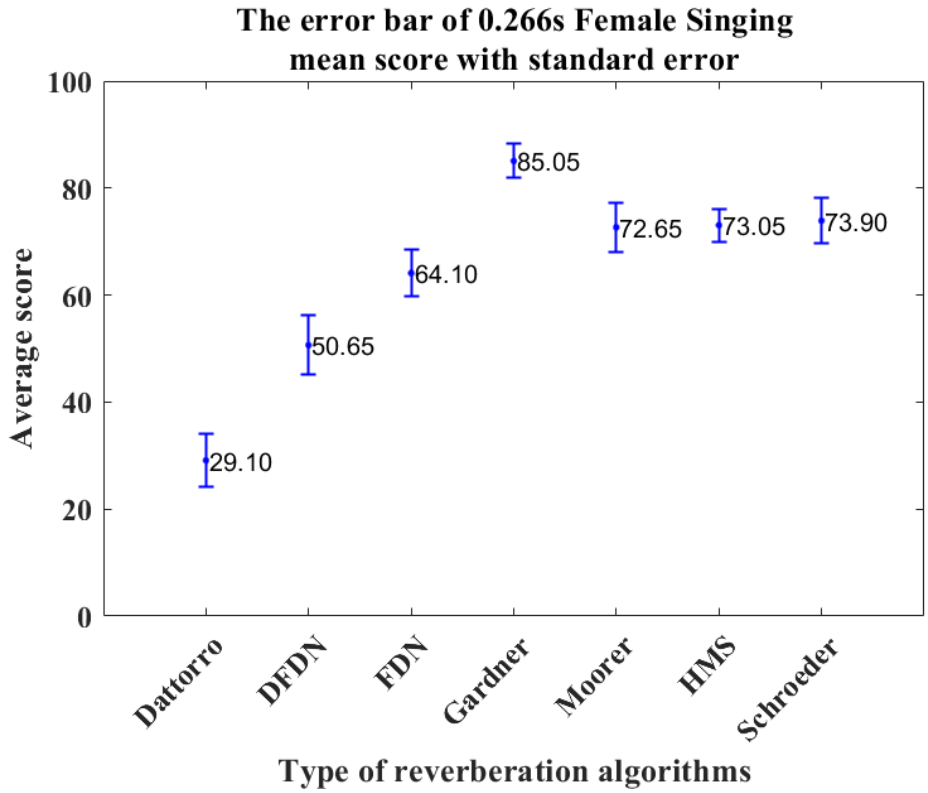

Box plot (0.95 Drum)
Score
100
80
60
40
20
0
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Ref
Anchor
Type of reverberation algorithm

(b)

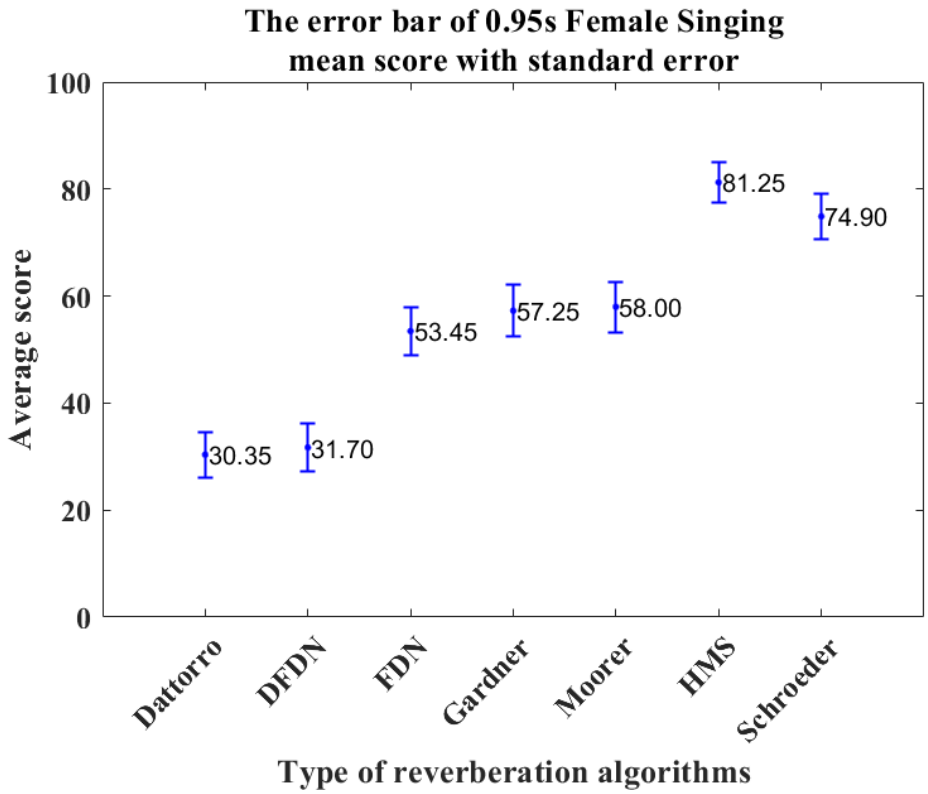


(c)

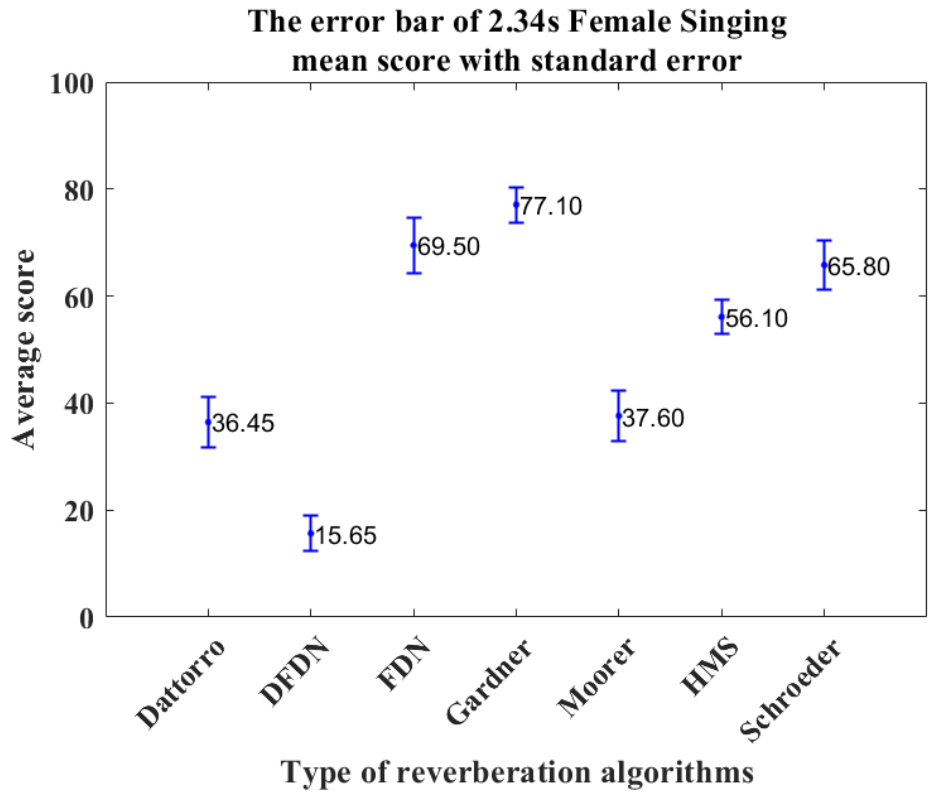
Figure B.12: Box plots of the scores of seven reverberation algorithms, reference and anchor simulating drum beat with 0.266 s, 0.95 s, and 2.34 s reverberation time. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s ('+' in figures presents outliers).


The error bar of 0.266s Female Singing
mean score with standard error
Average score
100
80
60
40
20
0
85.05
72.65
73.05
73.90
64.10
50.65
29.10
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(a)

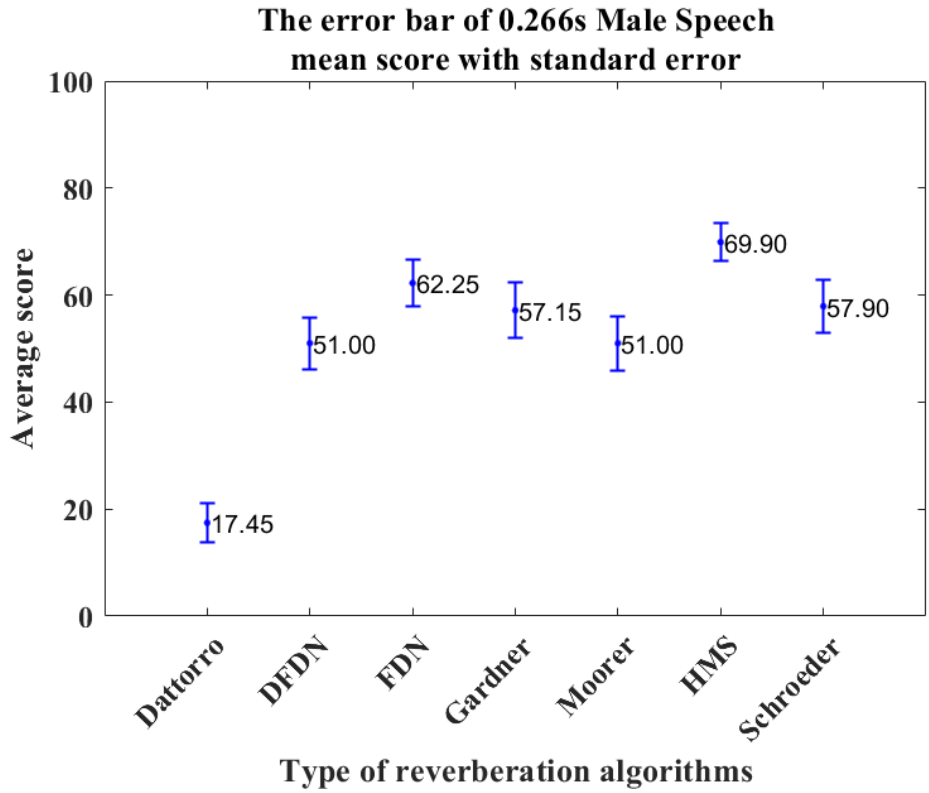

The error bar of 0.95s Female Singing
mean score with standard error
Average score
100
80
60
40
20
0
81.25
74.90
57.25
58.00
53.45
30.35
31.70
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(b)

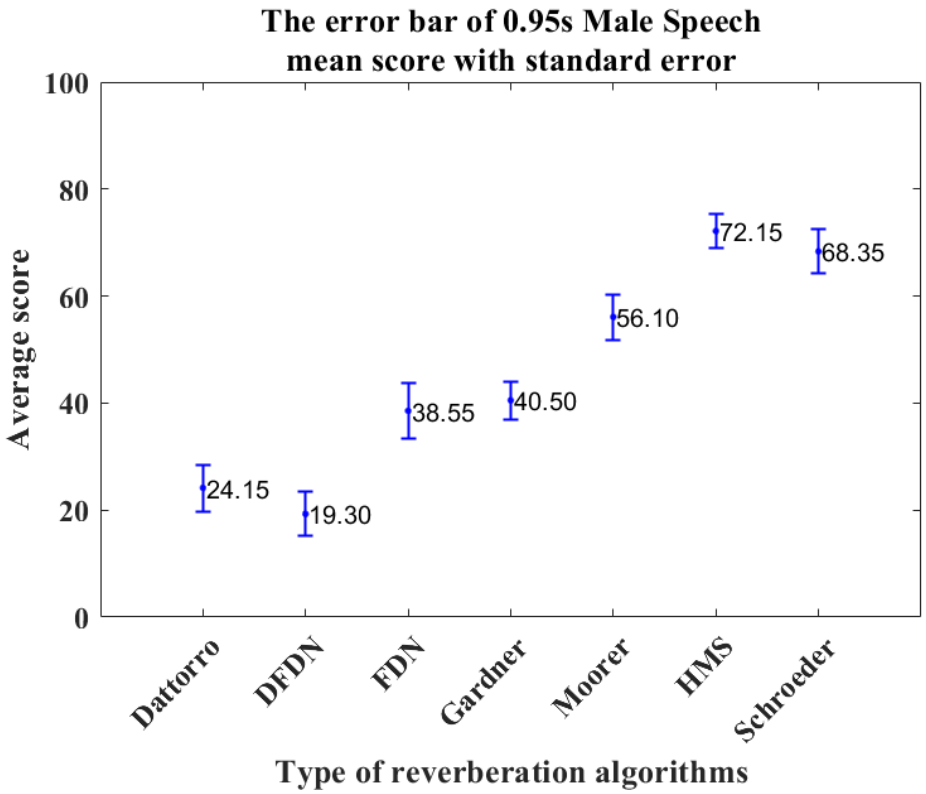


(c)

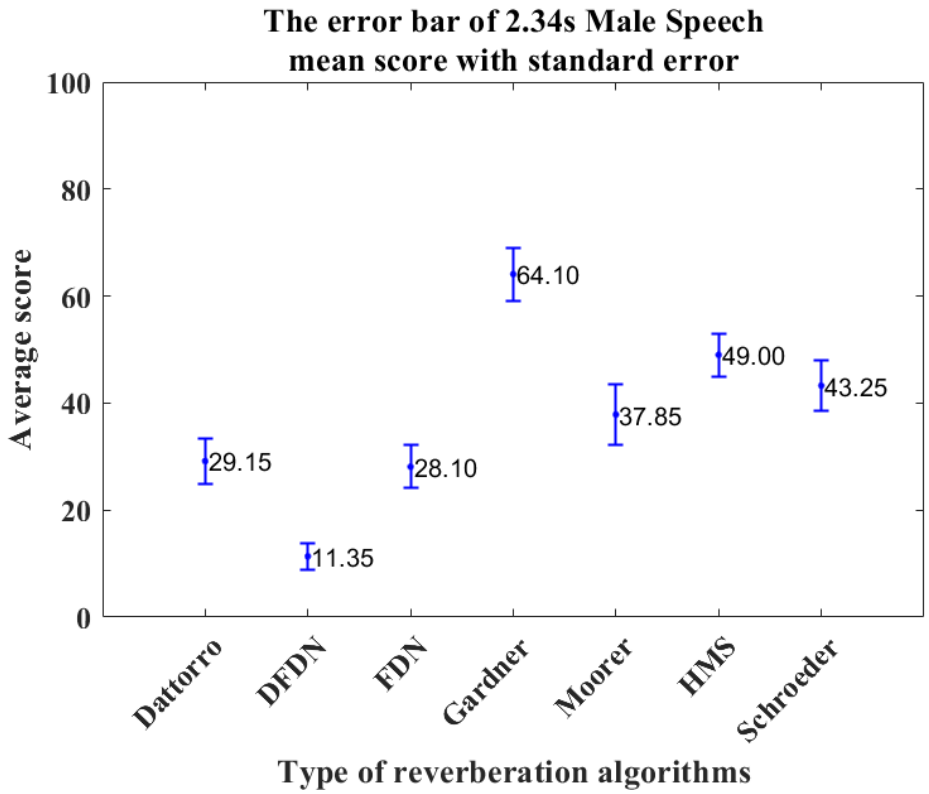
Figure B.13: The error bars of the mean score with standard error of female singing simulated by reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s.


The error bar of 0.266s Male Speech
mean score with standard error
Average score
100
80
60
40
20
0
17.45
51.00
62.25
57.15
51.00
69.90
57.90
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(a)

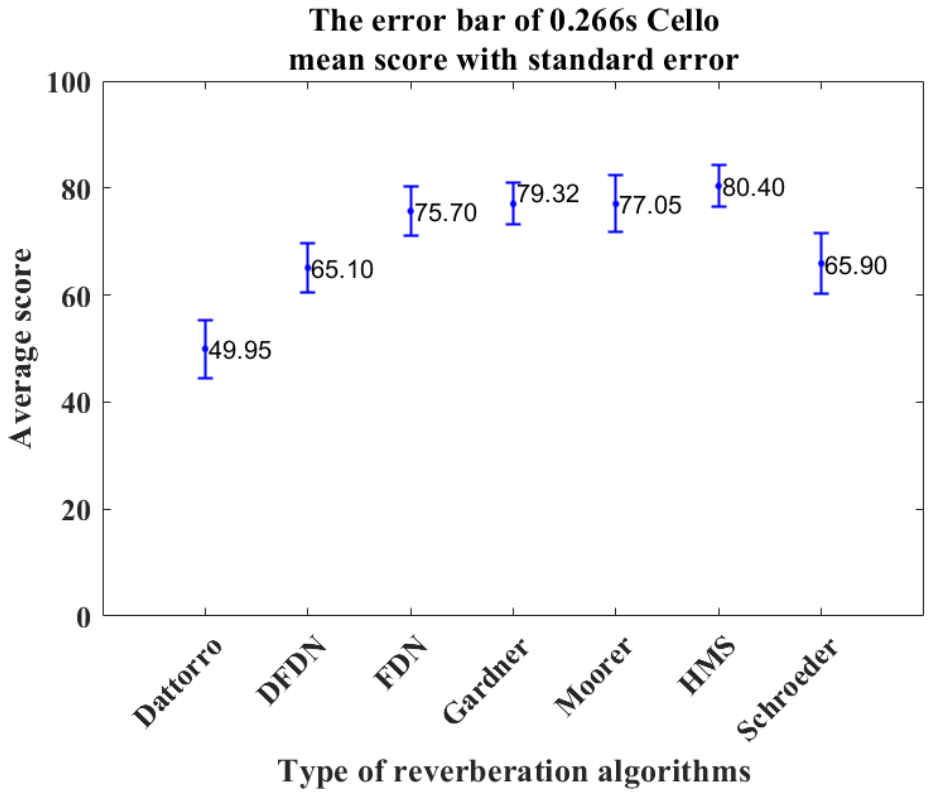

The error bar of 0.95s Male Speech
mean score with standard error
Average score
100
80
60
40
20
0
24.15
19.30
38.55
40.50
56.10
72.15
68.35
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(b)

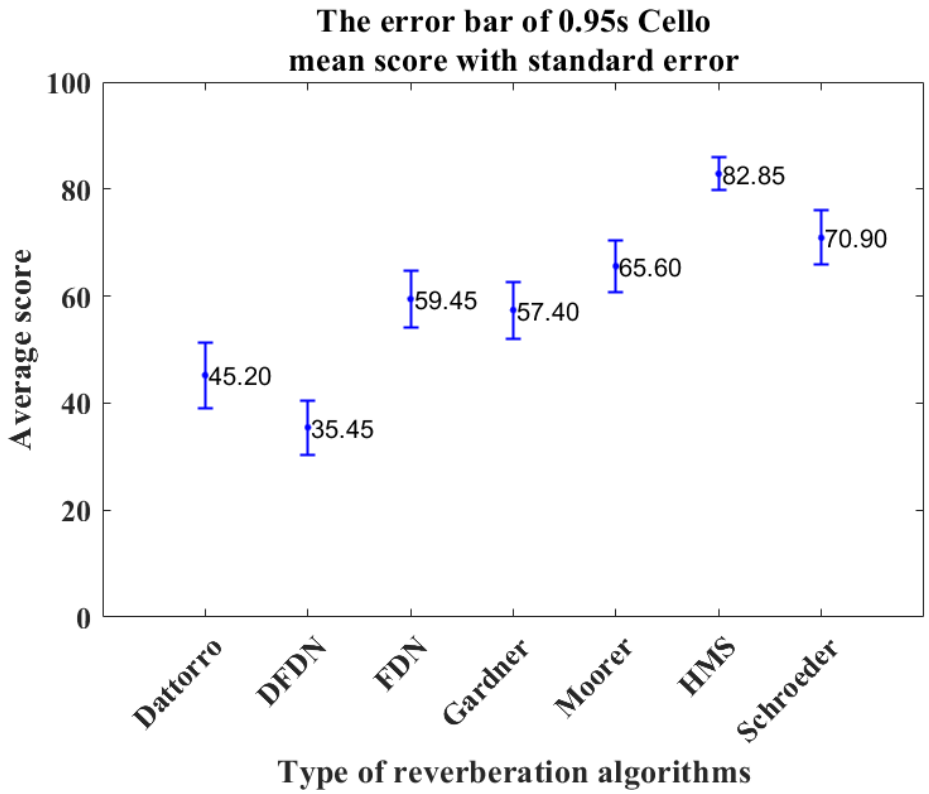


(c)

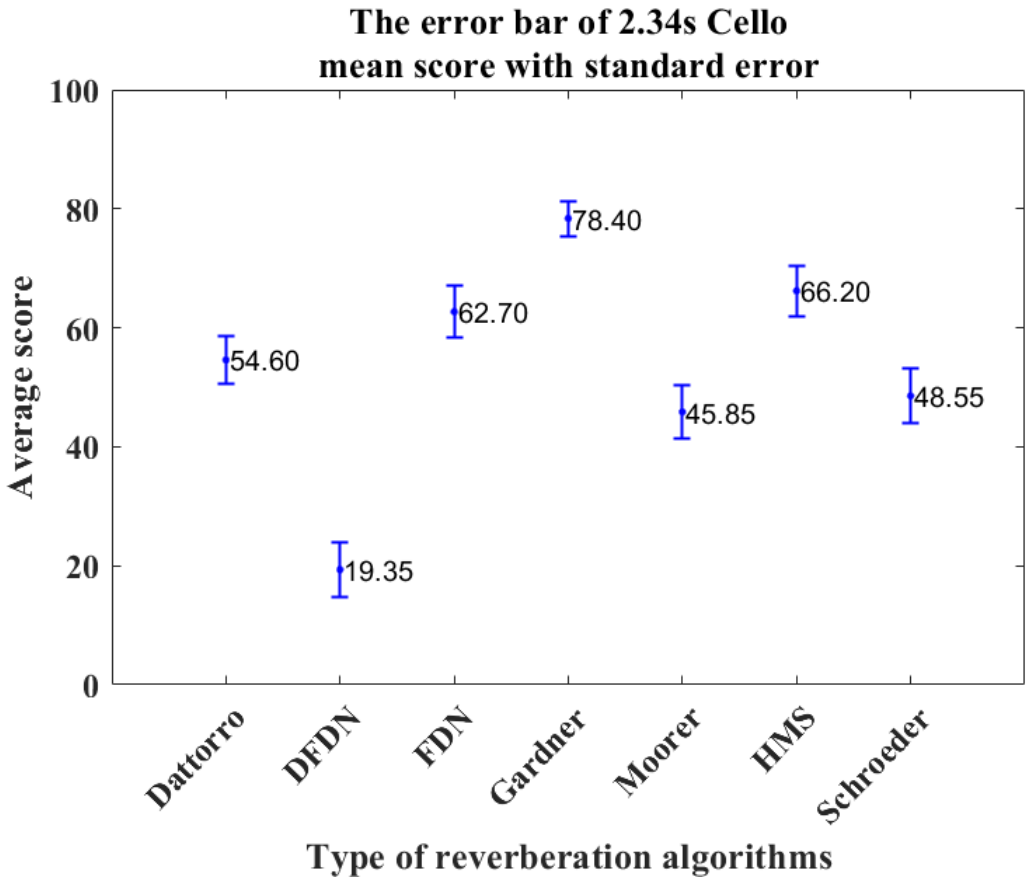
Figure B.14: The error bars of the mean score with standard error of male speech simulated by reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s.


The error bar of 0.266s Cello
mean score with standard error
Average score
100
80
60
40
20
0
49.95
65.10
75.70
79.32
77.05
80.40
65.90
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(a)

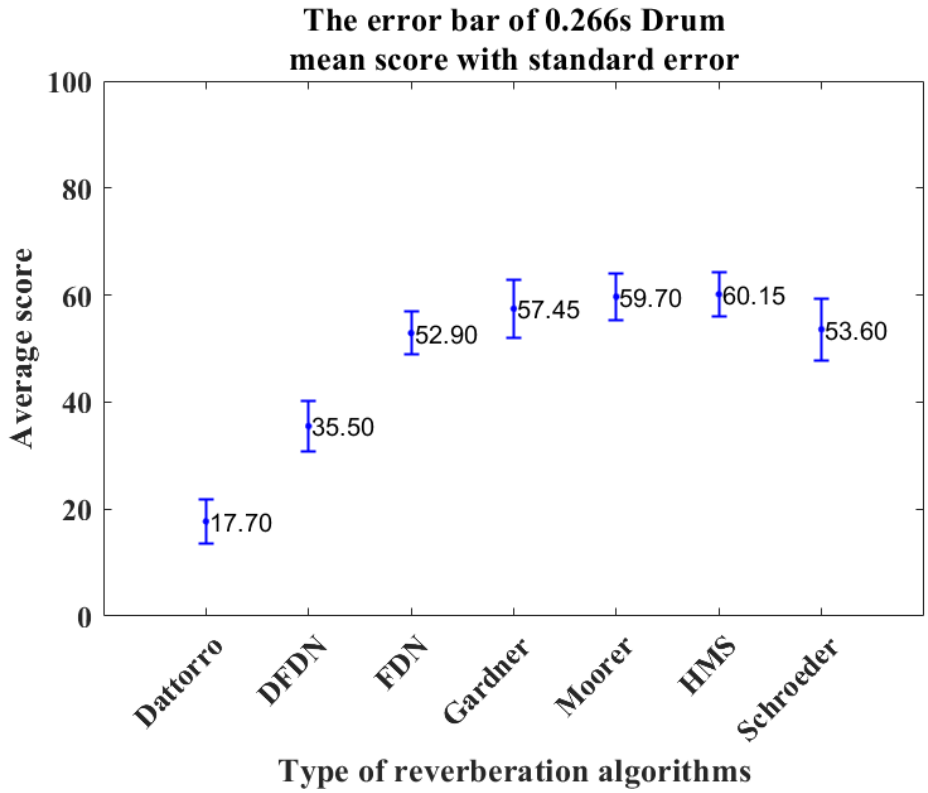

The error bar of 0.95s Cello
mean score with standard error
Average score
100
80
60
40
20
0
45.20
35.45
59.45
57.40
65.60
82.85
70.90
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(b)

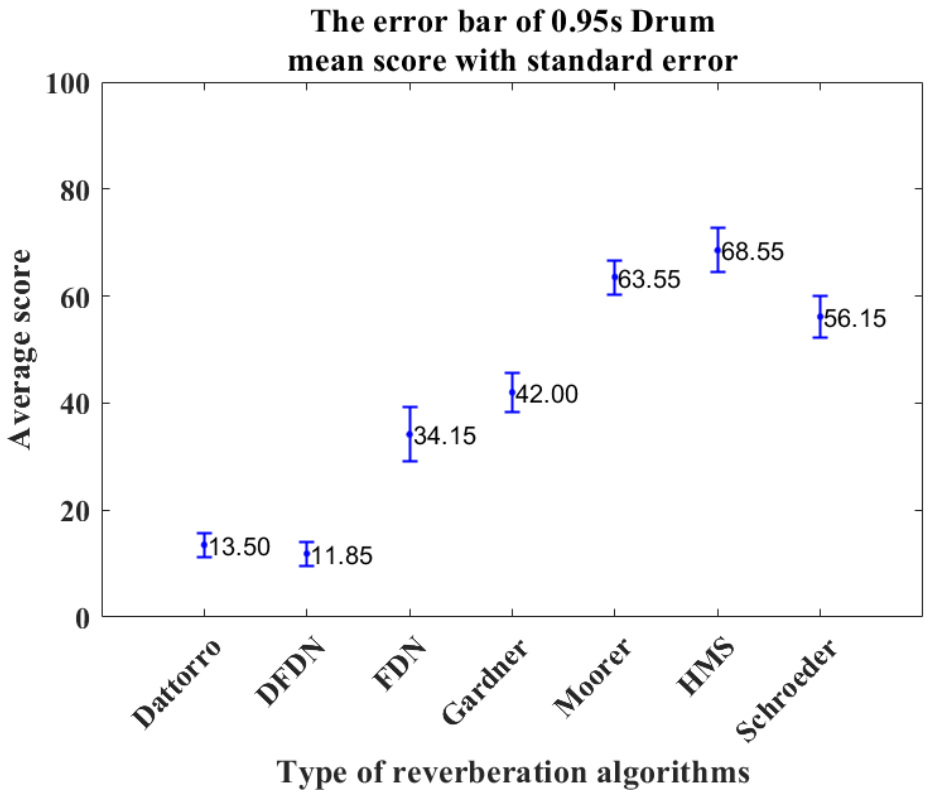


(c)

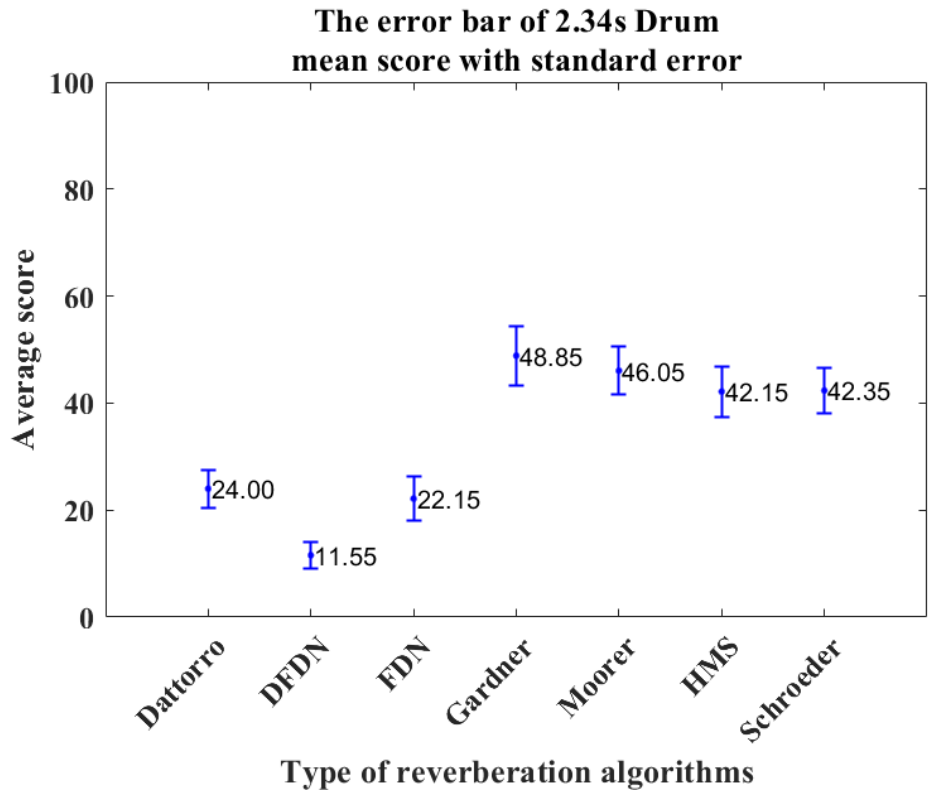
Figure B.15: The error bars of the mean score with standard error of cello piece simulated by reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s.


The error bar of 0.266s Drum
mean score with standard error
Average score
100
80
60
40
20
0
17.70
35.50
52.90
57.45
59.70
60.15
53.60
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(a)


The error bar of 0.95s Drum
mean score with standard error
Average score
100
80
60
40
20
0
13.50
11.85
34.15
42.00
63.55
68.55
56.15
Dattorro
DFDN
FDN
Gardner
Moorer
HMS
Schroeder
Type of reverberation algorithms

(b)



(c)

Figure B.16: The error bars of the mean score with standard error of drum beat simulated by reverberation algorithms under short (0.266 s), medium (0.95 s), and long (2.34 s) reverberation times. (**a**) 0.266 s. (**b**) 0.95 s. (**c**) 2.34 s.

C

Index of Accompanying Materials

The accompanying materials folder is laid out as follows:

Listening Test Documents:

This folder contains all documents for the listening tests conducted in this study.

- **Consent Form Parameters:** Consent form for the listening test presented in Chapter 3.
- **Consent Form Algorithms:** Consent form for the listening test presented in Chapter 4.
- **Information Sheet Parameters:** Information sheet for the listening test presented in Chapter 3
- **Information Sheet Algorithms:** Information sheet for the listening test presented in Chapter 4.
- **Protocol Parameters:** Protocol for the listening test presented in Chapter 3.
- **Protocol Algorithms:** Protocol for the listening test presented in Chapter 4.
- **Ethics Application Parameters:** Ethics application for the listening test

presented in Chapter 3 with approval code Mi111120.

- **Ethics Application Algorithms:** Ethics application for the listening test presented in Chapter 4 with approval code Mi070821.

Listening Test Stimuli:

This folder contains all stimuli for the listening tests conducted in this study.

- **Parameter Evaluation:** Folder containing the stimuli for the staircase listening test in Chapter 3.

- **Algorithm Evaluation:** Folder containing the stimuli for the MUSHRA listening test in Chapter 4. Twelve sub-folders are included, corresponding three reverberation times (0.266s, 0.95s and 2.34s) and four audio samples (Female Singing, Male Speech, Cello and Drum).

MATLAB Code:

This folder contains main code for the implementation of the dynamic reverberation algorithm proposed in this thesis.

- **Parameter Evaluation:** Folder containing the MATLAB script of the staircase listening test application and auxiliary audio samples for Chapter 3.

- **Algorithm Evaluation:** Folder containing the MATLAB scripts of reverberation algorithms for Chapter 4. Eight sub-folders are included, corresponding the implementation of seven reverberation algorithms and convolution of measured impulse responses.

- **HMSG Plugin:** Folder containing the MATLAB scripts of the HMSG reverb plug-in for Chapter 5. 'HMSGreverbPlugin.m' is the main function.

After running the main function, it need to input command 'audioTest-Bench(HMSGreverbPlugin)' in the command space of MATLAB to debug, visualize, and configure the audio plug-in.

References

[1] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[2] H. Hacihabiboglu, E. De Sena, Z. Cvetkovic, J. Johnston, and J. O. Smith III, "Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics," *IEEE Signal Processing Magazine*, vol. 34, no. 3, pp. 36–54, 2017.

[3] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, "Sonic interactions in virtual reality: State of the art, current challenges, and future directions," *IEEE computer graphics and applications*, vol. 38, no. 2, pp. 31–43, 2018.

[4] I. Engel and L. Picinali, "Reverberation and its binaural reproduction: The trade-off between computational efficiency and perceived quality," 2022.

[5] "Reverberation and acoustics." `https://www.askmattrab.com/notes/622-reverberation-and-acoustics`. [Accessed November 20, 2023].

[6] H. Hacıhabiboğlu and F. Murtagh, "Perceptual simplification for model-based binaural room auralisation," *Applied Acoustics*, vol. 69, no. 8, pp. 715–727, 2008.

[7] D. Dziwis, S. Zimmermann, T. Lübeck, J. M. Arend, D. Bau, and C. Pörschmann, "Machine learning-based room classification for selecting binaural room impulse responses in augmented reality applications," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–8, IEEE, 2021.

[8] S. Liu and D. Manocha, *Sound synthesis, propagation, and rendering*. Morgan & Claypool Publishers, 2022.

[9] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, 2015.

[10] A. Southern, J. Wells, and D. Murphy, "Rendering walk-through auralisations using wave-based acoustical models," in *2009 17th European Signal Processing Conference*, pp. 715–719, IEEE, 2009.

[11] S. J. Schlecht and E. A. Habets, "Accurate reverberation time control in feedback delay networks," *Proc. Digital Audio Effects (DAFx-17), Edinburgh, UK*, pp. 337–344, 2017.

[12] J. M. Shen, *Real-Time Audio Reverberation for Virtual Room Acoustics*. PhD thesis, University of Maryland, College Park, 2020.

[13] N. Agus, H. Anderson, J.-M. Chen, S. Lui, and D. Herremans, "Minimally simple binaural room modeling using a single feedback delay network," *Journal of the Audio Engineering Society*, vol. 66, no. 10, pp. 791–807, 2018.

[14] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[15] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Audio Engineering Society Convention 90*, Audio Engineering Society, 1991.

[16] S. J. Schlecht and E. A. Habets, "On lossless feedback delay networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1554–1564, 2016.

[17] S. J. Schlecht and E. A. Habets, "Feedback delay networks: Echo density and mixing time," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 374–383, 2016.

[18] S. J. Schlecht and E. A. Habets, "Sign-agnostic matrix design for spatial artificial reverberation with feedback delay networks," in *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*, Audio Engineering Society, 2018.

[19] B. Alary, A. Politis, S. Schlecht, and V. Välimäki, "Directional feedback delay network," *Journal of the Audio Engineering Society*, vol. 67, no. 10, pp. 752–762, 2019.

[20] O. Das and J. S. Abel, "Grouped feedback delay networks for modeling of coupled spaces," *Journal of the Audio Engineering Society*, vol. 69, no. 7/8, pp. 486–496, 2021.

[21] J.-M. Jot and K. S. Lee, "Augmented reality headphone environment rendering," in *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2016.

[22] F. Conway and N. Paterson, "Spatial audio and reverberation in an augmented reality game sound design," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, Audio Engineering Society, 2010.

[23] E. T. Chourdakis and J. D. Reiss, "A machine-learning approach to application of intelligent artificial reverberation," *Journal of the Audio Engineering Society*, 2017.

[24] A. Harma, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, and H. Nironen, "Techniques and applications of wearable augmented reality audio," in *Audio Engineering Society Convention 114*, Audio Engineering Society, 2003.

[25] C. Yeoward, R. Shukla, R. Stewart, M. Sandler, and J. D. Reiss, "Real-time binaural room modelling for augmented reality applications," *Journal of the Audio Engineering Society*, vol. 69, no. 11, pp. 818–833, 2021.

[26] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, "Scattering delay network: An interactive reverberator for computer games," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, Audio Engineering Society, 2011.

[27] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1478–1492, 2015.

[28] S. Djordjevic, H. Hacihabiboglu, Z. Cvetkovic, and E. De Sena, "Evaluation of the perceived naturalness of artificial reverberation algorithms," in *Audio Engineering Society Convention 148*, Audio Engineering Society, 2020.

[29] "Longitudinal wave." `https://www.123rf.com/photo_47669971_sound-waves-propagation-design.html`. [Accessed February 3, 2023].

[30] P. Haughton, "Sound waves," in *Acoustics for Audiologists*, pp. 43–88, Brill, 2002.

[31] R. Sauerheber, "Light and sound speed mechanisms, relative velocities, simultaneity, and special relativity," *Relative Velocities, Simultaneity, and Special Relativity (April 1, 2024)*, 2024.

[32] C. J. Linder, "University physics students' conceptualizations of factors affecting the speed of sound propagation," *International Journal of Science Education*, vol. 15, no. 6, pp. 655–662, 1993.

[33] B. ERGHELEGIU, M. A. SANDU, and D. IORDAN, "The influence of temperature on sound waves.," *Scientific Papers. Series E. Land Reclamation, Earth Observation & Surveying, Environmental Engineering*, vol. 13, 2024.

[34] A. Goodwin and J. Trusler, "6 speed of sound," *Exp. Thermodyn*, vol. 6, pp. 237–323, 2003.

[35] J. Newman and J. Newman, "Sound," *Physics of the Life Sciences*, pp. 1–28, 2008.

[36] K. Siedenburg, C. Saitis, and S. McAdams, "The present, past, and future of timbre research," *TIMBRE: acoustics, perception, and cognition*, pp. 1–19, 2019.

[37] S. Uppenkamp and M. Röhl, "Human auditory neuroimaging of intensity and loudness," *Hearing Research*, vol. 307, pp. 65–73, 2014.

[38] R. Burkard, "Sound pressure level measurement and spectral analysis of brief acoustic transients," *Electroencephalography and clinical Neurophysiology*, vol. 57, no. 1, pp. 83–91, 1984.

[39] B. Berglund, T. Lindvall, *et al.*, "Community noise," 1995.

[40] A. D. B. Honours, "Acoustic trauma: Bioeffects of sound," *Applied Biophysics, Aether Research Laboratory*, 1994.

[41] "What is sound?." `https://svantek.com/academy/what-is-sound/`. [Accessed November 28, 2023].

[42] S. McAdams, "Musical timbre perception," *The psychology of music*, pp. 35–67, 2013.

[43] C. Sujatha, "Fundamentals of acoustics," in *Vibration, Acoustics and Strain Measurement: Theory and Experiments*, pp. 161–217, Springer, 2023.

[44] "Sound fields: Free versus diffuse field, near versus far field." `https://community.sw.siemens.com/s/article/sound-fields-free-versus-diffuse-field-near-versus-far-field`. [Accessed December 11, 2023].

[45] E. F. Ray, "Industrial noise series, part iv, modeling sound propagation," *June*, vol. 16, p. 2010, 2010.

[46] M. Harrison, *Vehicle refinement: controlling noise and vibration in road vehicles*. Elsevier, 2004.

[47] B. Basuki and M. Palupi, "Sound energy approach in the use of lrad as a bird deterrent device in sam ratulangi manado international airport, north sulawesi, indonesia," in *IOP Conference Series: Materials Science and Engineering*, vol. 943, p. 012053, IOP Publishing, 2020.

[48] C. H. Hansen, "Fundamentals of acoustics," *Occupational Exposure to Noise: Evaluation, Prevention and Control. World Health Organization*, vol. 1, no. 3, pp. 23–52, 2001.

[49] T. Tanaka and M. Otani, "An isotropic sound field model composed of a finite number of plane waves," *Acoustical Science and Technology*, vol. 44, no. 4, pp. 317–327, 2023.

[50] D. Siano, M. Viscardi, and M. Panza, "Experimental acoustic measurements in far field and near field conditions: characterization of a beauty engine cover," *Recent Advances in Fluid Mechanics and Thermal Engineering*, pp. 50–57, 2014.

[51] I. ISO, "3745, acoustics—determination of sound power levels of noise sources—precision methods for anechoic and semi-anechoic rooms," *International Organization for Standardization, Geneva, Switzerland*, 2003.

[52] R. J. Buelow, "The design considerations of an anechoic chamber," tech. rep., SAE Technical Paper, 1999.

[53] C. Hak and R. Wenmaekers, "The impact of sound control room acoustics on the perceived acoustics of a diffuse field recording," *WSEAS Transactions on Signal processing*, vol. 6, no. 4, pp. 175–185, 2010.

[54] F. Rumsey, *Spatial audio*. Taylor & Francis, 2012.

[55] A. Roginska and P. Geluso, *Immersive sound: the art and science of binaural and multi-channel audio*. Taylor & Francis, 2017.

[56] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[57] T. Potisk and D. Svenšek, "Head-related transfer function," *Ljubljana, Slovenia*, 2015.

[58] L. Sun, X. Zhong, and W. Yost, "Dynamic binaural sound source localization with interaural time difference cues: Artificial listeners," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 2226–2226, 2015.

[59] F. Stevens, *Strategies for Environmental Sound Measurement, Modelling, and Evaluation*. PhD thesis, University of York, 2018.

[60] T. McKenzie, *High Frequency Reproduction in Binaural Ambisonic Rendering*. PhD thesis, University of York, 2019.

[61] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[62] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, 1992.

[63] H. Gamper, *Audio augmented reality in telecommunication*. PhD thesis, Graz University of Technology, 2010.

[64] H. Wallach, "On sound localization," *The Journal of the Acoustical Society of America*, vol. 10, no. 4, pp. 270–274, 1939.

[65] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[66] T. Fischer, M. Caversaccio, and W. Wimmer, "A front-back confusion metric in horizontal sound localization: The fbc score," in *ACM Symposium on Applied Perception 2020*, pp. 1–5, 2020.

[67] T. R. Letowski and S. T. Letowski, "Auditory spatial perception: Auditory localization," tech. rep., ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING, 2012.

[68] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *The Journal of the Acoustical Society of America*, vol. 42, no. 2, pp. 480–488, 1967.

[69] D. Howard and J. Angus, *Acoustics and psychoacoustics*. Routledge, 2013.

[70] C. T. Choi and Y.-H. Lee, "A review of stimulating strategies for cochlear implants," *Cochlear implant research updates*, vol. 16, pp. 77–90, 2012.

[71] A. Sheikh, K. Shabbir, A. Imtiaz, *et al.*, "Structure and physiology of human ear involved in hearing," in *Auditory System-Function and Disorders*, IntechOpen, 2022.

[72] P. W. Alberti, "The anatomy and physiology of the ear and hearing," *Occupational exposure to noise: Evaluation, prevention, and control*, pp. 53–62, 2001.

[73] C. Freigang, N. Richter, R. Rübsamen, and A. A. Ludwig, "Age-related changes in sound localisation ability," *Cell and tissue research*, vol. 361, pp. 371–386, 2015.

[74] X. Zhong, B. Xie, and H. Glotin, "Head-related transfer functions and virtual auditory display," *Soundscape Semiotics-Localization and Categorization*, vol. 1, 2014.

[75] V. Lemaire, F. Clerot, S. Busson, R. Nicol, and V. Choqueuse, "Individualized HRTFs from few measurements: a statistical learning approach," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 4, pp. 2041–2046, IEEE, 2005.

[76] C. Jenny, C. Reuter, *et al.*, "Usability of individualized head-related transfer functions in virtual reality: Empirical study with perceptual attributes in sagittal plane sound localization," *JMIR serious games*, vol. 8, no. 3, p. e17576, 2020.

[77] M. La Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 4, pp. 322–336, 2000.

[78] A. McKeag and D. S. McGrath, "Sound field format to binaural decoder with head tracking," in *Audio engineering society convention 6r*, Audio Engineering Society, 1996.

[79] M. Schutte, *Aspects of room acoustics, vision and motion in the human auditory perception of space*. PhD thesis, lmu, 2021.

[80] W. G. Gardner, "Reverberation algorithms," in *Applications of digital signal processing to audio and acoustics*, pp. 85–131, Springer, 1998.

[81] W. C. Sabine and M. D. Egan, "Collected papers on acoustics," 1994.

[82] J. Bradley, "A new look at acoustical criteria for classrooms," in *Proceedings of Inter-noise*, pp. 1–9, 2009.

[83] P. Stitt, E. Hendrickx, J. Messonnier, and B. F. Katz, "The role of head tracking in binaural rendering," in *29th Tonmeistertagung-VDT International Convention*, pp. 1–5, 2016.

[84] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Efficient conversion of xy surround sound content to binaural head-tracked form for hrtf-enabled playback," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 1, pp. I–21, IEEE, 2007.

[85] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich, "3d binaural sound reproduction using a virtual ambisonic approach," in *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS'03. 2003*, pp. 174–178, IEEE, 2003.

[86] M.-V. Laitinen, T. Pihlajamäki, S. Lösler, and V. Pulkki, "Influence of resolution of head tracking in synthesis of binaural audio," in *Audio Engineering Society Convention 132*, Audio Engineering Society, 2012.

[87] E. Shotter, *Absolute auditory object localization.* PhD thesis, Loughborough University, 1997.

[88] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel, "More than 50 years of artificial reverberation," in *Audio engineering society conference: 60th international conference: dreams (dereverberation and reverberation of audio, music, and speech)*, Audio Engineering Society, 2016.

[89] H. Mi, G. Kearney, and H. Daffern, "Impact thresholds of parameters of binaural room impulse responses (brirs) on perceptual reverberation," *Applied Sciences*, vol. 12, no. 6, p. 2823, 2022.

[90] R. Stewart and M. Sandler, "Statistical measures of early reflections of room impulse responses," in *Proc. of the 10th int. conference on digital audio effects (DAFx-07), Bordeaux, France*, pp. 59–62, 2007.

[91] H. Kuttruff, "A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries," *The Journal of the Acoustical Society of America*, vol. 98, no. 1, pp. 288–293, 1995.

[92] G. Defrance and J.-D. Polack, "Measuring the mixing time in auditoria," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3499, 2008.

[93] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," in *Audio Engineering Society Convention 128*, Audio Engineering Society, 2010.

[94] A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *Journal of the Audio Engineering Society*, vol. 60, no. 1/2, pp. 54–62, 2012.

[95] N. Prodi, M. Pellegatti, and C. Visentin, "Effects of type of early reflection, clarity of speech, reverberation and diffuse noise on the spatial perception of a speech source and its intelligibility," *The Journal of the Acoustical Society of America*, vol. 151, no. 5, pp. 3522–3534, 2022.

[96] K. H. Kuttruff, "Auralization of impulse responses modeled on the basis of ray-tracing results," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 876–880, 1993.

[97] H. Gölzer and M. Kleinschmidt, "Importance of early and late reflections for automatic speech recognition in reverberant environments," *Elektronische Sprachsignalverarbeitung (ESSV)*, 2003.

[98] A. Warzybok, J. Rennies, T. Brand, S. Doclo, and B. Kollmeier, "Effects of spatial and temporal integration of a single early reflection on speech intelligibility," *The Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 269–282, 2013.

[99] D. Griesinger, "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," *Acta Acustica united with Acustica*, vol. 83, no. 4, pp. 721–731, 1997.

[100] T. Lokki and J. Pätynen, "Auditory spatial impression in concert halls," *The Technology of Binaural Understanding*, pp. 173–202, 2020.

[101] J. Paulus, C. Uhle, and J. Herre, "Perceived level of late reverberation in speech and music," in *Audio Engineering Society Convention 130*, Audio Engineering Society, 2011.

[102] D. T. Murphy, "Digital waveguide mesh topologies in room acoustics modelling.," 2005.

[103] L. L. Beranek, "Concert hall acoustics—1992," *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 1–39, 1992.

[104] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of reverberation in small rooms: a literature study," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, 2014.

[105] F. Rumsey, "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm," *Journal of the Audio Engineering Society*, vol. 50, no. 9, pp. 651–666, 2002.

[106] L. L. Beranek, *Concert halls and opera houses: music, acoustics, and architecture*, vol. 2. Springer, 2004.

[107] J. R. Hyde, "Discussion of the relation between initial time delay gap (itdg) and acoustical intimacy: Leo beranek's final thoughts on the subject, documented," in *Acoustics*, vol. 1, pp. 561–569, Multidisciplinary Digital Publishing Institute, 2019.

[108] L. Beranek, *Concert halls and opera houses: music, acoustics, and architecture*. Springer Science & Business Media, 2012.

[109] S. Werner and S. Füg, "Controlled auditory distance perception using binaural headphone reproduction–evaluation via listening tests," in *Proceedings of the 27th Tonmeistertagung, VDT International Convention, Cologne, Germany*, pp. 22–25, 2012.

[110] J. Hyon and D. Jeong, "Variable acoustics in performance venues-a review," *The Journal of the Acoustical Society of Korea*, vol. 40, no. 6, pp. 626–648, 2021.

[111] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 6, pp. 1187–1188, 1965.

[112] J. Bitzer, D. Extra, S. Fischer, and U. Simmer, "Artificial reverberation: Comparing algorithms by using monaural analysis tools," in *Audio Engineering Society Convention 121*, Audio Engineering Society, 2006.

[113] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 2, pp. 221–224, IEEE Computer Society, 1992.

[114] P. S. L. R. F. Impulse and M. Steimel, "Implementation of a hybrid reverb algorithm," 2019.

[115] S. R. Bistafa and J. S. Bradley, "Predicting reverberation times in a simulated classroom," *The Journal of the Acoustical Society of America*, vol. 108, no. 4, pp. 1721–1731, 2000.

[116] L. L. Beranek, "Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1399–1410, 2006.

[117] W. B. Ford, *Studies on divergent series and summability and the asymptotic developments of functions defined by Maclaurin series*, vol. 143. American Mathematical Soc., 1960.

[118] A. Nowoświat and M. Olechowska, "Investigation studies on the application of reverberation time," *Archives of acoustics*, vol. 41, no. 1, pp. 15–26, 2015.

[119] M. Hodgson, "Experimental evaluation of the accuracy of the sabine and eyring theories in the case of non-low surface absorption," *The Journal of the Acoustical Society of America*, vol. 94, no. 2, pp. 835–840, 1993.

[120] I. ISO, "3382: Acoustics-measurement of room acoustic parameters-part 2: Reverberation time in ordinary rooms. 2008," *Geneva: ISO*.

[121] T. J. Peltonen *et al.*, "A multichannel measurement system for room acoustics analysis," Master's thesis, 2000.

[122] N. Toma, M. D. Topa, V. Popescu, and E. Szopos, "Comparative performance analysis of artificial reverberation algorithms," in *2006 IEEE International Conference on Automation, Quality and Testing, Robotics*, vol. 1, pp. 138–142, IEEE, 2006.

[123] P. Fausti and A. Farina, "Acoustic measurements in opera houses: comparison between different techniques and equipment," *Journal of Sound and Vibration*, vol. 232, no. 1, pp. 213–229, 2000.

[124] C. Hak and R. Wenmaekers, "Room in room acoustics: using convolutions to find the impact of a listening room on recording acoustics," in *2013 International Symposium on Room Acoustics (ISRA 2013)*, pp. 1–9, 2013.

[125] H. Chen, T. D. Abhayapala, P. N. Samarasinghe, and W. Zhang, "Direct-to-reverberant energy ratio estimation using a first-order microphone," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 226–237, 2016.

[126] G. v. Békésy, "Über die entstehung der entfernungsempfindung beim hören.," *Akustische Zeitschrift*, 1938.

[127] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, 2008.

[128] J. Pätynen, B. F. Katz, and T. Lokki, "Investigations on the balloon as an impulse source," *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. EL27–EL33, 2011.

[129] J. A. Kemp, *Theoretical and experimental study of wave propagation in brass musical instruments*. PhD thesis, University of Edinburgh, 2002.

[130] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio engineering society convention 108*, Audio Engineering Society, 2000.

[131] M. Holters, T. Corbach, and U. Zölzer, "Impulse response measurement techniques and their applicability in the real world," in *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09)*, pp. 108–112, 2009.

[132] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *Journal of the Audio Engineering Society*, vol. 37, no. 6, pp. 419–444, 1989.

[133] F. Policardi, "MLS and Sine-Sweep measurements," *Università di Bologna, Italia ELEKTROTEHNIŠKI VESTNIK*, vol. 78, no. 3, pp. 91–95, 2011.

[134] M. Guski and M. Vorländer, "Impulsive noise detection in sweep measurements," *Acta Acustica united with Acustica*, vol. 101, no. 4, pp. 723–730, 2015.

[135] S. Siltanen, T. Lokki, and L. Savioja, "Rays or waves? understanding the strengths and weaknesses of computational room acoustics modeling techniques," in *Proceedings of the International Symposium on Room Acoustics, ISRA*, pp. 29–31, 2010.

[136] S. B. Shelley, *Diffuse boundary modelling in the digital waveguide mesh*. PhD thesis, University of York, 2007.

[137] M. Vorländer, "Computer simulations in room acoustics: Concepts and uncertainties," *The Journal of the Acoustical Society of America*, vol. 133, no. 3, pp. 1203–1213, 2013.

[138] S. Hassani and S. Hassani, "Dirac delta function," *Mathematical Methods: For Students of Physics and Related Fields*, pp. 139–170, 2009.

[139] D. Baowan, B. J. Cox, T. A. Hilder, J. M. Hill, and N. Thamwattana, *Modelling and mechanics of carbon-based nanostructured materials*. William Andrew, 2017.

[140] S. C. Dutta Roy, "Some fundamental issues related to the impulse function," *IETE Journal of Education*, vol. 57, no. 1, pp. 2–8, 2016.

[141] J. R. Beltrán and F. A. Beltrán, "Matlab implementation of reverberation algorithms," *Journal of New Music Research*, vol. 31, no. 2, pp. 153–161, 2002.

[142] K. Prawda, S. Willemsen, S. Serafin, and V. Välimäki, "Flexible real-time reverberation synthesis with accurate parameter control," in *23rd International Conference on Digital Audio Effects*, pp. 16–23, 2020.

[143] B. Wang and J. Saniie, "Learning fir filter coefficients from data for speech-music separation," in *2020 IEEE International Conference on Electro Information Technology (EIT)*, pp. 245–248, IEEE, 2020.

[144] R. Oshana, *DSP for Embedded and Real-time Systems*. Elsevier, 2012.

[145] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on advances in signal processing*, vol. 2009, pp. 1–10, 2009.

[146] P. Stade, B. Bernschütz, and M. Rühl, "A spatial audio impulse response compilation captured at the wdr broadcast studios," in *27th Tonmeistertagung-VDT International Convention*, pp. 551–567, 2012.

[147] G. Davis, S. Andre, I. Munoz, and N. Peters, "Perceptual evaluation of personalized BRIRs and headphone compensation," in *Audio Engineering Society Conference: 2019 AES INTERNATIONAL CONFERENCE ON HEADPHONE TECHNOLOGY*, Audio Engineering Society, 2019.

[148] G. M. Naylor, "Odeon—another hybrid room acoustical model," *Applied Acoustics*, vol. 38, no. 2-4, pp. 131–143, 1993.

[149] "KU 100." `https://en-de.neumann.com/ku-100`. [Accessed August 4, 2022].

[150] "G.R.A.S. KEMAR." `https://www.grasacoustics.com/products/head-torso-simulators-kemar/product/733-45bb`. [Accessed August 4, 2022].

[151] T. Leclère, M. Lavandier, and F. Perrin, "On the externalization of sound sources with headphones without reference to a real source," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2309–2320, 2019.

[152] H. G. Hassager, F. Gran, and T. Dau, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2992–3000, 2016.

[153] Z. Jiang, J. Sang, C. Zheng, and X. Li, "The effect of pinna filtering in binaural transfer functions on externalization in a reverberant environment," *Applied Acoustics*, vol. 164, p. 107257, 2020.

[154] J. Catic, S. Santurette, and T. Dau, "The role of reverberation-related binaural cues in the externalization of speech," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 1154–1167, 2015.

[155] J. Catic, S. Santurette, J. M. Buchholz, F. Gran, and T. Dau, "The effect of interaural-level-difference fluctuations on the externalization of sound," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1232–1241, 2013.

[156] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 823–828, 2010.

[157] Y. Pang, M. Sun, X. Jiang, and X. Li, "Convolution in convolution for network in network," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 5, pp. 1587–1597, 2017.

[158] S. W. Smith *et al.*, "The scientist and engineer's guide to digital signal processing," 1997.

[159] S. Kantorik, "Convolution reverb & impulse responses," 2014.

[160] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[161] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Nature, 2020.

[162] G. Kearney, *Auditory scene synthesis using virtual acoustic recording and reproduction*. PhD thesis, Trinity College Dublin, 2010.

[163] M. Blau, A. Budnik, and S. van de Par, "Assessment of perceptual attributes of classroom acoustics: Real versus simulated room," *Proceedings of the Institute of Acoustics*, vol. 40, no. Pt 3, pp. 556–564, 2018.

[164] M. Blau, A. Budnik, M. Fallahi, H. Steffens, S. D. Ewert, and S. Van de Par, "Toward realistic binaural auralizations–perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario," *Acta Acustica*, vol. 5, p. 8, 2021.

[165] T. Lokki and V. Pulkki, "Evaluation of geometry-based parametric auralization," in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, Audio Engineering Society, 2002.

[166] J.-M. Jot, "Efficient models for reverberation and distance rendering in computer music and virtual audio reality," in *ICMC*, Citeseer, 1997.

[167] L. P. Huelsman, "Analog electrical filters," 2003.

[168] A. A. Kumar, *Digital signal processing*. PHI Learning Pvt. Ltd., 2014.

[169] M. Slaney, "A review of filter design," 1989.

[170] J. Corey, *Audio production and critical listening: Technical ear training*. Taylor & Francis, 2016.

[171] W. G. Gardner, *The virtual acoustic room*. PhD thesis, Massachusetts Institute of Technology, 1992.

[172] L. Maheswari and M. Anand, *Analog electronics*. PHI Learning Pvt. Ltd., 2009.

[173] P. A. Regalia, S. K. Mitra, and P. Vaidyanathan, "The digital all-pass filter: A versatile signal processing building block," *Proceedings of the IEEE*, vol. 76, no. 1, pp. 19–37, 1988.

[174] D. Davis and E. Patronis, *Sound system engineering*. Routledge, 2012.

[175] J. A. Moorer, "About this reverberation business," *Computer music journal*, pp. 13–28, 1979.

[176] H. Mikelson, "Implementing the gardner reverbs in csound," in *The Csound book: perspectives in software synthesis, sound design, signal processing, and programming*, pp. 483–492, 2000.

[177] M. Ermann, *Architectural acoustics illustrated*. John Wiley & Sons, 2015.

[178] M. R. Schroeder, "Natural sounding artificial reverberation," in *Audio Engineering Society Convention 13*, Audio Engineering Society, 1961.

[179] J. Dattorro, "Effect design, part 1: Reverberator and other filters," *Journal of the Audio Engineering Society*, vol. 45, no. 9, pp. 660–684, 1997.

[180] U. Zölzer, *Digital audio signal processing*. John Wiley & Sons, 2022.

[181] J. P. C. P. C. Neves, *Musically-Informed Adaptive Audio Reverberation*. PhD thesis, Universidade do Porto (Portugal), 2017.

[182] D. Schröder, *Physically based real-time auralization of interactive virtual environments*, vol. 11. Logos Verlag Berlin GmbH, 2011.

[183] H. Järveläinen and M. Karjalainen, "Reverberation modeling using velvet noise," in *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*, Audio Engineering Society, 2007.

[184] N. Vandenberg, "Desc9115: Digital audio systems-written assignment ii," 2011.

[185] N. Toma, M. Topa, and E. Szopos, "Aspects of reverberation algorithms," in *International Symposium on Signals, Circuits and Systems, 2005. ISSCS 2005.*, vol. 2, pp. 577–580, IEEE, 2005.

[186] P. Kubinec, O. Ondrácek, M. Hagara, A. Fibich, and T. Bagala, "Reverberator's late reflections parameters calculation," in *2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–4, IEEE, 2018.

[187] S. R. Bistafa and J. S. Bradley, "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics," *The Journal of the Acoustical Society of America*, vol. 107, no. 2, pp. 861–875, 2000.

[188] G. Ballou, *Handbook for sound engineers*. Taylor & Francis, 2013.

[189] H. Giesbrecht, W. McFarland, T. Perry, and M. McGuire, "Algorithmic reverberation," 2009.

[190] W. G. Gardner, "3d audio and acoustic environment modeling," *Wave Arts, Inc*, vol. 99, 1999.

[191] A. A. de Lima, F. P. Freeland, P. A. Esquef, L. W. Biscainho, B. C. Bispo, R. A. de Jesus, S. L. Netto, R. W. Schafer, A. Said, B. Lee, *et al.*, "Reverberation assessment in audioband speech signals for telepresence systems.," in *SIGMAP*, pp. 257–262, 2008.

[192] A. Politis and D. Poirier-Quinot, "Jsambisonics: A web audio library for interactive spatial sound processing on the web," in *Interactive Audio Systems Symposium*, 2016.

[193] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[194] R. H. Hardin and N. J. Sloane, "Mclaren's improved snub cube and other new spherical designs in three dimensions," *Discrete & Computational Geometry*, vol. 15, pp. 429–441, 1996.

[195] R. Nicol, "Sound field," in *Immersive Sound*, pp. 276–310, Routledge, 2017.

[196] J. Vilkamo, "Spatial sound reproduction with frequency band processing of b-format audio signals," *Helsinki University of Technology*, 2008.

[197] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE computer graphics and applications*, vol. 21, no. 6, pp. 34–47, 2001.

[198] A. Haj-Bolouri, "The experience of immersive virtual reality: A phenomenology inspired inquiry," *Communications of the Association for Information Systems*, vol. 52, no. 1, pp. 782–814, 2023.

[199] P. Larsson, A. Väljamäe, D. Västfjäll, A. Tajadura-Jiménez, and M. Kleiner, "Auditory-induced presence in mixed reality environments and related technology," *The engineering of mixed reality systems*, pp. 143–163, 2010.

[200] M. Harju *et al.*, "Exploring narrative possibilities of audio augmented reality with six degrees of freedom," Master's thesis, 2021.

[201] M. Gospodarek, *Acoustic and Perceptual Factors Affecting Plausibility in Sound Design for Audio Augmented Reality Experiences*. PhD thesis, New York University, 2024.

[202] N. Corporal, *Big Picture, Big Sound: Investigating Headphone and Loudspeaker Frameworks to Enhance Users' Sense of Presence and Emotional State in Virtual Reality*. PhD thesis, Queensland University of Technology, 2024.

[203] V. Hohmann, R. Paluch, M. Krueger, M. Meis, and G. Grimm, "The virtual reality lab: Realization and application of virtual sound environments," *Ear and Hearing*, vol. 41, pp. 31S–38S, 2020.

[204] J. d. Rincón, *Music technology*. Cambridge University Press, 2012.

[205] G. Tanev and A. Božinovski, "Virtual studio technology inside music production," in *International Conference on ICT Innovations*, pp. 231–241, Springer, 2013.

[206] N. Peters, J. Choi, and H. Lei, "Matching artificial reverb settings to unknown room recordings: A recommendation system for reverb plugins," in *Audio Engineering Society Convention 133*, Audio Engineering Society, 2012.

[207] B. S. Storli, "The virtual space: The impact of time-based signal processing," Master's thesis, 2022.

[208] L. McCormack, A. Politis, T. McKenzie, C. Hold, and V. Pulkki, "Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple Ambisonic receivers," *Journal of the Audio Engineering Society*, vol. 70, no. 5, pp. 355–372, 2022.

[209] J. van der Hooft, M. T. Vega, T. Wauters, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "From capturing to rendering: Volumetric media delivery with six degrees of freedom," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 49–55, 2020.

[210] "What are the "6 Degrees of Freedom"?." `https://industrial-ia.com/what-are-the-6-degrees-of-freedom-6dof-explained/#:~:text=The%206%20degrees%20of%20freedom,pitch%2C%20roll%2C%20or%20yaw.` [Accessed April 29, 2023].

[211] B. Alary and V. Välimäki, "A method for capturing and reproducing directional reverberation in six degrees of freedom," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–8, IEEE, 2021.

[212] T. McKenzie, L. McCormack, and C. Hold, "Dataset of spatial room impulse responses in a variable acoustics room for six degrees-of-freedom rendering and analysis," *arXiv preprint arXiv:2111.11882*, 2021.

[213] J. M. Arend, S. V. A. Garí, C. Schissler, F. Klein, and P. W. Robinson, "Six-degrees-of-freedom parametric spatial audio based on one monaural room impulse response," *Journal of the Audio Engineering Society*, vol. 69, no. 7/8, pp. 557–575, 2021.

[214] S. Bech and N. Zacharov, *Perceptual audio evaluation-Theory, method and application*. John Wiley & Sons, 2007.

[215] D. de la Prida, A. Pedrero, L. A. Azpicueta-Ruiz, and M. Á. Navacerrada, "Listening tests in room acoustics: Comparison of overall difference protocols regarding operational power," *Applied Acoustics*, vol. 182, p. 108186, 2021.

[216] V. Hongisto, D. Oliva, and L. Rekola, "Subjective and objective rating of the sound insulation of residential building façades against road traffic noise," *The Journal of the Acoustical Society of America*, vol. 144, no. 2, pp. 1100–1112, 2018.

[217] D. de la Prida, A. Pedrero, M. Á. Navacerrada, and A. Diaz-Chyla, "Methodology for the subjective evaluation of airborne sound insulation through 2-AC and Thurstonian models," *Applied Acoustics*, vol. 157, p. 107011, 2020.

[218] V. Chmelík, M. Rychtáriková, H. Müllner, K. Jambrošić, L. Zelem, J. Benklewski, and C. Glorieux, "Methodology for development of airborne sound insulation descriptor valid for light-weight and masonry walls," *Applied Acoustics*, vol. 160, p. 107144, 2020.

[219] C. Guastavino, B. F. Katz, J.-D. Polack, D. J. Levitin, and D. Dubois, "Ecological validity of soundscape reproduction," *Acta Acustica united with Acustica*, vol. 91, no. 2, pp. 333–341, 2005.

[220] A. S. Sudarsono, Y. W. Lam, and W. J. Davies, "The effect of sound level on perception of reproduced soundscapes," *Applied acoustics*, vol. 110, pp. 53–60, 2016.

[221] D. de la Prida, A. Pedrero, M. Á. Navacerrada, and C. Díaz, "Relationship between the geometric profile of the city and the subjective perception of urban soundscapes," *Applied Acoustics*, vol. 149, pp. 74–84, 2019.

[222] C. Xu, T. Oberman, F. Aletta, H. Tong, and J. Kang, "Ecological validity of immersive virtual reality (ivr) techniques for the perception of urban sound environments," in *Acoustics*, vol. 3, pp. 11–24, MDPI, 2020.

[223] A. Gonzalez, M. Ferrer, M. De Diego, G. Pinero, and J. Garcia-Bonito, "Sound quality of low-frequency and car engine noises after active noise control," *Journal of Sound and Vibration*, vol. 265, no. 3, pp. 663–679, 2003.

[224] A. J. Torija, S. Roberts, R. Woodward, I. H. Flindell, A. R. McKenzie, and R. H. Self, "On the assessment of subjective response to tonal content of contemporary aircraft noise," *Applied Acoustics*, vol. 146, pp. 190–203, 2019.

[225] F. Martellotta, "The just noticeable difference of center time and clarity index in large reverberant spaces," *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 654–663, 2010.

[226] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences," *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 551–562, 2016.

[227] B. N. Postma and B. F. Katz, "Perceptive and objective evaluation of calibrated room acoustic simulation auralizations," *The Journal of the Acoustical Society of America*, vol. 140, no. 6, pp. 4326–4337, 2016.

[228] F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2746–2760, 2019.

[229] T. J. Cox, W. J. Davies, and Y. W. Lam, "The sensitivity of listeners to early sound field changes in auditoria," *Acta Acustica united with Acustica*, vol. 79, no. 1, pp. 27–41, 1993.

[230] G. A. Soulodre and J. S. Bradley, "Subjective evaluation of new room acoustic measures," *The Journal of the Acoustical Society of America*, vol. 98, no. 1, pp. 294–301, 1995.

[231] J. S. Bradley, R. Reich, and S. Norcross, "A just noticeable difference in C50 for speech," *Applied Acoustics*, vol. 58, no. 2, pp. 99–108, 1999.

[232] T. Okano, "Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections," *The Journal of the Acoustical Society of America*, vol. 111, no. 1, pp. 217–229, 2002.

[233] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2110–2117, 2002.

[234] I. B. Witew, G. K. Behler, and M. Vorländer, "About just noticeable differences for aspects of spatial impressions in concert halls," *Acoustical science and technology*, vol. 26, no. 2, pp. 185–192, 2005.

[235] M. C. Vigeant and R. D. Celmer, "Effect of experimental design on the results of clarity-index just-noticeable-difference listening tests," in *Proceedings of 20th International Congress on Acoustics, ICA 2010*, pp. 23–27, 2010.

[236] S. Klockgether and S. van de Par, "Just noticeable differences of spatial cues in echoic and anechoic acoustical environments," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. EL352–EL357, 2016.

[237] A. Lindau and S. Weinzierl, "Assessing the plausibility of virtual acoustic environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810, 2012.

[238] R. R. Torres, U. P. Svensson, and M. Kleiner, "Computation of edge diffraction for more accurate room acoustics auralization," *The Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 600–610, 2001.

[239] A. Southern, D. Murphy, T. Lokki, and L. Savioja, "The perceptual effects of dispersion error on room acoustic model auralization," in *Proc. Forum Acusticum, Aalborg, Denmark*, pp. 1553–1558, 2011.

[240] D. Ackermann, C. Böhm, F. Brinkmann, and S. Weinzierl, "The acoustical effect of musicians' movements during musical performances," *Acta Acustica united with Acustica*, vol. 105, no. 2, pp. 356–367, 2019.

[241] H. Steffens, S. van de Par, and S. D. Ewert, *Perceptual relevance of speaker directivity modelling in virtual rooms*. Universitätsbibliothek der RWTH Aachen, 2019.

[242] M. R. Schroeder, D. Gottlob, and K. Siebrasse, "Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters," *The Journal of the Acoustical Society of America*, vol. 56, no. 4, pp. 1195–1201, 1974.

[243] Y. Ando, "Calculation of subjective preference at each seat in a concert hall," *The Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 873–887, 1983.

[244] Y.-J. Choi and F. R. Fricke, "A comparison of subjective assessments of recorded music and computer simulated auralizations in two auditoria," *Acta acustica united with acustica*, vol. 92, no. 4, pp. 604–611, 2006.

[245] M. S. Lawless and M. C. Vigeant, "Effects of test method and participant musical training on preference ratings of stimuli with different reverberation times," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 2258–2272, 2017.

[246] N. Zacharov, *Sensory evaluation of sound*. CRC Press, 2018.

[247] G. V. Civille and J. Seltsam, "Sensory evaluation methods applied to sound quality," *Noise Control Engineering Journal*, vol. 51, no. 4, pp. 262–270, 2003.

[248] H. acoustics, "Conducting listening tests." `https://cdn.head-acoustics.com/fileadmin/data/global/Application-Notes/SVP/Listening-Tests-06.2011.pdf`. [Accessed August 11, 2022].

[249] J. Boley and M. Lester, "Statistical analysis of abx results using signal detection theory," in *Audio Engineering Society Convention 127*, Audio Engineering Society, 2009.

[250] B. C. Moore, *An introduction to the psychology of hearing*. Brill, 2012.

[251] T. N. Cornsweet, "The staircase-method in psychophysics," *The American journal of psychology*, vol. 75, no. 3, pp. 485–491, 1962.

[252] R. M. Rose, D. Y. Teller, and P. Rendleman, "Statistical properties of staircase estimates," *Perception & Psychophysics*, vol. 8, no. 4, pp. 199–204, 1970.

[253] H. Levitt, "Transformed up-down methods in psychoacoustics," *The Journal of the Acoustical society of America*, vol. 49, no. 2B, pp. 467–477, 1971.

[254] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, "webmushra—a comprehensive framework for web-based listening tests," *Journal of Open Research Software*, vol. 6, no. 1, 2018.

[255] B. Series, "Method for the subjective assessment of intermediate quality level of audio systems," *International Telecommunication Union Radiocommunication Assembly*, 2014.

[256] A. LABS, "webMUSHRA." `https://www.audiolabs-erlangen.de/resources/webMUSHRA`. [Accessed August 12, 2022].

[257] N. Schinkel-Bielefeld, "Quantifying sequential context effects in subjective quality evaluation," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 269–274, IEEE, 2014.

[258] M. Rychtáriková and M. Horvat, "Developing a methodology for performing listening tests related to building acoustics," 2013.

[259] J. A. Kelly and L. H. Levy, "The discriminability of concepts differentiated by means of the semantic differential," *Educational and Psychological Measurement*, vol. 21, no. 1, pp. 53–58, 1961.

[260] J. Kang and M. Zhang, "Semantic differential analysis of the soundscape in urban open public spaces," *Building and environment*, vol. 45, no. 1, pp. 150–157, 2010.

[261] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[262] V. W. Berger and Y. Zhou, "Kolmogorov–smirnov test: Overview," *Wiley statsref: Statistics reference online*, 2014.

[263] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Annals of cardiac anaesthesia*, vol. 22, no. 1, pp. 67–72, 2019.

[264] B. B. Schultz, "Levene's test for relative variation," *Systematic Zoology*, vol. 34, no. 4, pp. 449–456, 1985.

[265] H. Arsham and M. Lovric, "Bartlett's test.," *International encyclopedia of statistical science*, vol. 1, pp. 87–88, 2011.

[266] P. E. McKight and J. Najab, "Kruskal-wallis test," *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.

[267] A. Chen and W. Zhu, "Revisiting the assumptions for inferential statistical analyses: A conceptual guide," *Quest*, vol. 53, no. 4, pp. 418–439, 2001.

[268] A. Hilton and R. A. Armstrong, "Statnote 6: post-hoc anova tests," *Microbiologist*, vol. 2006, pp. 34–36, 2006.

[269] B. Ci and R.-O. Rule, "Confidence intervals," *Lancet*, vol. 1, no. 8531, pp. 494–7, 1987.

[270] R. Väänänen *et al.*, "Efficient modeling and simulation of room reverberation," Master's thesis, 1997.

[271] H. Kuttruff, "Room acoustics spon press," *London, UK,*, 2000.

[272] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pp. 99–102, IEEE, 2001.

[273] D. A. Brodén, K. Paridari, and L. Nordström, "Matlab applications to generate synthetic electricity load profiles of office buildings and detached houses," in *2017 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*, pp. 1–6, IEEE, 2017.

[274] D. F. Williamson, R. A. Parker, and J. S. Kendrick, "The box plot: a simple visual method to interpret data," *Annals of internal medicine*, vol. 110, no. 11, pp. 916–921, 1989.

[275] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *The American Statistician*, vol. 32, no. 1, pp. 12–16, 1978.

[276] N. K. Srinivasan, M. Stansell, and F. J. Gallun, "The role of early and late reflections on spatial release from masking: Effects of age and hearing loss," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. EL185–EL191, 2017.

[277] R. EBU-Recommendation, "Loudness normalisation and permitted maximum level of audio signals," *European Broadcasting Union*, 2011.

[278] B. Series, "ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level," 2011.

[279] N. R. Council *et al.*, "Hearing loss: Determining eligibility for social security benefits," 2004.

[280] C. Dodge and T. A. Jerse, *Computer music: synthesis, composition, and performance*. Macmillan Library Reference, 1985.

[281] S. Thavam and M. Dietz, "Smallest perceivable interaural time differences," *The Journal of the Acoustical Society of America*, vol. 145, no. 1, pp. 458–468, 2019.

[282] P. A. Regalia and M. K. Sanjit, "Kronecker products, unitary matrices and signal processing applications," *SIAM review*, vol. 31, no. 4, pp. 586–613, 1989.

[283] H. Zhang, F. Ding, *et al.*, "On the kronecker products and their applications," *Journal of Applied Mathematics*, vol. 2013, 2013.

[284] K. Shanthi and N. Nagarajan, "Memory based hardware efficient implementation of fir filters," *International Review on Computer and Software*, vol. 8, no. 7, pp. 1718–1726, 2013.

[285] A. O. M. Salih, "Audio noise reduction using low pass filters," *Open Access Library Journal*, vol. 4, no. 11, pp. 1–7, 2017.

[286] S. J. Schlecht, *Feedback Delay Networks in Artificial Reverberation and Reverberation Enhancement*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2018.

[287] C. M. Harris, "Absorption of sound in air versus humidity and temperature," *The Journal of the Acoustical Society of America*, vol. 40, no. 1, pp. 148–159, 1966.

[288] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A perceptual evaluation of individual and non-individual HRTFs: A case study of the SADIE II database," *Applied Sciences*, vol. 8, no. 11, p. 2029, 2018.

[289] R. D. Patterson, "Auditory filter shapes derived with noise stimuli," *The Journal of the Acoustical Society of America*, vol. 59, no. 3, pp. 640–654, 1976.

[290] S. Pearce, "Audio spatialisation for headphones - impulse response dataset." `http://doi.org/10.5281/zenodo.4780815`, 2021. [Accessed December 18, 2022].

[291] G. Kearney, M. Gorzel, F. Boland, and H. Rice, "Depth perception in interactive virtual acoustic environments using higher order ambisonic soundfields," in *Proccedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, 2010.

[292] M. Gorzel, G. Kearney, A. Foteinou, S. Hoare, and S. Shelley, "Lady Chapel, St Albans Cathedral, Openair, Audiolab, University of York.." `https://www.openair.hosted.york.ac.uk/?page_id=595`, 2010. [Accessed December 18, 2022].

[293] M. Melnyk, A. Kernyskyy, and M. Lobur, "Comparison of methods for measuring reverberation time," *The National University of Lithuanian Politician's Wizard. Computer design systems. Theory and Practice*, no. 908, pp. 11–17, 2018.

[294] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, 1995.

[295] "xcorr." `https://uk.mathworks.com/help/matlab/ref/xcorr.html`. [Accessed January 14, 2023].

[296] A. ISO, "Measurement of room acoustic parameterspart 1," *ISO Std*, 2009.

[297] S. Girón, M. Galindo, and T. Gómez-Gómez, "Assessment of the subjective perception of reverberation in spanish cathedrals," *Building and Environment*, vol. 171, p. 106656, 2020.

[298] I. Frissen, B. F. Katz, and C. Guastavino, "Effect of sound source stimuli on the perception of reverberation in large volumes," in *Auditory Display*, pp. 358–376, Springer, 2009.

[299] C. Leider, *Digital audio workstation*. McGraw-Hill New York, 2004.

[300] "What Are DAWs, Audio Plugins, and MIDI Controllers?." [Accessed November 21, 2022].

[301] C. Borß, "A vst reverberation effect plugin based on synthetic room impulse responses," in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-09), Como, Italy*, 2009.

[302] E. Ly and J. Villegas, "Generating artificial reverberation via genetic algorithms for real-time applications," *Entropy*, vol. 22, no. 11, p. 1309, 2020.

[303] M. Ducceschipaq and C. J. Webbpbq, "Plate reverberation: Towards the development of a real-time physical model for the working musician," 2016.

[304] "reverberator." `https://uk.mathworks.com/help/audio/ref/reverberator-system-object.html`. [Accessed December 18, 2022].

[305] J. Dattorro, "Effect design, part 2: Delay line modulation and chorus," *Journal of the Audio engineering Society*, vol. 45, no. 10, pp. 764–788, 1997.

[306] "Audio I/O: Buffering, Latency, and Throughput." `https://uk.mathworks.com/help/releases/R2020a/audio/gs/audio-io-buffering-latency-and-throughput.html#mw_3afa0c65-b451-4b3e-a368-5a894d6251b5`. [Accessed December 20, 2022].

[307] "Audio plugins in matlab." `https://uk.mathworks.com/help/audio/gs/audio-plugins-in-matlab.html`. [Accessed December 20, 2022].

[308] "rceps." `https://uk.mathworks.com/help/signal/ref/rceps.html`. [Accessed January 11, 2023].

[309] "Filter." `https://uk.mathworks.com/help/matlab/ref/filter.html`. [Accessed December 28, 2022].

[310] "Measure performance of streaming real-time audio algorithms." https://uk.mathworks.com/help/releases/R2020a/audio/examples/measure-performance-of-streaming-real-time-audio-algorithms.html. [Accessed December 28, 2022].

[311] "dsp.iirfilter." https://uk.mathworks.com/help/dsp/ref/dsp.iirfilter-system-object.html. [Accessed December 28, 2022].

[312] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *2009 16th International Conference on Digital Signal Processing*, pp. 1–5, IEEE, 2009.

[313] R. Stewart and M. Sandler, "Database of omnidirectional and B-format room impulse responses," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 165–168, IEEE, 2010.

[314] "Ambisonic binaural decoding." https://uk.mathworks.com/help/audio/ug/ambisonic-binaural-decoding.html. [Accessed January 9, 2023].

[315] "HRTF Database Acoustics Research Institute." https://projects.ari.oeaw.ac.at/research/experimental_audiology/hrtf/database/hrtfItEARI.html. [Accessed January 9, 2023].

[316] F. Grond and P. Lecomte, "Higher order ambisonics for supercollider," in *Linux audio conference*, 2017.

[317] W. Reichardt and W. Schmidt, "Die wahrnehmbarkeit der veranderung von schallfeldparametern bei der darbietung von musik," *Acustica*, vol. 18, no. 5, pp. 274–282, 1967.