
Implementation Of A Hybrid Reverb Algorithm

Parameterizing Synthetic Late Reverberation From Impulse
Responses

Master Thesis
Maximilian Steimel

Aalborg University
Sound and Music Computing



AALBORG UNIVERSITY

STUDENT REPORT

School of Information and Communication
Technology

Niels Jernes Vej 12
DK-9220 Aalborg Ø
<http://sict.aau.dk>

Title:
Implementation Of A Hybrid Reverb
Algorithm

Theme:
Master Thesis

Project Period:
Spring Semester 2019

Project Group:
Individual

Participant(s):
Maximilian Steimel

Supervisor(s):
Cumhur Erkut
Chris Tuttle

Copies: 1

Page Numbers: 78

Date of Completion:
August 15, 2019

Abstract:

Hybrid Reverb describes the combination of convolution reverb and a synthetic reverb structure to create artificial reverberation. This report describes recent methods in the field. An implementation of a hybrid reverberator is described, using state-of-the-art methods by combining convolution reverb for modelling early reflections and a filter structure for synthetic late reverb, whose coefficients are parameterized from room impulse responses, using Energy Decay Relief analysis and Least Squares optimization.

The implementation is evaluated through signal comparison, a subjective listening test, and an objective evaluation of perceived audio quality. The goal of the evaluation was to find out if the implemented method performs better than a more rudimentary hybrid reverberation method. The results show only a marginal difference in favor for the implementation, while laying the foundation for future work and possible alternatives to the investigated structure.

Contents

Preface	3
1 Introduction	5
2 Background	7
2.1 Natural Occurrence and Perception of Reverb	7
2.1.1 Direct Sound	8
2.1.2 Early Reflections	8
2.1.3 Late Reverberation	9
2.1.4 Absorption	9
2.1.5 Diffusion	9
2.2 Artificial Reverberation	10
2.2.1 Convolution Reverb	10
2.2.2 Feedback Delay Network	11
2.2.3 Hybrid Approaches	14
3 Hybrid Reverb Methods	17
3.1 Analysis of Reverberant Systems	17
3.1.1 RIR Partitioning and Truncation	18
3.1.2 Late Reverberation Decay Analysis	20
3.2 Parameterization of FDN Reverb	23
3.2.1 Delay Times	23
3.2.2 Absorptive Filters	24
3.2.3 Proportional Parametric Equalizers	24
3.2.4 Parametric Filter Optimization	28
4 Hybrid Reverb Algorithm	33
4.1 Design Overview	33
4.2 Implementation	34
4.2.1 Early Reflections	34
4.2.2 Spectral Decay Analysis	35
4.2.3 Synthetic Late Reverb	37

Contents	1
4.2.4 FDN Filter Parametrization & Optimization	41
4.2.5 Mixing Early and Late Reverb	43
5 Evaluation	45
5.1 Signal Comparison	45
5.1.1 T_{60} Comparison	45
5.1.2 Time Domain Comparison	46
5.1.3 Spectral Comparison	48
5.1.4 Remarks	50
5.2 Subjective Listening Test	50
5.2.1 Choice of RIR and Stimuli	50
5.2.2 Test Design	51
5.2.3 Participants	52
5.2.4 Setup	52
5.2.5 Issues	52
5.3 PEAQ	52
6 Results and Discussion	55
6.1 Listening Test results	55
6.1.1 smallRoom	55
6.1.2 mediumChamber	56
6.1.3 naveCathedral	57
6.2 PEAQ Results	58
6.3 Discussion	58
6.3.1 Perspectives On Listening Test Design	60
7 Conclusion	61
7.1 Future Work	62
Bibliography	63
A Listening Test Results: One-Way ANOVA	69
A.1 smallRoom	69
A.2 mediumChamber	70
A.3 naveCathedral	70
B Listening Test Results: Comments On Samples	71
B.1 smallRoom	71
B.2 mediumChamber	73
B.3 naveCathedral	75

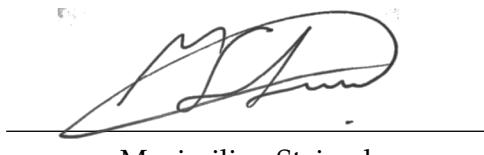
Preface

This thesis was conceived as the final project of the Sound and Music Computing Master's program at Aalborg University, Copenhagen. The work was carried out between April 1st and 15th August 2019.

The report investigates the implementation of a hybrid reverb, combining various methods of signal analysis and artificial reverberation. The thesis project was conducted at *Ableton AG* in Berlin, Germany, and was hosted and counselled through the product development team. The implementation was evaluated through signal comparison, a subjective listening test (APE), and an objective test of perceived audio quality (PEAQ).

The author would like to thank his supervisor Cumhur Erkut for his extensive guidance throughout the whole thesis period. Special thanks go to each member of the *FA Sound* for their relentless counselling, knowledge and support. Lastly, the author would like to thank Dr.-Ing. Sebastian Schlecht for providing additional knowledge and annotations about his work.

Aalborg University, August 15, 2019



Maximilian Steimel
<msteim17@student.aau.dk>

Chapter 1

Introduction

Since its first introduction to the digital domain, artificial reverberation holds the promise to recreate the sound of any possible or impossible space. It is an essential tool for evoking depth in music and film production, active acoustic enhancement or simply as a creative tool for the creation of music and sound art. The main strength of digital artificial reverb lies in the possibilities that come with its parametrization, as this allows the user to tweak a reverb device to his needs and allow for a great range of different spaces. However, if one is to dive into the vast realm of reverb algorithms that are known to the public since the creation of the first digital reverb by Manfred Schroeder in 1962 [1], one quickly realizes that there are a lot of algorithms known that model a certain range of spaces but none of them can do it all.

A method that surpasses this problem elegantly in terms of the sonic outcome, is the powerful technique of convolution reverb, which enables to recreate the sound of literally any real space or artificial reverberant system, as all aspects that characterize the room (when time invariant) are faithfully stored in an impulse response. However, using an impulse response for reverberation is merely like taking a snapshot of a space or reverb device, as very little can be changed about the deeper characteristics of the retrieved impulse response afterwards.

The solution seems to lie in combining convolution and synthetic reverb methods to create a more potent reverberator that will provide the best of both worlds and that can faithfully recreate a desired space by maintaining all the flexibility that comes with its possibilities of parametrization. This essentially forms the concept of the hybrid reverberator.

The goal of this thesis is to investigate and implement a method that enables to extract the important characteristics of a room impulse response and faithfully apply it on an algorithmic reverb structure that can recreate the sound, while also taking advantage of the strengths of convolution reverb. In an ideal case, this

keeps the flexibility of further user-facing parametrization, while leading to a high quality sonic outcome.

The Background chapter investigates on the necessary methods for synthetic reverb, followed by a chapter that revises state-of-the-art methods of hybrid reverberation, in order to justify the implementation.

The Implementation chapter documents the implementation of a hybrid reverb, using statistical analysis to retrieve a truncated impulse response and employs convolution for the modelling of the early reflections part, while completing the late reverberation with a generic synthetic reverb structure whose parameters are optimized from a previously analyzed room impulse response. The whole implementation is carried out using the *MATLAB* language.

The outcome of the implementation is evaluated using a subjective listening test, as well as general signal comparison to the original impulse response and an objective test of audio quality. Previous considerations regarding the test design, as well as a thorough presentation and discussion of the results is given. The report is completed by concluding on the discussed results and suggesting future improvements. Further data from the listening test can be found in the appendices A and B.

The outcome of this thesis hopefully fuels future investigations and laid the foundation for a possible embedding of the algorithm into feasible real-time product.

Chapter 2

Background

2.1 Natural Occurrence and Perception of Reverb

In general, reverb can be seen as the result of sound waves that propagate within a room from a sound source and get reflected between walls and objects that comprise the room. The sound travels at a speed of approximately 344 m/s at a room temperature of 21 °C [2]. Therefore the individual reflections will increase in density while decaying in amplitude, over a certain time period, until all kinetic energy has been absorbed by the room. The totality of all reflections eventually lead to a dense audible texture, which might still contain the original sonic characteristics and parts of the frequency content of the initial sound, and this is perceived as the resulting sound of reverberation.

For the sake of modelling one can break down this process in to three parts.

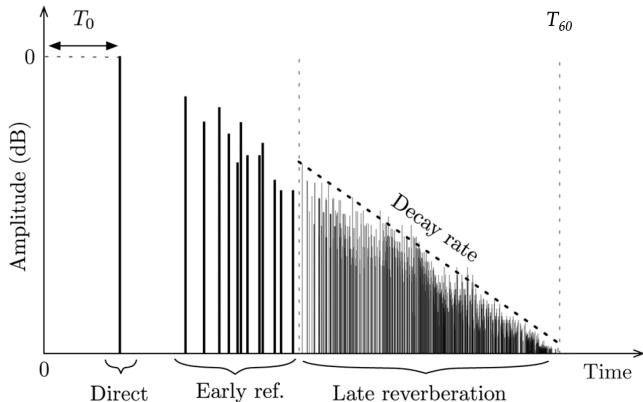


Figure 2.1: Schematic figure of a Room Impulse Response, as in [3]. The individual repetitions of the initial sound are represented as bars on the time axis, whose length represents their amplitude in dB.

As can be seen in figure 2.1 the direct sound sound, denoted as T_0 is initially followed by a phase of sparse first-, second- and eventually third-order reflections which are considered the early reflections. The progression is completed by a decaying tail of late reverberation, at which point the reflections become so dense that they can't be individually distinguished, and are perceived as one dense texture. The whole length of the progression is denoted as T_{60} , which is the time it takes for the overall reverberation to decay -60dB in amplitude [4].

The different phases of this physical process and related concepts are discussed more in-depth as follows.

2.1.1 Direct Sound

The direct sound is the direct wave-emission that originates from the sound source and reaches the listener without having undergone any form of reflection. Depending on the distance to the listener, it would however undergone some kind of attenuation in its intensity. This is calculated as

$$I_{direct} = \frac{QW_{source}}{4\pi r^2}, \quad (2.1)$$

where I depicts intensity of the direct sound that reaches the listener in [W/m^2], Q is the directivity factor, W is the power of the source, and r is the distance in m . The difference in intensity and timing between the direct sound and the early reflections are essential to give the listener an idea about the distance to the source and the rough dimensions of the room [5, 6].

2.1.2 Early Reflections

The early reflections make up the first- second- and sometimes third-order reflections of the initial sound that will reach the listener as reflections from walls and objects in the room. These reflections will therefore differ in timing, intensity and direction, and might even have changed in frequency content, due to possible absorption which the material of the walls, objects and air itself impose on the reflected sound.

In this way, the early reflections do also have an effect on the tonal qualities of the reverberated sound, as the perceived timbre might even be altered, depending on how the reflections occur [6]. Some ambiguity seems to exist in the definition of the transition-point that separates the early reflections from the diffuse late reverberation. However, it seems to be generally agreed on, that the early reflections phase ends roughly between 80 to 100 ms after the initial sound [7].

2.1.3 Late Reverberation

The late reverberation is the phase of the reverb, where the individual reflections have become so dense that they are just perceived as one dense texture that still increases in density. In general it can be said that the development of the late reverb tail follows an exponential decay, while the reflections are distributed in a stochastic manner. [8, 9]. As the diffused reflections in this phase will reach the listener from all sides, there is no perceived direction to this part of the reverb sound anymore. The actual decay time of the late reverberation is determined through the size of the room and the degree of absorption that the walls and objects in it impose on the diffused sound.

Therefore, the decay-behaviour of the late reverberation offers a major cue about the dimensions and the materials of the room [10]. Ideally, the late reverberation decays exponentially and exhibits no coloration whatsoever, this is however rarely the case in nature. Depending on the architecture of the given room, some reflections might be repelled back and forth between two boundaries, leading to an audible periodicity that gives rise to phenomena like 'flutter echo' or 'standing waves' [6].

If these periodic repetitions occur fast enough (i.e. faster than 30ms), they will be perceived as a coloration of the reverberated sound, leading to an attenuation or amplification of certain frequencies, perceived as resonances that may change the spectral content of the reverb tail [4].

2.1.4 Absorption

Absorption refers to the attenuation of high frequencies that occur while sound waves are propagating through a space. The attenuation is caused both by the air itself, and by materials of the walls and objects that comprise the room, essentially through frictional processes that diminish the kinetic energy of the sound that is propagating. The absorption has a strong effect on the frequency-dependant decay time of the room, and the resulting auditory impression, which inherits information about the material that the reverberant room is treated with [11].

2.1.5 Diffusion

Diffusion is considered the phenomena that occurs when the sonic energy distributes evenly all around the space. Ideally, diffusion is characterized through a perfect smearing of the densening reflections while no particular frequency is reinforced through any resonance [4]. Although this property is hardly achieved in perfect form in real environments, it is possible with an artificial reverberator and is achieved through delay line modulation or through a distinct distribution of delay times.

2.2 Artificial Reverberation

This section is a brief introduction to the methods in artificial reverberation that are deemed relevant for the background of this thesis.

2.2.1 Convolution Reverb

Probably the most straight forward way to reverberate a signal, is to convolve it with the impulse response of a reverberant space or system. The impulse response itself can be obtained by measuring the computational acoustic model [12]. In the case of a real room, using a balloon pop [13], pistol shot [14], sine sweep [15] or a pseudo-random noise sequence [16], have been discussed.

As a general paradigm in convolution reverb, it is assumed that the desired system is Linear and Time Invariant (LTI) and can therefore completely determined through an Finite Impulse Response (FIR) filter. By convolving the impulse response of a space or reverberant system with some input, the output that is generated will therefore sound just like it would have been passed through the room or system itself.

The convolution operation can be defined as

$$y[n] = x[n] * h[n], \quad (2.2)$$

where $x[n]$ is the input-signal, $h[x]$ is an FIR filter completely determined through an impulse response and $*$ is the convolution operation.

The convolution summation as a direct-form FIR filter is then given as

$$y[n] = \sum_{m=0}^{N-1} h[m]x[n-m]. \quad (2.3)$$

As the direct computation will require the calculation of the DFT, spectral multiplication, and IDFT between each input and output sample, it leads to a delay that prohibits a real-time application. Several solutions have been suggested to this issue in the form of block-wise convolution, where the operation is done on larger blocks of the input in the frequency domain, either on segments of varying [17, 18, 19] or fixed length [20].

An advantage of convolution reverb is that it faithfully replicates all nuances and spatial characteristics of a given room, assuming that the impulse response has been recorded correctly. As the early reflections are of great perceptual importance for the spatial impression of a given space [21, 22], this is a very desirable property to have. Timing, direction and absorption of the early reflections, which would otherwise need to be taken care of with various filters, are already included in the desired response that is being convoluted.

Its disadvantage however lies in the limited possibilities to further parameterize this type of artificial reverb. Although the overall decay time might be influenced by applying a global amplitude envelope on the impulse response, the frequency characteristics can not be changed to a great extend.

2.2.2 Feedback Delay Network

In this section, the general structure of the Feedback-Delay-Network (FDN) is briefly described and elaborated on.

Origin

Artificial reverb was first introduced into the digital domain by Manfred Schroeder and Ben Logan. They suggested an algorithm architecture that should consist of recursive delay filters, i.e. feedback comb filters as the basic units [23].

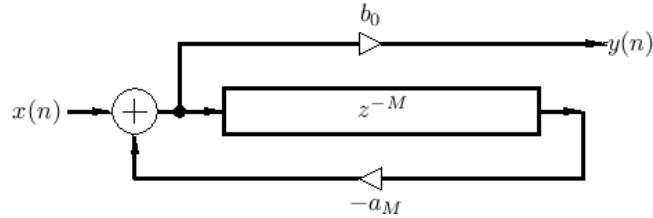


Figure 2.2: A feedback comb filter, where z^{-M} depicts the order of the delay, $-a_M$ is the negative feedback gain and b_0 is a gain scalar.

As the proposed recursive filter produces a recurring repetition of the incoming signal, which decays in amplitude over time, the result can be seen as a crude approximation to the many reflections that occur in a real space over time, if several of these filters are used in parallel. Schroeders topology uses parallel comb filters to increase echo density, which are fed into a chain of serial all-pass filters in order to further diffuse the signal [1].

Since this model only pays attention to the late and diffuse part of reverberation, it was later extended by Moorer who proposes an FIR filter in the form of a tapped delay line, in order to model the early reflections part of the reverb as discrete echoes [24].

He also suggested to add a low-pass filter into the feedback-path of the comb filter, which enables the emulation of the high-frequency absorption that occurs in the air, due to an increasing attenuation of high frequencies with each iteration. Further algorithms that build on these basic concepts have been introduced, by Dattorro [25], Dahl and Jot [26]. A more complete conglomeration of the matter can be found in [3].

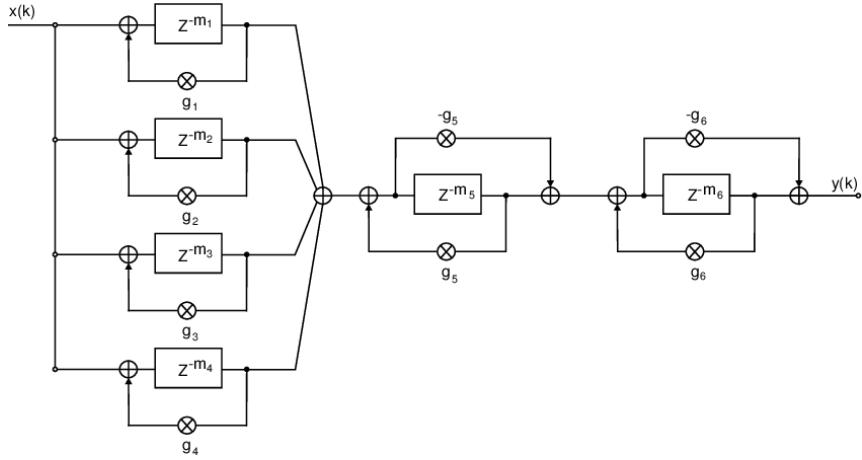


Figure 2.3: Schroeders reverberator-design, as proposed in [1]

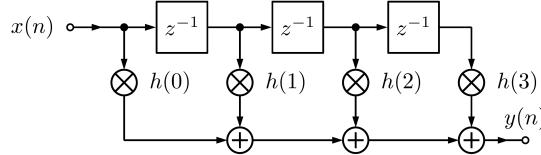


Figure 2.4: A tapped delay line, where outputs are derived after several discrete versions of the delayed signal.

Overview and Structure

A FDN can be generalized as a vectorized comb filter, where a single delay line is replaced by a diagonal delayline matrix, and where each delay line in the system is connected through a feedback matrix. This cross-coupling promises an actual buildup of echo density over time and was first suggested by Gerzon [27, 28], as it was expected to yield better results than the linear structure of parallel comb-filters and serial all-pass filters of Schroeders reverberator [23, 1], where the echo density would not increase after a certain period of time or would simply not become so dense at all.

In order to implement a multichannel reverberator, the concept was later revisited by Stautner and Puckette [29], who also introduced the idea of using a 4×4 Hadamard feedback matrix for cross-coupling, which maintains general stability.

A generalized FDN design is suggested by Jot, which extends the method to any recursive delay network having a unitary feedback matrix [30]. Jot also adapts the idea from Schroeder and Moorer [1, 24] of using absorptive filters to attenuate the higher frequencies within the reverberation.

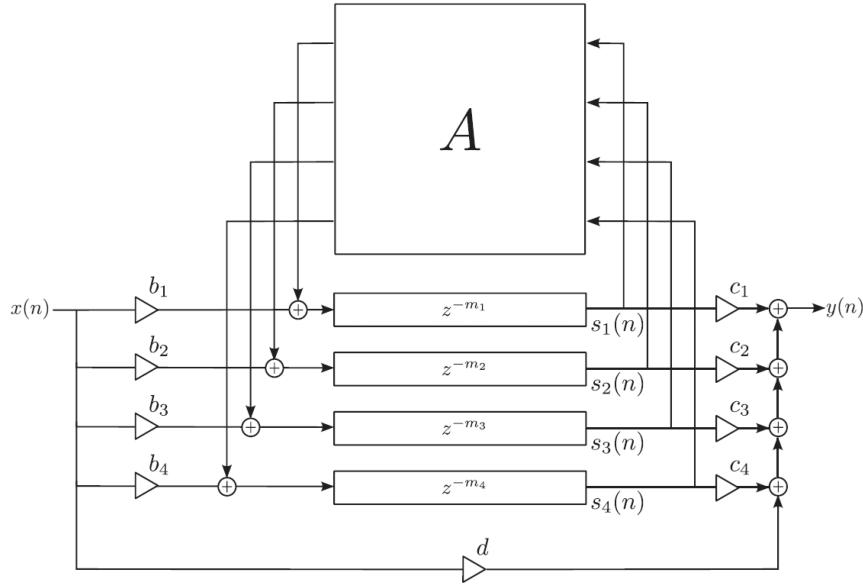


Figure 2.5: Example of a generic 4-channel FDN as suggested in [29], where b and c are gain coefficients, z^{-m} depicts a delay of order m and A is the feedback matrix .

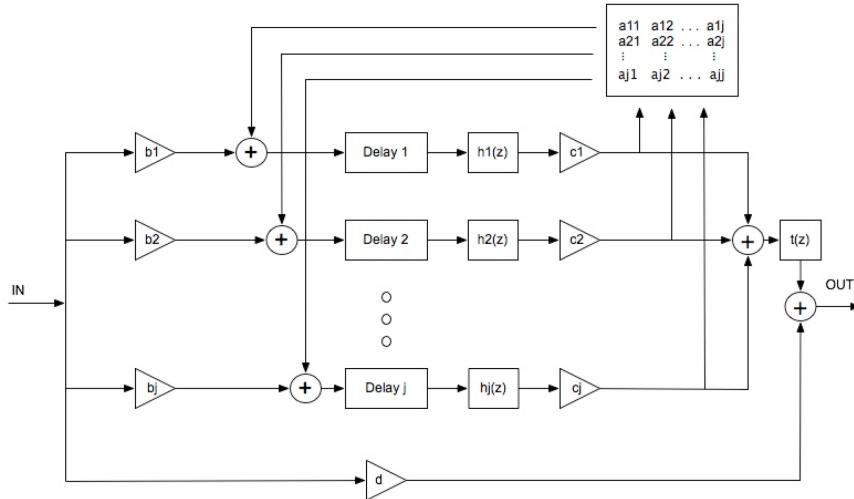


Figure 2.6: Jot's FDN generalization, where $h(z)$ is an absorptive filter, and $t(z)$ is a tone correction filter [30].

This enables the elimination of the typical ringing noise that occurs within any FDN at longer decay times. Furthermore, this also enables to control the decay time of the reverberator as a function of frequency [29].

Matrices

Depending on the dimension of a given FDN, the interconnection of a lot of recursive filters can quickly lead to stability problems. Ideally, a stable FDN is achieved by using a matrix structure that will lead to a lossless impulse response if no absorptive filters are involved. This dictates that the matrix at hand is unitary, as well as its determinant being equal to zero, as all eigenvalues within the matrix have unit-modulus. One matrix that fulfills this requirement is the above mentioned permuted Hadamard matrix

$$\mathbf{A} = g \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \quad (2.4)$$

where g is the global feedback-gain coefficient while $|g| < \frac{1}{\sqrt{2}}$.

If an FDN of greater order is to be realized, a matrix of related dimensions with the same properties is desirable. An elegant solution is proposed by Jot in the form of the embedded $N = 4$ Householder matrix [30].

In this case, the following matrix

$$\mathbf{A}_4 = \frac{1}{2} \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}, \quad (2.5)$$

is simply embedded for each entry in a matrix of identical structure

$$\mathbf{A}_{16} = \frac{1}{2} \begin{bmatrix} \mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 \\ -\mathbf{A}_4 & \mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 \\ -\mathbf{A}_4 & -\mathbf{A}_4 & \mathbf{A}_4 & -\mathbf{A}_4 \\ -\mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 & \mathbf{A}_4 \end{bmatrix}. \quad (2.6)$$

Although it makes sense to start with a lossless prototype while designing an FDN using a fixed and unilossless matrix, several examples of time varying and closely unilossless matrices have recently been proposed by Schlecht and Habets, in order to enhance the perceptual quality of the reverb tail [[schlecht_time_varying_2015](#)] and to enable direction-dependant decay times in a spatial scenario [31].

2.2.3 Hybrid Approaches

The general idea of a hybrid reverberator, is to combine parts of the previous mentioned methods, in order to achieve a product that contains both the perceptual accuracy of convolution reverb, and the parametric flexibility of algorithmic

reverberation. Moorer was the first to mention the significance of modelling the early reflections when constructing an artificial reverberator, while modelling them through the use of a tapped delay line, where each delay time and amplitude is calculated, using the image-source method [24].

The notion of combining the strengths of convolution and algorithmic reverb was later risen by Jot [32], who suggests to parameterize an FDN from an impulse response, by analyzing the Energy Decay Relief (EDR) of the IR, to then make an FDN match that curve by adjusting its filter coefficients accordingly. The first actual hybrid reverberator that employs convolution for the early reflections and recursive filters for the late reverberation was established by Browne [33], however using a fixed truncation point for the early reflections and no matrix-based FDN. In order to partition the early reflections that have been captured within the IR more faithfully, Stewart and Murphy revisit both Jot's and Browne's method resulting in an approach to first truncate the early reflections portion of the impulse response, and later add it to a properly parameterized late FDN reverberation through parallel processing, which then forms the complete hybrid reverberator [34].

A comparable approach is also conducted by Abel for modelling plate reverberation [35], while an alternative approach to the FDN but also in a hybrid context is implemented by Lee and others, in order match a hybrid reverberator to the IRs of an EMT 140 plate reverberator. Here the late reverb portion is approximated through an IIR filter structure that is based on small segmented FIR parts of decaying noise [36].

Another approach using modal reverberation is also implemented by Abel , where a modal decomposition is first performed on the RIR and the late reverberation is re-synthesized by modelling these modes through a parallel bank of resonant filters [37].

The relevant underlying methods for the work of this thesis, are discussed more in-depth in the following chapter.

Chapter 3

Hybrid Reverb Methods

A Hybrid Reverberator can be realized by using only the segment of an RIR that contains the early reflections phase, and then complete the reverberation with an algorithmic reverb that will faithfully replicate the late reverberation tail which was originally found in the RIR. This will require at least two steps of analysis on the RIR at hand, first to enable a precise separation of the early reflections from the RIR, and secondly to faithfully extract the decay behaviour of the late reverb phase. The analysis results can then be used to parameterize a filter-bank-based algorithmic reverb.

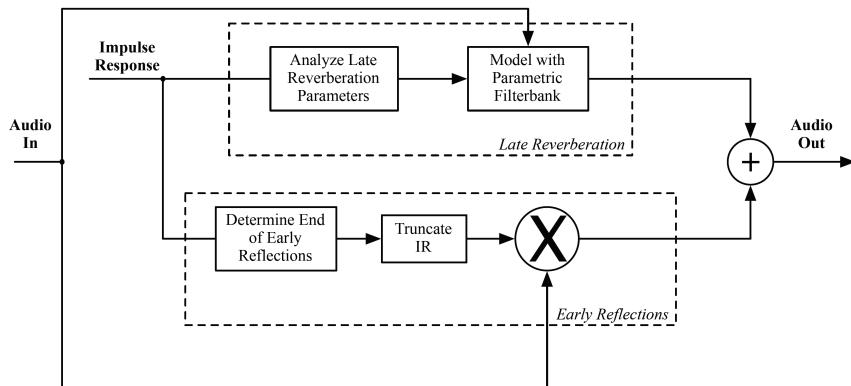


Figure 3.1: High-level algorithm sketch of a hybrid reverberator, as in [38].

3.1 Analysis of Reverberant Systems

This paragraph introduces relevant analysis methods for both separating early reflections and late reverb from each other within a given RIR, and methods for the description and analysis of the late reverberation.

3.1.1 RIR Partitioning and Truncation

In the context of a hybrid reverb framework, it is desirable to separate the early reflections from an RIR, and truncate it accordingly, to then later mix it with the output of the algorithm that reproduces the late part of the reverberation.

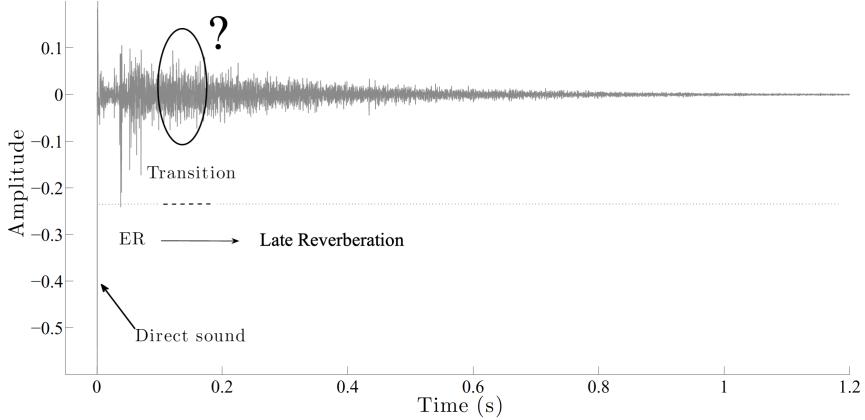


Figure 3.2: Amplitude plot of a generic RIR with a truncation point in question, as in [39].

A first solution to this problem is proposed by Stewart and Sandler [38] who assume the approximately Gaussian distribution of reflections captured in a RIR and that early reflections tend to occur below one third of the standard deviation from the mean. Here, the individual reflections are just treated like a group of samples:

$$\sigma = \sqrt{E(x^2) - (E(x))^2} \quad (3.1)$$

where $E(x)$ is the expected value of x . The authors therefore choose a Gaussian estimator approach which employs a 20ms window that loops through the RIR on a sample-by-sample basis to calculate the mean and standard deviation. Within a normal distribution it is assumed that two thirds of the observed samples occur between $-\sigma$ and σ , while one third lie between $[-\infty; -\sigma]$ and $[\sigma; +\infty]$. Determining the ratio $R(n.h)$ between the samples that lie within and outside that distribution is then computed as:

$$R(n.h) = \sum_{n=0}^{N-1} \frac{N_{x(n.h)}[-\sigma; \sigma]}{N_{x(n.h)}[-\infty; -\sigma] + N_{x(n.h)}[\sigma; +\infty]}, \quad (3.2)$$

where $N_{x(n.h)}$ denotes the number of samples of $x(n.h)$ within the observation, and h is a hop size window. As early reflections tend to have more samples within the

standard deviation than the late reflections, the approach is useful. Once the standard deviation reaches $R(n.h) \geq 2$, the time point beyond this window is chosen as the end of the early reflections.

Primavera [40] proposes a method to instead look at the skewness and kurtosis through the use of a Jarque-Bera test. The test is performed on a windowed portion of the impulse response, windowed through a 24ms rectangular window on a sample-by-sample basis. The Jarque-Bera test is defined as

$$JB = \frac{n}{6} (s^2 + \frac{1}{4}(k - 3)^2), \quad (3.3)$$

where s is the skewness given as

$$s = \frac{E(x - \mu)^3}{\sigma} \quad (3.4)$$

while E is the expected value, μ denotes the mean and σ denotes the standard deviation, and where k is the kurtosis defined as

$$k = \frac{E(x - \mu)^4}{\sigma^4} - 3 \quad (3.5)$$

and n is the number of observed samples. The point where the hereby determined coefficient JB tends to zero is then set as the time frame in the RIR where the early reflections end.

Defrance and Polack [39] look at the issue from the perspective of the phase instead by assuming that the phase of the reflections that are captured within a RIR will change faster over time and therefore lead to a randomization of the phase, as the reflections reach a certain density. The assumption arises from the fact that the more sonic energy is reflected, absorbed and scattered by boundaries and materials in a room, the more the occurring reflections will change in modulus and phase [41]. In order to determine how the phase changes over time, Defrance and Polack implement an application of the Fourier Transform that is performed on an extensible window of variable width (labeled XFT), which enables to observe the evolution of the unwrapped phase at each point in time, depending on the past of the signal. It is defined as

$$Y(w, v) = \sum_{n=0}^{wN-1} y(n) \cdot e^{-2j\pi nv}, \quad (3.6)$$

where $y(n)$ is the signal to be analyzed, N is the number of samples the analysis is performed on and w is a hop size window. In order to maintain a constant frequency resolution, the here defined XFT on the same number of samples at

each time frame. By calculating the mean regression error between the unwrapped phase and its linear regression, the development of the phase can then be plotted with regards to each time frame, enabling to determine the truncation point. Once this error passes a certain predetermined threshold, this point can be used as the truncation point for the early reflections (e.g. where the regression trend changes from linear to random).

3.1.2 Late Reverberation Decay Analysis

T₆₀

The simplest measure that characterizes a reverberant room or system, is the time that it takes for the sonic power that remains in the system after the ceasing of the direct sound to decay -60 dB . The measure was first introduced by Wallace Sabine [42], in order to determine the reverberation time in concert halls, and for real rooms it is composed as

$$T = \frac{V}{A} \cdot 0.161 \text{ s m}^{-1}, \quad (3.7)$$

where T is the reverberation time, V is the total volume of the room, A is the total absorption area and 0.161 is an empirical constant inversely dependent on the speed of sound, in units of m/s .

A more generalized way to determine T_{60} is to measure the amplitude of the reverb in a room or system by recording the change of its overall amplitude over time after the initial sound, and then determine the point where that amplitude reaches the -60 dB threshold, as is sketched in figure 3.3.

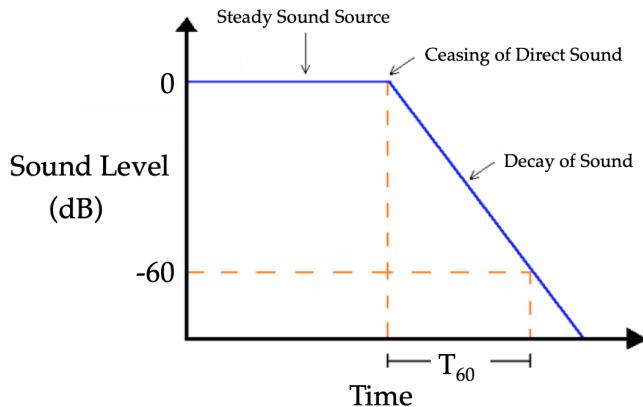


Figure 3.3: Schematic sketch of a reverb amplitude progression.

Energy Decay Curve

The Energy Decay Curve (EDC) was established by Schroeder [43] as a metric to determine the remaining energy in the impulse response of a reverberant system for any point in time. If a reverberant system is excited with a random noise signal, multiple measurements would need to be conducted, as the random nature of the excitation signal would lead to onsets that would not always contain exactly the same energy. In order to obtain the curve that would precisely reflect the overall decay progression of energy, many measurements and averaging between the results would be necessary. The EDC solves this issue by computing a backwards integration of the obtained impulse response

$$EDC_h(t) = \int_t^{\infty} h^2(\tau)d\tau, \quad (3.8)$$

where $h(t)$ is the given impulse response.

Although it enables to determine the overall decay time curve precisely, the EDC is not sufficient for determining the decay of energy for specific frequency bands.

ISO 3382-1:2009

The ISO standard 3382-1:2009 for the measurement of room acoustic parameters [44] actually employs a related approach, where instead the T_{30} , denoted as the time it takes for the reverberation to decay $-30dB$, is determined by first calculating a backwards integration of the squared impulse response

$$E(t) = \int_t^{\infty} p^2(\tau)d\tau = \int_{-\infty}^t p^2(\tau)d(-\tau) \quad (3.9)$$

where p denotes the sound pressure of the impulse response as a function of time, E is the energy of the decay curve as a function of time, and t denotes time.

This integration is done both for a subband-filtered version of the impulse response at 500Hz and 1000Hz. A line that fits the resulting decay curves for both filtered versions is then determined using least-squares, in order to determine the resulting reverberation time for both subbands, and the global reverberation time found by averaging the line between both bands.

Energy Decay Relief

An improvement to Schroeder's EDC method is suggested by Jot [8] in the form of the Energy Decay Relief (EDR). Instead of just using only the time-domain energy of the impulse response, the EDR is calculated on the time-frequency representation of an impulse response, and is defined as

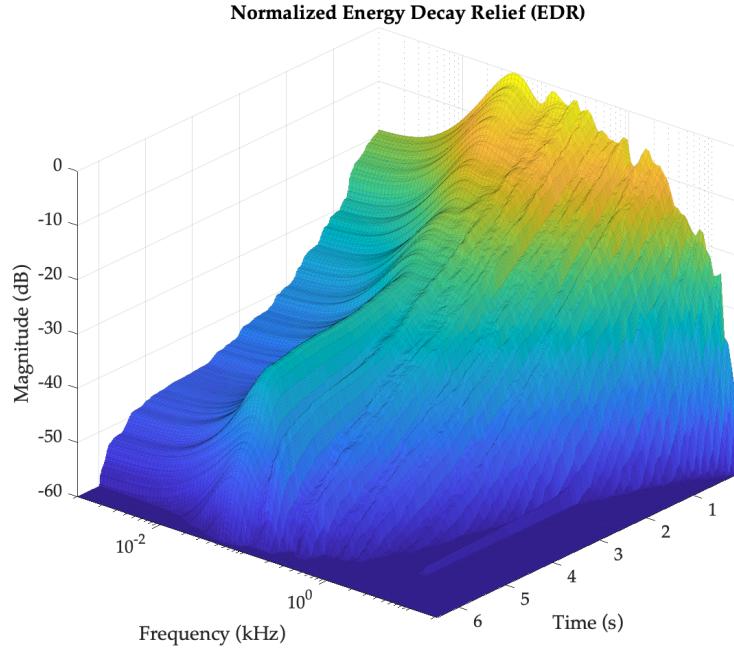


Figure 3.4: EDR plot of a room impulse response that was obtained in a cathedral, calculated for 100 logarithmically spaced bands.

$$EDR_{h(t,f)} = \int_t^{+\infty} p_h(\tau, f) d\tau, \quad (3.10)$$

where $p_h(\tau, f)$ is an energetic time-frequency representation of the impulse response. In the digital domain, this can be computed using the Short-Time Fourier Transform (STFT) as:

$$EDR(t_n, f_k) = \sum_{m=n}^M |H(m, k)|^2, \quad (3.11)$$

where $H(m, k)$ is bin k of the STFT at time-frame m , and M is the total number of time frames. In this way, the EDR enables to determine the total amount of energy in an impulse response at time $t_n = nT$ for a frequency band that is centred around $f_k = kf_s/N$, where N is the FFT length [45].

3.2 Parameterization of FDN Reverb

3.2.1 Delay Times

Initializing an FDN with adequate delay times for each delay line is essential, as the delay times are responsible for both reaching an accurate amount of echo density over time, and for maintaining an approximately colourless frequency distribution in the reverberation, without an absorptive filter in effect. In essence, this means that the reverb texture will ideally not contain any "flutter", meaning an irregular and "grainy" uneven amplitude modulation of the resulting impulse response related to what Schlecht coins *clustering* [46]; and also no "ringing", which occurs as a distinct emphasis of certain frequencies, which are occurring as resonances in the frequency spectrum of the reverberation[45].

Mean Free Path Approach

One possible guideline which however only makes sense if the rough measures of the room that is modelled through the FDN are actually known, is the measure of the mean free path which can be a guide for the average delay line length. The mean free path can be calculated after Sabine [42] as

$$\bar{d} = 4 \frac{V}{S} \quad (\text{mean free path}), \quad (3.12)$$

where V denotes total room volume and S denotes the total surface area. The average delay line length M_i can then be computed as

$$\frac{\bar{d}}{cT} = \frac{1}{N} \sum_{i=1}^N M_i, \quad (3.13)$$

where c is the speed of sound and T is the sampling period [45]. However, in a scenario where it is desired to parameterize the FDN from meaningful information that is retrieved from an impulse response, this approach is not useful.

Prime Numbers

Another guideline which was first found by Schroeder, is to use delay-times that are mutually prime, so that the prime factorization of the individual delay times in each delay line exhibit no common factors [23]. Expressed in a simplified way, these constraints will allow the individual delay lines to always fill in each other's gaps with the recurring repetitions of the incoming signal, and smoothly increase the echo density with not too many individual echoes piling up at certain points in time.

Furthermore, when it is required to scale the delay line lengths in a real-time scenario, it is useful to take the delay line length as an integer power of a distinct prime number p_i :

$$\hat{M}_i = p_i^{m_i} \quad (3.14)$$

so that the multiplicity of the prime p_i is then $m_i \geq 1$.

Mode Density

Another important requirement is inherited through the mode density of the FDN as a lossless system. As the total amount of allocated delay also determines the total order of the FDN and therefore the number of poles on the unit circle, while the individual delay times determine the distribution of the poles, only an even distribution of the poles can ensure that the reverberation will not exhibit any resonances, which would to the aforementioned 'ringing' [45]. If M is the number of poles on the unit circle, the required mode-density is given as

$$M/f_s = M \text{Tper Hz.} \quad (3.15)$$

where f_s is the samplerate. Schroeders guideline for a sufficient mode density in the frequency domain scales inversely proportional to the reverberation time T_{60} and is then given as [23]

$$M \geq 0.15T_{60}f_s. \quad (3.16)$$

3.2.2 Absorptive Filters

The simplest method to control the decay of an FDN as a function of frequency, is the introduction of low-pass filters into the feedback path of a lossless FDN. The method has its origin in Moorer's introduction of low-pass-comb filters [24], and was adapted to the use in FDNs by Jot [30, 32], Gardner [47] and Dahl [26].

3.2.3 Proportional Parametric Equalizers

A more elaborate solution to the use of absorptive filters, is the application of Proportional Parametric Equalizers as suggested by Jot [48]. Just as with simpler filter designs, these filters can be introduced at any point within a loss less prototype of an FDN, where the energy is preserved if no absorptive filter is in effect. Jots method enables the emulation of a more flexible and precise absorption where the decay time of the reverb can be adjusted for multiple frequency bands and by adjusting them using a dB scale. In the context of an FDN, the requirement for such a filter is that its negative dB gain $G_i(f)$ for any frequency f is proportional to the unit delay length τ_i , divided by the reverberation decay time $Tr(f)$:

$$G_i(f) = 20 \log_{10} |g_i(f)| = -60 \frac{\tau_i}{Tr(f)}. \quad (3.17)$$

This requirement is fulfilled by Jot's modification of the parametric equalizer structure originally proposed by Regalia and Mitra in [49]. Regalia and Mitra's filter is given by the transfer function:

$$H(z) = \frac{1}{2}[(1+k) + (1-k)A(z)] \quad (3.18)$$

where k is a gain coefficient ranging from -12 to 12 dB and $A(z)$ is a first- or second-order all-pass filter. Depending on the filter structure that is substituted for $A(z)$, either a shelving- or a presence-filter can be realized. Cascading several of these filters with Jot's modification within the feedback path of an FDN, provides a proper control of the frequency-dependent decay, as their dB gains can be adjusted proportionally to some target frequency-dependent decay time $Tr(f)$ [48].

The following paragraphs briefly introduce these different filter types and the constraints that Jots proportionality-modification impose on them.

Proportional Shelving Equalizer

A first-order low-shelving filter that has the transition frequency ω and gain k is proposed by Regalia and Mitra and substitutes

$$A(z) = -\frac{a + z^{-1}}{az^{-1} + 1}$$

while

$$a = \frac{t-1}{t+1}; t = \tan(\omega/2).$$

As noted by Jot [48] and as can be observed in figure 3.5, the magnitude response for boosting and attenuating when adjusting k is not symmetric, which can lead to a hardly predictable response when multiple of the same filters are cascaded.

In order to solve this issue, Jot modifies this filter by introducing a constraint on the filter coefficient a :

$$a = \frac{r-1}{r+1}; r = \frac{t}{\sqrt{k}} = \frac{\tan(\omega/2)}{\sqrt{k}} \quad (3.19)$$

which makes the feedback and feed forward coefficient a dependent on both the transition frequency ω and the gain coefficient k [48].

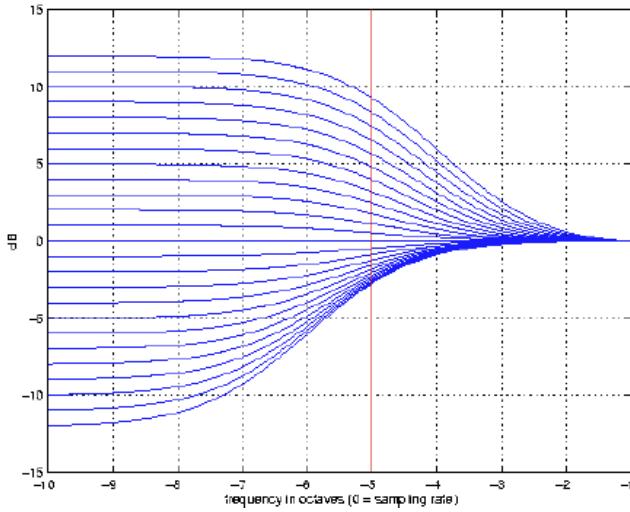


Figure 3.5: Frequency response of the 1st order shelving filter as proposed by Regalia and Mitra. Here gain k is adjusted while keeping frequency ω constant [48].

If the modified transfer function $A(z)$ is plugged into the original filter equation 3.18 by Regalia and Mitra [49], the proportional shelving equalizer can be simplified as

$$\tilde{H}_1(z) = \frac{(t\sqrt{k} + 1) + (t\sqrt{k} - 1)z^{-1}}{(t/\sqrt{k}) + 1 + (t/\sqrt{k} - 1)z^{-1}}. \quad (3.20)$$

An observation of the resulting magnitude response in figure 3.6 yields that Jot's proportionality constraints lead to a symmetric response when boosting or attenuating. Additionally it can be seen, that between -10 and 10 dB, the magnitude response is self similar.

Proportional Presence Equalizer

In order to realize a Proportional Presence Equalizer which has the frequency bandwidth ω and gain k at center frequency f , $A(z)$ in equation 3.18 is substituted for

$$A(z) = \frac{a + c(1-a)z^{-1} - z^{-2}}{az^{-2} + c(1-a)z^{-1} - 1} \quad (3.21)$$

while

$$a = \frac{t-1}{t+1}; t = \tan(\omega/2); c = \cos(f). \quad (3.22)$$

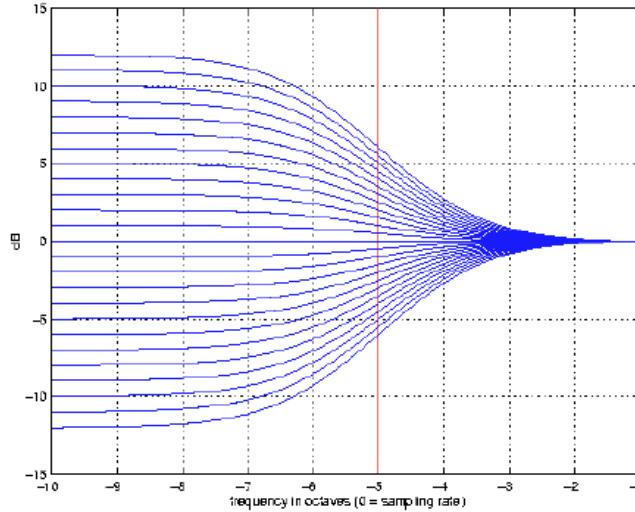


Figure 3.6: Frequency response of the modified shelving filter. [48]

Also here, an asymmetric response between boosting and attenuating can be observed in figure 3.7.

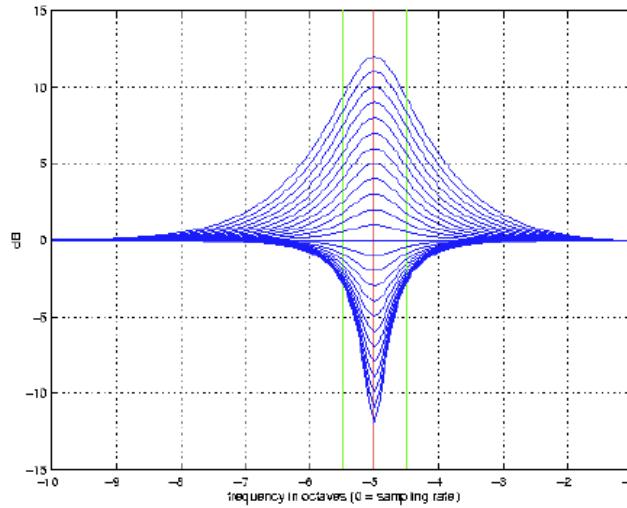


Figure 3.7: Frequency response of the 2nd order presence filter as proposed by Regalia and Mitra. Here gain k is adjusted while keeping bandwidth ω and frequency f constant [48].

Jots modification [48] is then given by imposing the same constraints that are given in 3.19 on equation 3.21. Through simplifying the resulting term, the transfer function $\tilde{H}_2(z)$ of the Proportional Presence Equalizer becomes

$$\tilde{H}_2(z) = \frac{(t\sqrt{k} + 1) - 2cz^{-1} - (t\sqrt{k} - 1)z^{-2}}{(t/\sqrt{k} + 1) - 2cz^{-1} - (t/\sqrt{k} - 1)z^{-2}}. \quad (3.23)$$

In the same way, it can be observed that this also leads to a symmetric response for the presence filter while boosting or attenuating, as can be seen in figure 3.8, and also here self similarity between -10 and 10 dB is observable.

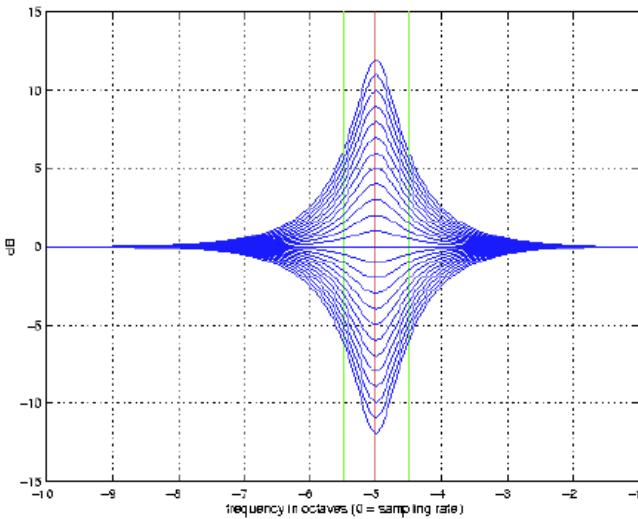


Figure 3.8: Frequency response of the modified presence filter [48].

These properties make it possible to cascade several filters of the same type within the feedback path of a FDN and to faithfully adjust the frequency-dependent decay time very precisely, while maintaining a stable system, because the magnitude response of one filter will only marginally leak into its neighbour within the range between -10 and 10 . Additionally, the self similarity property makes it possible to predict the filters behaviour in a range between -10 and $10dB$, by just considering the change within $1dB$.

3.2.4 Parametric Filter Optimization

A method proposed by Schlecht [50] suggests a optimization method for designing the parameters for the previously introduced proportional parametric filters in the context of an FDN, according to a given target reverberation time for different frequency bands. This requires that the EDR analysis introduced in chapter 3.1.2 is sub-band filtered and therefore provides the T_{60} values for each sub-band whose center frequency corresponds to the center or cutoff frequency of each filter in Schlecht's parametric filter-chain.

Schlecht formulates a 2nd order low-shelf filter $H_{LS,g}(z)$ and high-shelf filter $H_{HS,g}(z)$ respectively as

$$H_{LS,g}(z) = g^{1/2} \frac{p_0 + p_1 z^{-1} + p_2 z^{-2}}{q_0 + q_1 z^{-1} + q_2 z^{-2}} \quad (3.24)$$

$$H_{HS,g}(z) = g / H_{LS,g}(z), \quad (3.25)$$

with

$$\begin{aligned} p_0 &= g^{1/2} \Omega^2 + \sqrt{2} \Omega^{1/4} + 1 \\ p_1 &= 2g^{1/2} \Omega^2 - 2 \\ p_2 &= g^{1/2} \Omega^2 - \sqrt{2} \Omega g^{1/4} + 1 \\ q_0 &= g^{1/2} + \sqrt{2} \Omega g^{1/4} + \Omega^2 \\ q_1 &= 2g^{1/2} \Omega^2 - 2 \\ q_2 &= g^{1/2} - \sqrt{2} \Omega g^{1/4} + \Omega^2, \end{aligned} \quad (3.26)$$

where g depicts the gain which lies at DC ($\omega = 0$) for the low-shelf case $H_{LS,g}(z)$ and at Nyquist frequency ($\omega = fs/2$) for the high-shelf case $H_{HS,g}(z)$; and where $\Omega = \tan \omega_c/2$ and ω_c is the cutoff frequency in radians.

Although this is not the same shelving filter as introduced in [48], it still obeys to the proportionality property, as the same constraints are introduced.

The 2nd order peak-notch filter $H_{PN,g}(z)$ is formulated by Schlecht in a slightly altered way, which is however mathematically equivalent to Jot's formulation in [48]:

$$H_{PN,g}(z) = g^{1/2} \frac{p_0 + p_1 z^{-1} + p_2 z^{-2}}{q_0 + q_1 z^{-1} + q_2 z^{-2}} \quad (3.27)$$

with

$$\begin{aligned} p_0 &= g^{1/2} + g \tan(B/2) \\ p_1 &= -2g^{1/2} \cos(\omega_c) \\ p_2 &= g^{1/2} - g \tan(B/2) \\ q_0 &= g^{1/2} + \tan(B/2) \\ q_1 &= -2g^{1/2} \cos(\omega_c) \\ q_2 &= g^{1/2} + \tan(B/2), \end{aligned} \quad (3.28)$$

where B is the bandwidth of the filter which can be scaled using the quality factor Q through the relation $B = \frac{\omega_c}{Q}$.

Schlecht suggests to construct the complete parametric equalizer by cascading a low-shelf, 8 peak-notch filters and a high-shelf filter, whose cutoff or center frequencies ω_c are spaced logarithmic and have fixed bandwidth- and center/cutoff

frequency for each filter. In this way the magnitude response of the equalizer is only altered through the scaling of the command gains \mathbf{g} , which are given as

$$\begin{aligned}\mathbf{g} &= [g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8]^T \\ \gamma &= 20 \log_{10}(\mathbf{g}),\end{aligned}\tag{3.29}$$

where \log_{10} is applied on each vector element and $(\cdot)^T$ denotes transposition.

As the proportional parametric filters will have an attenuating effect on the total reverb decay which increases with every feedback loop cycle, Schlecht relates the *attenuation-per-sample* $\delta(\omega)$ [50] in dB to the frequency-dependent reverberation time T_{60} as

$$\delta(\omega) = -60 \frac{1}{fs T_{60}(\omega)},\tag{3.30}$$

where fs is the sampling frequency.

The given design paradigm to achieve a target reverberation time is then

$$\alpha(\omega) \approx m\delta(\omega)\tag{3.31}$$

or

$$\alpha(\omega) = 20 \log_{10}|A(\omega)|,\tag{3.32}$$

where m denotes the total amount of delay of the FDN delay lines in samples, and where $|A(\omega)|$ denotes the magnitude response of the proportional parametric filter.

Schlecht formulates a constrained Linear Least-Squares problem in order to approximate the magnitude of the overall parametric filter structure:

$$\gamma MLS = \arg_{\gamma} \min \|B\gamma - \tau\|_2^2\tag{3.33}$$

where B is an interaction matrix containing the sampled magnitude responses for each filter within the parametric filter structure

$$B = 20 \log_{10}|[\mathbf{g}', H_{1,g'}(\omega_p), \dots, H_{L,g'}(\omega_p)]|,\tag{3.34}$$

at a prototype gain specified by \mathbf{g}' and K specifies the spacing of the sampled frequencies, and where

$$\tau = m\delta(\omega_p),\tag{3.35}$$

is a vector containing the frequency-dependent target T_{60} values.

Furthermore, the constraints are given by limiting the range of the command gains γ as

$$-10 \leq \gamma l \leq 10 \text{ for } 1 \leq l \leq L. \quad (3.36)$$

Schlecht additionally provides a problem formulation to optimize the command gains of the filters directly on the resulting reverberation time T_{60} , which is formulated as a constrained Nonlinear Least-Squares problem. The error formulation to be minimized on the resulting reverberation time is then given as

$$\left\| \frac{1}{\alpha} - \frac{1}{m\delta} \right\|_2^2 \quad (3.37)$$

and the objective function γTLS (" T_{60} Least Squares") is given as

$$\gamma TLS = \arg_{\gamma} \min \left\| \frac{1}{B\gamma} - \frac{1}{\tau} \right\|_2^2 = \sum_{k=1}^K \left(\frac{1}{(B\gamma)_k} - \frac{1}{\tau_k} \right)^2. \quad (3.38)$$

Minimizing function 3.33 or 3.38 for γ will in turn, allow to find the vector that contains the right command gain parameters for each filter, which in turn yield the desired target T_{60} for each of the corresponding frequency bands. In principle, this method fulfills the goal of approximating the frequency-dependent decay of a RIR with an FDN reverb. More details on the implementation of this procedure are given in the following chapter.

Chapter 4

Hybrid Reverb Algorithm

This chapter documents both the considerations that led to the final design of the implemented algorithm, and documents the implementation process itself. The goal was to create a hybrid reverb algorithm that is able to use the perceptually important portion of a RIR and complete the reverberation with a synthetic late reverberation algorithm, that will match the decay of the RIR sufficiently well. Alongside the other building blocks it was therefore necessary to prototype a dedicated algorithm for synthetic late reverberation, with properties that allow a sufficient parametrization of the synthetic reverb part, inspired by existing methods and recently published improvements in the field.

4.1 Design Overview

Based on the existing methods presented in section 2.2.3 and chapter 3, the implementation was decided to be comprised of the following processing steps:

1. Early Reflections Section
 - (a) RIR Analysis + Truncation
 - (b) Convolution Processing
2. RIR Analysis Section
 - (a) EDR Analysis
 - (b) Sub-Band T60 Determination
3. Synthetic Late Reverb Section
 - (a) FDN Filter Optimization
 - i. Probe Filters to retrieve Prototype Magnitude Response

- ii. Optimize Filter gains according to target T_{60} times.
- (b) FDN Processing of Input

4. Mixdown of Signals

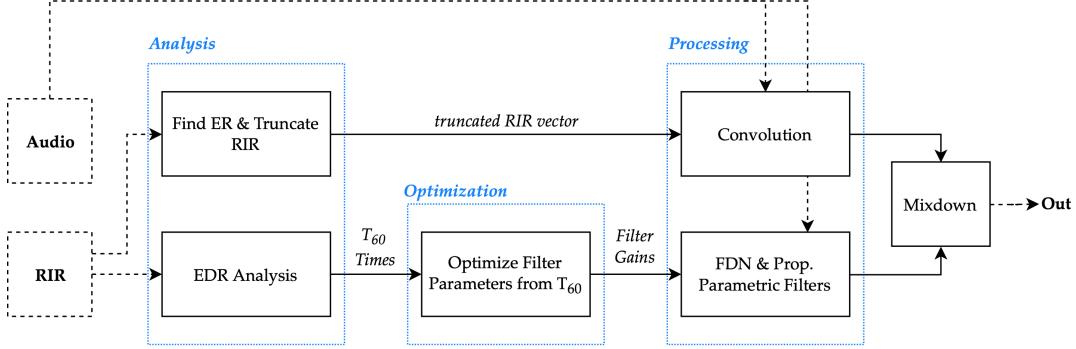


Figure 4.1: High level block diagram of the Hybrid Reverb implementation which sketches out the flow of data.

4.2 Implementation

The implementation was done using the *MATLAB*¹ programming environment with the help of both the *Signal Processing Toolbox*² and *IoSR Toolbox*³ for reverberation decay analysis and DSP processing tasks, and the *Optimization Toolbox*⁴ for the filter optimization in 4.2.4. All operations were done using a global sample rate of 44100 Hz.

4.2.1 Early Reflections

RIR Truncation

The early reflections portion was intended to be modelled through convolution of the input signal with a truncated version of a RIR. In order to find the correct point where the early reflections transited to the late reverberation, the method by Stewart [38], based on standard deviation was utilized. A requirement for the algorithm needed to finding the truncation point was that it should be both sufficiently

¹<https://www.mathworks.com/>

²<https://www.mathworks.com/products/signal.html>

³<https://github.com/IoSR-Surrey/MatlabToolbox>

⁴<https://www.mathworks.com/products/optimization.html>

precise and would require very little computation time. Although other methods were investigated on, this one was found to always determine a perceptually correct transition point, while guaranteeing an almost imminent computation.

Analysis Process

The process itself was implemented by windowing through portions of the RIR at hand with a 20 ms window of rectangular shape with no overlap, and calculating the mean and standard deviation within each window. This window size was chosen according to [7] and is necessary as a smaller window size would not provide a sufficient amount of samples to perform the respective calculations on, and would therefore lead to results regarding both the mean and standard deviation which are inaccurate and not representative. A larger window size however would be prone to determine a truncation point which is too early or too late and would therefore neglect perceptually important information of the RIR, or lead to an early reflections portion which is too long and not representative.

The actual truncation point was found by calculating the percentage of samples within each window that lie outside a standard deviation of 1. A threshold of 30% was set, determining that the window containing at least 30% of samples outside the standard deviation would be the window indicating the correct truncation point, and the RIR could be truncated after that window, according to [7]. In order to not introduce any additional overtones into the RIR, it was truncated by applying the second half of a Hann-window of 32 samples size after the truncation point, in order to smooth it out towards zero energy.

In the case of a stereo RIR, the analysis was only performed on the left channel of the file, and the truncation point was found accordingly.

Convolution Processing

The input sound was processed, by convoluting the truncated RIR and the input signal, using MATLAB's *conv* function.

4.2.2 Spectral Decay Analysis

As the goal for the late reverberation parametrization of the hybrid reverb was to model the frequency-dependant decay of the RIR as closely as possible, an analysis method was required that would faithfully reflect the decay behaviour of the later reverb of an RIR for a specified set of frequency bands. It was found that the method of the Energy Decay Relief (EDR), as described in 3.1.2 would be suitable for this task, as it allows an observation of the decay progression for specified frequencies and therefore a sufficiently precise determination of T_{60} times for each frequency band.

EDR Analysis Procedure

The EDR was implemented by employing a windowed Short-Time Fourier Transform (STFT), using a sliding 30ms window with 75% overlap. In this way it was ensured to capture a sufficient frequency resolution within each window, while maintaining a sufficient time resolution, in order to calculate the EDR precisely. First it was tried to calculate the EDR for 100 logarithmically spaced frequency bands.

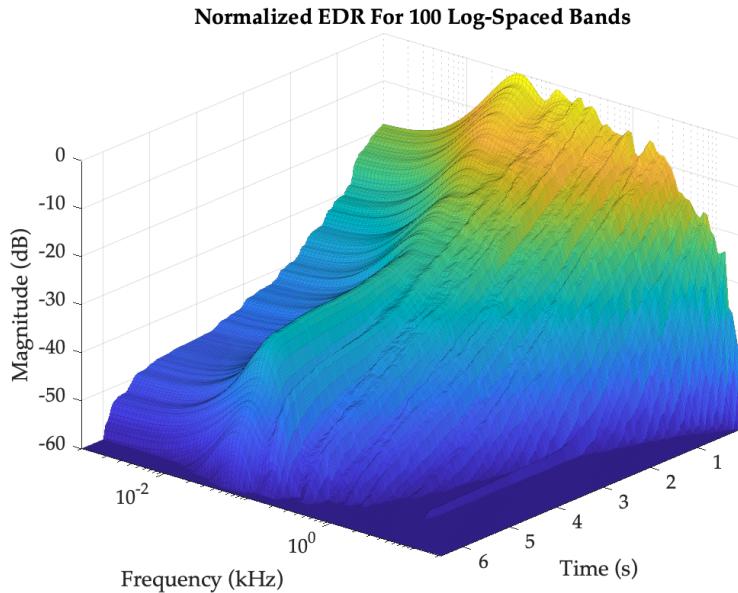


Figure 4.2: EDR plot of the *naveCathedral* RIR, calculated for 100 logarithmically spaced frequency bands.

Although this results in a plot that provides more information about the frequency content of the RIR over time, the resulting data contains a lot of information that might not be of use in the later optimization process that is needed for the FDN parametrization described in 4.2.4. It was therefore decided to only calculate the EDR for the specific bands: 43, 62.5, 125, 250, 500, 1000, 2000, 4000, 8000 and 11360 Hz.

It is to be mentioned that although frequencies 62.5 to 8000 Hz are derived from octave bands, the frequencies 43 Hz and 11360 were chosen for different reasons. A justification for the choice of these frequencies is related to the optimization procedure, and is elaborated on in 4.2.4.

In order to only reflect the meaningful range necessary to determine the actual T_{60} values, the data was normalized to 0dB.

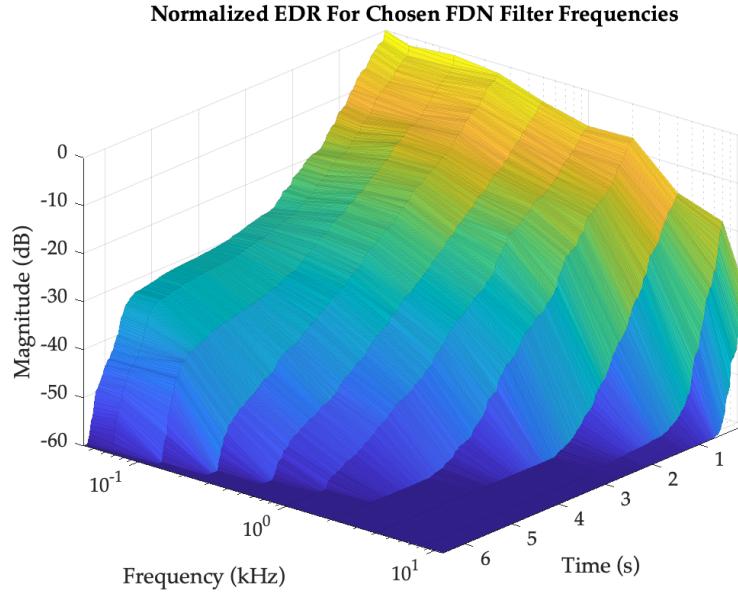


Figure 4.3: EDR plot of the *nave-cathedral* RIR, calculated only for the specified bands.

Sub-Band T_{60} Determination

The actual T_{60} time was determined by finding the point in time where the plotted EDR graph crosses the -60dB threshold for each sub band frequency. This method was suitable because the samples used had a sufficiently good Signal-To-Noise Ratio (SNR). However in a case where the dynamic range of the RIR would have contained too much noise, it would have been more suitable to find the point in time where the graph crosses the -30dB threshold, in order to retrieve the corresponding T_{30} time, and double the retrieved value afterwards, in order determine the T_{60} for each desired band.

It is to be mentioned that, although the hybrid reverberator can be used with stereo RIRs, the EDR analysis was always performed solely on a mono version of the RIR at hand, since the FDN that would later model the reverberation models both channels of the late reverberation, and a stereo image is only created by retrieving the FDN channels in a certain way that enables a stereo listening.

4.2.3 Synthetic Late Reverb

For the late reverberation modelling, a structure was needed that would allow to parametrize the resulting frequency-dependent reverberation sufficiently, while no other structural properties would alter the sound in any way. Additionally it was desired to employ an algorithm structure whose resulting echo density would

increase over time, as established in 2.2.2. For both reasons it was decided to use a Feedback Delay Network (FDN), which was chosen to be lossless, besides the absorptive filters that would be introduced in each feedback path.

FDN Structure

The FDN was chosen to consist of 16 delay lines that are interconnected through a feedback matrix. Each delay line contained a proportional parametric filter bank as suggested in [50] which is placed in the feedback path between the corresponding delay element and the feedback matrix.

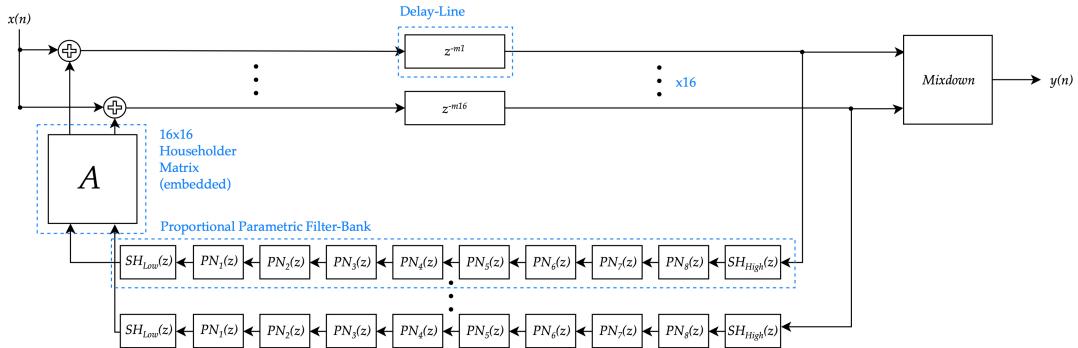


Figure 4.4: Structure of the implemented FDN design.

Matrix

All feedback matrices established in 2.2.2 were implemented for the purpose of investigation. In the end, the chosen matrix for the FDN was a Householder matrix. This decision was made based on empirical listening to the lossless FDN prototype with different matrices, using an impulse as an input signal. It was found that the Householder matrix led to the most even distribution of audible echoes, while also yielding the most "colourless" reverb-tail.

For the case of the 16×16 FDN, the Householder matrix was found as suggested by Jot in [30] through recursive embedding a 4×4 Householder matrix as established in equation 2.6.

Delay Times

Since the delay-times were meant to be hard coded into the FDN and not parameterized afterwards, they were chosen to make the FDN reach sufficient echo density, while not exhibiting any particular resonances. The following delay-times were used in the end:

Delay Line	Delay in ms	Delay in samples @ 44.1
M1	10	441
M2	11.6356	513
M3	13.4567	593
M4	16.734501	738
M5	20.186199	890
M6	25.741699	1135
M7	31.469299	1388
M8	38.294399	1689
M9	46.6838	2059
M10	55.456699	2446
M11	65.175499	2874
M12	76.824303	3388
M13	88.562302	3906
M14	101.278	4466
M15	115.397003	5089
M16	130.501999	5755

Absorptive Filters

As an elaborate absorptive filter solution, it was decided to introduce a Proportional Parametric filter bank as suggested by Jot in [48] and adapted by Schlecht in [50] into the feedback path of each delay line within the FDN. The filter bank was comprised of a 2nd order low- and high shelf filter as established in chapter 3.2.4, using the filter equations 3.24 and 3.25 respectively; as well as 8 2nd order Peak-Notch filters, based on the filter equation 3.27.

The filters where cascaded as can be seen in figure 4.4, and were initialised with the fixed frequencies: 43, 62.5, 125, 250, 500, 1000, 2000, 4000, 8000 and 11360 Hz. These frequencies are identical to the frequencies used for the EDR calculation in 4.2.2, as the purpose of the filter bank is to set the frequency-dependent decay within the FDN reverberator for exactly these frequency bands. A justification for the frequency values follows in section 4.2.4.

FDN Mixdown

As the 16 delay lines of the FDN enable to retrieve 16 uncorrelated outputs from the system, these outputs needed to be mixed down to stereo in a meaningful way, so that the reverberation would spread evenly across the stereo panorama. First it was tried to sum the first 8 FDN channels to the left stereo output and the remaining channels to the right output without further treatment, but this led to a clearly audible phase cancellation between these channels, and to an undesirable auditory

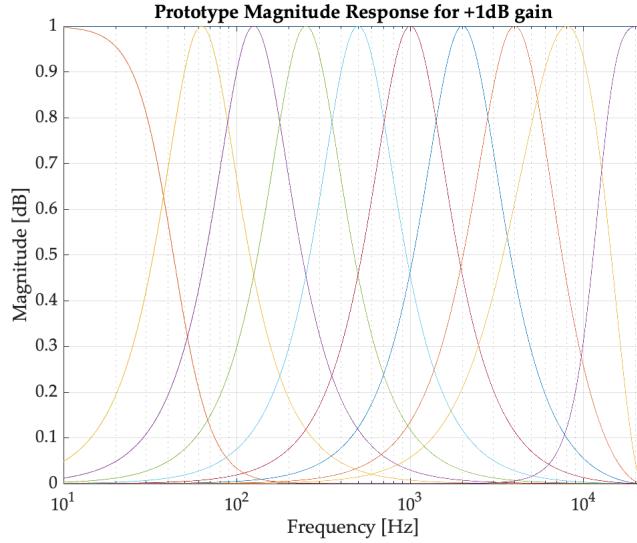


Figure 4.5: Magnitude response of the Proportional Parametric filter bank at +1dB prototype gain.

impression in general. In order to achieve better results, the FDN outputs were mixed down using an Inter Aural Cross Correlation (IACC) matrix, as suggested by Gardner in [51].

The process was implemented by first mixing down the 16 FDN channels to two uncorrelated outputs, by multiplying the mixing matrix:

$$\mathbf{M} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \end{bmatrix}, \quad (4.1)$$

onto the array that holds the 16 channels of audio data:

$$\mathbf{y}_{\text{mixed}}(n) = \mathbf{M}\mathbf{y}_{\text{FDN16}}(n). \quad (4.2)$$

The mixed array was then processed as

$$\begin{aligned} y_L(n) &= \cos(\theta)y_{1,\text{mixed}}(n) + \sin(\theta)y_{2,\text{mixed}}(n) \\ y_R(n) &= \cos(\theta)y_{2,\text{mixed}}(n) + \sin(\theta)y_{1,\text{mixed}}(n) \end{aligned} \quad (4.3)$$

where $y_{1,\text{mixed}}(n)$ and $y_{2,\text{mixed}}(n)$ are the first and second row of the mixed array of uncorrelated outputs respectively, $y_L(n)$ and $y_R(n)$ are the resulting stereo outputs, and θ relates to the IACC coefficient as

$$\theta = \arcsin(IACC)/2; \quad (4.4)$$

where $IACC$ is a number between 0 and 1. For this implementation, an IACC coefficient of 0.5 was chosen.

4.2.4 FDN Filter Parametrization & Optimization

The implemented optimization process is based on Schlecht's optimization approach [50] which adapts Jot's FDN filter model [48] and offers two feasible methods to optimize the gain parameters of the filters, as established in 3.2.4.

It is to be mentioned that since all optimization processes described in [50] are only taking into consideration a FDN with a single delay line, each parametric filter bank within the implemented 16x16 FDN was optimized separately, as each of them yield a different delay time. Additionally to the gain of each filter within the filter bank per delay line, there is an additional broadband gain which is applied on all gains within one delay line as a global scalar. This scalar is also subject to the later optimization.

Interaction Matrix

The interaction matrix B of equation 3.34 was retrieved based on Schlecht's approach in [50], by sampling the magnitude response of the filter-bank at a resolution of $K = 100$ log-spaced control frequencies.

The resulting matrix holds a prototype representation of the magnitude response of each filter within the parametric filter bank. Since the self similarity property discussed in 3.2.3 allows to represent the magnitude response of the filters within a range of -10 to 10 dB through the response of $1dB$ prototype gain, this is possible, as pointed out in [50]. However, this also means that the follow up optimization process can only provide meaningful results when using the constraints in equation 3.36.

As can be observed in figure 4.6 it was found that using octave band related frequencies for the low- and high-shelf filter would produce to big of a gap between the magnitude response of these filters and the peak/notch filters within the interaction matrix.

By instead choosing the frequencies 43 and 11360 Hz for the lowest and highest positioned filter, this effect was avoided, as can be seen in figure 4.7.

T₆₀ Target Matrix

The target vector τ as established in equation 3.35 was retrieved by calculating the attenuation per sample according to equation 3.30 for each T_{60} value that was retrieved from the EDR analysis. Since each delay line within the FDN holds a different delay time, this needed to be done for each delay line separately, resulting in 16 target vectors which were then gathered into a matrix, forming the final target matrix.

Since the EDR analysis was only performed for the 10 center frequencies that correspond the frequencies of the parametric filters within each FDN delay line, this matrix needed to be "stretched out" by interpolating between the values in order to

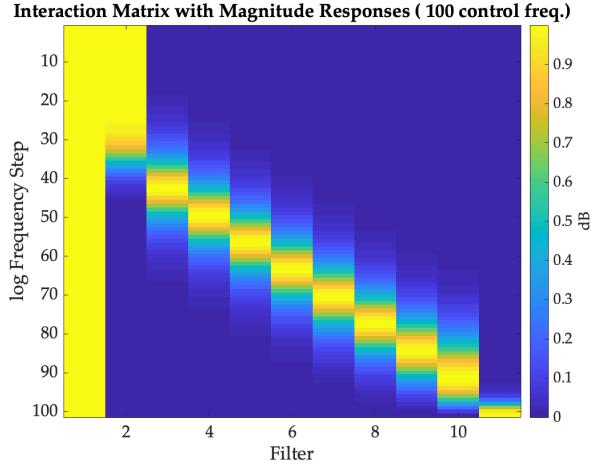


Figure 4.6: Interaction matrix representing the response of each filter within the filter bank at 1 dB gain, using only octave bands as center frequencies.

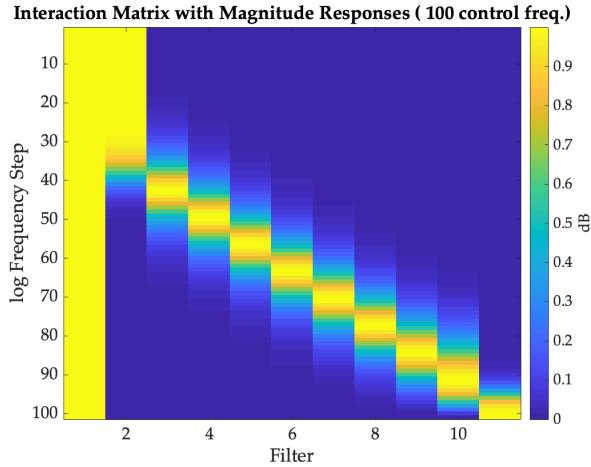


Figure 4.7: Interaction matrix representing the response of each filter within the filter bank at 1 dB gain, using altered frequencies for the shelving filters

provide the attenuation-per-sample times for 100 frequencies instead, so that both the target matrix τ and the interaction matrix B would have the same dimensions. Another solution to this problem would have been to retrieve the EDR at 100 frequencies straight away. However this led to a considerable increase of the EDR computation time and to less reliable results in the later optimization procedure.

Optimization on Magnitude Response (MLSCon)

The optimization using the Linear Least-Squares solution established in equation 3.33 was implemented by supplying the interaction matrix B the target τ and the constraints of 3.36 to the *lsqlin* function in *MATLAB*, using default options. The resulting vector γ offers both the gains for each filter within one delay line, as well as the global gain that is applied on all gains within the delay line.

Optimization on target T_{60} Times (TLSCon)

For the optimization via minimization of the Nonlinear Least-Square solution, the *fmincon* function was employed. Additionally to the interaction matrix B the target τ and the constraints of 3.36, the function was also supplied with the gradient in the form of the first derivative of the function in equation 3.38 according to [50], which is given as:

$$\frac{\partial}{\partial \gamma_i} \left\| \frac{1}{B\gamma} - \frac{1}{\tau} \right\|_2^2 = 2 \sum_{k=1}^K \left(\frac{1}{(B\gamma)_k} - \frac{1}{\tau_k} \right) \left(\frac{-B_{ki}}{(B\gamma)_k^2} \right) \quad (4.5)$$

4.2.5 Mixing Early and Late Reverb

In order to mix the outputs of the convolution and the modelled FDN reverberation meaningfully and without energy loss between both reverbs over time, both sounds were cross-faded using the first and second half of a 64ms Hann window as a cross-fading curve.

Chapter 5

Evaluation

5.1 Signal Comparison

In order to investigate on the performance of the implementation when modelling different types of RIRs, three different RIRs were used for testing¹:

1. **smallRoom** - An impulse response retrieved from the *Small Room* preset of the *Bricasti BM7* digital hardware reverberator, which is the recreation of a small and bright sounding room.
2. **mediumChamber** - An impulse response retrieved from the *Medium Chamber* preset of the *Bricasti BM7* digital hardware reverberator, which is the recreation of a medium sized concert chamber.
3. **naveCathedral** - A real-world recorded RIR that was done within one nave of the *Cathedral Church of Our Lady and St Philip Howard* in Arundel, Great Britain.

5.1.1 T_{60} Comparison

The global T_{60} times of the modelled RIRs were calculated according to the ISO standard 3382-1:2009 as specified in 3.1.2.

RIR	Original RIR	MLSCon Modelled RIR	TLSCon Modelled RIR
smallRoom	0.7962	0.4857	0.3465
mediumChamber	1.2851	1.1558	0.8969
naveCathedral	6.3880	2.7214	2.6898

¹All RIRs were taken from the impulse response library of the *Convolution Reverb Pro*, which is part of the *Ableton Live* software (<https://www.ableton.com/en/packs/convolution-reverb/>), in order to ensure a sufficient quality standard.

The results shown in the presented table indicate that both methods are undershooting the global T_{60} of the original RIR severely in the case of the *naveCathedral* and the *smallRoom* RIR. In the case of the *mediumChamber* RIR, the target T_{60} is much better approximated, at least for the MLSCon method. In general it can be seen that besides the general tendency of both methods to undershoot the global T_{60} , the MLSCon method outperforms the TLSCon method.

5.1.2 Time Domain Comparison

Taking a look at the plot of the overall progression of amplitude over time shows that all modelled RIRs are exhibiting a clear tendency to actually overshoot the signal according to their overall signal energy. This is especially evident for the *mediumChamber* RIR in figure 5.1 and even more for the *naveCathedral* in figure 5.2

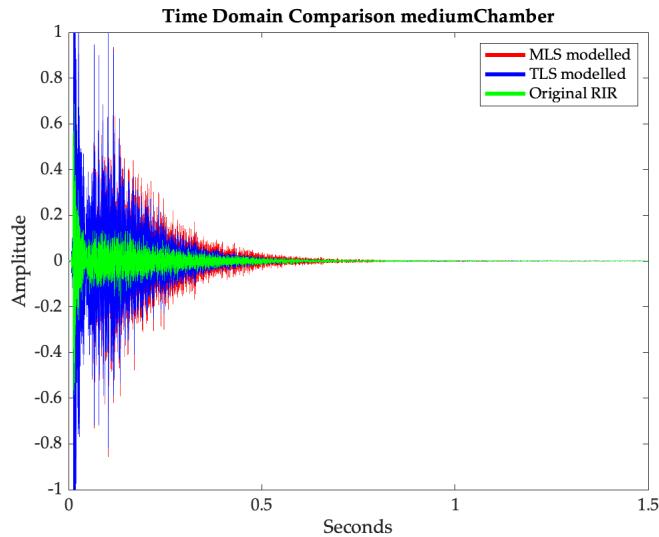


Figure 5.1: Amplitude plot of the *mediumChamber* RIR, comparing the MLSCon (Linear) and the TLSCon (Nonlinear) optimization results.

As can be observed or the *smallRoom* RIR in figure 5.3, both modelling methods overshoot less in comparison to the previous examples, however the difference is still clearly evident.

In general it can be observed that in all cases, the MLSCon method led to a slightly higher amplitude profile, although the difference is only marginally visible. As a positive remark it can be seen that, although the modelled RIRs are exhibiting a clear overshoot in amplitude in all cases, the shape of the decay curves of the original are still matched. This is clearly visible by comparing e.g. *mediumChamber* in figure 5.1 where the decay progression of the original RIR and the modelled

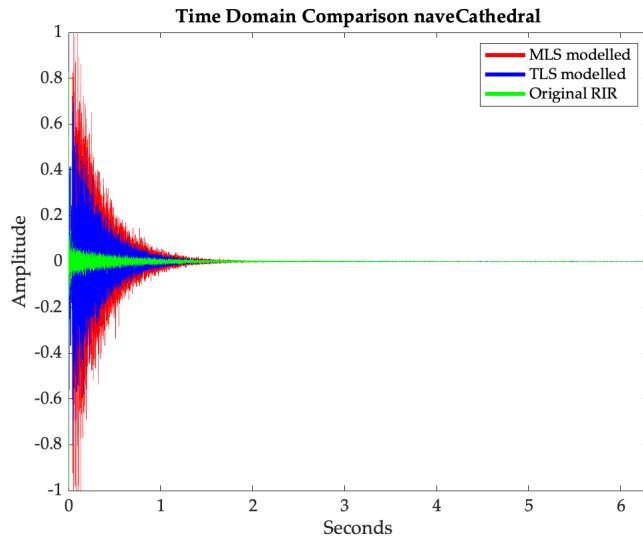


Figure 5.2: Amplitude plot of the *naveCathedral* RIR, comparing the MLSCon (Linear) and the TLSCon (Nonlinear) optimization results.

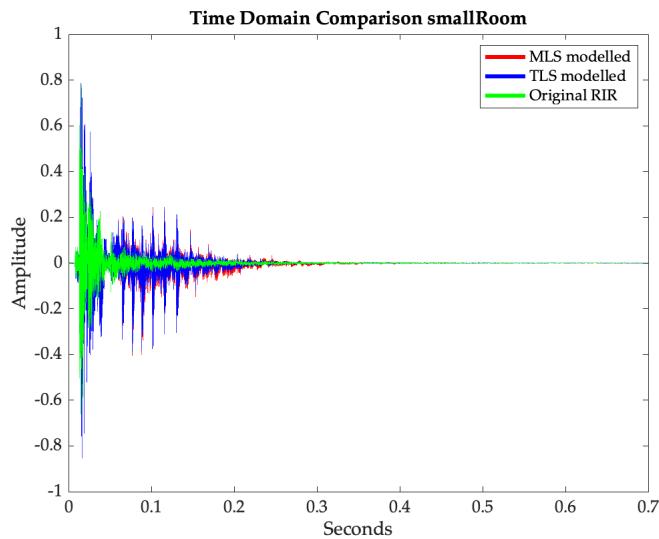


Figure 5.3: Amplitude plot of the *smallRoom* RIR, comparing the MLSCon (Linear) and the TLSCon (Nonlinear) optimization results.

RIRs exhibit the same irregularities and peaks, to the *naveCathedral* plot in figure 5.2 where the decay curves of both the original and the modelled case are rather smooth.

5.1.3 Spectral Comparison

The general observation that the overall amplitude profile of the modelled RIRs is overshoot compared to the original RIR in any case, while the global T_{60} is undershot, seems contradicting at first. This can however be backed by looking at the spectral content of the decay.

The following spectral plots have been calculated at a window size of 2048 samples.

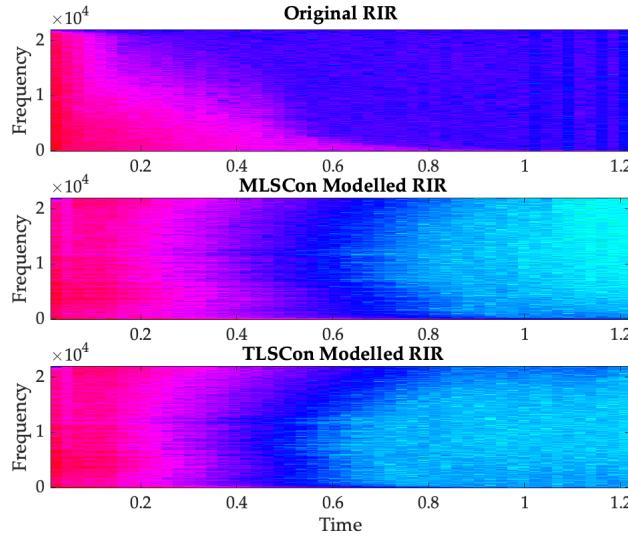


Figure 5.4: Spectrogram plot of the *smallRoom* RIR, comparing the MLSCon (Linear) and the TLSCon (Nonlinear) optimization results.

Figures 5.4, 5.5 and 5.6 show that all modelled RIRs exhibit more energy in the high frequency range from the start, and also the higher frequencies take much longer to decay, compared to the original RIRs. This also offers an explanation why the overall energy of the modelled RIRs is higher, as the initial energy is too high in all frequencies.

However, the spectral plots also exhibit a clear discrepancy between the MLSCon and the TLSCon method, being that for the TLSCon method the higher frequencies have a strong tendency of having a longer decay time than the mid frequencies, sometimes even as long as the low frequency range. This is especially evident for the *smallRoom* plot in figure 5.4 and the *mediumChamber* in figure 5.5, where the decay of the high frequencies is almost as long, or even longer than the lower frequencies. This is an undesirable outcome, as the original RIRs always have clear order of decay, where the highest frequencies decay the fastest. The closest fit to the original RIR frequency profile can actually be observed for the MLSCon version of the *naveCathedral* RIR in figure 5.6.

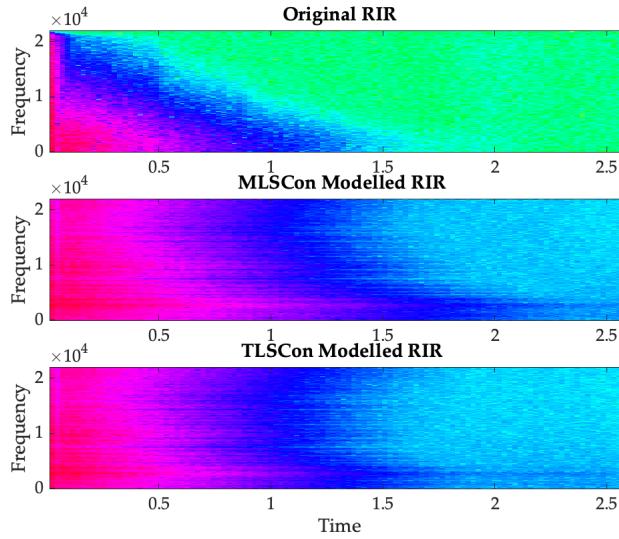


Figure 5.5: Spectrogram plot of the *mediumChamber* RIR, comparing the MLSCon (Linear) and the TLSCon (Nonlinear) optimization results.

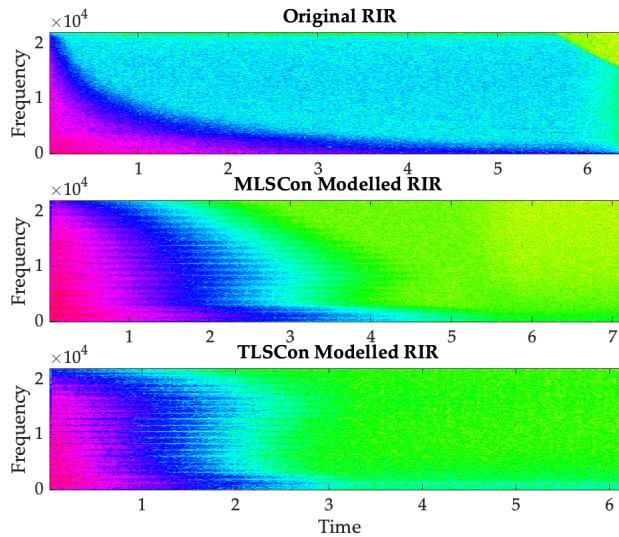


Figure 5.6: Spectrogram plot of the *naveCathedral* RIR, comparing the MLSCon (Linear) and the TLSCon (Nonlinear) optimization results.

The spectral examination shows that with regards to the spectral content, the MLSCon method outperforms the TLSCon method, although just marginally. In general, both modelled RIRs for each example still show a considerable difference

to the original.

5.1.4 Remarks

As both the differences in global T_{60} times, as well as the spectral decay profile between MLSCon and TLSCon method showed that the MLSCon method is modelling the original RIRs slightly better in all cases, it was decided to only use the modelled RIRs retrieved through the MLSCon method for further testing.

5.2 Subjective Listening Test

A listening test was carried out in order to investigate how well the implemented algorithm performs when modelling a certain RIR. It was decided to conduct an APE test, based on the webAudioEvaluation Toolbox by De Man and others [52]. This test form was chosen because it was developed for a related purpose in [53] and was found to enable the participant to assess multiple samples against each other, while also comparing them to a global reference. The Multiple Stimuli With Hidden Reference and Anchor (MUSHRA) [54] was also considered but was decided against, as it only assesses each stimulus to the reference and offers a scale with lower resolution than APE.

In the chosen setup, the implementation of this work should be compared against the more rudimentary hybrid reverb implementation of Stewart [34], in order to determine if the parametric filters in combination with an optimization procedure yield a considerable improvement in comparison to an older method.

5.2.1 Choice of RIR and Stimuli

The stimuli for the test were generated by reverberating 4 different sounds that were chosen so that they would contain a clear transient alongside some rich material in different frequency ranges but with rather short decays. This was expected to excite each reverb structure with meaningful frequencies and to give the participants enough time after ceasing of the direct sound to listen to the actual reverb tail. The following sounds were reverberated at 50% Wet setting:

1. **Drum** - A dry recording of a single hit on a snare drum.
2. **Guitar** - A riff of short strokes of an electric guitar, which was recorded clean without any additional distortion or effect.
3. **Impulse** - The RIR in raw form, only containing a Dirac impulse as an input signal.
4. **Vox** - A dry recording of a male voice, speaking the words "the mind is space."

5.2.2 Test Design

These sounds were reverberated both with the implemented hybrid reverb and Stewart's algorithm from [34]. The APE test scheme also requires to include a hidden reference and hidden anchor. For the hidden reference, the 4 test sounds were reverberated with the original RIRs through convolution. The anchor on the other hand is required to clearly exhibit the worst sound quality and was therefore generated by reverberating each sound using a Schroeder reverb [1] with randomized delay-times and a decay time that would roughly match each of the original RIRs. Pilot tests revealed that participants would find the test confusing when not having an exposed reference available. Therefore, additionally to the hidden reference the participant was presented with an exposed reference in the form of a clickable button which would play the sound at hand convoluted with the original RIR.

All reverb samples were presented in mono, so that the participant would judge the quality of the presented reverbs primarily based on the frequency content of the reverb tail, and the results would not be influenced due to differences in stereo width. The test was divided into 12 parts, since the 3 different room types (*smallRoom*, *mediumChamber*, *naveCathedral*) were tested each with 4 different stimuli. On each page it was only one room type being presented and one sound being reverberated through the 4 different methods: anchor (Schroeder reverb), reference (convolution), the implemented hybrid reverb and Stewart's algorithm.

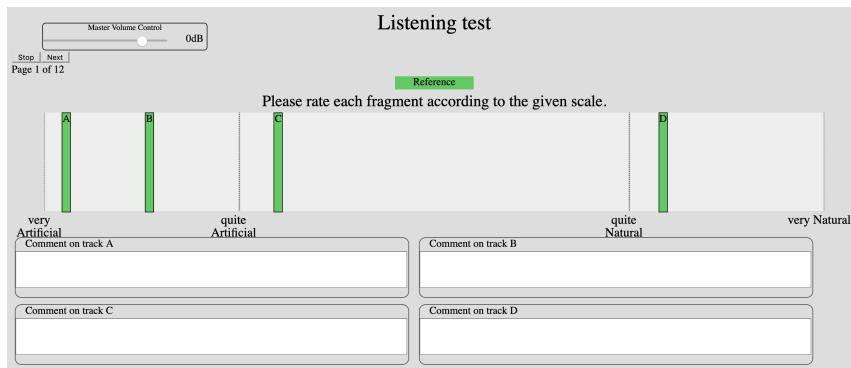


Figure 5.7: The design of the APE test.

The design of the APE test for each page can be seen in 5.7. The participant would be presented with the 4 samples, each represented through green bars that would play the sound when clicked, and that should be dragged around on a scale ranging from "very Artificial" to "very Natural", in order to rate each sample with respect to the exposed reference. Additionally, participants could leave comments for each sound on the page, which was however not mandatory. The participant

was not allowed to proceed to the next page, before the reference was played and each fragment on the scale was played and moved at least once. The initial position and labels of all samples, as well as the order of testing pages was randomized for each participant, so that the resulting data would not be skewed because of a rigid testing order.

5.2.3 Participants

The test was conducted on 21 participants, aged between 26 and 45 years; of the product development division at Ableton AG office in Berlin, Germany. All participants had experience in the development of audio software and were in good health at the day of testing.

5.2.4 Setup

The test was conducted locally on a *Apple MacBook Pro Retina late 2014*, running *Chrome* and using a *RME Fireface UCX* audio interface in combination with a pair of *Beyerdynamic DT 770* headphones as a listening apparatus.

5.2.5 Issues

Pilot tests revealed the aforementioned issue that the test was confusing when no exposed reference was offered. Additionally, it was first considered to use the attribute *Different/Similar to the reference* for the rating scale. However, this made participants only separate the hidden reference from the rest while neglecting the difference within test samples. This was found to be an undesirable effect which would not fully take advantage of the APE design. It was therefore decided to let the participants judge the presented samples according to the attribute of "naturality" instead, with respect to the exposed reference.

5.3 PEAQ

In addition to the subjective listening test evaluation, it was decided to perform a Perceptual Evaluation Of Audio Quality (PEAQ) test [55], using the simplified algorithm by Kabal [56]. The framework uses a psycho-acoustic approach by modelling the human hearing and enables to compare a test signal to a given reference. The output variables of the used model are mapped through a neural network, resulting into an objective difference grade (ODG) which reaches from 0 "imperceptible" to -4 "very annoying". The full range of these values are explained in table 5.1.

Ideally the reference at hand contains the best version of the signal in question, and was decided to be the original RIR in this case. Only the actual modelled

Descriptor	ODG value
Imperceptible	0
Perceptible, but not annoying	-1
Slightly annoying	-2
Annoying	-3
Very annoying	-4

Table 5.1: Table representing the descriptors of the ODG scale used in the PEAQ test.

impulse responses of the implemented hybrid reverb and the Stewart implementation for each of the room types *smallRoom*, *mediumChamber* and *naveCathedral* were compared to respective reference RIRs. In this way an objective measure of the subjective difference between the implementation and Stewart's algorithm could be established. All samples were upsampled to a samplerate of 48.000Hz in order to be processed through the framework.

Chapter 6

Results and Discussion

The following chapter presents the results of the tests and further discussion.

6.1 Listening Test results

In this section the statistical results of the ratings for each modelled RIR and each of the 4 stimuli type will be discussed. The null hypothesis for the evaluation was that there is no perceivable difference between the different methods of modelling the RIR. In order to validate this a one-way Analysis of Variance (ANOVA) was performed on the retrieved test data per modelled RIR, using an $\alpha = 0.05$. The results can be seen in A, yielding the following p-values: $P_{smallRoom} = 1,29274e - 16$, $P_{mediumChamber} = 1,33528e - 16$, $P_{naveCathedral} = 9,97952e - 22$, indicating that the differences in ratings are significant enough to reject the null-hypothesis.

6.1.1 smallRoom

The box plot in figure 6.1 shows that the anchor was mostly identified for the case of the *Drum* with a mean rating of 0.18 and even more evidently for the *Impulse* sample with a consistent mean rating below 0.1. However it was not easily identified for the *Guitar* with a mean of 0.42 and the *Vox* sample with a mean of 0.32, both exhibiting skewness towards higher values and a very high standard deviation. The plot also shows that the reference was clearly rated the highest in most cases, exhibiting a mean rating between 0.85 and 0.92 for all stimuli, although the standard deviation for the *Drum* and the *Impulse* is rather high.

The plot only shows a clearly higher mean rating for the implemented method in comparison to Stewart's algorithm when listening to the reverb with the *Vox* sample, exhibiting a mean of 0.58. For the *Guitar*, the mean score is actually higher 0.62, whereas the standard deviation is rather large, indicating some ambiguity in the ratings. For the *Drum* sample only a marginal difference in favor for the im-

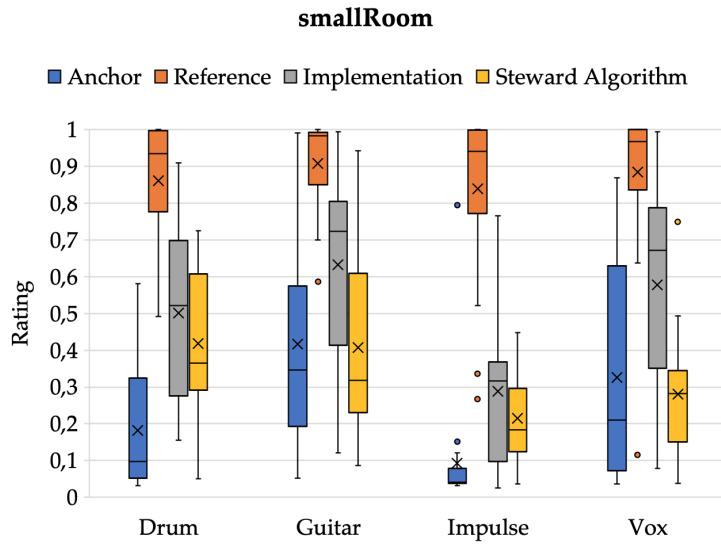


Figure 6.1: Boxplot showing the ratings from the APE test for the *smallRoom* RIR.

plementation is visible and the *Impulse* signal finally led to the consistently lowest rating for both hybrid reverb methods.

In general a significant favorization for the implementation can only be observed for the *Vox* sample. The *Impulse* sample exhibited the lowest rating for both methods and a significant distance to the reference, suggesting that its quality was generally perceived as bad.

6.1.2 mediumChamber

The anchor was clearly identified for the *Drum* sample with a mean of 0.08 and *Impulse* sample with a mean below 0.01 and exceptional narrow distribution. It was however not rated as low for the *Guitar* with a mean of 0.32 and the *Vox* sample with a mean of 0.025, while both samples exhibit a rather high standard deviation and skewness towards higher values, indicating that participants had problems hearing clear differences between anchor and the tested implementations. The reference was rated highest in most cases, with the *Drum* always clearly exposing the reference.

A significant favorization for the implementation can only be seen for the *Impulse* sample with a mean rating of 0.51, compared to the Stewart algorithm with a rating of 0.27 and narrow distribution of the Stewart ratings. However, only a marginal difference is evident for the rest of the samples. It is visible that the implementation led to a generally broader distribution than the Stewart algorithm, indicating that participants could not hear a clear difference between both methods.

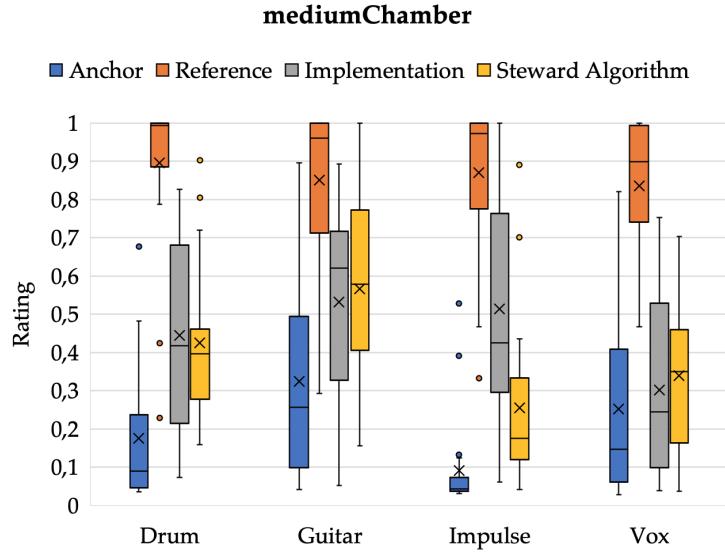


Figure 6.2: Boxplot showing the ratings from the APE test for the *mediumChamber* RIR.

when listening to all stimuli except the *Impulse*.

6.1.3 naveCathedral

The anchor was clearly identified for the *Drum* and *Impulse*, receiving consistent ratings below 0.2. Although *Guitar* and *Vox* show a mean below 0.2, both samples exhibit a skewness towards higher ratings. The reference was identified for all stimuli, although the rating range is the largest for the *Guitar* sample, exhibiting a lot of data points below the mean.

Regarding the difference between implementation and Stewart's algorithm it can be observed that there was no significant difference perceived between both methods for the *Drum* and the *Guitar* sample. In fact the latter exhibits the largest ambiguity in rating for this RIR, as the range of ratings is the largest for both methods and the reference. This could indicate that the *Guitar* sample was actually making it very hard to rate this RIR at all.

The clearest difference in favor for the implementation can be observed for the mean of the *Impulse* as an input signal, scoring a mean of about 0.45 compared to Stewart's algorithm with 0.25. The *Vox* sample exhibits a marginal higher mean in favour for the implementation, whose mean is almost the same as for the *Impulse*.

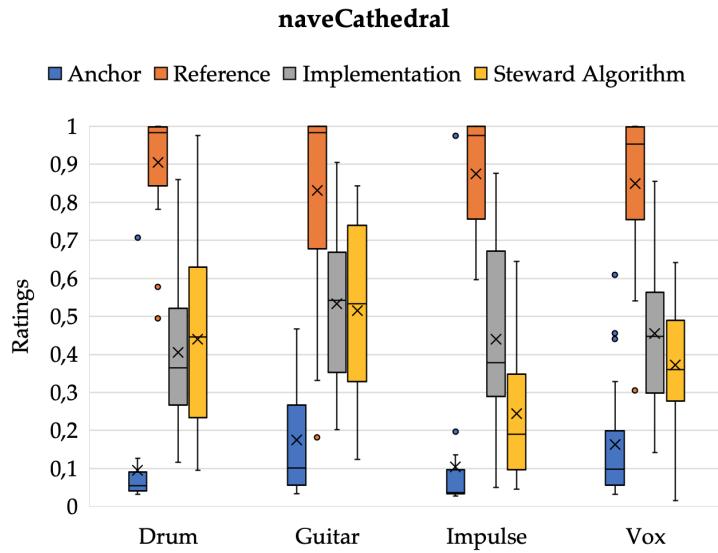


Figure 6.3: Boxplot showing the ratings from the APE test for the *naveCathedral* RIR.

6.2 PEAQ Results

The results of the PEAQ test can be seen in table 6.1 and exhibit a rather low rating in all cases. Although the results are slightly in favor for the implementation in the case of the *smallRoom* and the *mediumChamber* RIR, and slightly better for the Stewart algorithm in the *naveCathedral* case, all lie within the lowest quarter of the ODG scale between "annoying" and "very annoying". Only the *smallRoom* RIR scored better than the other modelled RIRs, receiving the rating "annoying".

	Implementation	Stewart Algorithm
smallRoom	-3.0960	-3.1649
mediumChamber	-3.7912	-3.8945
naveCathedral	-3.9075	-3.8949

Table 6.1: Table representing the resulting ODG values of the PEAQ test.

6.3 Discussion

The listening test results show that although the null hypothesis could be rejected, there can not be a clear preference be observed for the implementation in comparison to the more rudimentary implementation by Stewart for every stimuli. In

fact, quite a big degree of ambiguity can be observed. A possible explanation for ambiguity for the *smallRoom* could be the rather short decay of the reverb samples in all cases, as it does not give a lot of time to listen to and judge the reverb. Furthermore the *Impulse* sample clearly revealed that participants did not find both hybrid reverb methods particularly well performing in the case of the *smallRoom* RIR. The *mediumChamber* case exhibited the largest ambiguity in ratings for all stimuli, except for the *Impulse* sample which led to a clear preference of the implementation. A similar effect was observed for the *naceCathedral* case, where the *Impulse* stimulus led to the clearest preference for the implementation. As the Dirac impulse is often considered the most sonically revealing way of listening to a reverb, as it leads to an impulse response that fully characterizes the system, it makes sense that it was also giving the most distinguished results.

However, the rather large distance of the ratings for the implementation to the reference in most cases show, that the investigated hybrid reverb does not perform sufficiently well when modelling the investigated RIRs in general. This finding is backed through the results of the PEAQ test in section 6.2, which classifies all modelled RIRs in a range between "annoying" and "very annoying" on the ODG scale.

A possible explanation for this result could be given by looking at the signal comparison in section 5.1. It shows that the implementation is not modelling the tested RIRs sufficiently well, exhibiting not only shorter T_{60} times but also an over-emphasis of higher frequencies which clearly deviate from the original RIRs and leads to an unnatural reverberation overall. However it can already be seen that the decay of the energy- and frequency-profile of the original RIRs is at least approximated, leaving room for improvements. A possible way of improving on this would be to retain the higher frequencies in the starting phase of the reverberation in a meaningful way, so that the reverberation tail would start off with less energy in the higher frequencies.

The problem of undershooting the T_{60} times of the target RIRs on the other hand could be explained with the fact that Schlecht's applied optimization procedure in [50] does not take into account what happens when multiple delay-lines whose parametric filter banks are all optimized individually, are interconnected in a matrix. As the matrix will make the signal of one delay-line eventually travel through all other delay-lines and therefore undergo further absorption in a shorter time-frame, the overall decay of the reverberation is shortening. A possible solution to this could be to use a diagonal matrix, although this would limit the degree of density the late reverberation can reach. Further elaboration on possible improvements are given in 7.1.

6.3.1 Perspectives On Listening Test Design

Assessing the subjective quality of reverberation is not an easy task in general. Several pilot tests before the final test revealed that the test design needs to be clear enough so that participants are not confused by what is asked of them. Furthermore, when exciting the same reverb with different sounds that all have different frequency content, the sonic outcome can differ severely, which can be seen in the difference in results for the different samples when assessing the same target RIR. Using a scale with descriptors ranging between *very artifical* and *very natural* might have added in to the problem when assessing the reverb with the *Impulse* sample, as listening to the impulse response is the most coherent way to listen to a reverb, but also the most unnatural way.

Chapter 7

Conclusion

This report describes the implementation of a hybrid reverb algorithm that uses statistical analysis of room impulse responses in combination with convolution for the modelling of early reflections and a synthetic model for the late reverberation which is optimized from the time frequency analysis of a room impulse response, using the *MATLAB* programming environment.

Background theory for the natural occurrence of reverberation and the canonical structures needed to produce artificial reverberation in the digital domain was covered. Furthermore, several related methods for the analysis of reverberant systems, statistical analysis of room impulse responses and synthetic reverb parameter optimization were covered.

A hybrid reverb design was suggested and implemented, using statistical analysis and the Energy Decay Relief method for the analysis of impulse responses, and both a Linear and Nonlinear Least Squares optimization procedure for the optimization of the synthetic reverb parameters from room impulse responses.

The performance of the implementation was evaluated by modelling various room impulse responses and assessing the resulting signals. Furthermore, a subjective listening test in the form of an APE test was conducted, which compares the resulting modelled room impulse responses of the implementation to a more rudimentary hybrid reverb method. To complement the subjective evaluation, an objective evaluation was conducted using the PEAQ test method to assess the quality of the modelled impulse responses with respect to the originals.

The results show a minor preference for some stimuli towards the implementation in comparison to the rudimentary hybrid reverb algorithm in the listening test, but the results are weakly correlated. Both the results of the listening test and the PEAQ evaluation show that the quality of the modelled room impulse responses is overall not satisfactory in relation to the original impulse responses, leaving room for possible improvements in the future.

7.1 Future Work

A possible improvement to the algorithm could be to introduce a spectral correction filter before the input of the FDN structure in the late reverb structure, similar to the tonal correction filter proposed by Jot in [32], which has the same structure as the parametric filter bank within the FDN. Parametrizing this filter-bank correctly would allow to pre-filtered the input to the FDN so that the frequency profile of the late reverberation starts more accurately, according to the frequency profile of the original room impulse response that is being modelled. Essentially, this spectral correction filter bank could be parameterized using the magnitude profile of the first couple of STFT frames of the original impulse response.

Furthermore, taking into account the fact that the delay lines are interconnected within a matrix, a possible solution could be to use a diagonal matrix and re-evaluate the results.

Another improvement arises from the perspective of the delay times within the FDN. A fixed set of delay times was used in this implementation, which was only justified because they led to colourless results in the lossless case of the FDN. As an FDN is an IIR system and changing the delay times will change the frequency characteristics of the FDN, a meaningful method to parameterize these delay times could also be a way to match a given room impulse response in a better way. A possible strategy for this approach could be given in Schlecht's recently published work regarding the modal decomposition on FDNs in [57].

Bibliography

- [1] Manfred R. Schroeder. "Natural sounding artificial reverberation". In: *Journal of the Audio Engineering Society* 10.3 (1962), pp. 219–223.
- [2] Anders Gade. "Acoustics in halls for speech and music". In: *Springer handbook of acoustics*. Springer, 2007, pp. 301–350.
- [3] Vesa Valimaki et al. "Fifty Years of Artificial Reverberation". In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.5 (July 2012), pp. 1421–1448.
- [4] Heinrich Kuttruff. "Reverberation and steady-state energy density". en. In: *Room Acoustics, Sixth Edition*. 6th ed. London & New York: CRC Press, Sept. 2016, pp. 103–124.
- [5] David R Moore and Andrew J King. "Auditory perception: The near and far of sound localization". In: *Current Biology* 9.10 (May 1999), pp. 361–363.
- [6] David M. Howard, Jamie Angus, and Jamie Angus. "Chapter 1: Introduction to Sound". en. In: *Acoustics and Psychoacoustics*. 5th ed. Routledge, June 2017.
- [7] Rebecca Stewart. "Hybrid Convolution And Filterbank Artificial Reverberation Algorithm Using Statistical Analysis And Synthesis". en. MA thesis. York, England: The University of York, 2006.
- [8] Jean-Marc Jot, Laurent Cerveau, and Olivier Warusfel. "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model". English. In: Audio Engineering Society, Sept. 1997.
- [9] Barry A. Blesser. "An Interdisciplinary Synthesis of Reverberation Viewpoints". In: *Journal of the Audio Engineering Society. Audio Engineering Society* 49 (Oct. 2001), pp. 867–903.
- [10] Leo Beranek. "Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption". In: *The Journal of the Acoustical Society of America* 120 (Oct. 2006), pp. 1399–410.
- [11] Heinrich Kuttruff. "Sound absorption and sound absorbers". In: *Room acoustics*. 6th ed. London & New York: CRC Press, 2016, pp. 125–154.

- [12] Swen Müller and Paulo Massarani. "Transfer-Function Measurement with Sweeps". English. In: *Journal of the Audio Engineering Society* 49.6 (June 2001), pp. 443–471.
- [13] Jukka Pätynen, Brian F.G. Katz, and Tapio Lokki. "Investigations on the balloon as an impulse source". en. In: *The Journal of the Acoustical Society of America* 129.1 (Jan. 2011), pp. 27–33.
- [14] David Griesinger. "Beyond MLS-Occupied Hall Measurement with FFT Techniques". English. In: Audio Engineering Society, Nov. 1996.
- [15] Angelo Farina. "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique". English. In: Audio Engineering Society, Feb. 2000.
- [16] E. Mommertz and S. Müller. "Measuring impulse responses with digitally pre-emphasized pseudorandom noise derived from maximum-length sequences". In: *Applied Acoustics* 44.3 (Jan. 1995), pp. 195–214.
- [17] William G. Gardner. "Efficient Convolution without Input-Output Delay". English. In: *Journal of the Audio Engineering Society* 43.3 (Mar. 1995), pp. 127–136.
- [18] Andrew Reilly and David McGrath. "Convolution Processing for Realistic Reverberation". English. In: Audio Engineering Society, Feb. 1995.
- [19] Frank Wefers. *Partitioned convolution algorithms for real-time auralization*. en. Aachener Beiträge zur technischen Akustik Band 20. Berlin: Logos Verlag Berlin GmbH, 2015.
- [20] Barry D. Kulp. "Digital Equalization Using Fourier Transform Techniques". English. In: Audio Engineering Society, Nov. 1988.
- [21] M. Barron. "The subjective effects of first reflections in concert halls—The need for lateral reflections". In: *Journal of Sound and Vibration* 15.4 (Apr. 1971), pp. 475–494.
- [22] Michael F. E. Barron. "The effect of early reflections on subjective acoustical quality in concert halls". en. phd. University of Southampton, Jan. 1974.
- [23] Manfred R. Schroeder and Benjamin F. Logan. "'Colorless' Artificial Reverberation". English. In: *Journal of the Audio Engineering Society* 9.3 (July 1961), pp. 192–197.
- [24] James A. Moorer. "About This Reverberation Business". In: *Computer Music Journal* 3.2 (June 1979), p. 13.
- [25] Jon Dattorro. "Effect design, part 1: Reverberator and other filters". In: *Journal of the Audio Engineering Society* 45.9 (1997), pp. 660–684.

- [26] Luke Dahl and Jean-Marc Jot. "A Reverberator Based On Absorbent All-Pass Filters". en. In: *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)* (2000), p. 6.
- [27] M. A. Gerzon. "Synthetic stereo reverberation, Part I". In: *Studio Sound* 13 (1971), pp. 632–635.
- [28] M. A. Gerzon. "Synthetic stereo reverberation, Part II". In: *Studio Sound* 14 (1972), pp. 24–28.
- [29] John Stautner and Miller Puckette. "Designing Multi-Channel Reverberators". en. In: *Computer Music Journal* 6.1 (1982), p. 52.
- [30] Jean-Marc Jot and Antoine Chaigne. "Digital Delay Networks for Designing Artificial Reverberators". English. In: Audio Engineering Society, Feb. 1991.
- [31] Sebastian J Schlecht and Emanuël A P Habets. "Sign-Agnostic Matrix Design for Spatial Artificial Reverberation with Feedback Delay Networks". en. In: (2018), p. 11.
- [32] J. Jot. "An analysis/synthesis approach to real-time artificial reverberation". In: *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 2. Mar. 1992, 221–224 vol.2.
- [33] Sean Browne. "Hybrid Reverberation Algorithm Using Truncated Impulse Response Convolution And Recursive Filtering". MA thesis. Coral Gables, Florida: University of Miami, 2001.
- [34] Rebecca Stewart and Damian Murphy. "A Hybrid Artificial Reverberation Algorithm". en. In: (2007), p. 9.
- [35] Jonathan S. Abel, David P. Berners, and Aaron Greenblatt. "An Emulation of the EMT 140 Plate Reverberator Using a Hybrid Reverberator Structure". English. In: Audio Engineering Society, Oct. 2009.
- [36] Keun Sup Lee, Nicholas J Bryan, and Jonathan S Abel. "Approximating Measured Reverberation Using A Hybrid Fixed/Switched Convolution Structure". en. In: (2010), p. 4.
- [37] Jonathan S. Abel, Sean Coffin, and Kyle Spratt. "A Modal Architecture for Artificial Reverberation with Application to Room Acoustics Modeling". English. In: Audio Engineering Society, Oct. 2014.
- [38] Rebecca Stewart and Mark Sandler. "Statistical Measures of Early Reflections of Room Impulse Responses". en. In: (2007), p. 4.
- [39] Guillaume Defrance and Jean-Dominique Polack. "Measuring the mixing time in auditoria". en. In: *The Journal of the Acoustical Society of America* 123.5 (May 2008), pp. 3499–3499.

- [40] A. Primavera et al. "Objective and Subjective Investigation on a Novel Method for Digital Reverberator Parameters Estimation". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.2 (Feb. 2014), pp. 441–452.
- [41] Lothar Cremer and Helmut A. Müller. *Principles and applications of room acoustics*. en. Applied Science, 1982.
- [42] Wallace Clement Sabine. "Reverberation". English. In: *Collected Papers On Acoustics*. Cambridge: Harvard University Press, 1923, pp. 3–68.
- [43] M. R. Schroeder. "New Method of Measuring Reverberation Time". In: *The Journal of the Acoustical Society of America* 37.3 (Mar. 1965), pp. 409–412.
- [44] "Acoustics - Measurement of room acoustic parameters - Part 1: Performance Spaces". Pat. ISO 3382-1:2009. July 2009.
- [45] Julius Orion Smith. *Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects*. en. W3K Publishing, 2010.
- [46] Sebastian J. Schlecht and Emanuel A. P. Habets. "Feedback Delay Networks: Echo Density and Mixing Time". en. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25.2 (Feb. 2017), pp. 374–383.
- [47] William G. Gardner. "A real-time multichannel room simulator". en. In: *The Journal of the Acoustical Society of America* 92.4 (Oct. 1992), pp. 2395–2395.
- [48] Jean-Marc Jot. "Proportional Parametric Equalizers - Application to Digital Reverberation and Environmental Audio Processing". English. In: Audio Engineering Society, Oct. 2015.
- [49] P. Regalia and S. Mitra. "Tunable digital frequency response equalization filters". en. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35.1 (Jan. 1987), pp. 118–120.
- [50] Sebastian J Schlecht and Emanuël A P Habets. "Accurate Reverberation Time Control in Feedback Delay Networks". en. In: (2017), p. 8.
- [51] William G. Gardner. "Reverberation Algorithms". en. In: *Applications of digital signal processing to audio and acoustics*. Boston: Kluwer, 1998, pp. 85–131.
- [52] Nicholas Jillings, Brecht De Man, and Reiss Joshua D. "Web Audio Evaluation Tool: A browser-based listening test environment". In: July 2015.
- [53] Brecht de Man, Kirk McNally, and Joshua Reiss. "Perceptual Evaluation and Analysis of Reverberation in Multitrack Music Production". en. In: *Journal of the Audio Engineering Society* 65.1/2 (Feb. 2017), pp. 108–116.
- [54] "Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems". en. In: *R BS*. (Oct. 2015), p. 36.

- [55] Dinu Câmpeanu and Andrei Câmpeanu. "PEAQ – An Objective Method to Assess the Perceptual Quality of Audio Compressed Files". en. In: *Proceedings of the International Symposium on System Theory* (2005), pp. 487–492.
- [56] P. Kabal. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. Lab Technical Report. Dept. Electrical & Computer Engineering, McGill University, 2002.
- [57] Sebastian J. Schlecht and Emanuël A. P. Habets. "Modal Decomposition of Feedback Delay Networks". en. In: *arXiv:1901.08865 [cs]* (Jan. 2019).

Appendix A

Listening Test Results: One-Way ANOVA

A.1 smallRoom

SUMMARY					
Groups	Count	Sum	Average	Variance	
Anchor	21	3,819517966	0,181881808	0,031336296	
Reference	21	18,06270213	0,860128673	0,03087992	
Implementation	21	10,53477028	0,501655728	0,054769646	
Stewart	21	8,785398897	0,418352328	0,038019677	

ANOVA						
Source of Var.	SS	df	MS	F	P-value	F crit
Between Groups	4,981205899	3	1,660401966	42,84755167	1,29274E-16	2,718784982
Within Groups	3,100110791	80	0,038751385			
Total	8,08131669	83				

A.2 mediumChamber

SUMMARY					
Groups	Count	Sum	Average	Variance	
Anchor	21	3,702215802	0,176295991	0,031451557	
Reference	21	18,81078831	0,895751824	0,040396646	
Implementation	21	9,331275186	0,444346437	0,066008452	
Stewart	21	8,944299066	0,425919003	0,038264946	
ANOVA					
Source of Var.	SS	df	MS	F	P-value F crit
Between Groups	5,652300471	3	1,884100157	42,79089336	1,33528E-16 2,718784
Within Groups	3,522432011	80	0,0440304		
Total	9,174732483	83			

A.3 naveCathedral

SUMMARY					
Groups	Count	Sum	Average	Variance	
Anchor	21	2,001443657	0,095306841	0,020374082	
Reference	21	19,00891582	0,905186468	0,020804749	
Implementation	21	8,527039107	0,406049481	0,041164144	
Stewart	21	9,236757683	0,439845604	0,057865495	
ANOVA					
Source of Var.	SS	df	MS	F	P-value F crit
Between Groups	7,024473681	3	2,341491227	66,80027864	9,97952E-22 2,718784
Within Groups	2,804169413	80	0,035052118		
Total	9,828643094	83			

Appendix B

Listening Test Results: Comments On Samples

B.1 smallRoom

- Anchor

- Drum:
 - "strong echoes"
 - "quite artificial"
 - "too distinct echoes in the tail"
- Guitar:
 - "Sounds like a spring reverb."
 - "strong echoes"
 - "natural"
 - "Sounds more like delay than reverb "
- Impulse:
 - "digital, like an 8 bit machine gone wrong"
 - "very artificial"
- Vox:
 - "Wobbly tail makes it very artificial"
 - "ringing echo tail"
 - "artificial"

- Reference

- Drum:

- "closest to reference"
- "very natural"
- "sounds the same"

– Guitar:

- "Sounds like it's actually being recorded in a semi-reflective room."

– Impulse:

- "reference"
- "very natural"

– Vox:

- "Copy of reference."
- "closest to reference"
- "very natural"

• **Implementation**

– Drum:

- "comb filter"
- "very natural"
- "very bassy"

– Guitar:

- "'Warmer' than A (more low-end carries through)."
- "a bit boxy, bass boost"
- "natural"

– Impulse:

- "Interesting reverb, not very close to the reference"
- "very artificial"

– Vox:

- "Somehow gets bayy at the beginning while the high pitched part comes in afterward."
- "bass boost"
- "lots of bass"
- "very natural"

• **Steward Algorithm**

– Drum:

- "washy, truncated"

- "very artificial"
- "too long and noisy"

– Guitar:

- "Sounds almost like it was recorded in a large hall, but I notice some 'reflective' patterns that seem unreal."
- "a bit washed out"
- "artificial"
- "Quite long tail compared to the reference"

– Impulse:

- "truncated"
- "very artificial"

– Vox:

- "Same as before, long tail makes it sound a little artificial but not really"
- "washy and hissy"
- "artificial"

B.2 mediumChamber

- Anchor

– Drum:

- "Again, the wobble on the reverb tail makes it sound extremely artificial"
- "strong echoes"
- "very artificial"
- "Too discrete in the late echos"

– Guitar:

- "strong echoes in tail"
- "very natural"

– Impulse:

- "Obviously very affected."
- "broken"
- "very artificial"
- "Very grainy"

– Vox:

- "Normal at the beginning, but the tail has a "wobbly" characteristic that doesn't sound natural at all."

- "strong echoes in tail"
- "natural"

• **Reference**

– Drum:

- "Sounds like a copy of the reference"
- "closest to reference"
- "natural"
- "sounds the same"

– Guitar:

- "closest to reference"
- "very natural"

– Impulse:

- "Identical to the reference"
- "closest to reference"
- "very natural"

– Vox:

- "Sounds like an exact copy of the reference."
- "closest to reference"
- "quite natural"

• **Implementation**

– Drum:

- "Again, the reverb sounds higher pitched than the source material. I'm trying to envision a physical space where this would happen, and the best I can think of is inside a plastic pipe? Somehow?"
- "strong high resonance"
- "very artificial"
- "Wrong frequency/time"

– Guitar:

- "high resonance"
- "very natural"

– Impulse:

- "Sounds like a small room."
- "high resonance"
- "lots of high harmonics interesting."

- "natural"
- "Quite metallic"

– Vox:

- "Reverb seems very high pitched, which makes it not sound normal. Almost sounds like hissing."
- "resonant highs"
- "artificial"

• **Steward Algorithm**

– Drum:

- "Extra-long tail makes it sound little unnatural, but I guess it could be in a giant concrete hall or something."
- "truncated, undamped"
- "natural"
- "too much noise in the late"

– Guitar:

- "a bit hissy"
- "artificial"

– Impulse:

- "Sounds 'airy' and full of artifacts"
- "sounds like a cheap digital reverb"
- "gameboy yeah"
- "quite artificial"
- "Hissy but close to natural"

– Vox:

- "It generally sounds fine, but what makes it sound artificial is how long the high-end tail lingers. It decays so slowly that it sounds artificial."
- "a bit washy"
- "artificial"

B.3 naveCathedral

• **Anchor**

– Drum:

- "machine gun echoes"
- "artificial"

- "Delay much?"

– Guitar:

- "shatter verb is usually the term used for such a reverb."
- "Extremely noticeable artifacts."
- "strong echoes."
- "very artificial."

– Impulse:

- "broken"
- "horrible"
- "very artificial"
- "Super grainy, metallic and more like delay."

– Vox:

- "Obviously artificial, but I like it :)"
- "jarring echo in tail, cool sound but not natural at all"
- "very artificial"
- "too grainy"

• Reference

– Drum:

- "closest to reference"
- "very natural"

– Guitar:

- "it aint bad but a unlikely large room (cathedral)."
- "Unrealistic tail sustain, but otherwise sounds quite natural."
- "closest to reference"
- "very natural"

– Impulse:

- "Copy of original."
- "closest to reference."
- "very natural"

– Vox:

- "closest to reference"
- "very natural"
- "same"

• Implementation

- Drum:
 - "comb filter"
 - "natural"
- Guitar:
 - "the most natural one for my taste"
 - "More realistic tail sustain than D, sounds like it was recorded far away from the source sound."
 - "thinner than the reference"
 - "artificial"
- Impulse:
 - "resonances"
 - "high artifacts, choruslike"
 - "artificial"
 - "Quite a lot of high frequencies compared to reference."
- Vox:
 - "Very long tail based on the 'sss' sound in the vocal track sounds unrealistic."
 - "artificial"

• Steward Algorithm

- Drum:
 - "hissy truncated hall"
 - "natural"
 - "End so abruptly."
- Guitar:
 - "quite a lot of high freq content in the reverb"
 - "Artifacts which seem to 'hang' in the upper register nicely simulate a realistic space."
 - "smaller space than the reference."
 - "quite natural"
- Impulse:
 - "hissy, truncated"
 - "gameboy like, i love it (but does not sound like reference at all)."
 - "artificial"
 - "Ends so abruptly."
- Vox:

- *"Very long tail based on the 'sss' sound in the vocal track sounds unrealistic."*
- *"artificial"*
- *"tilted towards highs."*