A Binaural Reverberator

reverb VST plugins

Master thesis Report

Matteo Girardi

Aalborg University Copenhagen
Sound and Music Computing

This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) kpathsea version 6.2.2
Last modified on: *Friday 2nd June, 2017* at *03:04*

AALBORG UNIVERSITY
STUDENT REPORT

Sound and Music Computing
Aalborg University Copenhagen
http://www.aau.dk

Title:
A Binaural Reverberator: reverb VST plugins

Theme:
Artificial reverberation

Project Period:
Spring Semester 2017

Project Group:
000

Participant(s):
Matteo Girardi

Supervisor(s):
Stefania Serafin
Smilen Dimitrov

Copies: 1

Page Numbers: 51

Date of Completion:
Friday 2^{nd} June, 2017

Abstract:

The following dissertation presents two VST plugins which implement a well-known reverberation algorithm called *Feedback Delay Network*. One of which has been slightly modified by using an HRTF model in order to spatialise the early reflections. These plugins should be considered as prototypes and they have been implemented using MATLAB and Audio System Toolbox. In order to understand which plugins yield a better sound quality an experiment has been designed and collected data has been analysed.

Contents

Preface

The following dissertation presents the final project of the Sound and Music Computing Master's programme at Aalborg University of Copenhagen. The research field concerns digital artificial reverberation which is a relevant topic of SMC current research and a well-known DSP topic having several techniques and methods. Therefore, such topic is highly significant for personal experience and knowledge and it is considered relevant for background and expertise. The student would like to thank Stefania Serafin and Smilen Dimitrov for guidance, valuable input and wise supervision. Finally the student would like to thank his family and friends who supported him throughout the entire Master's programme period.

Aalborg University, Friday 2nd June, 2017

Matteo Girardi
<mgirar15@student.aau.dk>

Chapter 1

Introduction

"Acoustics really blossomed in the 19th century [1]."

Room modelling has been an active research field [2] since the beginning of the 20th century and Wallace Clement Sabine was one of the first to carry out pioneering studies on acoustics of rooms with his publication about reverberation [3]. Reverberation is an important property of sound since it carries acoustic information, thus, it defines the sound quality [4, p. 151] and it conveys a sense of the space. Analog and digital methods have been used to develop reverberation effects and initially such effects were only implemented using analog technology. A famous instrument like the Hammond Organ was designed including a spring reverberator, which is still widely used for guitar amplifiers. From the 1980s, digital electronic has slowly taken over analog technology and currently digital technology is predominand [2]; hybrid methods exist as well. One of the first digital reverberators is the Lexicon Delta T-101 [5] [6], and such company is one of the leading manufacters of digital reverberation effects. Even though dedicated reverberation devices are still widely used in audio production, e.g. Lexicon PCM96 stereo reverb[1], reverberation effects running on personal computer, such as VST plugins, are an important branch of the audio production industry. For example, Wave[2] is one on the world's leading developer of audio plugins for professional audio productions and have many reverb plugins among its products[3]; there are several others audio plugins and signal processing companies competing in such market as well, e.g. izotope[4], mcdsp[5], arturia[6] . A recent trend of artificial reverberation methods is *virtual analog* which simulates "vintage" analog and electromechanical

[1] `https://lexiconpro.com/en-US/products/pcm96`
[2] `http://www.waves.com/`
[3] `https://goo.gl/b95tiL`
[4] `https://www.izotope.com/`
[5] `http://mcdsp.com/`
[6] `https://www.arturia.com/`

reverberation unit by software [7] [8] [9] [10].

1.1 Context

This report will present the final project of Sound and Music Computing Master's programme at Aalborg University of Copenhagen. The main theme is digital artificial reverberation and two VST plugins have been implemented. The project have been supervised by Stefania Serafin and Smilen Dimitrov.

1.2 Motivation and Goal

Reverberation effects try to model reflections generated by a room or a concert hall, therefore, how the perception of a sound source is influenced by an acoustic environment. Usually reverberation plugins do not consider how sound is perceived by a listener since the effect of torso, head and pinnas is not taken into consideration. It is true that a reverberator should not consider such effect since most music is produced for loudspeakers, and not for headphones. However, the fundamental motivation of this project is to investigate if a reverberation plugin which takes into account spatialization of early reflections by an Head-Related transfer function could yield an improved sound image quality. This idea is suggest by [11, p. 124]. Moreover, in headphone audio the addition of reverberation lacking of early reflections is usually perceived inside the listener's head, hence including such set of reflections helps to externalize the sound image [12] [13] [14]. A downside of such implementation is that music and sound processed by this modified reverberation plugin should be listen to only through headphones.

From a SMC student point of view this topic is an important knowledge to conquer and to master, since reverberation effects are one of the most widely used effects in audio production, music, film and virtual environment applications [15]. Therefore, gaining important knowledge on audio plugins and effects for music production is considerd by the student relevant to his experties. Lastly, extending his background and experience to this field may yield to possible job positions in the near future. The project is to be considered as first step towards future developments, since the plugins are merely prototypes built to test the main idea. Therefore, the student plans to implement these plugins using low-level programming languages, i.e. C++/C, and a second test should be run on sound engineers during a post production session using such plugins. Other topics have been considered such as Wavefield synthesis, hearing health care, project/collaboration with a company and augmented audio application for mobile. Even though those topics are fascinating, the topic of artificial reverberation has been chosen because the student have previously gained experience in the those topics considered at the beginning of the project. Moreover, since the student's background concerns sound engineering and music production, developing and implementing VST plugins lured the student into artificial reverberation.

1.3 Project overview

The following essay will discuss digital artificial reverberation. During this project two reverb VST plugins have been implemented using a well-known DSP technique known as *feedback delay networks* (FDN, from now on); one of which consists of a slightly modified FDN by using an HRTF model [16] in order to spatialise early reflections. This project came up while reading literature about reverberation, in particular a book [11] inspired the student to pursue the project. This essay is addressed to sound engineers and music producers who wish to have a proper reverberation tool during post-production session. The reverberation plugins may be useful for students since they have been developed using a recent MATLAB toolbox called Audio System Toolbox[7], which makes quite easy to prototype VST plugins. First of all this research project has been carried out by reading and reviewing concepts of reverberation. Second of all, the literature has been reviewed and the current researches have been investigated, thus defining the state of the art. Then the feedback delay network technique has been studied as well as the Head-Related Transfer Function. Once relevant knowledge has been acquired, the reverberation plugins have been implemented starting from basic DSP techniques such as digital filters in order to develop the building blocks of the FDN algorithm. Lastly, tests have been carried out and, at a later time, data have been analised. The rest of this report is organized as follows. Chapter 2 summarizes reverberation concepts, reviews the state of the art of digital artificial reverberation, starting from early researches back in the 50s. Chpater 3 discusses the design and implementation process which has been done in MATLAB programming language. Chapter 4 will analise the results and discuss them. Chapter 5 presents the main conclusions.

[7] Audio System Toolbox

Chapter 2

State of the Art

2.1 Reverberation

> *"W.C. Sabine is generally considered to be the father of architectural acoustics [1]."*

Reverberation is defined as the perception of spaciousness in the sounds and any acoustic environments produce a natural reverberation: concert halls [17] [18], forests [19], city streets [20] [21] or a mountain range [22] have their own distinctive and particular reverberation characteristics. It is a constant presence in our daily life and it is important for synthesized music as well as for audio recordings, since its presence is often preferred for most sound. Musicians know quite well the effect that room acoustic has on sound, e.g. a musical piece played in two halls can yield to completely different experiences, since those halls have different reverberant characteristics. Therefore, reverberation may influence the performance; tempo and dynamics may have to be adapted to that acoustic environment [23] and music without reverberation sounds dry and lifeless [24]. On the other hand, too much reverberation may cause a performance to be muddy and unintelligible. Reverberation is essentially a set of many reflections generated by sound waves colliding surfaces, which disperse the sound. This phenomenon enriches sound by overlapping it with its reflections [23]. Such reflections play a part in determining the "colour" of the sound, thus a change in timbre. Reverberation depends on certain factors such as: volume and dimension of the space, type, shape and number of surfaces. A sound source has a direct path and an indirect path: the direct path is defined as the shortest way from a sound source to the listener, the indirect path consists of multiples delayed and attenuated copies of the original sound, which take longer paths by reflecting off the walls, ceiling, floor and objects [15]. As a sound wave travels is all directions it gets absorbed, reflected, delayed and attenuated, due to air, surfaces and objects, and the amplitude of each reflection is inversely proportional to the distance traveled, frequencies content of each reflection is also modified due to the directivity of the sound source and due to the material absorption of the reflecting surfaces. Direct sound and indirect sound blend together giving what we call reverberation. Indirect sound can be further divided into two parts: *early reflections* and *late reverberation*. Early reflections arrive on a much shorter time scale, shorlty after the direct sound and they are not perceived separately as human hearing integrates them with the direct sound. These reflections include the first-order (one bounce) reflections and the second-order (two bounce) reflections [25]. Due to the precedence effect, their individual directions are not perceived. While the direct path carries information on the direction and the position of the source, early reflections conveys important information about room's shape, size, reflecting surfaces composition and contribute to the perception of the sound color. After these early reflections, another set of reflections arrive to the listener, the *late reflections*. These reflections have a high density; they are randomly distributed, usually decay exponencially, and give rise to *diffuse reverberation*. Such set of reflections gives more cues of the room's size, as well as the distance of the sound source. The time moment between early reflections and late reflections is called mixing time. The reverberation time, often denoted RT_{60} measures the time that it takes for a sound pressure level or intensity to decay by 60 dB; it depends on the volume of the room and the nature of its reflective surfaces.

$$RT_{60} = 0.164 \times V/A. \tag{2.1}$$

Where V is the room volume in cubic meter and A is the total absorption of the room's surfaces in metric sabins [1]. Usually, small rooms have a smaller reverberation time than larger rooms where sound waves travel on a longer distance, although acoustic treatment and other factors can influence it. For example, concert halls have reverberation times around 1.5 and 2 seconds. Highly reverberat environment, such as Cathedrals, may have reverberation times of more than 3 seconds. As a sound wave reflects off a surface, some of its energy is lost and all materials absorb acoustic energy to some extent. Hard and solid surfaces reflect sound very efficiently, whereas soft surfaces are very absorbant. Other measurements correlated with the perception of reverberation are the frequency dependence of the reverberation time, the time delay between the arrival of the direct sound and the early reflections, and the rate of buildup of the echo density. Low frequencies are the last to fade, however materials may affect the reflection of frequencies. The amount of time between direct sound and early reflections varies quite a lot depending on the acoustic environment; a delay greater than 50 ms can result in distinct echoes, whereas a delay smaller than 5 ms can contribute to a listener's perception that the space is small. A delay ranging from 10 ms to 20 ms is found in most good halls [23]. The reate at which the echoes reach the listener depends on the volume of the room and it is roughly proportional to the square root of the room's volume. Small spaces are characterized by a rapid buildup of echo density.

2.2 Artificial Reverberation

During the second half of the 20th century, there has been extensive researches into techniques and methods for simulationg natural reverberation; engineers tried to invent electronic devices capable of simulating the long terms of sound propagations in enclosures. A reverberator either software or hardware, can be thought as a filter which tries to emulate the impulse response of the space to be simulated. Natural room/hall ambience is an important feature of recorded music and having control over the parameters that determine the characteristics of the reverberation gives the sound engineers the ability to shape the perception of spaciousness. However, this was not the case until the invention of artificial reverberation for music broadcasting and recording in the 1920s [2]. Since then several methods and techniques have been proposed, such as:

- Chamber reverberation
- Tape delays
- Spring reverberation
- Plate reverberation
- Digital reverberation

Close micing and a damped studio environment produced a dry sound that lacked the concert hall acoustics desired for music performance, therefore, an early reverberation technique involved specially constructed echo chamber where the dry recorded signal was sent by a loudspeaker, meanwhile several microphones were placed so as to get the artificial reverberated sound [26]. The dry sound and processed sound were later added together, hence, recordings made in a small absorbent studio sounds as if they had been made in a concert hall [27]. Several electromechanical reverberation devices have been developed, including tape delays [27], spring reverberation [28] and plate reverberation [29]. Even though such devices and techniques produce a high-quality reverberation, their use is limited to sound recording studio, not easy to use, impossible to transport and they may vary from unit to unit. These limitations and the importance of reverberation in recorded music has resulted in the creation of artificial reverberators. A recent field of application for artificial reverberation is virtual environment, where simultaing room acoustics is critical for producing a convincing immersive experience [24].

2.3 Digital Reverberation

> "Almost every bit of audio that we hear from recordings, radio, television, and movies has had artificial reverberation added [24]."

As suggested by [2], reverberation algorithms can be grouped into three categories:

- delay network
- convolutional
- computational acoustic

Delay networks are based on comb and allpass filters; the input is delayed, filtered and fed back along a number of path. Convolutional methods consist of a recorded or estimated impulse response of an acoustic space which is convolved with the inputh signal. These two categories are often employed to produce a desired perceptual or artistic effect. Computational acoustic simulates the acoustic energy propagation in the modeled goemetry and generally find application in acoustic desing and analysis scenarios.

2.4 Delay Network Methods

The idea of artificial reverberation based on digital signal processing was first introduced by Schroeder [30] in the early 1960s. Schroeder proposed a reverberator based on comb and allpass filters. Those filter are considered the building blocks for digital audio signal processing systems, and are extensively used for reverberation effects.

2.4.1 Comb Filter

There are two basic comb-filter types, *feedforward* and *feedback* which can be both regardred as computational model of echoes.

Feedforward Comb Filter

A *feedforward comb filter* consists of delay line whose input is fed forward to the output and can be depicted as follows:

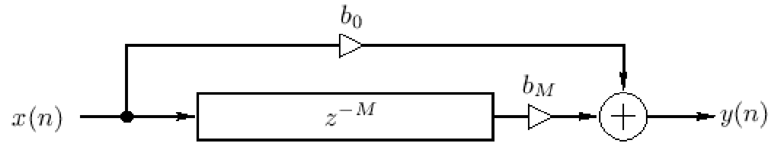

Figure 2.1: Feedforward Comb Filter

The difference equation for the feedback comb filter:

$$y(n) = b_0 x(n) + b_M x(n - M) \tag{2.2}$$

By setting $b_0 = 1$ and $b_M = g$, an echo simulator is implemented. Therefore, it is a computational physical model of a single discrete echo.

Feedback Comb Filter

A *feedback comb filter* consists of a delay line whose output is fed back to the input.

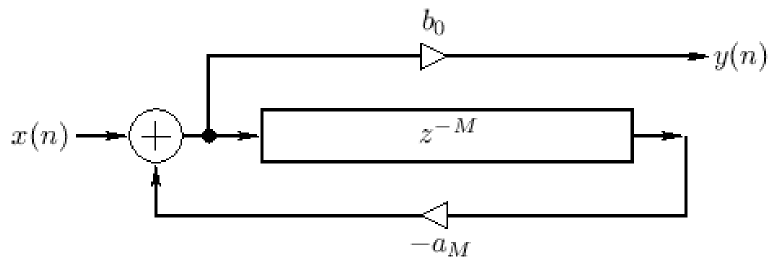

Figure 2.2: Feedback Comb Filter

The difference equation describing a feedback comb filter is given by:

$$y(n) = b_0 x(n) - a_M y(n-M) \tag{2.3}$$

This particular filter can be regarded as a computational physical model of a series of echoes, exponencially decaying and uniformly spaced in time. In order to guarantee stability the coefficient a_M must be less than 1 in magnitude.

$$|a_M| < 1 \tag{2.4}$$

Otherwise each echo will be louder than the previous, producing a never-ending, growing series of echoes [11]. Sometimes the output signal is taken from the end of the delay line instead of the beginning, in which case the difference equation becomes:

$$y(n) = b_M x(n-M) - a_M y(n-M) \tag{2.5}$$

2.4.2 Allpass filter

Another important block of digital audio signal processing system is the *allpass filter*. It is called "allpass" because all frequencies are "passed" and its frequency response is 1 at each frequency, hence having a gain of 1 at all frequencies. This particular filter is extensively used in the fields of artificial reverberation and digital effects [11]. An allpass filter is basically a combination of a feedforward comb filter and a feedback comb filter having the feedforward coefficient being negative of the feedback coefficient.

The difference equation describing an allpass filter is given by:

$$y(n) = b_0 x(n) + x(n-M) - a_M y(n-M) \tag{2.6}$$

and its transfer function is:

$$H(z) = \frac{b_0 + z^{-M}}{1 + a_M z^{-M}} \tag{2.7}$$

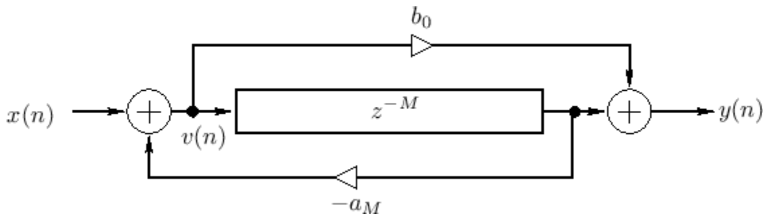

Figure 2.3: Allpass Filter

2.4.3 Classic reverb structures

In the early 1960s Manfred Schroeder and Ben Logan [31] [30] proposed the first digital reverberation algorithms. They introduced the digital allpass filter which produces a series of decaying echoes, but mainatined an overall "colorless" spectrum [2]. Schroeder reverberator is based on recursive comb filter and delay-based allpass filters as computational structures suitable for the inexpensive simulation of complex patterns of echoes. In particular, the allpass filter based on the recursive delay line has the form:

$$y(n) = -g \cdot x(n) + x(n-m) + g \cdot y(n-m) \tag{2.8}$$

where m is the length of the delay in samples and it yields to a dense impulse response and a flat frequency response. Such filter is a standard component used in almost all the artificial reverberators designed up to now. Schroeder proposed a nested allpass structure in order to control the reverber wet/dry mix. He also suggested additional structures for simultaing early reflections using a sparse FIR filter [32]. Early reflections have a great importance in the perception of the acoustic space. Such set of reflections can be implemented using a *Tapped Delay Line* (TDL) which is a delay line with multiple reading points that are weighted and summed together to provide a single output. In 1979 Moorer presented is paper about reverberation [33]. Moorer did extensive experimentations on structures for artificial reverberation and enhanced the Schroeder's structures relating some basic computational structurs such as tapped-delay line for early reflections simulation, comb and allpass filter with the physical behavior of actual rooms. Moreover, the g coefficient is substituted with a lowpass filter in order to simulate air absorption [34].

2.4.4 Feedback Delay Networks

Feedback Delay Network (FDN) were first introduced by Gerzon [35], who proposed an "orthogonal matrix feedback reverberation unit". Feedback comb filters were know to be a computational physical model of echoes, however individually they yielded poor quality reverberation. Hence having several such filters could sound good when cross-coupled [11]. FDN structure for artificial reverberation is

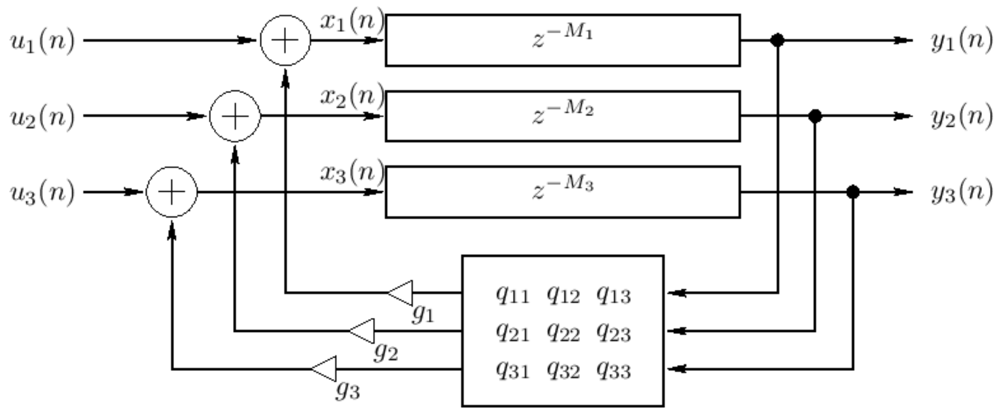

Figure 2.4: Feedback Delay Networks

based on delay lines interconnected in a feedback loop by means of a matrix and can be regarded as a vector generalization of the recursive comb filter:

$$y(n) = x(n-m) + g \cdot y(n-m) \tag{2.9}$$

The m-sample delay line is replaced by a bunch of delay lines of different length and the feedback gain g by a feedback matrix $\mathbf{G}$. More specifically, the FDN structure is a *vector feedback comb filter* [2] with N feedback "channels" which is obtained by replacing the delay line with a diagonal delay matrix, and replacing the feedback gain g by the product of a diagonal matrix Γ times an orthogonal matrix $\mathbf{Q}$. An important part of a FDN structure is the orthogonal feedback matrix which strongly affects the quality of the reverberation, particularly the smoothness of the decaying sound. Stautner and Puckette [36] suggested a specific four-channel FDN reverberator having a feedback matrix as follows:

$$A = g\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \tag{2.10}$$

which is a special form of a 4×4 Hadamard matrix. The "mixing matrix" provides diffusion by "scattering" energy amongst the N channels, basically it increases the density of the late reverberation. Another important part of a FDN reverberator is the delay-line length which should be ideally mutually prime. An improved FDN algorithm has been proposed by Jot [37] [38] who developed a systematic FDN allowing largely independent setting of reverberation time in different frequency bands. Jot's FDN reverberators are presently considered to be among the best choices for high-quality artificial reverberation [11].

The inner loop calculations of the Jot's FDN expressed as:

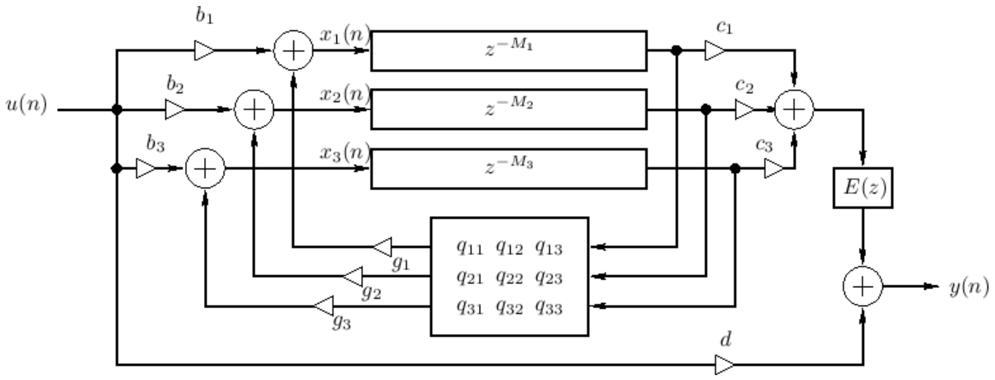

Figure 2.5: A feedback delay network structure proposed for artificial reverberation by Jot [37]

$$\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{bmatrix} = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & g_{33} \end{bmatrix} \begin{bmatrix} x_1(n-M_1) \\ x_2(n-M_2) \\ x_3(n-M_3) \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} u(n) \tag{2.11}$$

and the loop output given by:

$$v(n) = \begin{bmatrix} c_1 c_2 c_3 \end{bmatrix} \begin{bmatrix} x_1(n-M_1) \\ x_2(n-M_2) \\ x_3(n-M_3) \end{bmatrix} \tag{2.12}$$

In order to achieve frequency-dependent decay control, the g_i coefficients can be replaced by low-order digital filters. An additional low-order filter $E(z)$ is applied to the non-direct signal. This filter is called a “tonal correction” filter by Jot, and it serves to equalize modal energy irrespective of the reverberation time in each band. More recently, the FDN concept has been recently extended by Sena [39] incorporating frequendy-dependent wall absortion and directivity of sources and receivers (microphones).

2.5 Spatial Hearing

As a sound wave reaches a listener’s ears many information of the surroundings are processed by the auditory system. In fact, a listener is able to determine the location, the distance and the spatial extents of sound sources, as well as some characteristics of rooms. The auditory system uses several different cues for locating sound sources, such as time and level differences between both ears as well as spectral information. Hence, by comparing the information of both ears humans can have a quite clear perception of the surroundings. Therefore, an audio engineer can artificially simulate such listening capabilities in order to process almost any sound sources for modelling a simulation of a real scenario. Knowing how the body filters sound is important in reproducing binaural sound. As previously stated, sounds propagate from a source to the listener and they are widely modified by the environment. The physical and geometric characteristics of rooms are overlaid on the sound signal arriving to the listener's ears, yielding to reverberation. Three-dimensional sound has a central importance for vitual reality systems.

2.5.1 Head-Related Transfer Function

A sound signal in both ears will be different from the original sound signal and from each other. A transfer function from a sound source to the ear canals is called Head-related transfer function (HRTF) and it is a function used in acoustics that characterizes how a particular ear (left or right) receives sound from a point in space. HRTF is dependent on the direction of a sound source related to the listener, it yields temporal and spectral differences between left and right ear canals. Since ears are located on different sides of the skull, the arrival times of a sound signal vary with direction. The skull casts an acoustic shadow on the far-most ear respect to the sound source. Such shadow is most prominent at frequencies above 2kHz and below 800 Hz has no effect. Other part of the body such as torso, shoulders and pinnae have an effect on sound as well. Therefore, it is well-know that different body parts modify the spectrum of the sound that reaches the ear drums. These changes are captured by the HRTF. Such function varies in a complex way with azimuth, elevation, range and frequency, and it varies significantly from person-to-person.

2.5.2 Duda and Brown's HRTF model

A well-known HRTF model has been proposed by Brown and Duda [16]. Their model is a simple, effective and efficient example for synthesizing binaural sound from a monaural source. Having separate modules, the model simulates vertical as well as horizontal and externalization effects. Additionally, the parameters in the model can be adjusted to fit a particular individual's characteristics.

Each model's component correspond to major structural parts of the body and the external environment. Thus, such model is a higly simplified representation of some very complex phenomena. Duda and Brown's goal was not to faithfully simulate physical process, but to provide the simplest customizable sytem that is capable of producing strong impression of all the spatial dimension. As shown in figure 2.6 a monaural input feeds the head and the shoulder model, and the room model. The head and shoulder model are summed up together and feed the pinna model which produces elevation effects. Finally, the room model's output is added to provide range effects. Hence, each component of the model effects at least one of the three spatial dimensions.

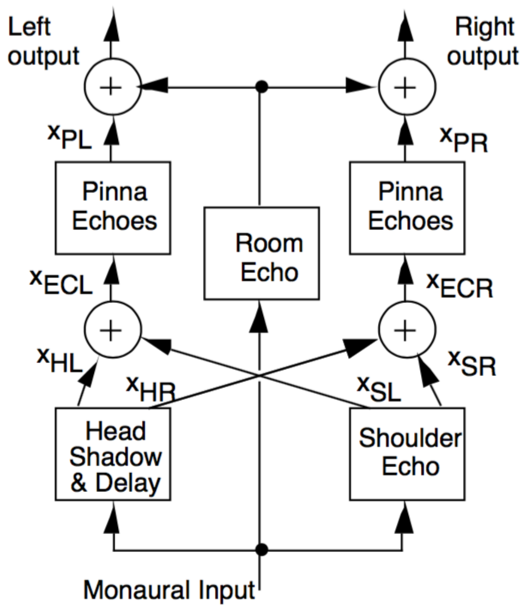

Figure 2.6: Components of Duda and Brown's HRTF model [16].

The Head model

As sound waves strike the head diffraction occurs leading to the sound being delayed and "shadowed" at the most far ear. If the head's shape is approximated by a sphere of radius a, Woodworth's formulas [40] provide an accurate estimate of the time delay.

$$T_L(\theta) = \frac{a + a\theta}{c} \tag{2.13}$$

$$T_R(\theta) = \frac{a - a\sin\theta}{c} \tag{2.14}$$

$T_L(\theta)$ represent the difference between the time that the incident wave strikes the head and the time that it reaches the left ear. $T_R(\theta)$ corresponds to the time difference for the right ear. Let c be the speed of sound. The head shadow effect is introduced by the simple one-pole/one-zero transfer function:

$$H(s,\theta) = \frac{\alpha(\theta)s + \beta}{s + \beta}, \; where \; \beta = \frac{2c}{a} \tag{2.15}$$

The coefficient $\alpha(\theta)$ range from 0 to 2 and it shifts the position of the zero as the azimuth changes. If $\alpha = 0$, sound arrives directly opposite the ear, hence, maximum head shadow. If $\alpha = 2$, there is a 6-dB boost at high frequencies having the sound directly indident on the ear. If ears are placed diagonally across the head, Duda and Brown suggest:

$$\alpha_L(\theta) = 1 - \sin(\theta) \tag{2.16}$$

$$\alpha_R(\theta) = 1 + \sin(\theta) \tag{2.17}$$

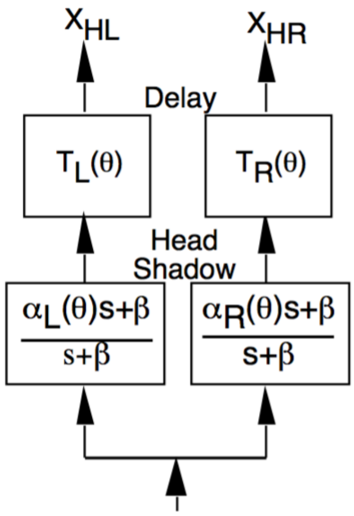

Figure 2.7: The head model proposed by [16].

The Pinna model

The high frequency content of a sound is affected by the pinna's shape providing elevation cues and some azimuth information. The pinna model is shown in Fig. 2.8.

where the ρ_k are the reflection coefficients and the τ_k are the time delays of the kth event of a total of n. Brown and Duda [16] showed that 5 events were enough to represent the pinna response and that it was convenient to use constant values for the amplitudes ρ_k, independent of azimuth, elevation and the subject. The time delays seem to be properly approximated by the following formula:

$$\tau_k(\theta, \phi) = A_k \cos(\theta/2) \sin(D_k(90^\circ - \phi)) + B_k \tag{2.18}$$

In this equation, dependent on the azimuth and elevation, the A_k is an amplitude, B_k an offset and D_k is a scaling factor that should be adapted to the individual listener. In the following table one can see the values for the parameters used in the pinna model. Only one set of values for D_k in the plugins is considered.

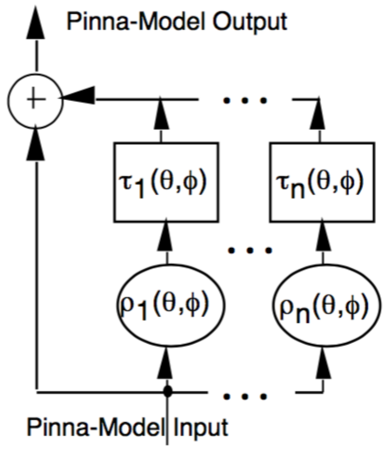

Figure 2.8: The pinna model [16].

Table 2.1: Pinna model coefficients

k	ρ	A_k	B_k	D_k
1	0.5	1	2	1
2	-1	5	4	.5
3	0.5	5	7	.5
4	-0.25	5	11	.5
5	0.25	5	13	.5

Following what Brown and Duda proposed, the shoulder model will not be considered [16]. The room model has been implemented using a FDN structure which will be discussed in chapter 3.

2.6 Related work

Carty and Lazzarini [41] presented two Csound[1,2] opcodes: *hrtfearly* and *hrtfreverb*. Those opcodes are binaural reverberation processors having accurate processing of early reflections and FDN approach for late diffuse field. Recent HRTF dynamic processing algorithms are used to allow dynamic direct sources and early reflections. The FDN model is a flexible binaural processing unit and it considers interaural coherence providing an efficient and robust late reverberation model. Carty and Lazzarini employs two approches. The first one consists of interpolating HRTF magnitudes directly, the second involves phase interpolation. Such

[1] `http://csound.github.io/`
[2] `http://www.csounds.com/`

approches allow phase changes and accurate low frequency interaural phase difference. In hrtfearly the image methods [42] is used for early reflections processing and phase truncation HRTF processing as well, which spatializes and moves the direct sound source and early reflections in accordance with the image model. The user can can choose the order of the early reflections, since HRTF processing can be costly. Other features are offered such as dynamic parameters of source and listener location, lowpass filter modelling the surface's response, three bands equalizer to allow multiband reflective surfaces, distance processing using interpolated delay line. The hrtfreverb and hrtfearly opcodes can be used seperately as well as together providing accurate source location and reverberation or more general binaural reverberator. In hrtfreverb a Jot's FDN model is used and it considers the parametric scenario, as well as independent early reflection processing. Other binaural reverberator are proposed by [43, 44, 45].

Chapter 3

Design and Implementation

In the following chapter the design and implementation of two VST plugins will be presented. The implementation of the two VST plugins has been carried out in MATLAB using a toolbox called *Audio System Toolbox* which makes easy to prototype and implement VST plugins.

3.1 MATLAB and Audio System Toolbox

Audio System Toolbox[1] provides algorithms and tools for the design, simulation, and desktop prototyping of audio processing systems. It includes libraries of audio processing algorithms, sources and measurements. It enables MATLAB developers to run audio processing algorithms on digital audio workstations (DAW) for testing, validation, and early prototyping and to generate VST plugins from MATLAB code. Users interfaces do not need to be design since Audio System Toolbox provides a defualt one. The Audio System Toolbox provides a gallery of open audioPlugin examples to use as reference. In order to test the implementation, such toolbox provides three useful commands:

- validateAudioPlugin "myAudioPlugin"
- audioTestBench "myAudioPlugin"
- generateAudioPlugin "myAudioPlugin"

The first command, *validateAudioPlugin*, generates and runs a Test Bench Procedure that exercises your audio plugin class. The second command, *audioTestBench*, let you test the audio plugin in real time providing a graphical interface through which you can develop, debug, and tune your audio plugin. It is possible to interact with properties of your audio plugin using associated parameter graphical

[1] `https://se.mathworks.com/products/audio-system.html`

widgets. The third command, *generateAudioPlugin*, generates a VST 2 audio plugin from a MATLAB ready to be used in your DAW. There are some consideration to keep in mind, such as:

- Your plugin must be compatible with MATLAB code generation.
- Your generated plugin must be compatible with DAW environments.

These commands are extremely useful in order to debug and test your plugin. Once the plugin has been generated and placed in the proper plugin folder, it is ready to be loaded in your DAW. Following a source-code example of a lowpass VST plugin developed using the Audio System Toolbox:

```
classdef myLPF < audioPlugin
    properties
        g = 0.1;
    end
    properties (Access = private)
        yLast = [0 0];
    end
    properties (Constant)
        PluginInterface = audioPluginInterface(...
            audioPluginParameter('g','DisplayName','LPF
               Coeff','Mapping',{'lin',0,1}));
    end
    methods
        function out = process(plugin, in)
            out = zeros(size(in));
            for i = 1:size(in,1)
                tmp = plugin.g*in(i,:) + (1-plugin.g)*
                   plugin.yLast;
                out(i,:) = tmp;
                plugin.yLast = tmp;
            end
        end
    end
end
```

Using the command *generateAudioPlugin -outdir myLPF*, the output is saved in the specified folder *-outdir*. Fig 3.1 shows the generated plugin of the source code.

Fig 3.2 shows a VST plugin loaded into Reaper.

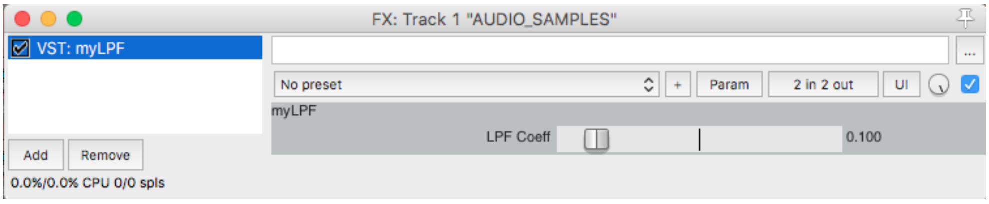

Figure 3.1: A plugin generated using Audio System Toolbox

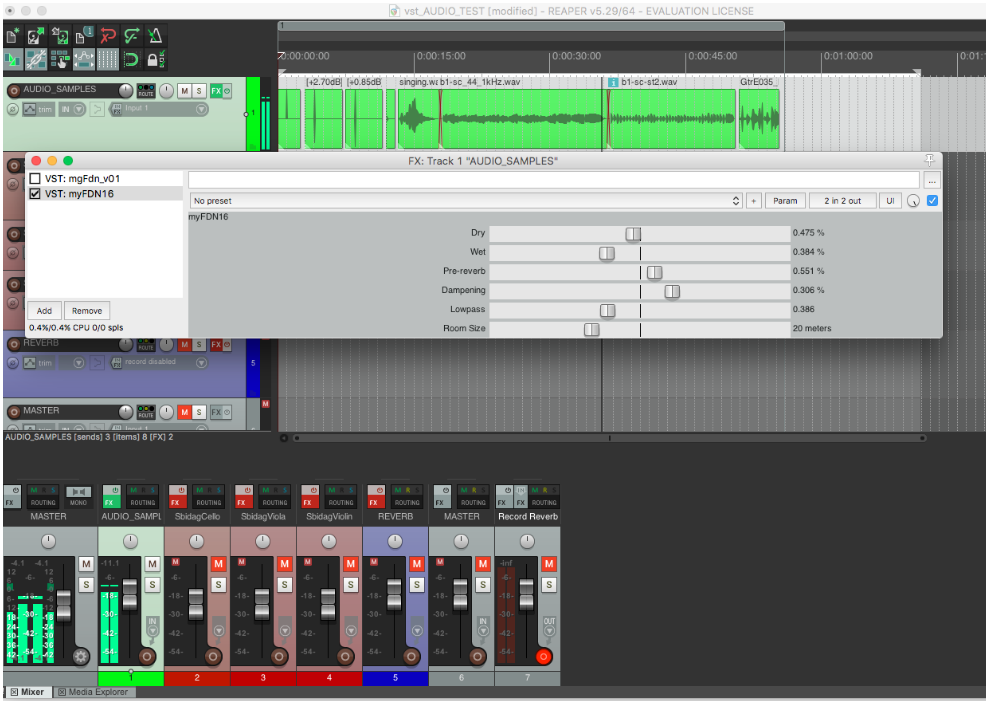

Figure 3.2: Reaper DAW and a generated VST plugin

3.2 Implementation details

In the early stage of development some components of the plugins has been implemented individually, as a seperate VST, in order to get familiar with the Audio System Toolbox; ordinary MATLAB implementation has been done as well. Both plugins consist of a tapped delay line for early reflection simulation, an FDN structure to simulate late reverberation and lowpass filters to simulate air absorption and surface reflections. One of these plugins has additional filters performing a head shadow effect and pinna delay effect. These HRTF components are based on the Brown and Duda paper on 3D-sound [16] which have been summurized earlier in Chapter 2.5.2. The student chose the Brown and Duda model because it is an efficient and simple model to implement. Additionaly, it was already used in a previous project. Other HRTF models have not been considered during the project, but may be considered in future improvements. The FDN structure is based on the one proposed by Jot [37] and it consists of 16 delay lines. The two VST plugins have been called:

- myFDN16
- mgFdn-v02

The first plugin implements a 16 delay line FDN structure with additional tapped delay lines for early reflections. The second plugin implements the same structure as the former and it has additional filters simulating head shadow and pinna delays. The basic core of a feedback delay network has been suggested by [34, p. 170] and extended by the student.

The impulse response of a room can be split into early reflections and a later, more diffuse reverberant tail. Several artificial reverberation models are based on this decomposition [2,8,11]. [41] they also used this approach.

3.2.1 First VST plugin: myFDN16

Fig 3.3 shows a very simplified graph of *myFDN16* structure

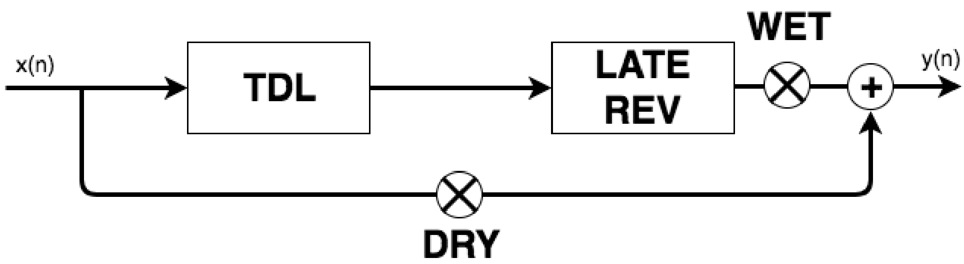

Figure 3.3: A simplified structure of the myFDN16 plugin.

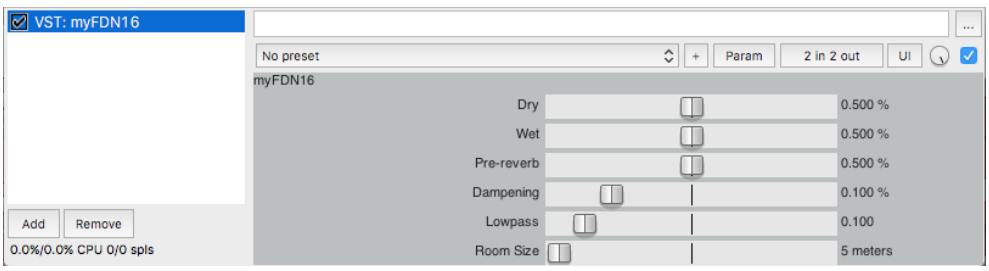

Figure 3.4: VST plugin of a 16 delay lines FDN

As shown in Fig 3.4 the *myFDN16* plugin has six sliders which control different parameters, such as:

- *Dry*
  It controls the amount of dry signal.
- *Wet*
  It controls the amount of wet signal.
- *Pre-reverb*
  It controls the amount of early reflections.
- *Dampening*
  It controls how much reflective the room is.
- *Lowpass*
  It controls the air absorption.
- *Room size*
  It controls the size of the room.

The TDL part is discussed in 3.2.3 and the FDN structure is discussed in 3.2.4.

3.2.2 Second VST plugin: mgFdn-v02

Fig 3.5 shows a very simplified structure of $mgFdn-v02$ structure

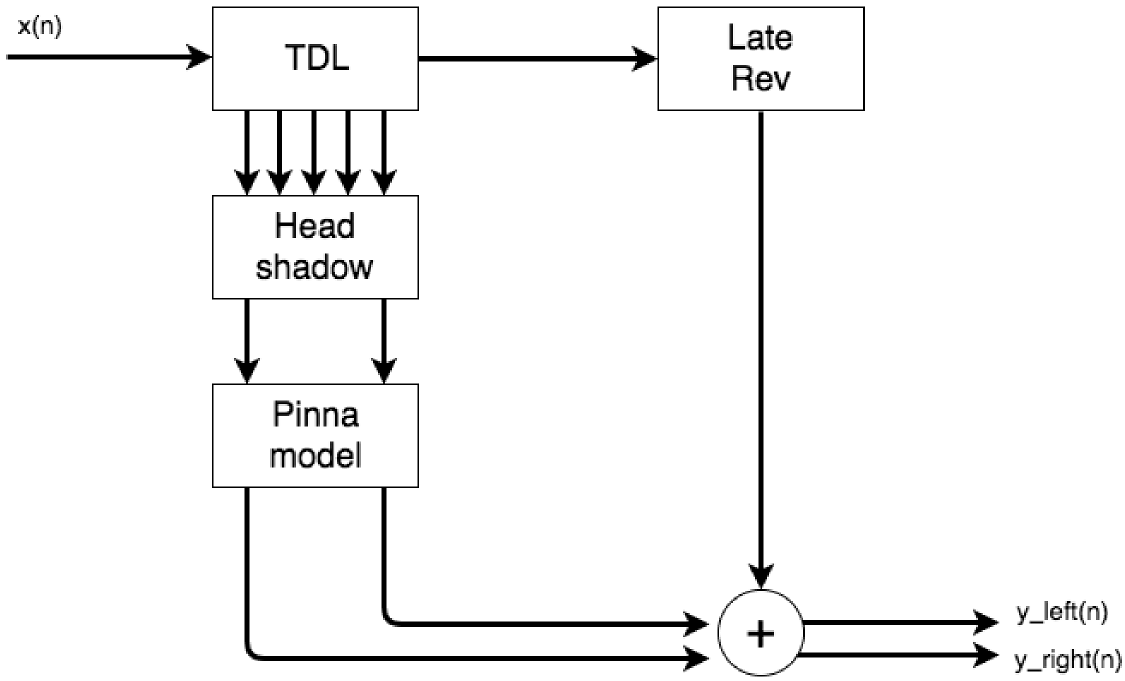

Figure 3.5: A simplified structure of the mgFdn-v02 plugin.

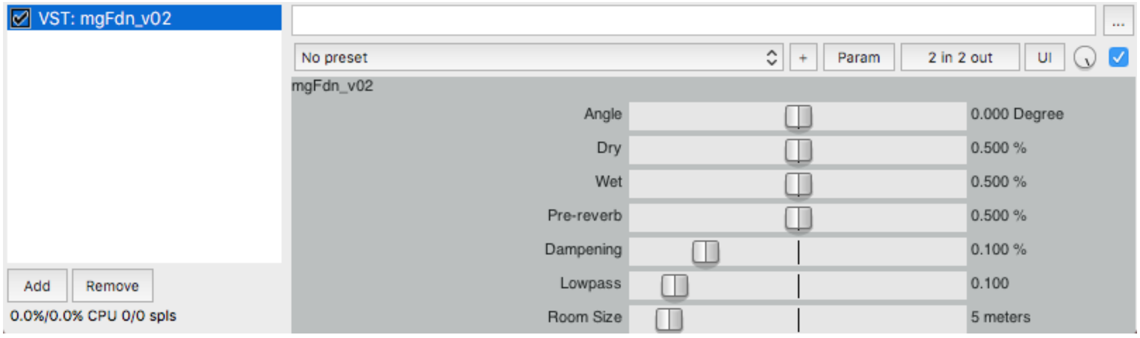

Figure 3.6: VST plugin of a 16 delay lines FDN and HRTF

As shown in Fig 3.6 the *mgFdn−v02* plugin has seven sliders which control different parameters, such as:

- *Angle*
  It controls the direction on the horizontal plane.
- *Dry*
  It controls the amount of dry signal.
- *Wet*
  It controls the amount of wet signal.
- *Pre-reverb*
  It controls the amount of early reflections.
- *Dampening*
  It controls how much reflective the room is.
- *Lowpass*
  It controls the air absorption.
- *Room size*
  It controls the size of the room.

Following each part of the plugins is presented and discussed.

3.2.3 Early Reflections

The listener's perception of the listening-space shape is strongly influeced by early reflection [46]. Such set of reflections is often taken to be the first 100ms or so [33] and it is often implemented using a *tapped delay line* (TDL) [11]. A tapped delay line is basically a shorter delay line within a larger one. A "tap" extracts a signal output from somewhere within the delay line, scales it, and usually sum with other taps to form an output signal. A tap may be *interpolating* or *non-interpolating*. The latter extracts the signal at some fixed integer delay relative to the input. TDL are often used to simulate multiple echoes from the same source signal and they are extensively used in the field of artificial reverberation. A TDL can be seen as a general causal *Finite Impulse Response* (FIR) filter having a tap after every delay elemnt. It is said to be *causal* because the output *y(n)* may not depend on *"future"* inputs. The general *difference equation* for the Mth-order FIR filter is:

$$y(n) = b_0x(n) + b_1x(n-1) + b_2x(n-2) + b_3x(n-3) + \cdot\ \cdot\ \cdot + b_Mx(n-M) \quad (3.1)$$

and the *transfer function* is:

$$H(z) = b_0 + b_1z^{-1} + b_2z^{-2} + b_2z^{-3} + \cdot\ \cdot\ \cdot + b_Mz^{-M} = \sum_{m=0}^{M} b_mz^{-m} \quad (3.2)$$

Fig 3.7 shows an example of a TDL with two internal taps. The output signal

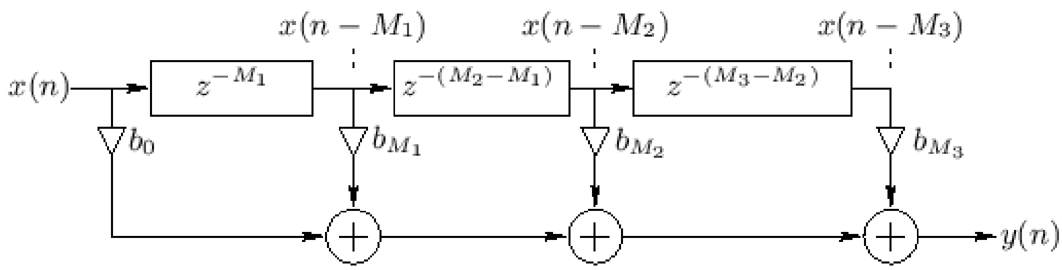

Figure 3.7: Tapped Delay Line (TDL)

is a linear combination of the input signal $x(n)$, the delay-line output $x(n-M_3)$, and the two tap signals $x(n-M_1)$ and $x(n-M_2)$. The difference equation of the TDL in Fig 3.7 is:

$$y(n) = b_0x(n) + b_{M_1}x(n-M_1) + b_{M_2}x(n-M_2) + b_{M_3}x(n-M_3) \quad (3.3)$$

The first plugin myFDN16 has a tapped delay line with 16 taps, while the second plugin mgFdn-v02 has only 6 taps. Probably the first plugin has too many early reflections. The weighted coefficients of the tapped delay line on both plugins are generated randomly using the MATLAB function *rand()* every time they are loaded into a DAW.

3.2.4 Late Reverberation

As stated earlier the late reverberation of both plugins is generate using a 16 delay lines FDN structure. Since one of the goal was to create a good quality reverberator, 16 delay lines seemed to be appropriate. Such number of delay lines is also used in the *Zita-Rev1* [11, p. 122]. In Fig 3.8 shows the FDN structure used in the plugins. The tonal correction filter is not considered.

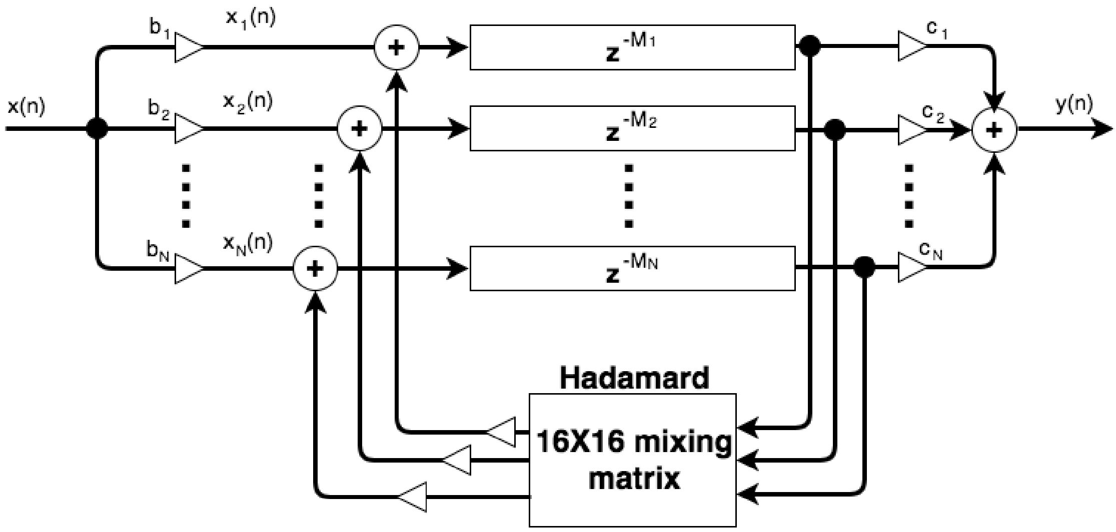

Figure 3.8: A generalized FDN model with $N = 16$ delay lines.

Following each component of the FDN structure is discussed.

3.2.5 Choice of Mixing Matrix

As suggested by [33], an "ideal" late reverberation impulse response should resemble exponentially decaying noise. When designing a reverberator it is a good practice to start with the "lossless case", e.g. an infinite reverberation time, and work on making the reverberator a good "noise generator". This starting point is referred to as "lossless prototype". Even though Stautner and Puckette [36] proposed the feedback matrix:

$$A = g\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \tag{3.4}$$

during the development of the plugins it has been used a Hadamard matrix for two reasons: first, it is used in the IRCAM spatialisateur [47]; second, informal test suggested that the Hadamard matrix yielded to better sound quality. Therefore, both plugins have a 16x16 feedback matrix. The parameter nammed *Dampening* controls the amount of feedback; it can be seen as the reflective properties of the room's surfaces. A second-order *Hadamard matrix* may be defined by:

$$\mathbf{H}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{3.5}$$

with higher order Hadamard matrices defined by recursive embedding, e.g.,

$$\mathbf{H}_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{H}_2 & \mathbf{H}_2 \\ \textbf{-}\mathbf{H}_2 & \mathbf{H}_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{3.6}$$

An $n \times n$ Hadamard matrix has the maximum possible determinant of any $n \times n$ complex matrix containing elements which are bounded by 1 in magnitude. This can be seen as an optimal *mixing and scattering* property of the matrix. Since the implementation has been done in MATLAB, generating a Hadamard matrix is straightforward by using the command *hadamard(N)*, where *N* is the matrix order and must be a power of 2.

3.2.6 Choice of Delay Lengths

As suggested by Schroeder and by [11, p. 111], the delay line lengths in an FDN are typically chosen to be *mutually prime*. That is, their prime factorization contain no common factors, hence, maximazing the number of samples that the lossless reverberator prototype must be run before the impulse response repeats. Since the two plugins have a GUI and the delay-line lengths need to be varied in real time it is useful to choose each delay-line length $\hat{M}_i$ as an integer power of a distinct prime number p_i:

$$\hat{M}_i = p_i^{m_i} \tag{3.7}$$

Using this method the delay-line lengths are always coprime, having no common factors other than 1. Therefore, it is possible to lengthen or shorten each delay line individually without affecting the mutually prime property. Having the desired delay-line lengths $\hat{M}_i$ arranged in ascending order

$$M_1 < M_2 < ... < M_N \tag{3.8}$$

and using the prime numbers in their natual order:

$$p_i \in \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, ...\} \tag{3.9}$$

then a good prime-power approcimations of $\hat{M}_i$ can be expected. Sine $M_i = p_i^{m_i} \implies log(M_i) = m_i log(p_i)$, an optimal choice of prime multiplicity m_i is

$$m_i = round\left[\frac{log(M_i)}{log(p_i)}\right] \tag{3.10}$$

where M_i is the desired length in samples. That is, m_i can simply be obtained by *rounding* $log(M_i)/log(p_i)$ to the nearest integer. This scheme is used in the two VST plugins to keep the 16 delay lines both variable and mutually prime. Following a MATLAB implementation to get the delay line lengths.

```
function m = prime_power_delays(fs,N,pathmin,pathmax)
    Np = N;
    i = [1:Np];
    prime =
       [2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,53,59, 61,
        67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109,
       113, 127, 131];

    % Approximate desired delay-line lengths using powers
       of distinct primes:
    c = 343; % soundspee;d in m/s at 20 degrees C for dry
       air
    dmin = fs*pathmin/c;
    dmax = fs*pathmax/c;
    dl = dmin * (dmax/dmin).^(i/(Np-1)); % desired delay in
        samples
    ppwr = floor(0.5 + log(dl)./log(prime(1:Np))); % best
       prime power
    m = prime(1:Np).^ppwr; % each delay a power of a
       distinct prime
end
```

where N is positive integer up to 16, *pathmin* is the minimum acoustic ray length in the reverberator (in meters) and textitpathmax is the maximum acoustic ray length (meters). The latter can be thought as the “room size”; however using such method there is no correlation between room size and delay line length. This approach has however a limitation which is a sudden change in the delay line length, hence, one can hear that the room size has increased without a smooth transition.

3.2.7 Air absorption simulation

In order to simulate air absorption a lowpass-feedback-comb filter has been implement having the difference equation as:

$$y(n) = \alpha x(n) + (1-\alpha)y(n-1) \tag{3.11}$$

and its transfer function:

$$H(z) = \frac{1}{1-\alpha z^{-M}} \tag{3.12}$$

During the implementation several informal test have been run regarding where to insert the lowpass filter. The student tried to insert the lowpass filter after the feedback matrix, however this approach yielded to an unpleasant and noisy feedback. Therefore, such lowpass filter has been inserted after each delay line.

3.2.8 Head shadow model

The head shadow model is the one proposed by Duda and Brown [16] and it is introduced by the simple one-pole/one-zero transfer function:

$$H(s,\theta) = \frac{\alpha(\theta)s + \beta}{s+\beta},\ where \beta = \frac{2c}{a} \tag{3.13}$$

Since it is an analog transfer function, it was derived to a digital version by appling a bilinear transform. By this, the following transfer function was obtained [2]:

$$H(z,\theta) = \frac{2\alpha(\theta) + T\beta + z^{-1}(-2\alpha(\theta) + T\beta)}{2 + T\beta + z^{-1}(-2 + T\beta)} = \frac{Y(z)}{X(z)} \tag{3.14}$$

And, hence, the following difference equation[3]:

$$Y[n] = \frac{a_0 X[n] + a_1 X[n-1] - b_1 Y[n-1]}{b_0} \tag{3.15}$$

where $a_0 = 2\alpha(\theta) + T\beta$ and $a_1 = -2\alpha(\theta) + T\beta$ as well as $b_0 = 2 + T\beta$ and $b_1 = -2 + T\beta$ are the filter coefficients.

3.2.9 Pinna delay model

As summarized in Chapter 2.5.2, the time delays are approximated by the following formula:

$$\tau_k(\theta,\phi) = A_k \cos(\theta/2)\sin(D_k(90^\circ - \phi)) + B_k \tag{3.16}$$

The following source code shows a function implementing the pinna delays as well as the head model delays.

```
function set.theta(plugin, val)
            plugin.theta = val;
            plugin.thetaRad = val*(pi/180);
            % -- Pinna Echoes
            plugin.NSamplesPM1L = floor(1*cos((plugin.
               thetaRad*-1)/2)*sin(1*(1.57-0))+2);
```

[2]For the derivation of Equation 3.14 Appendix A

[3]For the derivation of Equation 3.15 see Appendix A

```
            plugin.NSamplesPM2L = floor(5*cos((plugin.
               thetaRad*-1)/2)*sin(0.5*(1.57-0))+4);
            plugin.NSamplesPM3L = floor(5*cos((plugin.
               thetaRad*-1)/2)*sin(0.5*(1.57-0))+7);
            plugin.NSamplesPM4L = floor(5*cos((plugin.
               thetaRad*-1)/2)*sin(0.5*(1.57-0))+11);
            plugin.NSamplesPM5L = floor(5*cos((plugin.
               thetaRad*-1)/2)*sin(0.5*(1.57-0))+13);

            plugin.NSamplesPM1R = floor(1*cos(plugin.
               thetaRad/2)*sin(0.5*(1.57-0))+2);
            plugin.NSamplesPM2R = floor(5*cos(plugin.
               thetaRad/2)*sin(0.5*(1.57-0))+4);
            plugin.NSamplesPM3R = floor(5*cos(plugin.
               thetaRad/2)*sin(0.5*(1.57-0))+5);
            plugin.NSamplesPM4R = floor(5*cos(plugin.
               thetaRad/2)*sin(0.5*(1.57-0))+7);
            plugin.NSamplesPM5R = floor(5*cos(plugin.
               thetaRad/2)*sin(0.5*(1.57-0))+13);
            % -- Head Shadow
            plugin.NSamplesL = floor((plugin.a-plugin.a*sin
               (plugin.thetaRad)/343)*getSampleRate(plugin)
               )-4000;
            plugin.NSamplesR = floor((plugin.a+plugin.a*(
               plugin.thetaRad)/343)*getSampleRate(plugin))
               -4000;
        end
```

3.3 Conclusions

MATLAB and Audio System Toolbox provided a quite fast implementation of the algorithms. However, there are some tradeoffs to be taken into account. For example the GUI cannot the customized since MATLAB provides a defualt one. Additionaly, the command *audioTestBench* is quite unreliable when plugins are more complex as the code gets larger, hence it is used only when testing simple implementations. Moreover, the Audio System Toolbox is rather new hence communities and users are not diffused, and the documentation has still to grow. Nevertheless, such toolbox is suprisingly powerful since implementation of VST plugins can be quite straightforward and fast. It is worth to say that the efficiency of the implementation can be improved.

Chapter 4

Analysis

The following chapter presents test design and procedure employed for the experiment, as well as the statistical analysis of the collected data.

4.1 Testing

First of all, testing the sound quality of two VST plugins is not a trivial task since many different variables such as algorithms, equipments, subjective experience, test procedure and design, can vary greatly. Hence, it is not straightforward to determine which plugin sounds better. Additionaly, one can argue that in an audio post-producing scenario, the user will choose those plugins for different purposes. However, it is worth to design an experiment and attempt to analyse the data looking for results. Before going on with the design of the experiment, the student contacted privately an Audio Processing company, called *Dehumanizer*[1] which is specialized in developing VST plugins. The student knew such company since it gave a demonstration at Aalborg University Copenhagen in 2016 during an SMC colloquium. The student asked how to compare two similar VST plugins, what variables to test, how to design a test and what kind of participants are requested. Sad but true, the student got a quite unsatisfying answer, since no suggestions were given. Searching and surfing the web, the student found a quite interesting guide by ITU[2] which describes general methods for the subjective assessment of sound quality [48]. Such recommendation methods is based on Recommendation ITU-R BS.1116 – Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. These recommendation were not strictly followed since they are intended for small impairments in telecommunications and are not meant for audio post-productions effects. Nevertheless, the student found the guide pretty useful in order to test his

[1] `https://www.krotosaudio.com/`

[2] International Telecommunication Union

VST implementations. First, the audio quality is defined as the attribute including all aspects of the sound quality being assessed. It includes, but is not restricted to, such things as timbre, transparency, stereophonic imaging, spatial presentation, reverberance, echoes, harmonic distortions, quantisation noise, pops, clicks and background noise. In *Appendix 1* of [48] the main attributes are presented as well as in the essay Appendix B. Therefore, the student chose to design his own test using some recommendations and attributes provided by [48]. The main attributes found in [48] are divided in several sub-attributes which have been used as questions during the experiment. Since those sub-attributes are quite specific the student prepared a document with definitions in order to help the participants to understand and answer the questions properly. It is worth to say taht a different test approach was considered by the student and it consisted in a user test where the participants would have used the two VST plugins in a DAW. The task would have been to use the VSTs for post-produce a simple audio file. However, due to lack of expert participants this approach was not considered. The student believes that such approach could have been more appropriate for the essay research topic.

4.1.1 Test design

The student designed a test consisting of two listening tests where participants were asked to listen to two audio samples and then asnwer several questions about sound quality. The participants were all SMC and Medialogy students from Aalborg University Copenhagen and should be considered as non-expert since only few of them have experience in audio mixing and post-processing. In total 20 participants took part to the experiment. In order to verify such consideration some questions about general knowledge of sound, reverberation and audio post-production were asked as well. The audio samples were post-produced by the student using Reaper DAW. During the test the order of the two audio samples were randomized and eight question were asked. During the listening tests the same heaphones (SONY MDR-7506 Professional) were used for all participants. The participants were asked to judge several sound quality attributes based on a 5-grade Likert scale. Such scale is suggested by [48] and it is presented in Tab 4.1.

Table 4.1: 5-grade Likert scale

Quality	
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Following the fist part of the questionnaire is presented:

- Profession
- Age
- Define your expertise about reverberation
  newbie - expert (1-5)
- Define your expertise about mixing, recording and audio post-production
  newbie - expert (1-5)
- Degine your expertise about sound, music and acoustics
  newbie - expert (1-5)

Following the second part of the questionnaire is presented:

- Listen to the audio file
- Spatial impression
  Homogeneity of the spatial sound
  not homogeneous - homogeneous
- Stereo impression
  Directional Balance
  imprecise - precise
- Stereo impression
  Location accuracy
  imprecise - precise
- Stereo impression
  Sound image width
  narrow - wide
- Transparency
  sound source definition
  confused - distinct
- Sound colouring
  sound colour
  dark - brilliant
- Freedom from noise and distortions
  imperceptible disturbances - perceptible diturbances
- Main impression
  bad - excellent

- Any comments?

The second part of the questionnaire was repeated for the second audio file. The participants were asked to leave some comments and at the end of the quetionnaire their were asked to choose the audio file that they liked to most.

- Which audio file did you like the most?
- and why?

4.1.2 Audio Samples

The audio samples were post-produced by the student using the two VST plugins in Reaper DAW and they were not recorded by the student. The two audio samples consist of a classical trio (cello, viola and violin). These instruments were spread around using the panner for the *myFDN16* plugin and the angle parameter for the *mgFdn* so as to recreate a real concert scenario. To get a better results it would have been better to record some audio files in an anechoic chamber so as to have the most dry signal possible and record a trio as well. However, due to technical and time constraints the student decided to use pre-recorded audio files which were given by a colleague. Since the two plugins have several parameters, during the post-production section the student kept the same parameters for both plugins in order to get the most similar audio files as possible. However, audio post-production is not an easy task since it relies on knowledge and experience, hence, one can argue that the audio samples were not mixed and post-produce in a professional way. The audio files can be listen at the following links: mgFdn Binaural Reverberator VST, myFDN16 Reverberator VST.

4.1.3 Repeated-measures Analysis of Variance

The statistical test employed is is the *Repeated-measures Analysis of Variance*. Since observations are taken from the same group of subjects the student decided to employ such analysis. An advantage of a repeated-measures design is that each subject acts as his or her own control, and this can increase the ability to detect differences. Fig. 4.2 provides some basic statistics for the eight level on the indipendent variable.

From Tab. 4.2 it can be seen that, on average, the two plugins are very similar. The only attributes that differ are the *sound definition* and *sound colouring*. Hence, it can be supposed that the results will not show any difference between the two plugins. The results of Mauchly's sphericity test for each three effects in the model shows that the significance values have been violated and so the F-values should be corrected. Fig. 4.1 shows the results of ANOVA with corrected F-values.

The output is split into sections that refer to each of the effects in the model and the error terms associated with these effects. By looking at the significance values

Table 4.2: Descriptive Statistics

	Mean	Std. Deviation	N
Binaural reverberator (mgFdn-v02)			
Homogeneity of the spatial sound	3.55	.999	20
Directional Balance	3.75	1.293	20
Location Accuracy	4.05	.999	20
Sound image width	4.00	.973	20
Sound definition	4.25	.851	20
Sound Colouring	3.35	.933	20
Freedom	1.90	1.334	20
Main impresison	4.05	.999	20
Reverberator (myFDN16)			
Homogeneity of the spatial sound	3.25	.967	20
Directional Balance	3.50	1.100	20
Location Accuracy	3.45	1.099	20
Sound image width	4.00	.725	20
Sound definition	2.95	.999	20
Sound Colouring	2.05	.945	20
Freedom	2.15	1.309	20
Main impresison	3.65	1.040	20

it is clear that there is a significant effect of the type of VST used, a significant main effect of the type of attributes used and a significant interaction between these two variables. The first part of 4.1 tells us the effect of the audio effects used in the experiment. This effect tells us that if we ignore the type of attributes that was used, participants still rated the two plugins differently.

Table 4.3: Estimated Marginal Means

				95% confidence Interval
VST	Meand	Std. Error	Lower Bound	Upper Bound
1	3.613	.109	3.385	3.840
2	3.125	.153	2.805	3.445

From Fig. 4.2 it can be seen both plugins have a similar trend among the sound quality attributes and that the binaural reverberator slightly differs from the other reverberator.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
VST	Sphericity Assumed	19.013	1	19.013	5.623	.028
	Greenhouse-Geisser	19.013	1.000	19.013	5.623	.028
	Huynh-Feldt	19.013	1.000	19.013	5.623	.028
	Lower-bound	19.013	1.000	19.013	5.623	.028
Error(VST)	Sphericity Assumed	64.238	19	3.381		
	Greenhouse-Geisser	64.238	19.000	3.381		
	Huynh-Feldt	64.238	19.000	3.381		
	Lower-bound	64.238	19.000	3.381		
soundquality	Sphericity Assumed	125.938	7	17.991	17.490	.000
	Greenhouse-Geisser	125.938	3.407	36.965	17.490	.000
	Huynh-Feldt	125.938	4.241	29.692	17.490	.000
	Lower-bound	125.938	1.000	125.938	17.490	.001
Error(soundquality)	Sphericity Assumed	136.813	133	1.029		
	Greenhouse-Geisser	136.813	64.731	2.114		
	Huynh-Feldt	136.813	80.588	1.698		
	Lower-bound	136.813	19.000	7.201		
VST * soundquality	Sphericity Assumed	22.138	7	3.163	4.694	.000
	Greenhouse-Geisser	22.138	3.449	[illegible]	[illegible]	.003
	Huynh-Feldt	22.138	4.307	[illegible]	[illegible]	.001
	Lower-bound	22.138	1.000	22.138	4.694	.043
Error(VST*soundquality)	Sphericity Assumed	89.613	133	.674		
	Greenhouse-Geisser	89.613	65.525	1.368		
	Huynh-Feldt	89.613	81.826	1.095		
	Lower-bound	89.613	19.000	4.716		

Figure 4.1: Tests of Within-Subjects Effects

4.2 Results and Discussion

The results show that over all the binaural plugin is slightly better than the other reverberator plugin. However the student believes that such results is determinated by the post-production session, in other words, it depends on the way that the audio files have been mixed and post-produced.

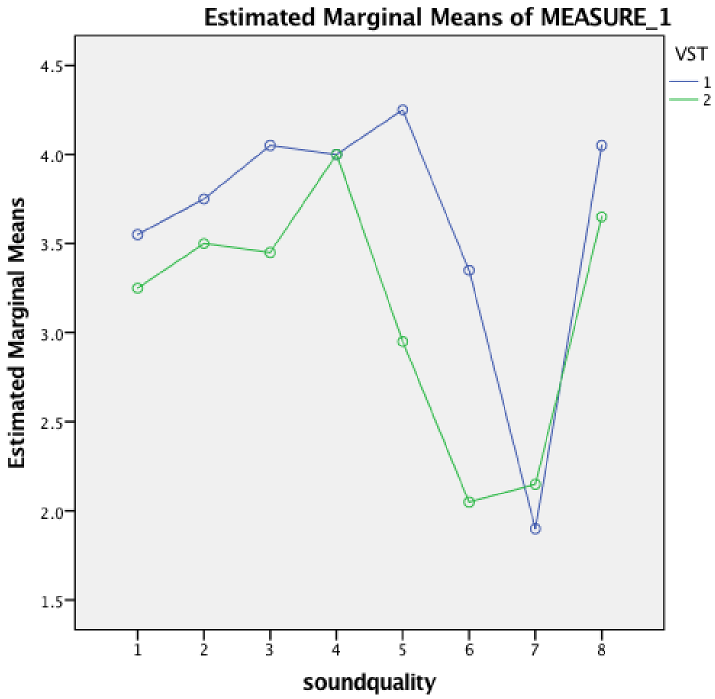

Figure 4.2: Estimated Marginal Means Graph. VST1 refers to the Binaural reverberator.

Chapter 5

Conclusions

In this essay two reverberator plugins have been presented, analysed and discussed. Following the results of the statistical test it can be concluded that the binaural reverberator may have a better sound quality than the non-binaural reverberator. The student expected such results since it is not trivial to test the sound quality of two plugins, and maybe it is quite unusual. Even though statistical analysis have been run, It is not confirmed such hypothesis since there are many variables which have not been taken into account or have been set on the side. For example, it might be that some bugs in the MATLAB source code will be found, or that another experiment should be arrange in orther to test the user experience of the two plugins.

5.1 Future work

Since these two plugins are merely prototype, many improvements can be implemented. First of all, interpolation for the delay line length should be implemented in order to have a smooth transition between room size. At this stage when the room size's slider is changed, the reverberation changes quite sharply. Additionaly all the delay line of the FDN algorithm should the varied in time so as to ensure a smooth decay [2]. Then it is necessary to implement the image source methods in order to calculate the direction of the early reflections. Lastly, a tonal correction filter should be implemented as well. The plugins should be implemented in a low-level programming language such as C/C++ in order to let the programmer have a better control over the algorithms, improve efficiency and customize the GUI since MATLAB does not provide such programming power.

Bibliography

[1] T. Rossing. *Springer Handbook of Acoustics*. Springer Handbook of Acoustics. Springer New York, 2007. ISBN: 9780387304465.

[2] V. Valimaki et al. "Fifty Years of Artificial Reverberation". In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.5 (2012), pp. 1421–1448. ISSN: 1558-7916. DOI: 10.1109/TASL.2012.2189567.

[3] Wallace Clement Sabine. "Reverberation". In: *Acoustics: Historical and Philosophical Development* (1900). ed. by R. D. Lindsay, Dowden, Hutchinson, and Ross, Stroudsburg, PA (1972).

[4] F.A. Everest and K. Pohlmann. *Master Handbook of Acoustics*. McGraw-Hill Education, 2009. ISBN: 9780071603331.

[5] Clyne Media. *Lexicon Celebrates 35 Years of Digital Audio*. Inc., press release. 2007.

[6] Barry A. Blesser and Francis F. Lee. "An Audio Delay System Using Digital Technology". In: *J. Audio Eng. Soc* 19.5 (1971), pp. 393–397. URL: http://www.aes.org/e-lib/browse.cfm?elib=2172.

[7] V. Valimaki et al. "Introduction to the Special Issue on Virtual Analog Audio Effects and Musical Instruments". In: *IEEE Transactions on Audio, Speech, and Language Processing* 18.4 (2010), pp. 713–714. ISSN: 1558-7916. DOI: 10.1109/TASL.2010.2046449.

[8] Julian Parker and Stefan Bilbao. "Spring Reverberation: A Physical Perspective". In: *The 12th International Conference on Digital Audio Effects (DAFx'09), Como, Italy, September 1-4, 2009*. 2009, verkkojulkaisu.

[9] Steinunn Arnardottir, Jonathan S. Abel, and Julius O. Smith III. "A Digital Model of the Echoplex Tape Delay". In: *Audio Engineering Society Convention 125*. 2008. URL: http://www.aes.org/e-lib/browse.cfm?elib=14800.

[10] Jonathan S. Abel et al. "Spring Reverb Emulation Using Dispersive Allpass Filters in a Waveguide Structure". In: *Audio Engineering Society Convention 121*. 2006. URL: http://www.aes.org/e-lib/browse.cfm?elib=13788.

[11] J.O. Smith. *Physical audio signal processing : for virtual musical instruments and audio effects*. W3K Publ., 2010. ISBN: 9780974560724.

[12] N. Sakamoto, T. Gotoh, and Y. Kimura. "On 'out-of-head localization' in headphone listening". eng. In: *Journal of the Audio Engineering Society* 24.9 (1976), pp. 710–16, 710–716. ISSN: 15494950, 00047554.

[13] Toni Liitola, Vesa Välimäki, and Instructors Tapani Ritoniemi. "Headphone Sound Externalization". eng. In: (2008). DOI: `10.1.1.107.4776`.

[14] V. Ralph Algazi and Richard O. Duda. "Headphone-Based Spatial Sound". eng. In: *Ieee Signal Processing Magazine* 28.1 (2011), pp. 33–42. ISSN: 15580792, 10535888. DOI: `10.1109/MSP.2010.938756`.

[15] J.D. Reiss and A. McPherson. *Audio Effects: Theory, Implementation and Application*. CRC Press, 2014. ISBN: 9781466560291.

[16] C. P. Brown and R. O. Duda. "An efficient HRTF model for 3-D sound". In: *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*. 1997, 4 pp.–. DOI: `10.1109/ASPAA.1997.625596`.

[17] Anders C. Gade. "Acoustics in Halls for Speech and Music". In: *Springer Handbook of Acoustics*. Ed. by Thomas D. Rossing. New York, NY: Springer New York, 2014, pp. 317–366. ISBN: 978-1-4939-0755-7. DOI: `10.1007/978-1-4939-0755-7_9`. URL: `http://dx.doi.org/10.1007/978-1-4939-0755-7_9`.

[18] D. Griesinger. "Concert Hall Acoustics and Audience Perception [Applications Corner]". In: *IEEE Signal Processing Magazine* 24.2 (2007), pp. 126–131. ISSN: 1053-5888. DOI: `10.1109/MSP.2007.323277`.

[19] Hiroyuki Sakai, Shin ichi Sato, and Yoichi Ando. "Orthogonal acoustical factors of sound fields in a forest compared with those in a concert hall". In: *The Journal of the Acoustical Society of America* 104.3 (1998), pp. 1491–1497. DOI: `10.1121/1.424360`. eprint: `http://dx.doi.org/10.1121/1.424360`. URL: `http://dx.doi.org/10.1121/1.424360`.

[20] Francis M. Wiener, Charles I. Malme, and Creighton M. Gogos. "Sound Propagation in Urban Areas". In: *The Journal of the Acoustical Society of America* 37.4 (1965), pp. 738–747. DOI: `10.1121/1.1909409`. eprint: `http://dx.doi.org/10.1121/1.1909409`. URL: `http://dx.doi.org/10.1121/1.1909409`.

[21] Judicaël Picaut. "Experimental study of sound propagation in a street". In: *The Journal of the Acoustical Society of America* 117.4 (2005), pp. 2417–2417. DOI: `10.1121/1.4786374`. eprint: `http://dx.doi.org/10.1121/1.4786374`. URL: `http://dx.doi.org/10.1121/1.4786374`.

[22] Reto Pieren and Jean Marc Wunderli. "Sound Reflections from Cliffs and Reverberation in an Alpine Valley". In: *Proceedings of the Forum Acusticum 2011, Aalborg, Denmark*. 2011.

[23] C. Dodge and T.A. Jerse. *Computer Music: Synthesis, Composition, and Performance*. Schirmer Books, 1997. ISBN: 9780028646824.

[24] M. Kahrs and K. Brandenburg. *Applications of Digital Signal Processing to Audio and Acoustics*. The Springer International Series in Engineering and Computer Science. Springer US, 2006. ISBN: 9780306470424.

[25] D.G. Loy. *Musimathics: The Mathematical Foundations of Music*. Musimathics v. 1. MIT Press, 2007. ISBN: 9780262122825.

[26] Michael Rettinger. "Reverberation Chambers for Broadcasting and Recording Studios". In: *J. Audio Eng. Soc* 5.1 (1957), pp. 18–22. URL: `http://www.aes.org/e-lib/browse.cfm?elib=285`.

[27] Lewis S. Goodfriend and John H. Beaumont. "The Development and Application of Synthetic Reverberation Systems". In: *J. Audio Eng. Soc* 7.4 (1959), pp. 228–234, 250. URL: `http://www.aes.org/e-lib/browse.cfm?elib=544`.

[28] H. Laurens. *Electrical musical instrument*. US Patent 2,230,836. 1941. URL: `https://www.google.com/patents/US2230836`.

[29] Kevin Arcas and Antoine Chaigne. "On the quality of plate reverberation". eng. In: *Applied Acoustics* 71.2 (2010), pp. 147–156. ISSN: 1872910x, 0003682x. DOI: `10.1016/j.apacoust.2009.07.013`.

[30] Manfred R. Schroeder. "Natural Sounding Artificial Reverberation". In: *Audio Engineering Society Convention 13*. 1961. URL: `http://www.aes.org/e-lib/browse.cfm?elib=343`.

[31] M. R. Schroeder. "Improved Quasi-Stereophony and "Colorless" Artificial Reverberation". In: *The Journal of the Acoustical Society of America* 33.8 (1961), pp. 1061–1064.

[32] M. R. Schroeder. "Digital Simulation of Sound Transmission in Reverberant Spaces". In: *The Journal of the Acoustical Society of America* 45.1 (1969), pp. 303–303. DOI: `10.1121/1.1971383`. eprint: `http://dx.doi.org/10.1121/1.1971383`. URL: `http://dx.doi.org/10.1121/1.1971383`.

[33] James A. Moorer. "About This Reverberation Business". In: *Computer Music Journal* 3.2 (1979), pp. 13–28. ISSN: 01489267, 15315169. URL: `http://www.jstor.org/stable/3680280`.

[34] U. Zölzer. *DAFX: Digital Audio Effects*. Wiley, 2011. ISBN: 9781119991304.

[35] M. A. Gerzon. *Synthetic stereo reverberation, part i and ii*. Studio Sound, vol. 13(I), 14(II), pp. 632-635(I), 24-28(II), 1971(I),1972(II).

[36] John Stautner and Miller Puckette. "Designing multi-channel reverberators". In: *Computer Music Journal* 6.1 (1982), pp. 52–65.

[37] Jean-Marc Jot and Antoine Chaigne. "Digital Delay Networks for Designing Artificial Reverberators". In: *Audio Engineering Society Convention 90*. 1991. URL: http://www.aes.org/e-lib/browse.cfm?elib=5663.

[38] J. M. Jot. *Etude et Réalisation d'un Spatialisatuer de Sons par Modèles Physiques et Perceptifs*. PhD thesis, French Telecom, Paris, Paris 92 E 019, written in French. 1992.

[39] Enzo De Sena, Huseyin Hacihabiboglu, and Zoran Cvetkovic. "Scattering delay network: An interactive reverberator for computer games". In: *Audio Engineering Society Conference: 41st International Conference: Audio for Games*. Audio Engineering Society. 2011.

[40] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997. ISBN: 9780262024136.

[41] Brian Carty and Victor Lazzarini. "Hrtfearly and Hrtfreverb: Flexible Binaural Reverberation Processing". In: *Proceedings of the 2010 International Computer Music Conference, ICMC 2010, New York, USA, 2010*. Michigan Publishing, 2010. URL: http://hdl.handle.net/2027/spo.bbp2372.2010.105.

[42] Jont B. Allen and David A. Berkley. "Image method for efficiently simulating small-room acoustics". In: *The Journal of the Acoustical Society of America* 65.4 (1979), pp. 943–950. DOI: 10.1121/1.382599. URL: http://dx.doi.org/10.1121/1.382599.

[43] Fritz Menzer and Christof Faller. "Binaural Reverberation Using a Modified Jot Reverberator with Frequency-Dependent Interaural Coherence Matching". In: *Audio Engineering Society Convention 126*. 2009. URL: http://www.aes.org/e-lib/browse.cfm?elib=14961.

[44] Fritz Menzer. "Binaural Reverberation Using Two Parallel Feedback Delay Networks". In: *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. 2010. URL: http://www.aes.org/e-lib/browse.cfm?elib=15526.

[45] Markus Noisternig et al. "A 3D Ambisonic Based Binaural Sound Reproduction System". In: *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. 2003. URL: http://www.aes.org/e-lib/browse.cfm?elib=12314.

[46] Gary Kendall and W.L. Martens. "Simulating the Cues of Spatial Hearing in Natural Environments". In: *Proceedings of the 1984 International Computer Music Conference, Paris*. Oct. 1984, pp. 111–126.

[47] J.-P Julien et al. "Spatilizer: A perceptual approach". In: *AES 94th Convention*. Vol. Preprint 3465 (B1-5). 1993, pp. 1–13.

[48] ITU-R Recommendation BS.1284-1. *General methods for the subjective assessmentof sound quality*. Tech. rep. Geneva, Switzerland: International Telecommunication Union, Dec. 2003.

Appendix A

Appendix A

A.1 Derivations

The analog transfer function has to be derived to the digital version by applying a bilinear transform:

$$s = \frac{2}{T}\frac{z-1}{z+1},\ \textit{where T is the sampling interval in seconds} \tag{A.1}$$

Applying the substitution in equation A.1 to equation **??**, the following filter transfer function in the digital domain is obtained:

$$H\left(z,\theta\right) = \frac{\alpha(\theta)(\frac{2}{T}\frac{z-1}{z+1}) + \beta}{(\frac{2}{T}\frac{z-1}{z+1}) + \beta} \tag{A.2}$$

The derivations followed to get equation 3.14 can be seen as follows

$$\begin{aligned}
H\left(z,\theta\right) &= \frac{\alpha(\theta)(\frac{2}{T}\frac{z-1}{z+1}) + \beta}{(\frac{2}{T}\frac{z-1}{z+1}) + \beta} = \frac{\frac{2\alpha(\theta)(z-1)}{T(z+1)} + \frac{T\beta(z+1)}{T(z+1)}}{\frac{2(z-1)}{T(z+1)} + \frac{T\beta(z+1)}{T(z+1)}} = \frac{\frac{2\alpha(\theta)(z-1)+T\beta(z+1)}{T(z+1)}}{\frac{2(z-1)+T\beta(z+1)}{T(z+1)}} \\
&= \frac{2\alpha(\theta)(z-1) + T\beta(z+1)}{2(z-1) + T\beta(z+1)} = \frac{z2\alpha(\theta)(1-z^{-1}) + zT\beta(1+z^{-1})}{z2(1-z^{-1}) + zT\beta(1+z^{-1})} \\
&= \frac{2\alpha(\theta)(1-z^{-1}) + T\beta(1+z^{-1})}{2(1-z^{-1}) + T\beta(1+z^{-1})} = \frac{2\alpha(\theta) - 2\alpha(\theta)z^{-1} + T\beta + T\beta z^{-1}}{2 - 2z^{-1} + T\beta + T\beta z^{-1}} \\
&= \frac{(2\alpha(\theta) + T\beta) - (2\alpha(\theta) + T\beta)z^{-1}}{(2 + T\beta) - (2 + T\beta)z^{-1}} = \frac{(2\alpha(\theta) + T\beta) + (-2\alpha(\theta) + T\beta)z^{-1}}{(2 + T\beta) + (-2 + T\beta)z^{-1}}
\end{aligned}$$

$$
\begin{aligned}
H(z,\theta) = \frac{Y(z)}{X(z)} &= \frac{a_0 + a_1 z^{-1}}{b_0 + b_1 z^{-1}} \\
&\iff Y(z)(b_0 + b_1 z^{-1}) = X(z)(a_0 + a_1 z^{-1}) \\
&\iff Y(z)b_0 + b_1 Y(z) z^{-1} = a_0 X(z) + a_1 X(z) z^{-1} \\
&\iff Y(z) = \frac{a_0 X(z) + a_1 X(z) z^{-1} - b_1 Y(z) z^{-1}}{b_0} \\
&\rightarrow Y[n] = \frac{a_0 X[n] + a_1 X[n-1] - b_1 Y[n-1]}{b_0}
\end{aligned}
$$

Appendix B

Appendix B

B.1 ITU attributes

Following the list of main attributes, sub-attributes and definition used during the experiment:

SPATIAL IMPRESSION
The performance appears to take place in an appropriate spatial environment.

- Homogeneity of the spatial sound:
  The subjective impression that the sound space is a homogeneous whole.

STEREO IMPRESSION
The sound image appears to have the correct and appropriate direction distribution of sound sources.

- Directional balance: The subjective impression that the sound sources within the sound image are placed in a way which makes the entire image balanced.
- Location accuracy: The subjective impression that all sound sources are accurately positioned in the sound image.
- Sound image width: The subjective impression of an appropriate width of the sound stage in the stereo sound field.

TRANSPARENCY
All details of performance can be clearly perceived.

- Sound source definition:
  The subjective impression that different instruments or voices sounding simultaneously can be identified and distinguished.

TIMBRE - SOUND COLOURING

Accurate portrayal of the different sound characteristics of sound source(s).

- Sound colour:
  The subjective impression of an appropriate sound for each source including all its characteristic harmonic elements.

FREEDOM FROM NOISE AND DISTORTIONS

- Absence of various disturbing phenomena such as electrical, acoustic noise, public noise, bit errors, distortions, etc.

MAIN IMPRESSION/APPRECIATION

- The subjective impression/appreciation of the recordings

Main attributes, sub-attributes and examples of common descriptive terms for the absolute assessment of sound quality in detail

Main attribute	Sub-attributes	Examples of common descriptive terms
1 Spatial impression		
The performance appears to take place in an appropriate spatial environment	Homogeneity of spatial sound Reverberance Acoustic balance Apparent room size Depth perspective Sound colour of reverberation	Room reverberate/dry Direct/indirect Large room/small room
2 Stereo impression		
The sound image appears to have the correct and appropriate direction distribution of sound sources	Directional balance Stability Sound image width Location accuracy	Wide/narrow Precise/imprecise
3 Transparency		
All details of performance can be clearly perceived	Sound source definition Time definition Intelligibility	Clear/muddy
4 Sound balance		
The individual sound sources appear to be properly balanced in the general sound image	Loudness balance Dynamic range	Sound source too loud/ too weak Sound compressed/natural
5 Timbre		
Accurate portrayal of the different sound Characteristics of sound source(s)	Sound colour Sound attack	Boomy/sharp Dark/light Warm/cold
6 Freedom from noise and distortions		
Absence of various disturbing phenomena such as electrical noise, acoustic noise, public noise, bit errors, distortions, etc.		Perceptible/imperceptible disturbances

7 Main impression

A subjective weighted average of the previous six attributes, taking into account the integrity of the total sound image and the interaction between the various parameters.