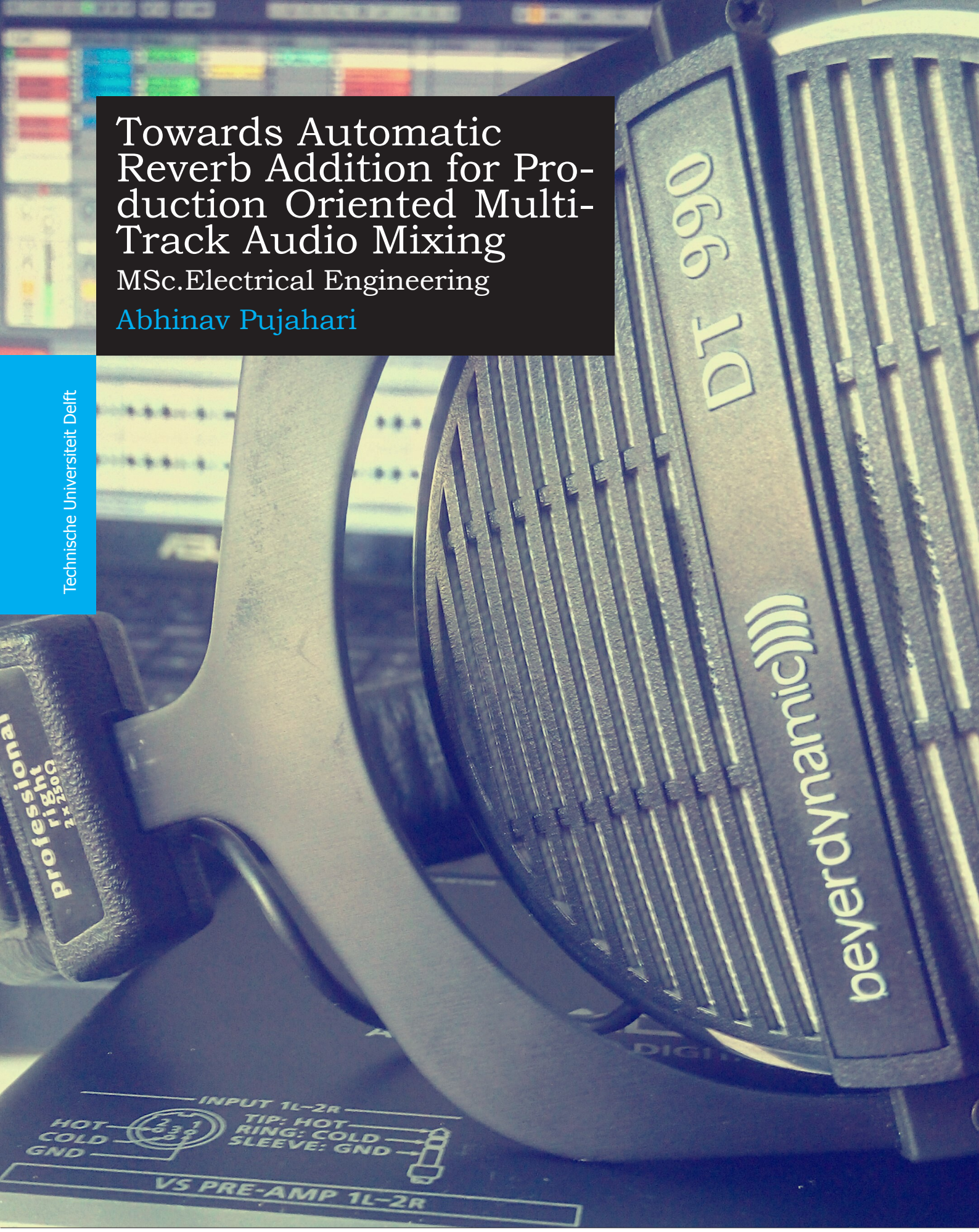
Towards Automatic Reverb Addition for Production Oriented Multi-Track Audio Mixing

MSc.Electrical Engineering

Abhinav Pujahari

Technische Universiteit Delft



TU Delft Delft University of Technology

Challenge the future

Towards Automatic Reverb Addition for Production Oriented Multi-Track Audio Mixing

MSc.Electrical Engineering

by

Abhinav Pujahari

in partial fulfillment of the requirements for the degree of

Master of Science
in Electrical Engineering

at the Delft University of Technology,
to be defended publicly on Wednesday, August 23, 2017 at 01:00 PM.

Supervisor:	Dr. ir. Cynthia Liem,	TU Delft
Thesis committee:	Prof. dr. Alan Hanjalic,	TU Delft
	Dr. ir. Joost Broekens,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

Preface

The inception of this thesis began with a deep personal interest in music and music mixing, and a desire to combine my personal interests with the academic skills gained at TU Delft. A labour of love two and a half years in the making, this thesis has required an immense amount of hard work and struggle. The nature of this thesis, being interdisciplinary in research required extensive ground work, survey and an analytic approach to compile and combine the information from multiple fields.

From surveying the current state of academic research on audio mixing in general and reverberation in particular, to reading books on studio practices and conversing with audio professionals to gain an insight into their methods and practices; this thesis represents a sincere attempt to analyze the problem, through every particular professional direction.

With the time taken and the struggles endured, this thesis is not merely representative of my academic skills and personal interests, but also of my personal growth from an adolescent to an adult. The process of compiling and creating this body of work has truly helped me to mature as a person and gain a deeper understanding of myself and my place in the modern information centered techno-capitalistic hierarchy.

This thesis would not be complete without the help, support and understanding of a number of people. Primarily, my daily supervisor Cynthia Liem, whose patience and understanding, along with her professional guidance, shaped this thesis from a general idea to a detailed and focused body of work. I will be eternally grateful to my parents, my elder sister and my brother-in-law for their emotional support and unending love. The persistence of their support, even through delays and struggles underlines their commitment to my success. I owe my gratitude to Academic counselors Jolien Kooijman and Eva de Haan at the EWI, and the Central Student Counselor John Staals, who helped keep me centered and focused on my work, and assisted me to conform with the rules and regulations of the university. Deep regards also go to friends Marlijn Helder and Regina Hoffman who acquainted me with professional producers and helped me with proofreading and improving my thesis respectively. Vivek Kannan and Istvan Deak exemplified the meaning of friendship with their constant support through catharsis and regular conversation. Acknowledgements also go to Jeroen and Yuri for their understanding and support.

Abhinav Pujahari
Delft, August 2017

Contents

1

Abstract

Sound spatialization is a natural, intuitive but sparsely researched topic in multi-track audio mixing. Although a lot of research has been devoted to the automatic fader gain settings, addition of dynamic range equalization and related effects, delay and Reverb have taken a backseat. The dichotomy in the artistic and engineering approaches to audio mixing have resulted in studio best practices not given their due with suitable algorithmic interpretations.
Due regard to studio practices along with a more holistic approach combining all the steps of audio mixing are especially necessary in the background of the exponential growth of bedroom studio producers and musicians, mixing and crafting their tracks personally. The additional growth in the availability of faster personal computing only fuels this trend.
This thesis attempts to be an exploratory foray into the addition of Reverb to production oriented multi-track mixing. Taking into account studio practices, 2 different algorithms are compared with a professionally mixed track and an unreverberated reference track. The results from hidden reference listening tests are analyzed to draw conclusions of the effectiveness of automatic methods of Reverb addition against the professionally mixed track. The results suggest that the current algorithms implemented are unable to reach the subjective perceptual quality of the professionally mixed track. However, some important conclusions are drawn from the theoretical and experimental research which provide clear guidelines for possible future implementations.

2

Introduction

This section introduces the topic, the central question of the thesis and provides a general overview of the structure of the thesis representing how the ideas and research is presented.

Music is an inherently temporal artistic experience; portraying the creativity of musicians optimally requires mixing of multiple sources (that might be multiple instruments or synthesized and processed tracks) to create a complete creative experience. Based on its temporal nature, the way the brain interprets and makes sense of music is tied to the physical processing characteristics of the human ear and our ability to naturally integrate different spectral scales to develop our perception. The work of Glasberg and Moore [1] on mathematically defining our perception of sound based on the biological processing characteristics of the human ear has directly led to a proliferation of research on audio synthesis and audio processing. Work on automatic audio mixing is a recent entrant to the world of audio processing, held back by the computational requirements necessary to be able to perform accurate real-time digital manipulations on multiple audio streams simultaneously.

Easier accessibility to increasing computational power afforded by more powerful personal computers has given way to an explosion in amateur home audio production. Professional audio software and digital audio workstations such as Ableton Live, Protools etc. enable even non-trained individuals to create high fidelity audio at home. These amateur producers rely on the quality of algorithmic processing, intuitive user interfaces and automatic effects provided by such software in order to compensate for their lack of professional production training. This creates a commercial niche for supplying automatic production software and algorithms which are reliable, mathematically robust and implementable in real time and most importantly, can create professional sounding results.

2.1. Motivation

The research on automatic mixing thus far has focused on fader gain setting [2–5], panning [6], filtering [7] and the like, with spatialization and Reverb addition taking a backseat. The work on Reverb addition is limited to recommendation of a particular plug-in [8, 9] or subjective control using user input [10]. The extent of this research space is sparse and it's goals and methodology disconnected from popular studio practices.

The need for Reverb addition is now exacerbated by the popularity of synthesized music which is not vocalized or played using physical instruments and is thus lacking in natural harmonics and spacial characteristics necessary to place the listener in a virtual physical space. This serves as abundant motivation for us to pursue a path to exploring and understanding studio practices better, and to implement them towards Reverb addition using robust mathematical algorithms.

2.2. Central Question and Objectives

This research aims to be an exploratory foray into automatic Reverb addition to multi-track production oriented audio. We aim to investigate the efficacy of multiple algorithms in adding Reverb to multi-track audio. The algorithms will be adapted from current research on automatic mixing that were used for other tasks in their original form. Adaptation of said algorithms will include changes that

account for the differences between the original function the algorithm served against the properties of added Reverb and also take into account popular studio practices. The results obtained will then be subjectively tested against each other and against a professionally mixed track using listening tests in order to assess their eventual capabilities.

The central question of this thesis is then, whether automatic Reverb addition algorithms can compete with a professional producer in terms of subjective production quality. Supplemental questions include discovering the reasons for the sparsity of research concerning Reverb and the lack of consideration to modern studio practices.

The objectives of this thesis include:

- Studying the history of Reverb and its use in modern production to better understand its importance, objectives of use and best studio practices involved.
- Compiling research that deals with Reverb and discuss their relevance to the current thesis and to the field of automatic music mixing.
- Designing an experimental set-up to form as the basis for the algorithmic interpretations of automatic Reverb addition.
- Adapt and implement pre-existing audio mixing algorithms to the addition of Reverb according to the experimental set-up designed.
- Set-up and conduct listening tests to evaluate the performance of the automatic Reverb addition algorithms against professionally mixed tracks.
- Analyze the results in order to explain their successes and discrepancies and define a path forward for automatic reverb addition.

2.3. Overview

The paper is organized according to the objectives listed above. Chapter 3 discusses the history and relevance of research in order to provide a basis for better understanding the place and need for this research. Chapter 4 covers current industry practices in the field of Audio Production; i.e; common practices used by studio professionals and producers. Chapter 5 discusses the state of current research regarding automatic mixing in general and Reverb in particular. Chapter 6 describes the common experimental set-up that will be manipulated by the algorithmic interpretations and explains the assumptions and effects used in order to create necessary music tracks required for input into said automatic Reverb implementations. Chapter 7 and 8 describe two different automatic Reverb adaptations based on current research on automatic mixing. Chapter 9 discusses the setup of the subjective listening tests necessary to compare the perception of tracks produced by professional producers versus the tracks created by our automatic implementations. Chapter 10 contains a discussion and analysis of the results obtained and includes a section on possible shortcomings. Chapter 11 concludes with a summation of the research and possible future work that can be attempted.

3

Reverb: History and Relevance

Reverb is a ubiquitous property of enclosed spaces, caused by the propagation of sound through the space and the multitude of reflections that bounce off the walls and structures in the room. Reverb is essential to create a sense of ambiance and depth, and its importance cannot be overstated in an age where sampled and dry synthesized sounds form a major part of music production. The following sections define reverberation from different viewpoints and lay the theoretical groundwork for adequately explaining and justifying the design of our experimental set-up and the assumptions used in our implementations in the following chapters.

3.1. Lexical Definition and Discernability

The simplest definition of Reverb and arguably the most accessible, comes from the Oxford English Dictionary (latin root): "be repeated several times as an echo". This defines the most essential nature of Reverb as consisting of reflections of a singular source. Accessibility doesn't necessarily lend itself to visibility, and the average music listener does not discern Reverb as a distinct effect, but rather as a part of the produced music itself. Fortunately, Reverberation is the only word in common parlance that distinctly and succinctly defines an auditory manifestation of a physical space.

The human auditory system breaks down a sound event into three independent components: identity of the source, its spatial position and an image of the space where the source and listener are located [11]. This contributes to Reverberation being able to unify individual stems in a production, since the distinct sound events are made to appear in the same spatial image. This also leads to the unique problem that well produced natural sounding Reverb is generally not noticed by the listener even though it melds the entire production together. Excessive and poorly produced Reverb becomes discernible due the sound artifacts it can create in a production.

Excessive reverberation can lead to loss of intelligibility and smothering of sound events, reducing the quality and palatability of produced music. Reverb is only reliably noticed by untrained individuals in large public spaces such as airports, movie theaters etc, but since this thesis specifically deals with Reverb in music production, this aspect will not be addressed.

3.2. Musical Evolution and Impact on Culture

Music has been prevalent in human communities since before the advent of written history. Being an important community bonding tool and in many cultures and a way to experience the divine; music has been used to describe and document humans' relationship to the world for centuries. Temples and churches, being the representations of the divine have played a key part in shaping a religious view through their acoustics. Architectural traditions influenced and shaped the development of their music before mathematical and statistical analysis made it possible for spaces to be designed based on the acoustics desired. Hard surfaced and large volume churches led to long reverberation times, generating an overwhelming sense of musical envelopment. Reverberation thus, conveyed a sense of God's omnipotent power in many cultures as it could impact and impress a believer without being seen [11]. Musical liturgy in Christianity has been attributed to the Roman emperor Constantine, who

constructed mammoth churches, modeled after Roman Basilicas in order to accommodate thousands of worshippers [11]. Many such instances now are important examples of architectural history (staircase at the Mayan pyramid at Chichen Itza can recreate the sound of the sacred quetzal bird through echoes of a hand clap; the size of St. Mark's cathedral in Venice afforded 2 different choirs at opposite ends of the church, leading to specific pieces of music using cori spezzati and antiphony), but they now also help us understand the relevance of music and the importance of Reverb in shaping cultural and religious traditions.

In the modern age, music has evolved from a solely shared, often religious experience to being more of an individual experience, with media players and associated technologies being marketed as personal devices. Religious liturgy and music still exist, but their consumption is far outweighed by the demand for commercial music. The consumption of music has also been democratized through on-demand mass broadcast mediums such as Youtube, Spotify, Soundcloud etc. This has all been made possible, first through mass broadcast mediums such as Radio and TV and more recently through digital mediums and accessibility of computational power. The march of technology has made creating and sharing commercial music, within the reach of anyone with a personal computer. Digital storage mediums and computers have also created digital studios and digital audio workstations which permit not only much superior production and audio quality, but also far finer control and complexity that would have been unimaginable to producers of the last century. Studio production has moved from magnetic tapes that restricted the number of stems to 5-8, to digital storage now, where productions with over a 100 stems are not unheard of. Since it is now possible to add a sensation of physical space to any piece of music with synthesized Reverb, this has led to an inversion of the earlier cultural practice of creating music for a specific physical space. Producers now have the flexibility to adapt both recorded and synthesized sections of music, with different spatial characteristics in the same production. This flexibility has afforded more creative choice to the producer and blurred the distinction in the responsibilities of a producer and an artist/musician.

3.3. Mathematical Analysis

The mathematical analysis of Reverb owes its existence to the need of architects and sound engineers to design public performance spaces with optimized acoustic characteristics. With the advent of the scientific method as a better way to model and understand a particular space's expected acoustics, sound design engineers became an integral part of the architectural process and composers were no longer constrained by the need to create music based on a space's peculiar acoustic characteristics. Cultural change in design has however, still retained its anthropological nature, with acoustic ideals being defined by both traditions and current science at the time. Case in point being the Boston Symphony hall, the design of which used statistics to compute the reverberation time, while also being helped by traditions and simple design accidents. Architectural acoustic science now depends on knowledge from statistical physics and psychoacoustics as well and music and architecture [11].

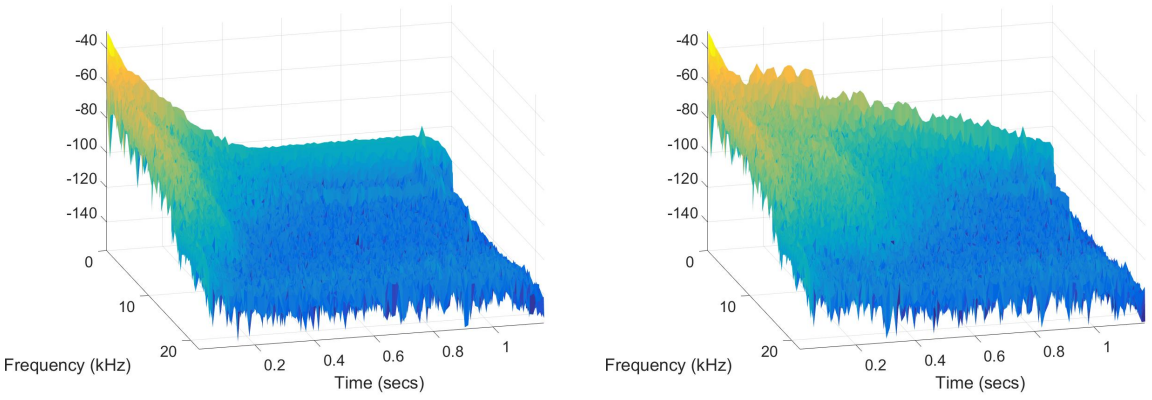


Figure 3.1: Spectrograms of a single snare hit, dry on the left and reverberated with a pre-delay of 100ms, Decay factor 0.200 and Send Fader gain 0.5. Portraying Reverb as being made up of early reflections (delayed echos) and late reflections

Reverberation can be thought of as a combination of 2 different sound events. The early reflections,

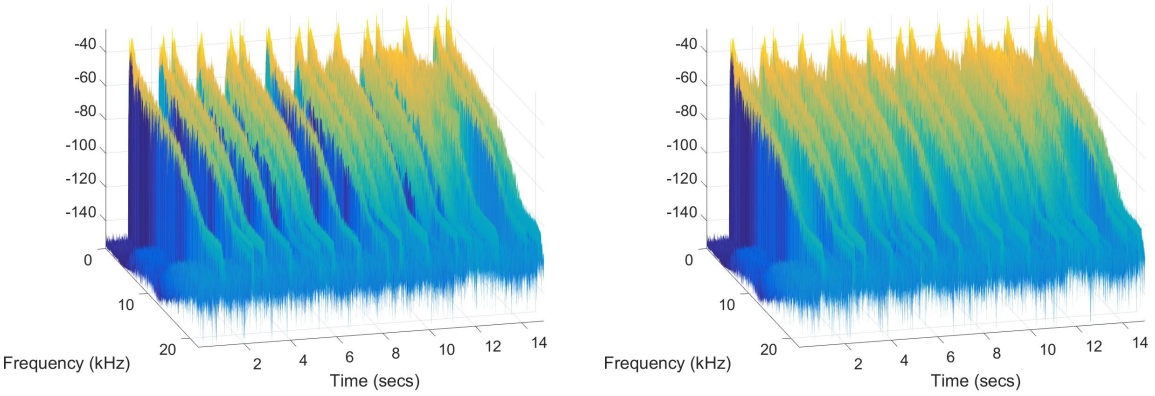


Figure 3.2: Spectrograms of 15sec recorded segments of Drums; dry on the left and reverberated with a pre delay of 100ms, Decay Factor 0.200 and Send Fader gain 0.5 on the right. Portraying the sound smearing and energy spreading properties of Reverb.

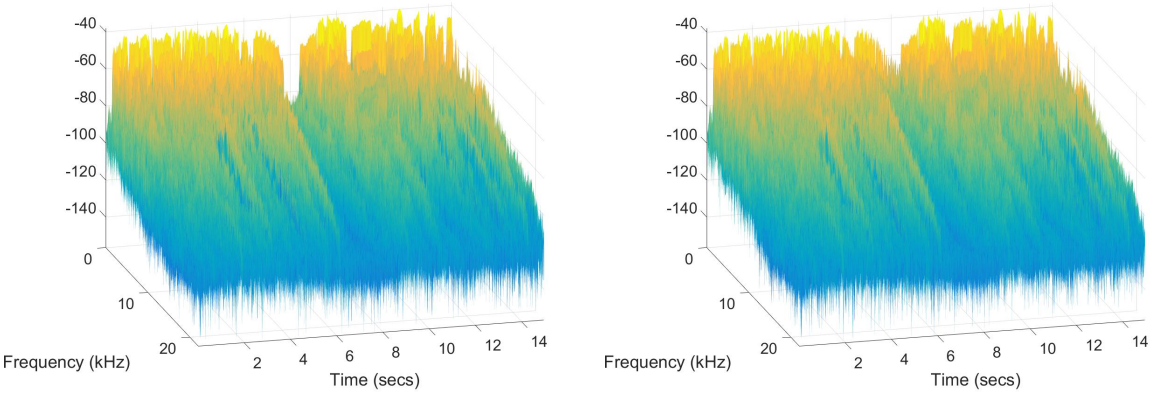


Figure 3.3: Spectrograms of 15sec recorded segments of Vocals; dry on the left and reverberated with a pre delay of 100ms, Decay Factor 0.200 and Send Fader gain 0.5 on the right. Portraying the energy spreading and timbre altering properties of Reverb.

which consist of lower energy delayed copies of the original sound (caused by direct reflections from nearby surfaces) and the late reverberation (caused by the melding of multiple reflections over multiple paths in a space). Figure 3.1 portrays the demarcation of reverb as these 2 different sonic events. Early reflections can clearly be observed at around 200ms, in the spectrogram on the right. Late reflections make up the linear decay of sound energy observed.

Reverberation tends to temporally spread a compact energy package. This temporal spreading can be assumed to be spectral power density conserving, but not phase or time conserving. Such an assumption is valid because although in the long term absorption from surfaces can remove energy, distinct audio sequences having the same power spectra sound indistinguishable when reverberated [12]. Figures 3.2 to 3.4 show audio spectrograms of audio samples (percussion, voice and piano) both before and after Reverb is added. The sound spreading behavior of Reverb is apparent, especially in the drums where the spaces of silence are filled with energy spillover from the previous percussion hits. In the vocals, the sound spreading behavior of Reverb works to change the timbre of the voice.

Perceptually recognizable signal detail is only preserved in the early reverberation, while late reverberation is indicative of a signal's power spectrum. The late reverberations destroy the phase relationships between overtones produced using a musical instrument and replace it with random phases [11]. The early reflections, which consist of direct reflections from a room's surfaces, can be modeled using simple delay lines. The complex nature of late reverberation has prompted a lot of research has focused on finding the parameters that can faithfully reproduce it.

Room dimensions and surface positions dictate the nature of standing waves and resonances that

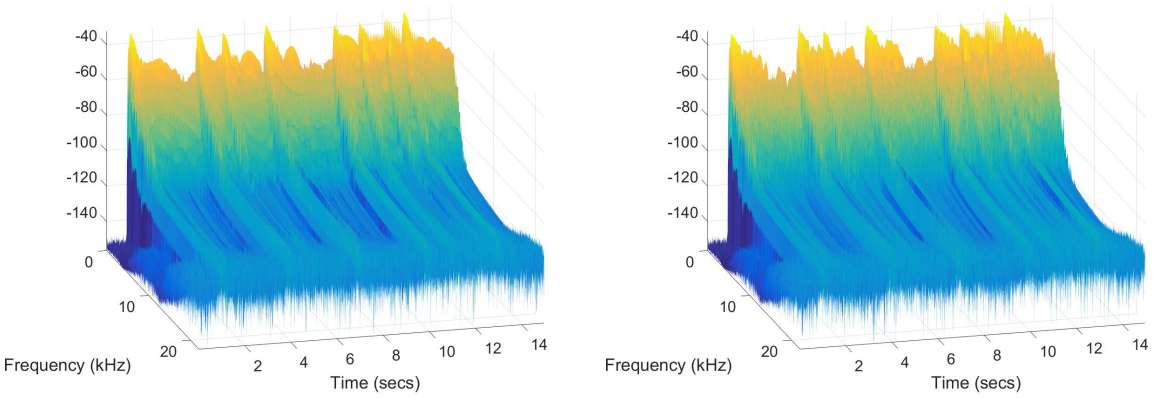


Figure 3.4: Spectrograms of 15sec recorded segments of Piano ; dry on the left and reverberated with a pre delay of 100ms, Decay Factor 0.200 and Send Fader gain 0.5 on the right. Portraying the energy spreading properties of Reverb. Early reflection peaks after every chord are apparent on the right.

can occur, that give Reverb distinctive characteristics that can help identify the particular space which shaped it. Such room resonances are defined using eigentones. An eigentone, when excited is the only frequency that can produce an output signal after the input signal has ceased. An eigenfrequency is the complex number value of an eigentone, the imaginary part defining the normal sine wave frequency and the real part defining the damping decay factor. Thus, the input spectrum determines which eigentones will be excited and the output spectrum is comprised of the sum of eigentone responses. Subsequently, the output spectrum can be rather distinctive for different spaces even with the same input audio spectrum. The source and listener locations can also change which eigentones can be excited, although for large performance spaces with complex characteristics, this can be fairly consistent.

Deterministic and Physical analysis Deterministic models using ray tracing and wave propagation physics are only possible for trivial geometries and are seriously constrained by the inordinate amount of computation and hence, time needed for resolving the solution. Their mathematical tediousness still limits their viability and use even with greater access to computational capacity. The complex nature of Reverb is also shaped due to the textural and absorption properties of a space's surfaces, which only helps to complicate the problem. Deterministic modeling is not just intractable but also seems futile since the human auditory system cannot perceive isolated components in high quality reverberation [11].

Simulations using scale models of performance spaces also remain intractable in spite of their intellectual attractiveness (straightforward execution with fewer assumptions) due to issues with acoustic medium densities in the physical case and the difficulty of adapting different metrics to a small scale model. Lower frequencies tend to show longer reverberation times than higher frequencies due to air and surface absorption. The surface properties of a space as mentioned in the last paragraph are difficult to adapt on a small scale and surface absorption properties also depend on the angle of incidence of the sound wave [13].

Recording impulse responses of real performance spaces has found traction, but their implementation in production assumes the acoustics of the performance space to be time invariant, which simply isn't true due to acoustics being influenced by the occupancy, temperature changes and other factors. Execution also requires convolution of the audio file with the recorded impulse response of the physical space, which carries a significant computational burden and can limit the number of stems that can be mixed in a single production before performance degradation sets in.

Statistical analysis Stochastic models have fared better with regards to academic attention and technological execution owing to the realistic production of reverberation effects with acceptable computational burden. They also seem more in tune with the perceptual characteristics of the human auditory system, since it does better at recognizing parameters of statistical ensembles rather than individual sound events.

The framing of Reverberation in statistical terms began with Sabine's important contribution [14] which was to frame the reverberation time as a function of the room's surface area (S), volume(V) and average surface absorption coefficient(α).

$$RT_{60} = 0.16\frac{V}{S\alpha} \tag{3.1}$$

This is possible due to the calculation of the mean free path of sound waves being dependent on the size of the room. This assumes that the physical process is stochastic and Ergodic, i.e; the collection of echoes and eigentones have a statistical regularity. Owing to these assumptions, this model only works when the room shape is sufficiently complex and the reverberation process displays short enough mixing times (time taken to reach statistically equal energy in all regions) to increase simultaneous echo arrivals to a certain minimum. The model still remains sufficiently accurate for audio engineers to analyze designs of performance spaces.

Moorer [15] showed that late reverberation can be modeled using exponentially decaying random noise as a synthetic impulse response. This is possible since the temporal spreading behavior of reverberation removes coherent phase relationship between overtones in music. The system ignored early reflections and the reverberation produced sounded sufficiently natural but was not identifiable as produced by a particular space. Schroeder [16] extended Sabine's fundamental approach to derive a set of statistical parameters characterizing the frequency response of the random impulse response. All of Schroeder's metrics are either directly or inversely proportional to reverberation time, which remains the only degree of freedom. This model also depends on assumptions of statistical regularity and the presence of a sufficiently large number of excited eigentones. This means, that the model is invalid at low frequencies in a real space owing to the lack of sufficient eigentone density. It is also not valid in the early part of the echo response due to low density of echoes prior to full mixing.

Temporal envelope statistics Temporal envelope perception by the human auditory system leads to the perception of modulation or regularity in the temporal envelope as reverberation artifacts. The auditory system demodulates the envelope and converts it into an internal auditory signal [11]. The auditory system adds a layer of non-linearity to the sensitivity in envelope variations. A sensitivity peak exists at 4Hz. Lower envelope rates than 4Hz are heard as wavering while higher rates sound rough and harsh. Perceptual experiments [17] have shown that this additional auditory processing mechanism that demodulates the envelope independently has a 1ms integration window. This has design implications since we can infer that any reverberation signal whose short term energy envelops is smooth when viewed through a 1ms window will not have a sense of roughness or harshness to it. This also affirms the need for a minimum echo density, since the signal would then have uniform Gaussian statistics according to the Central limit theorem, and hence show more of a smooth decay.

Perceptual characteristics: Artifacts observed, limits of reverberation When it comes to produced music, quality is paramount. So, it is important to characterize and understand the sound artifacts that can appear due to defects in the reverberation process. The different sound artifacts specific to the reverberation process, and their causes are enumerated below.

1. Spectral flutter: A direct result of the input signal's spectrum, can be caused due to interaction between neighboring eigentones and/or due to the input spectrum being narrow-band or spectrally pure that results in short term variability in the spectral balance.
2. Temporal flutter: This originates from defects during the sound spreading process, causing energy gaps or peaks in the short term energy envelope. Depending on the frequency of the temporal envelope variability, it can be perceived as discrete echoes, a coarse harshness or a tonal buzz.
3. Tail flutter: Caused due to non-uniform energy decay in the reverberation tail.

These effects can be better understood by discussing the limits of the reverberation process. Most artifacts occur because the spreading process is non-uniform or shows some defects, leading to insufficient complexity and echo density required to reach Gaussian statistics. This can occur either due to the input being unable to excite enough eigentones or the design of the reverberator itself (whether physical or digital) causing defects in the spreading process. Signal bandwidth is important to determine the spatial distribution of sound in a reverberation space. Signals with small bandwidth relative

to the reverberation time can cause a pronounced standing wave pattern. Spectrally pure tones can sometimes sound unreverberated because the output will also be a sine wave. Multiple overtones in music can help to mask individual spectral flutters as the composite signal is smooth compared to the flutter of a singular overtone. Audio signals containing only the fundamental note aren't typical in music and flutter caused due to spectral purity of the input is generally avoided when the input signal is of sufficient quality. Multi-track projects are also the norm in modern music production, and such an arrangement can also help meld and mask the defects that individual tracks might have produced when heard in isolation.

If the sound spreading process is periodic or too slow, causing energy gaps and peaks in the temporal waveform of the reverberated signal, it can lead to temporal and/or tail flutter. As explained above, perception of such flutter is a function of the human auditory and nervous system. Physical rooms with exceedingly simple rectangular layouts and lack of complex shapes and surfaces can cause such defects due to a greater chance of formation of standing waves, which form energy peaks in the decaying reverberation. In artificial reverberators, such defects can arise due to the lack of sufficient delay lines or faulty design that prohibits all the delay lines being filled.

Although spectral and temporal flutter arise from different processes, they often sound the same to an average listener. A short tone burst can show the effects of both types of flutter due to the reasons explained before. This facet is important to keep in mind while designing artificial reverberators since similar sounding artifacts might have completely different design solutions.

Ideally, the goals for artificially produced reverberation should be [11]:

1. The short term temporal energy envelope should be smooth when viewed through a 1ms window.
2. It should sound like appropriately filtered and shaped decaying pure white noise.
3. Coloration should be absent even at the end of the reverberation.
4. For a tone burst, the output power density spectrum should be similar to the input.
5. During the decay, the amplitude envelope should be random but with a constant spectral composition.

3.4. Technological Execution

The human cognitive process inherently places all sounds in an internal world picture; i.e, non-spatial hearing does not exist [18]. Reverb has been an inherent part of music and audio from time immemorial, purely because of the human auditory system's capability of 'placing' sound within an imaginary space. Technological advancement has given producers finer control of the aspects of reverberation, which makes the use of 'non-natural' sounding reverbs possible as a purely creative tool. Rather than being confined to the spatial aspects of their own studios, producers can now explore and create new spaces in which to place their tracks in. This subsection chronicles the evolution of artificial reverberation technology through the last century.

The very first way of adding Reverb to a mix included capturing the natural reverberation of the space the sound was recorded in and adding in as much, or as little of this ambiance track the producer felt was necessary. Producers would position musicians in a room with relation to how far and how loud they should appear in a mix and capture the ambiance using a set of microphones [19]. Using a pair of microphones, a stereo recording of the room ambiance can be created which manages to capture the highest complexity that no artificial Reverb would manage to create in real time. In-spite of the quality of Reverb captured in such a recording, the only aspect of freedom the producer has, is to modify the reflective nature of the surfaces using acoustic deadening or reflective panels. There is very limited scope of modifying the size of the room and the resultant reverberation time. It can cause serious problems if the final recording does not fit into the creative vision the producer has for the final mix, and individual Reverbs for particular tracks and/or instruments also can not be designed.

Specialized Reverb chambers were the next logical progression as these provided greater control over the reverberant characteristics of a room, without compromising on the quality of Reverb a real physical space can generate. Any enclosed space where a number of microphones are placed in order to record the sound emitted from a number of speakers can serve as a Reverb chamber. There have been instances of specially built Reverb chambers being used and also pre-existing spaces such as

bathrooms or staircases being utilized as improvised Reverb chambers [19]. All studios can be used as Reverb chambers as long as loudspeakers and microphone lines are present. Such an arrangement can provide greater freedom than a stereo recording of the ambiance during music recording sessions as the physical characteristics of the room can be changed using acoustic materials and, different positional arrangements used with regards to the placement of the loudspeakers and the microphones. Since every particular room has its own peculiar reverberant character, all the recordings made even with different recording and acoustic arrangements would sound very similar, with slightly different spectral coloration caused due to the alteration. Loudspeakers also are very different from real instruments in the way they produce and radiate sound. Large spaces with more complex reflective surfaces will also produce better reverberant characteristics when compared to smaller rooms, which have a propensity towards formation of standing waves.

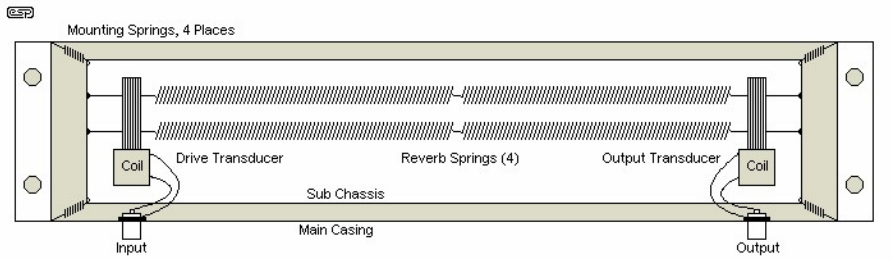


Figure 3.5: Schematic of a spring reverb unit, Courtesy of Elliot Sound Products [1]

Spring Reverb was possibly the first real 'portable' artificial reverberation device. Conceived by Bell Labs researchers trying to simulate delays that occur over long telephone lines but developed by engineers from the Hammond company in 1939 [19], spring Reverb can still be found in guitar amplifiers. Such a device uses a system of steel springs between electrical transducers to simulate reflections. An input transducer vibrates according to the input sound it is fed and in turn vibrates a spring whose motion is detected by an output transducer at the other end of the spring and converted back into an audio signal. Some parts of the vibrations bounce back and forth between the 2 ends, simulating the characteristic of multiple reflections that reverberation requires. Since the reflections only occur in a singular dimension (along a straight line) between the two ends of a spring; a spring reverb is not capable of simulating a rich natural sounding Reverb. They also come with limited controls and pre-delay and decay times are generally fixed. But, the highly colored and peculiar nature of a spring Reverb has found favor as a distinct effect and sound rather than being used to simulate natural Reverb.

Plate Reverb was the progression of a physical Reverb simulation system from one dimension to two dimensions. Being operationally similar to a spring Reverb, the vibrations are transmitted through a thin metal plate suspended in a wooden box. The vibrations travel through the plate, bounce off the edges and corners and are picked up by an output transducer. The first plate Reverb was the EMT 140, built by the German company EMT [19] in 1957. Weighing 600 pounds, as wide as a wall of a small room (2.4mX1.3m), these units were neither portable nor inexpensive. They did provide a bit more control over the decay time as compared to spring reverbs with a dedicated damping mechanism and the frequency response was more musical. The Reverb produced still did not resemble a natural Reverb, but akin to the use of spring Reverb, plate Reverb has found use more as an effect rather than to produce a certain ambiance.

Manfred Schroeder [20] is credited with the development of the first digital reverberation system in 1961. It consisted of a simple circuit with feedback assisted delay lines. The world's first commercial reverberation unit was produced by EMT with help from Dynatron in 1976, and was called the EMT 250. Offering basic controls like pre-delay and decay time, it could also produce effects such as delay and chorus. Many years of research and development aided by faster DSP chips with greater computational capability have birthed large complicated architectures seeking to simulate the rich complex nature of a natural Reverb. Companies tend to keep their architectures secret, and specific Reverb units are

liked and used for their distinctive sound by producers rather than their realism [21]. The commonality between most, if not all, architectures includes multiple delay lines to simulate thousands of complex reflections, filters and gains to simulate air and surface absorption and mappings that can change the behaviors of the gains and filters depending on the user's requirements. This flexibility offered by algorithmic Reverb to be able to tweak and control every aspect of the reverberation makes them powerful, versatile and popular with producers. But, architectures designed for greater realism can be both expensive in terms of value and computational cost.

Convolution Reverb arose from the simple idea of being able to sample the impulse response of a physical space and then being able to use it later to apply Reverb to any sound, even without access to the particular space. This technology has only seen fruition with the recent availability of faster DSP chips and computers that are capable of performing the millions of calculations required. The Reverb of a physical space is sampled using a pair of stereo microphones to record the room response, either to an impulse spike or a sine sweep. The recorded impulse response is then used to form a matrix that can be convoluted with the audio that needs to be reverberated. Although convolution Reverbs can create lifelike and natural results, they lack the flexibility required to alter the characteristics of the Reverb, since the positions of the speaker and the microphones, and the characteristics of the space are entrenched in the recording. This makes the convolution Reverb entirely dependent on the quality of the initial recording. Also suffering from a high computational burden, convolution Reverbs have found greater use in movie and film production, since lifelike emulation of specific physical spaces is more of a concern in that paradigm. The inflexibility, high computational cost and fixed nature of convolution Reverbs tend to constrain a producer's creativity, and hence algorithmic Reverbs are generally preferred [19] for music production.

As has been the theme of this thesis, amateur producers now consist of a viable market owing to the explosion in powerful personal computing. Digital audio workstations(DAW) like Ableton and FL studio now provide all the tools necessary for a producer to create a quality record. Real-time plugins for all sorts of effects are now both provided along with DAWs, and sold separately. Real-time Reverb plug-ins based on both algorithmic computation and convolution are available. They do require the user to be familiar with the options and controls in order to be able to create a suitable Reverb for the effect intended. A majority of these plug-ins also provide presets based on the size of the room or the kind of ambiance a producer might want to use. Although these provide a number of straightforward choices for the amateur producer, they also tend to curtail creativity and can cause problems with track cohesion and formation of audio artifacts if the preset parameters are not fine tuned further. These provide good starting points, but any producer striving for quality would need to adjust the settings of each such preset in order to craft the sound that best suits the mix.

This chapter helped to understand the evolution of Reverb as something natural and intuitive, a universal property of performance spaces to something desired and finely controllable. The importance of controllable synthesized Reverb in the age of increasingly dry electronically created music cannot be understated. The discussion on the technological evolution of synthesized Reverb segues appropriately to the next chapter, which discusses studio practices in general and the use of Reverb by producers in particular.

4

Studio Practices

Music mixing, arguably is as much of an art form as is music composition. In fact, without proper mixing and mastering, the details of multiple instruments in a large mix consisting of tens of different tracks, can become inaudible or worse, contribute negatively as noise. The job description of a music producer has evolved over the years, as a response to the technological innovations in recording and modifying sound [22]. Such an organic natural evolution has led to the development of multiple techniques to achieve similar end goals and contributes to the understanding of the process of production as being subjective and somewhat unsystematic. Understanding the end goals of music production however, does help us to find patterns in the process that enumerate the various steps necessary to transform raw audio recordings to marketable music.

This chapter details the current industry practices of production oriented music mixing and briefly explains the reasoning behind such practices. The further sections deal with the history of music production, the basics of music mixing and the application of Reverb respectively.

4.1. History of Music Production

The modern idea of music production generally refers to an individual responsible for recording, mixing and refining the sound of a particular music artist. This is a concept that has only recently been stabilized in the industry, especially after the advent of digital audio storage, reproduction and alteration. The possibilities that digital media provides with respect to having multiple takes, unlimited storage and a myriad of effects and refinements forced record companies to employ producers as specialists. Specialists whose role began to be seen as being technically creative and responsible for shaping the sound of an artist into a market ready success.

Before the advent of digital media storage and reproduction, producers were entrusted with multiple responsibilities. Bobby Owiniski demarcates the evolution of the modern music producer into multiple eras [22]. During the "early record label era", producers, limited by the technology of the age, acted as archivists rather than someone responsible for the creative direction of the music. This early era, beginning from about 1900, from when recorded music was turned into a business, required producers to be travelling entrepreneurs; scouting for talent and acting as technicians in the recording process.

The invention of magnetic tape recording was an important part of the "Mature Music Era" [22] which afforded producers more responsibility and control. The ability to use multiple takes and multiple tracks enabled producers to separate musical instruments, opening up a world of whole new possibilities. This is the point from when music producers began to have much more control over the creative process and were required to use their ingenuity to bring the best out of an artist in the studio. "The Independent era" brought about a revolution in the business aspect of music production, where producers went from being important but obscure record label employees to well known industry respected professionals being responsible for an artist's success [22]. The increase in their salaries and their treatment as independent employees hired by the artist, highlighted their importance in the music industry and their responsibility as technically creative professionals responsible for moulding a particular sound. Some producers such as Moby, Dr. Dre, Tony Brown and Dan Huff have been credited with either creating or changing the direction of a particular style of music.

4.2. The Process of Music Mixing

Music mixing; an important part of the responsibility of a music producer, has evolved with the use of multiple recording media, the quality of effects processors and the computation ability of modern mixing consoles. This section attempts to explain the modern music mixing process based on a digital audio workstation and the differences between the technical and creative responsibilities of the producer.

In short, the goal of music production in general and music mixing in particular is to combine individual tracks (which could be separate instruments or synthesized elements) into a coherent, pleasing and creative whole. This requires an intuitive understanding of the psycho-acoustic principles of how music is understood and processed by the brain. This includes moulding the sound to best use the listener´s attention and to creatively use each track to build up the possible emotions and ideas the artist wants to convey.
The production chain for recorded music involves, in sequence: Songwriting, Arranging, Recording and Editing, Mixing and finally mastering [19]. Modern music genres such as Dubstep, Electronic Dance Music etc use a lot of digitally synthesized music samples, which do not require recording in the traditional sense, and hence for these samples, recording is not part of the production chain. In the interest of brevity, this section will only deal with the Mixing aspect of the production chain.

4.2.1. The Concept of a Mix

It is essential to consider whether the individual elements or the whole mix itself holds more importance to the listener. This helps us decide whether we alter tracks in isolation or while listening to them in a mix perspective. There are a lot of tracks which might sound incomplete or lacking in isolation, but fit in perfectly with the rest of the mix. For example, high pass filtered solo vocals often sound 'flat' and unnatural when heard in isolation, but this process of high pass filtering is what helps to increase their definition in the overall mix.
Keeping this distinction in mind is especially important when considering Reverb, in order for us to understand its importance in creating a holistic sound-stage and a natural sounding virtual environment for the listener. Reverb not only helps us to 'place' particular instruments on a virtual sound-stage, but also helps to make individual tracks meld and sound natural together by shaping their temporal and frequency characteristics concurrently.

4.2.2. Approaches to Mixing

As mentioned in the previous section, the responsibilities of a producer and the process of mixing have evolved along with the technological advancements in the recording industry. This natural evolution has led to multiple approaches to achieve the same end means. Some of the more systematic approaches [19] used are described below. It is important to note that there are a myriad variations to the approaches given and the list is by no means exhaustive.

Serial Approach Starting from a few individual tracks, which are listened to in isolation and processed accordingly, more and more tracks are subsequently added and mixed. Although this helps to focus on the individual elements, it requires constant changes and runs the risk of making the mix too spectrally full and muddy.
This approach can include mixing tracks by their utility in the song i.e; the rhythm section first, then the harmony and then the melody or mixing tracks by the order of their importance to the genre of music, e.g; the beat being the most important in hip hop etc.

Parallel Approach This involves setting up a rough mix and approaching the mixing tasks one by one, i.e; setting up the faders, panning and effects processing etc. This approach has the advantage that the mix-perspective is retained, and none of the tracks are mixed in isolation. But, this approach can become difficult with increasing number of tracks and requires more focus and time to get the individual elements right.

Mixing by Section In many cases different sections of a song i.e; the chorus, intro, the verse, the bridge etc have different mixing requirements since different instruments need to be highlighted or

certain instrumental hooks need to be provided to keep the listener´s attention. Mixing by section is a useful approach in such a situation as the emotions that need to be conveyed by each section can be kept distinct from each other. This approach can be attempted either by chronological order (the intro first, then the verse, then the chorus etc) or by the order of importance (the chorus generally is the most important part of the song).
This approach is an order of magnitude higher than either the serial or parallel approach. The sections can be mixed either with individual tracks, or on a parallel basis depending on the producer´s preference.

4.2.3. Order of treatment in Mixing

There are a number of different effects processors that are used to prepare raw audio for mixing and to process audio further in order to make it fit better into the mix. The standard treatment order for audio is: Faders, Pan Pots, Processors, Modulation, Delay, Reverb and Automation [19]. These effects are described below.

Faders Faders constitute the level control for each individual track and are used to set the respective amplitude for each track based on its importance and its required audibility in the mix. The general approach to setting levels is determined first by their required audibility in the mix and then by the maximum level the audio can be set to before clipping sets in. Clipping is an audio waveform distortion caused by overdriving an amplifier, forcing it to produce a voltage or current beyond its design standard.

Pan Pots Pan pots are rotary dials either physical or virtual (in a digital audio workstation) that determine the balance of level between the left and right channels of a stereo mix for a particular track. There are certain guidelines that are followed when it comes to panning since it is an important component of virtual spatialization of the individual tracks. Hard panning is generally avoided, although not completely eliminated. Low frequency sources such as bass guitar or a kick drum are panned to the centre along with the main track (which is generally the lead vocals). Frequency balance between the left and right channels and monophonic compatibility must be maintained while panning different sources [23].
Low frequency sources are generally panned to the centre because the brain uses interaural time differences to localize low frequency sources, and not interaural level differences which are caused by a sound shadow forming due to absorption of sound waves by the head.

Processors The most important processing exercises are generally regarded to be equalization and filtering. Both of these treatments focus on either removing or enhancing certain specific spectral regions of each particular instrument or track. This is done in order to make space in the mix and to avoid multiple instruments occupying similar spectral regions, which can cause muddying of the mix and/or mask the details of those instruments.
Other important processing techniques include Compression(reduction of the dynamic range of a track), Gates(muting tracks below a certain amplitude), Duckers(temporarily reducing level) and Expanders(opposite of compression). Processing techniques like these are again, predominantly used to maintain spectral compatibility, to increase audibility of details and to eliminate noise.

Modulation Modulation refers to a set of specialized effects processing units which can lengthen of shorten sounds in a cyclic fashion, either temporally or spectrally. This can create effects such as Vibrato, Flanger, Chorus etc. Apart from being an important creative component, it helps to add depth to the mix and create an interesting background to prop up more important tracks.

Delay Delay is a simple but important effect that refers to an echo or repeat of the track being delayed. The most basic function is to delay the input signal by a set amount of time. This helps create a sensation of space, apart from being a creative effect that can fill the empty space in a mix. The delayed signal can be further processed to change its spectral characteristics, or be modulated in order to create a unique effect.

Automation Automation is not an effect by itself, but refers to specific processing actions that are automated (played back in sequence) while the final track is being bounced.

Reverb This processing effect being the focus of this thesis, is described in greater detail in the next section.

Figure 4.1: Interface of a modern DAW, courtesy of Ableton. Portrays the multiple slots (columns) on which audio files can be placed. Reverb emulators are usually placed on vertical columns labeled "Return" and fed through sends on individual track slots

4.3. Reverb in Audio Production

Crafting the spatial ambiance of a mix is incredibly challenging, but it is essential to elevate the nature of a dry recording by giving it character and space. Arguably, it is Reverb that creates a more cohesive mix out of individually recorded instruments and completes the production process. This section explains the necessity of Reverb in modern audio productions and nuances of its use that can help create a quality recording.

4.3.1. Uses in Production

Reverb is more flexible and versatile in its use than the average person's understanding of its utility. The broad uses of Reverb in audio production are enumerated below [19]. The distinction between necessary and creative uses are marked using either 'N' and/or 'C' in parentheses.

1. **Meld the instruments in a mix (N):** A mix can only sound cohesive when all the distinct instruments or tracks are given similar spatial characteristics. An ambiance Reverb is used to this effect and settings that are congruent to all the different tracks are used. Such a Reverb does not need to be heard explicitly, but felt through the tracks being well integrated and not sounding foreign to one another.

2. **Increase distinction of instruments (N):** Having disparate Reverbs on different instruments depending on their texture or timbre can help increase their discernability to the listener.

3. **Simulate Depth (N,C):** Reverb addition increases the perceived source to listener distance and can be used to 'place' different instruments in an imagined space, giving the mix an additional dimension, portraying an importance hierarchy and resolving masking.

4. **Enhance a mood (C):** Creative use of Reverb, especially the creation of imagined spaces can help the producer create a narrative, tell a story to achieve the required emotional affect.

5. **Livening Sounds (N,C):** Dry sounds i.e; sounds that have no reverberation in them do not sound natural, nor enjoyable. Such synthesized sounds are becoming more and more common in production owing to the popularity of electronic music and its influence. The use of Reverb in the right amount and the right kind, can elevate the perception of a recording in comparison to its dry version.

6. **Filling time gaps (N,C):** As discussed in the previous section, Reverb tends to spread a compact energy package and smear sound. Reverb can be used to fill the empty spaces in a mix and maintain continuity.

7. **Filling the stereo panorama (N,C):** Reverb on an instrument can be used as a distinct track panned either in relation to the dry instrument or disparate to it, in order to creatively construct a stereo panorama.

8. **Changing timbre of instruments (N):** The multitude of reflections from a quality reverberation can meld with the sound of the dry instrument to create a fuller, richer sound that is more palatable to the listener than the dry recorded instrument itself.

9. **Reconstruct decays and natural ambiance (N):** Instruments that are recorded in spaces with a colored or broken frequency responses need to be improved with reconstructed Reverb decays to produce a more natural sounding, non-colored recording.

10. **Resolve masking (N):** Excessive reverberation can cause masking due to smearing of sound, and increase in the temporal footprint of sound events. But when used in a nuanced manner, reverberation can be used to increase discernability between instruments. Instruments can be made more distinct with Reverb treatment by 'placing' them differently in a front-back fashion and expanding the temporal footprint of sound events so that they are heard when the masking instruments have ceased playing.

11. **Realistic stereo localization (N):** Reverbs can be added to mono signals in order to make them sound more realistic, extend their stereo width, make them clearer and better defined. A short Reverb with a realistic early reflection pattern is generally used for such an effect.

12. **Distinctive effects (C):** As mentioned before, plate and spring Reverbs are colored and not considered to be faithful or realistic. They have still found use as a distinctive effect due to their characteristic sound. General realistic Reverb can also be used as an effect to create movement, demarcate different sections and/or to create transitions.

It's important to understand where the use of Reverb is absolutely necessary vs where it is used creatively as an effect. Izhaki [19] in his very popular guide to audio mixing says that the pertinent question to ask is not "How natural is this reverb?" but "How beautiful?" and "How well does it fit into the mix?". Subjective creative quality is the paramount factor in creating a cohesive narrative throughout the production. This is a key insight since it implies that it is possible to create a quality production by being resourceful with limited computational capacity and substandard Reverb plug-ins; something that constrains most amateur producers.

4.3.2. Controllable parameters

The numerous plug-ins available today provide the producer with a myriad of controllable parameters which can be used to model the sound of the Reverb appropriately. A number of common and important controllable parameters are discussed further in this section. Some of the common controls found in modern plug-ins are:

1. **Direct Sound:** This refers to the dry, un-reverberated sound of the track itself.

2. **Pre-delay:** This refers to the time between the dry sound and arrival of the first reflection. This parameter can help us control the perception of the size of the room and the distance between the source and the listener (depth). The relative distance between the direct and the reflected sounds

gets smaller the further away the listener is from the source. This means that the pre-delay is inversely proportional to the distance between the source and listener.

3. **Decay time:** This refers to the time it takes for the amplitude of the reflections to decay by 60dB. Larger rooms have longer decay times owing to greater distances between surfaces. A longer decay time can also suggest highly reflective surfaces in the room.

4. **Room size:** This parameter is popular in the plug-ins due to it being straightforward in its implication. It is connected through the internal architecture to the variables controlling pre-delay and decay time. The larger the room size, the smaller the pre-delay and the longer the decay time. The smaller the room, the greater coloration that appears in the Reverb and the more vigorous the early reflection pattern.

5. **Density:** This parameter can control the density of reflections in both the early reflections and late reverberation (depending on the plug-in). Greater density in the early reflections suggests a small, highly reflective room and a greater density in the late reverberations suggests a larger room. Lower reflection density can help control timbre distortion of instruments and masking, especially for more important tracks in a mix (such as vocals) and textural sounds (such as guitars or organs), while higher reflection density for percussive sounds can help control flutter echo and metallic timbre caused due to the nature of such sounds.

6. **Diffusion:** This pertains to the scattering of sound and hence, a properly diffused field of sound implies greater scattering and a uniform frequency response. Often linked to the density parameter, diffusion can help control the degree to which echo density increases over time.

4.3.3. Methodology of Use

Modern DAWs are structured such that multiple tracks (which can be distinct instruments or sounds) are first added on individual slots, on which effects such as compression, equalization etc, can then be inserted to process the sounds as required. The effects are placed in a particular sequence, which was enumerated and detailed in a previous section. Reverb is generally one of the last effects to be applied and since its nature necessitates either the same Reverb parameters are applied to multiple tracks or at least, a high correlation between the Reverb parameters exists. A separate track is usually set up exclusively for the Reverb effect to which multiple tracks are fed using auxiliary sends.

Although Reverb can be inserted on the track slot in sequence without needing a separate send, such a structure is more computationally taxing since every track will have its own Reverb. This also increases the chances of ambiance collision due to different tracks sounding as if they were in dissimilar spaces. The number of different Reverbs used can vary depending on the needs of the production. A dry sounding production with fewer instruments and an upfront sound can be produced with just 2 Reverbs with varying parameters, while complex expansive productions with multiple instruments and varying sections may need a much larger number. Selecting and fine tuning a Reverb is a time consuming task, especially for amateur producers and quality can be compromised if the producer has to configure and arrange multiple Reverbs for every track rather than just a few that are applied to all the tracks [19]. Reverb inserts are better used for short term applications i.e, specialized Reverb in a particular section of a song.

In general Reverbs are placed on an auxiliary track fed by a post-fader send from the dry processed track. Such a Reverb is configured to provide only the wet sound and its contribution to the mix is determined by the send fader level of the dry track. It is important to control the aux-send fader level of the particular tracks separately to control their perceived depth, since depth is directly related to the ratio between the level of the dry track and the level of its Reverb. On a Reverb set up on a post fader send, increasing the level of the track will increase the dry sound bringing it forward, but also increase the level of Reverb, pushing the track back. Thus, it's important to use a separate aux-send fader to control the level of the Reverb and in turn the ratio between the dry and the wet sound.

There are other specialized set ups catering to particular contexts in production or simply producing peculiar effects. These include coupling delay and Reverb, sending a specially processed ghost track version of the dry sound to the reverb, instead of the dry track that feeds the mix bus, using multiple Reverbs panned distinctly on the same track etc. These specialized applications are beyond the scope of this project and will not be considered.

4.3.4. Possible problems

Often creative paradigms can seem limitless, with the same audio production being capable of being mixed, interpreted and portrayed by different producers in varying fashion. It is often the objective impairments that can bound the limits of such a creative space. The possible problems caused due to the improper use of Reverb are enumerated below [19]:

1. **Smear Definition:** As discussed in the mathematical analysis, Reverb smears sounds, which can cause loss of definition, localization and intelligibility. This often happens due to excessive reverberation in the form of insufficient damping and unnecessarily long reverberation time.

2. **Masking:** Reverbs can mask other sounds due to sound smearing and increase in the temporal footprint. This can be controlled by using appropriate reverberation parameters, or equalization and compression.

3. **Clutter:** This generally refers to the presence of too many concurrent sound events in a mix. In order to keep the listeners interested, it is important to maintain a certain sound stage with appropriate panoramic width and depth. Any use of Reverb that affects this arrangement and forces the listener to pay attention to multiple events at the same time can cause clutter.

4. **Timing:** This problem generally occurs with percussive instruments which rely on their timing for musical effect. Sound spreading can spoil the impression of rhythm and timing and thus judicious and limited use of Reverb on percussive instruments such as drums is critical.

5. **Change timbre:** As discussed before, melding of the multiple reflections produced by reverberation to the dry sound of an instrument can produce a richer fuller sound. But excessive reverberation or reverberation produced in rooms with a broken frequency response can lead to an unpalatable timbre. Further equalization and filtering would be necessary to correct such timbral deficiencies.

4.3.5. Reverb in relation to other effects

Reverb is important in isolation, but always works in conjunction with other effects in order to create a cohesive quality production. Reverb's relation to other effects, including certain overlaps in functionality are described below.

1. **Delay:** Delay is similar to Reverb in that it can lend a sense of depth and space to the audio to which it is applied, but it lacks the late reverberations that can fill the gaps and provide a more natural feel to the production. Delay can be used in lieu of Reverb but only for dry, sparse styles of music which would benefit from a clearer background and an upfront sound [24]. It is also essential to synchronize the delay times to the rhythm of the music without which, the reflections produced would create clutter and confusion.

2. **Panning:** Reverb is used along with panning in order to create a sound stage, to place different tracks and instruments in an imagined space. This helps in creating contrast between instruments, reducing masking and making the listener feel like a part of the space in which the music is arranged. Panning, in the strictest sense is used for left-right spacing and Reverb for front-back spacing. Both of these are required to create a realistic sound stage and an immersive experience. Sarroff and Bello [25] found that correlation between width(panning) and immersion (0.8745) was higher than between reverberation and immersion (0.5679), but the statistical relevance of both correlations suggests that panning and reverberation are used in conjunction with one another to create an immersive environment.

3. **Equalization and filtering:** Spectral balance is one of the goals of music production [19]. Every particular instrument occupies a certain space in the frequency spectrum and equalization and filtering is often used to shape the spectral responses of tracks in order to reduce spectral conflict and clutter. The reverberation responses of instruments are initially shaped by their own frequency response, but the late reverberation response, created by thousands of reflections should ideally be white noise [15]. This necessitates the use of equalization and filtering in order to shape reverberation to maintain the spectral balance in the mix. Excessive sound in one particular frequency space can lead to unpalatable coloration. Some Reverb plug-ins are provided

with built in equalizers [24] that can help shape the sound of the Reverb as required for proper spectral balance. A colored frequency response is undesirable when the track is heard in isolation, but the final goal of production is to ensure spectral balance in the entire mix, and this can be achieved even with colored frequency responses on individual instruments. Frequency response shaping can also be used for rudimentary depth simulation as cutting high and low frequencies makes an instrument sound farther, sounding somewhat like the delays heard on erstwhile long telephone lines [19].

4. **Compression and Gating:** Compression and gating can be used to alter the decay envelope of the Reverb for it to be non-linear and more controlled rather than an ideal linear decay. This is useful in case we want to eliminate non-audible parts of the decay, to maintain a stronger sense of rhythm while maintaining a natural sense of space and/or as a distinct effect. This is extremely useful for percussion instruments such as drums where sound events are punctuated by periods of silence where only the Reverb is heard. By using compression and gating to either hasten or cut off the decay, the Reverb will not interfere with the subsequent percussion events. This maintains the beat and sense of rhythm without causing clutter and masking.

4.3.6. Producer Preferences

There is a certain level of academic consensus in the scientific community for basing loudness equality of tracks [2, 4, 5, 26, 27] for music mixing optimization thus far, but there is no such analogue in the community of music producers. Techniques and procedures seems to be highly variable depending on the different genres the producer is known to work with and which particular studios he frequents for his mixing projects. There are conflicting approaches on setting up ambient microphones, dealing with leakage and preferences for either real rooms, hardware or software emulators [21]. These issues are further compounded by the use of ambiguous terminology in order to describe the characteristics of a mix; terms such as 'boomy', 'tight', 'thick', 'heavy' etc are often used, which lack any clear definition and are often correlated by different producers to different effects parameters.

Producers also tend to have a set way of working and mixing, something that has worked well for them in the past and makes their work-flow more streamlined. There also exist very strong preferences among producers for certain hardware emulators. This is purely human nature, working with things that are more familiar and generally avoiding risk. The type of equipment available at the particular studio and the frequency response of the room where mixing and mastering takes place also shapes the sound. Larger studios offer more equipment, dedicated Reverb chambers, a larger choice of hardware Reverb emulators etc. These choices, along with the level of familiarity the producer has with different equipment tend to create a certain sound every particular producer is known for. In the modern age, the layout of Digital Audio Workstations is also influencing the workflow and the process of recording and mixing, making things more streamlined and systematic.

Furthermore, cultural and societal trends have also driven an evolution in the approach with which Reverb is used in producing records. Many records in the 70s and 80s were awash with Reverb, but preferences have shifted toward more dry sounding productions in the modern age. This has been speculated to be caused by the use of more reflective and larger control rooms where mixing takes place [21]. Since such rooms tend to add a lot of Reverb of their own, producers tend to add less on the record. Some guidelines suggested these days is to have Reverbs inconspicuous, be used for ambiance and melding rather than as an effect and to keep decay times short [21].

One thing all producers agree on is to use their ear to listen for problems, artifacts and clutter. Preferences are only natural and do lend a guiding hand, but training and experience prevents mixing errors from occurring. It is also important to keep the artist's creative vision in mind to create the right effect and sound-stage intended for the listener. They understand that additional ambiance can only be added to a recording and the ambiance already present can not be removed. Thus, separate microphones are usually set up to record the room ambiance and reverberation characteristics, or producers rely entirely on Reverb emulators to add ambiance to their tracks. It is better to err on the side of caution and leave your options open, with additional ambiance tracks recorded rather than let ambiance leak into the dry recordings.

The reasons discussed above help us answer one of the supplemental questions to the central aim of this thesis, i.e; the lack of consideration to studio practices by the academic community. The level of variability in producer preferences and subjectivity in mix descriptions have made a scientific

characterization of the mixing process seem difficult and almost untenable. But, this has changed considerably with the advent of the digital age, and the audio production process being streamlined by the use of specific DAWs and compatible plug-ins. Roey Izhaki's book "Mixing Audio: Concepts Practices and Tools" [19], which is one of the most popular and cited books on Audio Mixing was first published in 2008. Access to computational power has grown by leaps and bounds since then, and the field is now ripe for faster development to serve the needs of amateur home and studio producers.

This chapter focused on understanding the music mixing process and how Reverb forms an essential part of it. The necessary and creative uses of Reverb were clearly detailed, and the conceptual and utilitarian understanding of Reverb from a producer's point of view was also discussed. The next chapter will discuss the present research that deals with Reverb and its automatic addition to produced music.

5

State of Current Research

The interest in automatic mixing research has been on the rise in correlation with an increase in accessible computation power, a deeper understanding of musical cognition, and an astronomical rise in home audio production and Internet music sharing. The field exists not only to provide for the growing consumer base of amateur audio producers, but also to supplement the fields of live audio presentation and sound design.
This chapter takes a detailed look at the current research space and attempts to filter and analyze research based on their utility for this thesis.

5.1. Approaches to Automatic Mixing

It is crucial to distinguish between an 'automatic' and an 'automated' process [6]. An automatic process involves an autonomous process that can be treated as a constrained rule problem, whose design regulates how the input signals are modified. The automated process on the other hand, as described in the previous chapter, regards processing actions that are played back in sequence as decided by the producer while the track is being recorded. Different distinct control architectures exist to control effects devices [28]:

1. **Direct user control device:** The simplest control architecture, these are non-adaptive and the signal processing is independent of signal content(e.g; fader).

2. **Auto-adaptive effects:** The control parameter here is formed using a feature extracted from the input signal (e.g; compressor).

3. **External-adaptive effect:** The control parameter is an extracted feature from a different channel to the one on which it has been applied. This can be either a feedforward (control feature from input) or a feedback(control feature from output) adaptive effect. Examples include ducking, side chain compression etc.

4. **Cross-adaptive effect:** "The control process is the result of the analysis of the content of each individual channel with respect to the other channels. The signal processing applied to each channel is dependent on the signal content of all channels involved." [28]

It is difficult to tackle the problem of audio mixing and production as a whole due to its recursive nature and the various disjoint linear and non-linear effects involved in its implementation. Researchers thus far have focused on automating one step of the mixing process at a time (i.e setting faders, panning, equalization and compression etc), using various mathematical and knowledge based methods. Based on algorithm implementation, approaches to automatic mixing can be broadly classified as [7]:

1. **Machine Learning:** In such a process, the system is initially trained using example content, which is used to extract features and data needed for application on the test tracks to be mixed [29]. This approach is limited by the sparse availability of multi-track content and copyright issues. It is also limiting from an artistic perspective, considering the full range of effects processing and

the variety of audio effects available are not exploited in such a system. Such a system, however can be used to replicate the feel or sound of a particular set of tracks.

2. **Grounded Theory:** The most resource intensive approach, this requires the use of studies (psycho-acoustic or perceptual evaluation) to determine listener preferences or psycho-acoustic phenomena which can be used to guide mix attributes [30]. Apart from being resource intensive, it can also suffer from too narrow a width of field to be considered a sufficient knowledge base for implementation of a system [7].

3. **Knowledge Engineering:** This approach; in contrast to grounded theory, assumes that the rules are known and they can readily be integrated into a system. This can involve studying best practices in the music production industry and coming up with algorithmic interpretations of the same. The drawbacks to this method are that the best practices are either not known or are highly variable. Although, some recent studies amalgamating information of best practices throw light on this matter [23]. The previous lack of knowledge on industry best practices in the academic setting has led to researchers building upon the field with a disconnect from what is generally practiced in the studio. This consequences of this approach will be explained in the next section.

Note that this section did not list approaches that deal with one effect at a time, but delves into the conceptual basis of the different approaches.

5.2. Musical Cognition and the need for automatic effects implementations

The state of current research in audio production and related fields owes a lot to the loudness models developed by Glasberg and Moore [1, 31]. The model serves as a de facto algorithmic ear, allowing us to calculate perceived loudness for both steady state and time varying sounds and most importantly, enables us to quantify the important psycho-acoustic effect of masking. Although a complete technical description of the same is out of the scope of this paper, an interested reader can refer to [32] for a good description.

This research has been a tremendous boost to the academic initiative and its popularity has led to a proliferation of research that uses the loudness model as a basis to mathematically explain the processing of sound by the biological system of the human ear and cochlea. The only real drawback of this method is the fact that the relative ease by which the loudness models are implemented leads to assumptions being made which are disjoint with their application in a studio environment. Case in point, being the assumption of equal partial loudness of all tracks in a mix. This forms the basis of the implementations in [2, 4, 5, 26, 27]. Although arguably intuitive, this assumption has been used to implement automatic mixing adaptations based on individual track gain optimization. Disproved through questionnaires to professionals [23], equal loudness/audibility of all tracks is never really implemented in the studio, especially with modern productions containing tens of different individual stems being mixed into a single track. The varying importance of the stems, change in focus and tenor through the different sections of the track, and the artistic choices necessary to convey a sense of motion in the narrative, require that the stems be processed differently, whether that be through effects implementations or application of different fader/gain values. Loudness is only a singular aspect of the more extensive and complicated art of audio production.

There is certainly more clarity regarding studio production practices with the advent of newer studies [23, 30, 33] on the topic. Researchers have been hampered due to the lack of knowledge of industry practices, but the explosion in accessible audio production, synthesized sound in electronic music and widespread Internet music sharing has incentivized them with a new market base for automatic mixing algorithms: the amateur home producers. Newer loudness descriptors more suited to musical signals have also been developed [34]. Although described in relation to broadcast standards and practices, these loudness descriptors could be more useful for constantly varying, high dynamic range musical signals. Further research on frequency analysis based on audio perception [35] can also help us derive features that provide a higher level of sophistication beyond loudness characterization.

In summary, the academic and research approach towards musical cognition and automatic music production has been held back due to a number of factors, such as:

1. The lack of knowledge in the academic realm regarding standard studio practices.
2. The wide variation in subjective descriptors used to define music production quality (i.e; boomy, tight, warm etc).
3. The use of untenable assumptions in order to develop mixing algorithms.
4. The lack of an objective measure of music quality and the wide variation in musical tastes inhibiting the development of such a measure.
5. The necessity of resource and time intensive listening tests for model validations.
6. Previous lack of usable computational power for real-time audio analysis of multiple stems for cross adaptive effects implementations.
7. Previous lack of a viable market or need for automatic mixing effects.

These reasons enumerated above help us answer the other supplemental question to the central aim of this thesis, i.e; the reasons behind the sparsity of research in automatic Reverb addition.

5.3. Research in Applied Reverb

Research in applied Reverb has been sparse at best, not just because of the broad challenges to relevant academic research described above, but also because more attention is given to processing that takes place earlier in the audio chain. Processing such as fader gain setting, compression and equalization bring about more easily discernible changes in the audio, and also are arguably the most important steps taken to make stems more suited for a cohesive mix. Reverb does not bring about as easily discernible changes in tracks, and its most important use stresses on making a mix sound "natural" and "cohesive" which are difficult to define or judge with mathematical measures. In the following section, the relevant papers along with their contributions and possible drawbacks are discussed.

5.3.1. Discussion of relevant papers

Quantitative characterization of perceptually relevant artifacts of synthetic reverberation using the earwig distribution Furlong and Dermat [36] described the characterization of perceptual reverberation artifacts using the Earwig Distribution. The Earwig distribution is a joint time-frequency distribution which incorporates accurate ear masking effects. This is achieved by using a frequency dependent smoothing kernel which simulates the masking behavior of the ear. Artifacts of perceptual reverberation, specifically 'graineyness' (noise like modulation of signal partials) and 'fluttering' (regular visible modulation of signal) are shown to be visible in the temporal excitation patterns obtained using the EWD (Earwig Distribution). The authors hope this finding could be used as a basis for an objective measure for such artifacts and consequently for an objective quality measure. This paper offers an objective measure that models the sound processing characteristics of the ear in a mathematically demonstrable way. But, in the words of the authors themselves: "The above analysis is for just one particular partial of one particular note of one particular instrument, but it is presented as a typical finding". The behavior of the excitation pattern and the consequent mixing of the multiple sounds in a mix, along with temporal and spectral masking will affect the visibility of such artifacts in the Temporal Excitation patterns, making them harder to detect. The ability to detect such artifacts will also be influenced by the other effects added further up the processing chain. For such a measure to be used reliably in a future objective measure of mix quality, its efficacy needs to be reliably demonstrated in a more realistic mix setting.

Towards a Computational model of Perceived Spaciousness in Recorded Music Sarroff and Bello[25] provide a comprehensive analysis and model of the spatial attributes discernible in music. Although not directly related to the automatic addition of Reverb to music, this work offers an important theoretical and practical base on which further implementations can be built. After a discussion

on what constitutes spaciousness in music, examined from multiple viewpoints (Acoustics, Sound Reproduction and Recorded Music); the authors describe a listening test wherein subjects were asked to rate 50 music samples based on 3 different intuitively characterized spatial attributes: width of source ensemble, extent of reverberation and extent of immersion. The experiments comprise results from both a controlled laboratory setting and an open online setting. The ratings are then processed to formulate an objective to subjective mapping function. Using a supervised machine learning approach, multiple audio features extracted using the MIR toolbox were tested for correlation with the spatial attributes and certain subset of features selected which best modelled the spatial attributes rated by the participants.
The discussion in the paper comprehensively lists the drawbacks of the computational model used. They are enumerated as follows: a) The attributes used were non-standard, unexamined in formal experiments and hence their semantic and perceptual validity can only be speculated. b) Only the side signal of the mid/side transformation of the audio was used for feature extraction, to minimize the data used for computation. This would have caused information loss and affected the computational model created. c) Only 2 spatial features were extracted from the music excerpts. d) An open online test environment exposes inherent variation related to different listening conditions and audio reproduction hardware used.
The model showed a relative absolute error (RAE) of 62.63, 67.20 and 64.36 respectively for the 'Width of Source Ensemble', 'Extent of Reverberation' and 'Extent of Immersion' respectively. These numbers are not tenable to a highly demanding, quality focused studio environment. In retrospect, with added knowledge of studio practices in adding spatial attributes to sound, parameters such as panning (left-right positioning), reverb/delay(front-back positioning, immersive environment) and fader gain values (simple front-back positioning) would have made better spatial attributes to be rated by the subjects. Although, the fact that the subjects rated music excerpts post production rather than during production (as is the context of this thesis) would have made a difference.

An important finding from this paper, especially in the context of overall spatialization can be the correlations found between the various attributes rated. Assuming, 'Extent of Reverberation' and 'Width of source ensemble' to be analogous to Reverb and Panning in the production sense, and 'extent of immersion' to be entirely subjective, based on the listener's perception of spatial immersion; the correlations found show width (panning) and immersion to be more correlated (R=0.8745) versus Reverberation and immersion (R=0.5679). Width (panning) and Extent of reverberation (Reverb) is shown to have a weaker correlation (R=0.3186). This shows, that even though in production, panning and reverb are treated as independent processing operations, their interaction in forming spatial attributes in the produced music leads to a more holistic aural experience that can only be achieved through the recursive actions of a producer (or a future automatic production algorithm). It also precludes these parameters being part of completely independent, sequential actions used to automatically produce raw recorded tracks.

Matching Artificial Reverb Settings to Unknown Room Recordings: a Recommendation System for Reverb Plugins Peters, Choi and Lei [8] describe a method to match artificial Reverb settings to provided room recordings by using a supervised machine learning approach. A Gaussian mixture model is trained on randomly chosen Mel Frequency Cepstral Coefficients extracted from audio files processed with particular Reverb presets. To provide a Reverb preset recommendation, MFCCs are again extracted from the test recording and a GMM computed. The preset is then found by computing likelihood ratios between the GMMs of the pre-recorded audio files and the test audio file. The Reverb preset with the highest likelihood ratio is then the chosen answer.

This system is not only targeted at amateur producers who want to save time while choosing a suitable Reverb preset for their recording, but also in other production scenarios where time is a critical factor and choosing a Reverb preset is a better option than building a Reverb from scratch. The system can only be as good as the Reverb plugin and presets used to train the model, limiting its viability for use in more professional productions. Training the model on multiple test files generated using varied genres and preset options would subsequently increase its computational complexity and data footprint. In their defence, the authors do state that the system is only designed to provide the closest match to a reference recording, and will need subsequent fine tuning of the suggested Reverb preset

as per the user's requirements.
One could also argue that this hinders the creativity of the artist, not just by providing an easy alternative to designing a Reverb from the ground up, but also circumventing the understanding of the use of Reverb as a tool for melding instruments and tracks together. This means that any and all Reverbs used on multiple tracks need to have a high degree of correlation to be able to contribute to increasing production quality. Consequently, such a system would do better in the hands of an experienced producer than a beginner, which is an inherent drawback to its intended function.

Learning to Control a Reverberator using Subjective Perceptual Descriptors Rafii and Pardo [10] describe a system that uses subjective descriptors (boomy, church-like etc) and a user-mediated rating mechanism in order to construct a reverberator that helps add spatial characteristics to music in a personalized manner. The system goes through a training phase where the user is asked to rate a series of audio samples generated by processing an audio file with the reverberator using a variety of different settings. The stereo reverberator, based on Moorer's ideas, is built using five parallel comb filters (to simulate the complex modal response of a space), with an all pass filter in series (to increase echo density without spectral coloration) and an added low pass filter (to simulate air and walls absorption). The delay and gain parameters of the different filters are conjugated in different combinations to build 5 different reverberation measures to control the system: Reverberation Time, Echo Density, Clarity, Central Time and Spectral Centroid. The collected user ratings are related to the five reverberation measures using linear regression, building a model which can predict the user rating given a reverberation impulse response. This controller built based on this mapping lets the user process the audio in terms of the subjective descriptor using a slider. This slider control changes all the five reverberation measures, based on the linear regression model. The experiment consisted of a training phase, where the user was asked to rate 60 audio examples using the subjective descriptors, the ratings used to build the model and the test phase, where the user was allowed to play with the reverberation controller to determine how well it worked. The users could then give their feedback regarding the potency of the system via a 0-10 numerical rating.

This system is clearly targeted at users who have no prior experience in mixing music and no idea of industry practices and jargon. Even though any previous knowledge of industry practices and previous understanding of the subjective descriptors used, might have biased the users during the training phase; the personalized nature of the system would still have produced results based on the user's understanding of the reverberation process. This is a strong point of strength for such a system, considering the wide variability in subjective descriptors used in the industry and the diversity in their understanding. Strong passionate preferences for a certain genre of music and/or a way of production would also be incorporated in the system, through the user's recorded answers during the training phase.
On the contrary, the assumption that a user's understanding of a subjective descriptor might remain the same through their widening understanding of music mixing or even over widely differing genres seems implausible. There is a risk of pigeonholing the user into a particular range of Reverb settings as determined by the singular test session. The validity of these speculations can only be tested over more full scale tests with enough variability in genres, music listening, varying mixing experience and listening environments. The reverberator as designed in the paper has enough variable range in order to produce both short and long Reverbs and be a workable generic model that can be adapted for use in different scenarios. But the variability of user consistency, system predictiveness and human ratings over different subjective descriptors (describing varying Reverb settings) suggest that methods more robust and accurate than linear regression should be explored for building the mappings.

More about this reverberation science: Perceptually good late reverberation Karjalainen and Järveläinen [37] focused on the psychoacoustic qualities of late reverberation, seeking to find the minimum requirements necessary (in terms of modal density, irregularity of time-domain response in each critical band, frequency resolution etc) needed for perceptually good late reverberation. Formal listening tests conducted by the authors showed that for all Reverb times, 1500 modes were enough to produce good quality Reverb as compared to a standard of 5000 modes. The number of modes could also be reduced by random physical placement and modulating modal frequencies. Using auditory modeling to evaluate Reverb quality was proposed, considering the then lack of shown artifacts.

This was later shown to be untrue by Furlong and Dermat [36]. A way to realize artificial reverb using parallel filters named "Modal Filter Reverb" is also proposed.

The emphasis placed on late reverberation as opposed to early reflections is important considering its undervalued role in research and to a limited extent, the studio as well. It is known that Reverb can be replaced with a cascade of well organized delays [19, 24]. But, what the delays lack is good quality late reverberation, which although being less discernible than other results from effects processing, plays a hugely important role in melding audio together and creating realism of spatial characteristics. Considering the importance of quality in a studio environment, more research is sorely needed in perceptual analysis of Reverb in general and late reverberation in particular. The research in this paper was restricted to monoaural audio samples, which augurs well for mono compatibility but could be critically limiting considering spatial characteristics are built using both Reverb and panning [25]. So, any test of quality in a monophonic setting can be argued to be unrealistic. In defence of the authors, the test subjects were specifically briefed to listen for metallic timbre caused due to spectral coloration of late reverberation, which does not affect spatial realism but only introduces noise in the audio stream. The parallel filter approach suggested is highly flexible and the meteoric rise in average computational capacity of the home PC makes it easily executable as well. Using auditory modeling to evaluate audio quality has still not been fully executed, but certain groundwork has certainly been done [35, 36, 38, 39].

Automatic adjustment of off-the-shelf reverberation effects The research by Heise, Hlatky and Lovisach [9] demonstrated matching the room impulse response of a reference recording with the audio produced using a VST reverberation plug in using multiple optimization strategies. The euclidean distance between the MFCC vectors of the 2 audio samples is minimized using Genetic Optimization, Nelder Mead Optimization, Brute-Force Nelder Mead Optimization and Particle Swarm Optimization. The results of the optimization along with audio samples created by professional sound engineers were used in a MUSHRA style listening test to validate the results. The listeners were able to identify the reference but the subjective ratings of the rest of the tracks remained close, with the subjects also remarking that the discernible differences were rather small.

The authors suggest the usefulness of this tool in comparing the sound quality of multiple effects plugins, although this seems of limited use to the amateur producer and of niche interest to a professional one. The suggestion of this tool being a measure to help use simpler DSPs in memory restricted systems also limits the possible quality achievable, which is of prime importance in any studio.
Where this tool could be useful is for content creators to recreate the feel of a physical space they like or match the feel of a song or recording that appeals to them. Especially video producers who lack the audio production skills or have limited time to do so. The optimization time and computational complexity is affected by the number of controllable parameters offered by a VST plug in, which can be highly variable. One of the plug ins used (Ambiance Reverb), has as much as 21 different parameters. The authors do touch upon the different grades of acoustic impact caused by the different parameters, but stick with random variation of the parameters during the optimization rather than one based on a sense of importance.

5.4. Relevance of Current Research

The body of research on automatic Reverb implementations thus far is sparse and lacking in a consensus on how to approach the problem, quantify it and define set objectives for optimization of important parameters.

On a general note, there needs to be research that clearly catalogues the correlation between and covariance of different Reverb parameters on the audio output, represented in a perceptually relevant time frequency representation. This would not only help identify a hierarchy of importance for the multitude of Reverb parameters used in the studio today but also help to recognize perceptually relevant features which can be used as a basis for optimization. A systematic study on the effect of Reverb on perceptually relevant spectral features can also help in creating algorithms that can adapt themselves to different instruments and/or genres. These efforts can be critical to streamlining the research efforts of multiple research groups and possibly lead to faster progress in the field.

This chapter included a discussion on the current research prevalent in the field which is relevant to

this thesis. Due to the lack of sufficient consensus on how to approach the particular issue of automating Reverb application to production oriented audio, we will be looking to adapt research that deals with automating other effects to the application of Reverb. The author feels this is a good starting point for an introductory foray into this endeavour. Such adaptations however, will require certain ground rules derived from our previous study and discussion on studio practices. The subsequent chapter deals with the development of an experimental set-up, details the assumptions made and defines the ground rules necessary for these adaptations.

6

Experimental Set-up

After the basis for understanding Reverb, its stochastic nature and its role in the mixing process discussed in the previous chapters, this chapter explains the experimental set-up and the assumptions used to create multiple algorithmic interpretations of potentially useful automatic Reverb addition implementations. Two different papers, approaching multi-track audio mixing from different viewpoints will be adapted and used to add Reverb to 5 different tracks of different genres. The common mixing set up used and the assumptions and rules that apply to both adaptations are further described in the following sections.

6.1. Assumptions

The assumptions used to create the algorithms and their veracity is explained below. These assumptions are based on the knowledge gained from our study of studio practices in Chapter 4.

1. **Sufficient reverberation quality can be achieved with digital emulators**: This assumption is true, owing to greater quality of software Reverb emulators and the greater choices in parameters afforded by them. Professional producers still swear by real rooms and dedicated Reverb chambers to add ambiance and effect to their recording, but for the amateur producer, digital Reverb emulators do offer sufficient quality.

2. **Creative quality is more important than objective realism**: This assumption is mostly true, its veracity depending upon the context. Completely false in situations where realism is desired, such as dialogue tracks for movies, but since this thesis deals with music production, that condition does not apply. As stated in the last section, the pertinent question producers should ask themselves while applying Reverb is "How good does it sound?" rather than "How realistic is it?" [19].

3. **2-3 Reverbs are sufficient for a single track**: This assumption is true. Mike Senior [24] suggests amateur producers use 2 Reverbs, a long and a short Reverb on 2 different auxiliary tracks, each fed by sends from all the tracks in the mix. The level of Reverb, long or short can then be controlled by the aux send level on each particular track. This is sufficient for both adding ambiance and varying the perceived depth of various instruments. This precludes the use of specialized Reverb on particular instruments on short sections to work as sonic hooks (to keep the listener's attention), to generate contrast and movement etc.

4. **Post and/or Pre-Reverb equalization is not required**: This assumption is not true, but used here for sake of simplicity in the implementations. Post Reverb equalization is recommended [21] in order to maintain spectral balance in the production and avoid artifacts and coloration. Pre-reverb equalization can be used to do the same and is found to be a common practice through producer interviews [23]. Sometimes an appropriate frequency shaped version of the track is fed to the Reverb instead of the actual version of the track fed to the master mix, for appropriate results. However, this is not attempted.

5. **The mixing process is sequential**: This assumption is also untrue, as explained in the industry practices chapter. The mixing process is circular rather than sequential and improvements are made by constantly listening and tweaking the processes and the effects applied. However, this arrangement is difficult to emulate algorithmically and since this study involves only Reverb, it is important to keep all the other effects added, constant. Thus, this assumption is needed to create a base of pre-processed tracks ready for algorithmic Reverb addition.

6. **Not all instruments/tracks are equal in importance**: This assumption is true. Producers often talk about the vocals being the most important track in a production [19, 21, 23], the track that is most noticed and evaluated by the listener, while the other instruments function as a harmonic basis. This track hierarchy is promulgated by assigning appropriate loudness (the greater the importance, the greater the loudness) and by simulating depth, using the ratio of direct sound to reverberant sound. The most important tracks are usually placed upfront and supporting tracks are placed in the background.

6.2. Rules of Mixing and Testing

This section details the primary rules used in order to create the tracks that are used as input to the multiple automatic Reverb addition systems and describes the commonalities between the different set-ups. The rules used are as explained below:

1. **5 tracks per song with set track hierarchy**: Every production/song will be made with 5 different tracks that either contain recordings of different instruments or recordings of multiple instruments from different microphones depending on the genre. Each track in every production will be given a level of importance based on the recording it contains. All excerpts of tracks used are obtained from Mike Senior's "Mixing Secrets" free multi-track download library[1]. Pre-edited excerpts provided in the multi-track library are used, however for the classical genre, the excerpt is obtained by snipping a 34 second segment starting at 40s into the tracks. Most of the multi-track productions obtained are provided with a large number of tracks, which are reduced to 5 tracks either by combining certain recordings and instruments or discarding certain tracks. The details of these operations are provided below.

2. **Effects applied before Reverb are the same for tracks input into different automatic implementations**: Processing and effects are applied in a set sequence i.e; Faders, Panning, and Processing effects with self referential, circular fine tuning improvements until we have input tracks that can be used on the automatic Reverb implementations. The Reverb for every track is panned to the same point in the panorama as the instrument it is applied to. The same input tracks are used on all the different automatic interpretations.
   A note on the importance of panning and fader control; as revealed in the previous 2 chapters, both of them work in tandem with Reverb to create an immersive environment and maintain track hierarchy respectively. Generally the most important track has the highest gain and is panned to the centre. Track hierarchy is first implemented through fader control (tracks with higher importance are accorded greater loudness) and panning guidelines and rules from Izhaki's "Mixing Audio" book [19] and Pestana's paper on studio best practices [23] are used. All the tracks used for the different productions are pre-processed (equalized, compressed and panned) using Audacity[2]. The equalization, although subjective and dependent on the spectral character of the individual tracks is based on the guidelines for equalization provided in Izhaki's "Mixing Audio" [19]. Compression is performed with the goal of obtaining maximum loudness before clipping occurs. The balance between the different tracks is set by ear and is based on the author's preference and experience. The tracks are ready for algorithmic input when the mix sounds spectrally balanced and pleasant with respect to inter-track loudness but lacks any ambient characteristics and sounds "dry" as a whole.

3. **Use of a single section in a song**: The excerpts obtained are pre-edited to contain a certain section of the song. These edits range anywhere from 17 to 35 seconds. This ensures that the parameter settings for the processing effects applied and the Reverb added by the automatic

[1] http://www.cambridge-mt.com/ms-mtk.htm
[2] http://www.audacityteam.org/

implementations will be constant throughout the segment used. This is a good baseline for testing our automatic Reverb implementations since it provides an architecture which can then be modified to be adaptive to temporal changes in the songs through the different sections and adjust itself accordingly.

4. **Multiple genres for different challenges**: 5 different genres of music will be used, each that requires a different approach to adding Reverb and thus providing a different challenge to each automatic implementation. This will allow us to test the adaptability of the implementations and find possible shortcomings in their design that degrades their performance in certain situations. The different genres used are:

   (a) **Classical**: The challenges with this genre lies in the way it is recorded. With multiple instruments and/or vocalists in the same room, placed and arranged according to the sound-stage desired, the entire performance is recorded through multiple microphones. There is a lot of leakage(recordings of instruments placed adjacent to the focus instrument) between microphones and each recording has a slightly different ambient character. The application of Reverb in this case needs to complement the ambient nature of the recordings themselves and to meld the different recordings together with an inconspicuous Reverb that does not cause ambiance collision.

   The production used is G.F.Handel's 'Non Lo Diro Col Labbro' performed by Asam Classical Soloists, recorded at a live chamber concert. The production is provided with 5 tracks, each representing recordings of all the instruments by diversely placed microphones.

   (b) **Drum & Bass**: This genre is assumed to be representative of the larger electronic dance music genre. Most modern electronic music productions are made with dry synthesized sounds, that have no spatial character of their own. Reverb application needs to be generous but judicious on all tracks except percussion, in order to create an immersive environment. Rhythm is what makes dance music enjoyable, and it is this rhythm provided by the percussion that is prone to clutter and confusion when reverberated excessively. Loss of rhythmic feel and clutter is an important inflection point to listen for.

   The production used is Skelpolu's 'Anomalous Weeping'; a drum & bass song provided with 11 different tracks. The tracks are reduced to 5 by combining the percussion tracks (Kick, Snare, Hi-hat and Cymbals), the synth tracks (Synth1, Synth2, Synth3, Synth4) and keeping the Bass, Piano and Harp tracks independent. All the percussion tracks and synths combined are pre-processed individually.

   (c) **Heavy Metal**: Heavy metal is based around highly distorted guitars which tend to fill the spectral field and can occupy the background without much Reverb application due to their nature. Excessive reverberation can also cause rhythmical confusion and timing errors. Drums used on such tracks are often described to be 'tight' and 'punchy' which suggests the spatial character of a very small room. Thus Reverb application here needs to be nominal and limited to its essential instrument melding nature.

   The production used is Dark Ride's 'Burning Bridges', provided with 18 tracks. The tracks are reduced to 5 by individually pre-processing and combining the percussion tracks (Kick, Snare up, Snare down, Hi-hat, Tom1, Tom2, Tom3 and Cymbals), electric guitar tracks (ElecGtr1 and ElecGtr2) and the backing vocals tracks (BackingVox1, BackingVox2 and BackingVox3). The bass track and the lead vocals tracks are kept independent of other instruments. 3 tracks with recordings of the drum room sound are discarded in order to keep the character of the instruments as dry as possible.

   (d) **Pop Music**: Pop music can be fairly varied in the type and number of instruments used, but modern pop music depends on electronically synthesized sounds or highly processed instruments working to form a basis to a golden voice. Pop music is marketed based on a particular artist's aura, and thus vocals are generally the most important track here. Reverb is critical here for depth application in order to bring the vocals centre and forward and to meld all the other tracks into the background to form a cohesive basis.

   The production used is Leaf's 'Come Around' provided with 13 tracks. Percussion tracks (Loop, Hat, Shaker), Guitar tracks (ElecGtr1, ElecGtr2, AcousticGtr and EbowGtr) and back-

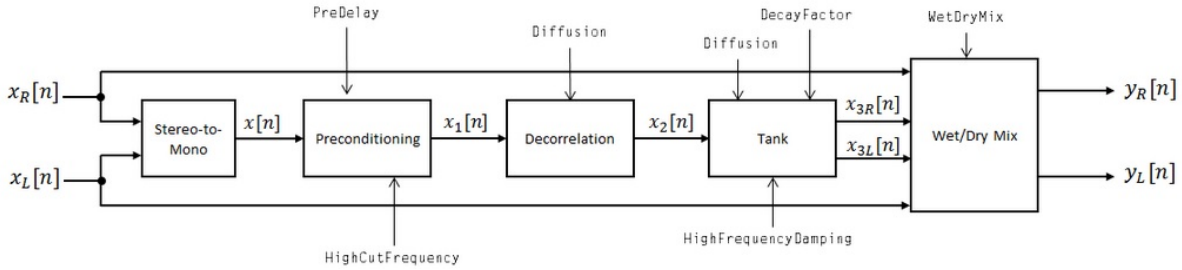


Figure 6.1: Schematic of Reverb Emulator used

ing vocals tracks (BackingVox1, BackingVox2, BackingVox3, BackingVox4) are individually pre-processed and combined. The bass and Lead vocals tracks are treated individually.

(e) **Singer/Songwriter Style Minimal recording**: This style of recording generally involves a single artist playing an instrument and singing at the same time. Other supporting tracks can be added during production that help to aid the creative vision of the artist. Such a sparse recording means that Reverb can be used to fill the stereo panorama and the empty spaces between sound events; this makes Reverb more noticeable in such a recording as compared to the other genres. Vocals still need to be given the highest priority which means pushing them forward with short Reverb times but also using the late reverberation to add timbre and highlight the overtones that give them a unique spectral character.

The production used is Irish songwriter Enda Reilly's 'Cur An Long Ag Seol'. Percussion tracks (DrumMic1, DrumMic2, Brushes), bass tracks (BassMic1, BassMic2), plucked instrument tracks (AcousticGtr, MandolinMic1, MandolinMic2) and bowed instrument tracks (Fiddle1, Fiddle2) are individually pre-processed and combined. The lead vocals track is treated independently.

5. **Use of headphones**: Headphones are used for all mixing and mastering tasks and will also used during listening tests in order to reduce the influence of the spatial character of the rooms where these events take place. This provides an important control aspect as the spectral nature of the recordings is maintained constant in all listening conditions. The headphones used are the Beyerdynamic DT 990; these studio reference semi-open headphones are very popular for use in studios around the world for tracking and music mixing.

6. **More emphasis on ambiance than effect**: The use of Reverb as an effect, to create otherworldly environments, to create sonic hooks etc, is not attempted as this is highly subjective with regards to both application and testing. Its use in ambiance creation and instrument melding is easier to judge more objectively, which can help us create better systems for automatic implementations.

7. **A single Reverb emulator is used**: All automatic implementations are designed in MATLAB to use the Reverberator function provided in MATLAB R2016b. Figure 6.1 shows the basic schematic of the Reverberation algorithm[3]. Figure 6.2 shows the parameters of the Reverberation block that can be varied to sculpt the character of the resulting Reverb.

8. **The same Reverb addition set-up is used**: 2 iterations of the same Reverb emulator are set up with varying parameters (short and long Reverb), and each track in the mix is connected to both iterations through a fader control. Automatic implementations manipulate the send fader values of all the tracks feeding these iterations and optimize them to specific values based upon the adaptations described in the next 2 chapters. The set-up is displayed in figure 6.3. The set-up used, is as described and recommended for amateur producers by producer Mike Senior in his article for Sound on Sound magazine [24]. The variables used for the long and short Reverb are tabulated in Table 6.1.

[3]https://nl.mathworks.com/help/audio/ref/reverberator-class.html

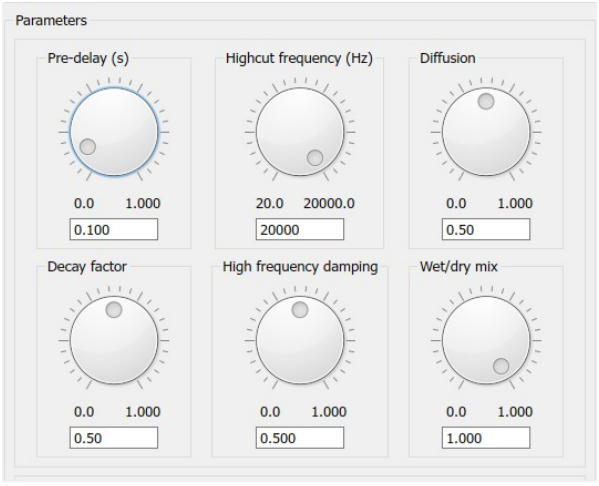


Figure 6.2: Variable parameters of the Reverberation Emulator

Table 6.1: Reverb Parameter Values

Parameter Values	Pre-Delay(s)	High Cut Frequency(Hz)	Diffusion	Decay Factor	High frequency Damping	Wet/dry Mix
Short Reverb	0.010	20,000	0.500	0.800	0.450	1
Long Reverb	0.100	20,000	0.500	0.200	0.350	1

With the assumptions, rules and set-ups that are common to both adaptations explained, the next 2 chapters will focus on describing the papers that are adapted to form specific algorithms for automatic addition of Reverberation to a set of edited and pre-processed tracks. Approaches based on mathematical optimization and spectral masking minimization are described, adapted and implemented. Any deviation from the rules and assumptions detailed in this chapter is clearly explained and justified.

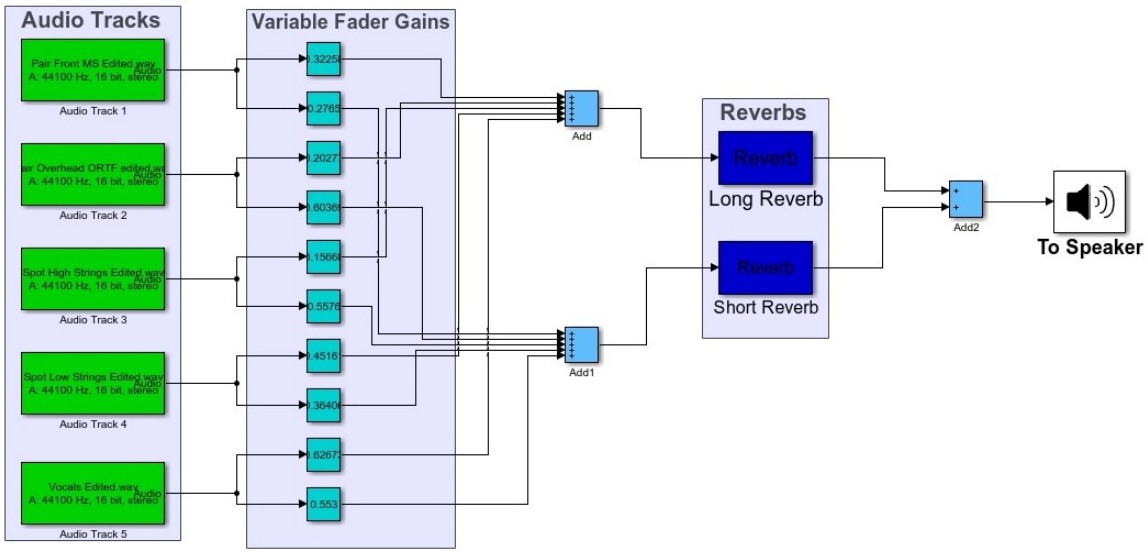


Figure 6.3: Additive reverb setup to be used, as routed in Simulink. Green blocks are audio tracks, Cyan blocks are send fader gains, Deep Blue blocks are Reverb blocks and Light Blue blocks are track addition operations

7

Mathematical Optimization

This approach is based on the paper "The Mathematics of Mixing" by Simpson, Terrell and Reiss [2]. The basis of the paper is a numerical optimization approach towards music mixing with the optimization being based on achieving equal partial loudness for all tracks. This assumption has been proven to be false through surveys and interviews of producers [23], but the mathematical architecture described for optimization is appropriately adapted and used with the goal of maintaining the loudness balance of the tracks set previously and preserving this track loudness hierarchy during the addition of Reverb.

7.1. Description

The mathematical reasoning used in the paper is defined henceforth. As used in the paper, scalar values are represented with lower case variables, vectors by bold lower case variables and matrices by upper case bold variables.

The process starts with a vector of gain variables.

$$\mathbf{g} = \begin{bmatrix} g_1 & g_2 & g_3 & \dots & g_n \end{bmatrix}^T \tag{7.1}$$

Where n denotes the number of tracks. These gain variables are multiplied by the audio streams, represented by a and this product is convoluted with the impulse response of the production system and room, denoted by **h**.

$$\mathbf{p}_i = (g_i a_i) * \mathbf{h} \tag{7.2}$$

The mixing engineer is modeled using the excitation pattern model of the human ear developed by Moore and Glasberg [32]. This consists of a linear filter, and filter banks in sequence that simulate the effects of the outer and middle ear and the frequency to place transformation that takes place in the basilar membrane, respectively. Further compression of the signal in each frequency band (simulates the cochlea), integration across frequency and smoothing, converts the intensity in the excitation pattern to perceptual units in terms of a loudness time series, represented here by $\mathbf{s}_i$.

$$\mathbf{E}_i = f(\mathbf{p}_i) \tag{7.3}$$

$$\mathbf{s}_i = c(\mathbf{E}_i) \tag{7.4}$$

Here f represents the function giving us the excitation pattern model, c represents compression, integration and smoothing to give us loudness.

When multiple sounds are heard concurrently, the interaction between their excitation patterns results in simultaneous masking. This is adapted into the formula to give us partial loudness which is the partial attribution of excitation between two simultaneous sounds.

$$\mathbf{s}_i = c\left(\mathbf{E}_i, f\left(\sum_{j=1, j\neq i}^{n} \mathbf{p}_j\right)\right) \tag{7.5}$$

The partial loudness obtained for each time series is converted into a single scalar loudness measure l_i by some form of averaging operation, and these measures from every component track are consolidated in a vector.

$$l_i = \mu(\mathbf{s}_i) \tag{7.6}$$

$$\mathbf{l} = \begin{bmatrix} l_1 & l_2 & l_3 & \dots & l_n \end{bmatrix}^T \tag{7.7}$$

These values of absolute partial loudness are then converted into loudness balance by using the mean loudness of all tracks as reference and converting the absolute values into dB.

$$b_i = 10log_{10}\left(\frac{l_i}{\frac{1}{n}\sum_{j=1}^{n} l_j}\right) \tag{7.8}$$

This gives us a consolidated vector containing the loudness balance of all the tracks.

$$\mathbf{b} = \begin{bmatrix} b_1 & b_2 & b_3 & \dots & b_n \end{bmatrix}^T \tag{7.9}$$

The authors then also use the overall loudness of the mix as a further constraint to make the mix description unique; since a mix defined by a specific loudness balance ratio can be produced for both a very loud and a very quiet mix, making the search space infinite. This overall loudness is obtained by applying the same functions to the summation of all audio signals as we did to the individual tracks.

This model can now be used to optimize a mix, based on a target balance and mix loudness. The excitation pattern model used, introduces non-linearity into the model which makes an analytical solution impossible, hence numerical optimization is more appropriate. The error vectors to be minimized are defined below.

$$e_i = b_i - b_{t_i} \tag{7.10}$$

$$e_m = l_m - l_{t_m} \tag{7.11}$$

$$\mathbf{e} = \begin{bmatrix} e_1 & e_2 & e_3 & \dots & e_n & e_m \end{bmatrix}^T \tag{7.12}$$

The subscript t identifies the target metric and l_m denotes the overall mix loudness.

The errors are a function of the fader gain vectors and the total error can be expressed as the sum of squared errors of **e**, denoted by e_T; the minimum of which can be found using an iterative nonlinear least squares optimization algorithm.

$$e_T(\mathbf{g}) = \sum_{i=1}^{n} e_i^2(\mathbf{g}) = \mathbf{e}(\mathbf{g})^T\mathbf{e}(\mathbf{g}) \tag{7.13}$$

The general form of the iterative numerical optimization is below, where the subscript q defines the iteration index, $\Delta_{\mathbf{g}}$ is the search direction and λ is the step size.

$$\mathbf{g}_{q+1} = \mathbf{g}_q + \lambda_q \Delta_{\mathbf{g}_q} \tag{7.14}$$

The search direction is found using the Gauss-Newton method, which uses normal equations to solve a linear least-squares problem at every iteration.

$$(\mathbf{J}_q^T\mathbf{J}_q)\Delta_{\mathbf{g}_q} = -\mathbf{J}_q^T\mathbf{e}_q \tag{7.15}$$

$$\Delta_{\mathbf{g}_q} = -(\mathbf{J}_q^T\mathbf{J}_q)^{-1}\mathbf{J}_q^T\mathbf{e}_q \tag{7.16}$$

$$\mathbf{J} = \begin{bmatrix} \frac{\delta e_1}{\delta g_1} & \frac{\delta e_1}{\delta g_2} & \dots & \frac{\delta e_1}{\delta g_n} \\ \frac{\delta e_2}{\delta g_1} & \frac{\delta e_2}{\delta g_2} & \dots & \frac{\delta e_2}{\delta g_n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \frac{\delta e_n}{\delta g_1} & \frac{\delta e_n}{\delta g_2} & \dots & \frac{\delta e_n}{\delta g_n} \\ \frac{\delta e_m}{\delta g_1} & \frac{\delta e_m}{\delta g_2} & \dots & \frac{\delta e_m}{\delta g_n} \end{bmatrix} \tag{7.17}$$

$\mathbf{J}_q$ is the Jacobian matrix as defined in equation 6.17 at iteration q and $(\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T$ is the *Moore-Penrose* pseudo-inverse of non-square matrix $\mathbf{J}$. Note that equation 7.16 has the same form as a linear least squares problem.

The authors also define a *sensitivity matrix* $\mathbf{S}$ that displays how sensitive each track and the overall mix are to changes in the fader gain values.

$$\mathbf{S} = \begin{bmatrix} \frac{\delta l_1}{\delta g_1} & \frac{\delta l_1}{\delta g_2} & \cdots & \frac{\delta l_1}{\delta g_n} \\ \frac{\delta l_2}{\delta g_1} & \frac{\delta l_2}{\delta g_2} & \cdots & \frac{\delta l_2}{\delta g_n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \frac{\delta l_n}{\delta g_1} & \frac{\delta l_n}{\delta g_2} & \cdots & \frac{\delta l_n}{\delta g_n} \\ \frac{\delta l_m}{\delta g_1} & \frac{\delta l_m}{\delta g_2} & \cdots & \frac{\delta l_m}{\delta g_n} \end{bmatrix} \tag{7.18}$$

The values of the diagonal elements in the sensitivity matrix are all positive, since raising the fader gain on a track will increase its loudness; however, the off diagonal elements are all negative or zero (except the last row, which represents the changes in overall loudness), since they represent the changes in loudness of a track in relation to the change in fader gain values of other tracks. The authors suggest that the values in these off diagonal elements show the interactions due to masking, and the negative values represent the decrease in loudness of tracks due to competing excitation caused by increase in loudness of the masking track. This is further proved by running the optimization with the off diagonal elements of the sensitivity matrix set to zero, in which case the algorithm takes three times as many iterations to converge.

7.1.1. Analysis

The strong points of this approach are underlined by the incorporation of auditory non-linearity using the loudness models of Moore and Glasberg [1]. An iterative numerical optimization approach that simulates changes in all tracks due to adjustments to a single track is also incredibly analogous to the human approach towards music mixing. The incorporation of ear masking effects is clearly explained through the sensitivity matrix.

The weak points in the approach are in the application of the loudness model. The particular averaging operation used to reduce the loudness time series to a singular loudness value is not specified. Such an averaging model will also not be able to account for the variations within tracks that occur with changes in the sections of a song. The target loudness balance used for optimization is a vector of zeros, i.e; equal loudness for all tracks. This assumption used for mixing optimization has been proven to be false through producer interviews and surveys [23]. A straightforward mathematical approach is always very attractive to engineers, however in such cases where human subjective judgment is necessary for evaluation, listening tests are necessary. This paper, even with all its flaws, gives us an excellent basis to build further models to mix music automatically.

7.2. Observations: How Reverb affects loudness

Changes in fader gain affect loudness directly, however, before we use the same mathematical basis to optimize for addition of Reverb, we must understand how Reverb addition affects the loudness of a track. This section catalogues the variation in loudness with increasing added Reverb. The observations are then used to justify certain design changes in the subsequent adaptation. A Reverb addition system similar to the set-up to be used for optimization is used to calculate the loudness characteristics of different tracks, with different applied levels of Reverb. A send fader controls the level of Reverb application by changing the level of the unreverberated track sent to a Reverb emulator. The set up is as shown in Figure 7.1.

The loudness of tracks with varying levels of added Reverb (send fader values 0.25,0.5,0.75 and 1) are calculated and plotted in Figure 7.2 to Figure 7.5 for multiple instruments. Tables 7.1 and 7.2 tabulate the mean short term loudness and mean momentary loudness respectively, of the different instruments. Figure 7.6 shows the variation of loudness of multiple instruments with increase in fader gain. The loudness is calculated using the loudness meter provided in MATLAB R2016b, based on EBU R 128 and ITU-R BS.1770-4 specifications.

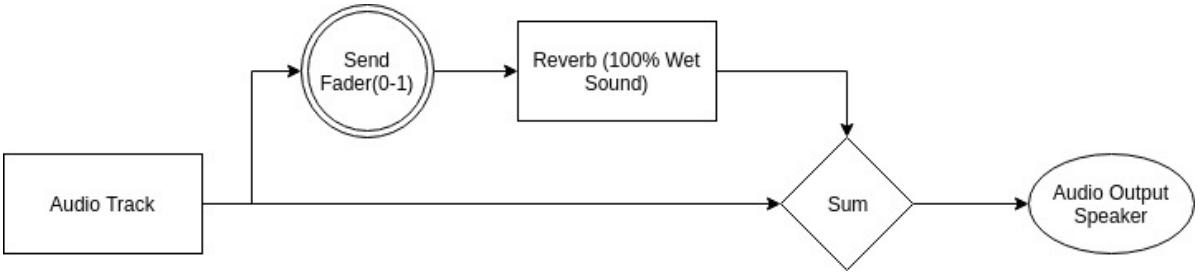


Figure 7.1: Reverb Set-up for Loudness Calculation

We will be using these broadcast standards of calculating loudness rather than Moore-Glasberg loudness since broadcast standards are universally defined for broadband audio and are far less computationally taxing than the calculation of Moore-Glasberg loudness, which makes them more suited to possible real time implementations of algorithms that include loudness calculation. Partial loudness of a particular track will be calculated by subtracting the combined loudness of the competing tracks from the total loudness. Although this is simplistic, such a calculation is quick and ideally suited to real time implementations.

Table 7.1 Mean Short Term Loudness (LUFS)

Instrument/Send Fader Level	0.25	0.5	0.75	1
Vocals	-25.8950	-25.5932	-25.1301	-24.5543
Piano	-23.1456	-22.6183	-21.9489	-21.2019
Electric Guitar	-18.7888	-18.5677	-18.1679	-17.6333
Drums	-16.5400	-16.2793	-15.8718	-15.3565

Table 7.2 Mean Momentary Loudness(LUFS)

Instrument/Send Fader Level	0.25	0.5	0.75	1
Vocals	-27.3121	-26.8333	-26.2102	-25.5156
Piano	-29.1061	-28.1937	-27.1268	-26.0451
Electric Guitar	-21.8622	-21.2918	-20.7291	-20.1071
Drums	-23.2877	-21.6205	-20.2834	-19.1586

The figures show that Reverb addition consistently increases the loudness of all the instruments. The sound smearing effect of Reverb is also clearly observed on the Drums and the Piano, which have distinct sound events spaced by silences in between. The added Reverb fills the space between the distinct sound events and increases the mean loudness of the tracks. This also leads to a greater gulf in the mean values of loudness observed for such instruments with varying added Reverb, as seen in tables 7.1 and 7.2. Continuously playing tracks like vocals and spectrally dense continuous tracks like distorted electric guitars show a lesser magnitude of change in their loudness compared to the Drums and the Piano. This behavior can be attributed to the lack of silence, both temporal and spectral, since it can heavily skew mean values to be lower in the long term.

Figure 7.6 clearly shows that the variation in mean loudness with increasing Reverb, is non-linear and dependent on the spectral and temporal nature of the instrument under consideration. Instruments that show a greater magnitude of changes to the mean loudness values will also show greater sensitivity to Reverb addition. These observations augur well for the implementation, since instruments with distinct sound events, especially percussion, need subtle application of Reverb, as excessive reverberation can cause clutter and loss of rhythm. The difference between the curves of momentary loudness and short term loudness values also shows that momentary loudness is better able to track individual sound events and shows a greater sensitivity to increase in added Reverb, especially for instruments like the drums and piano. Thus, we will use momentary loudness values for our subsequent implementation.

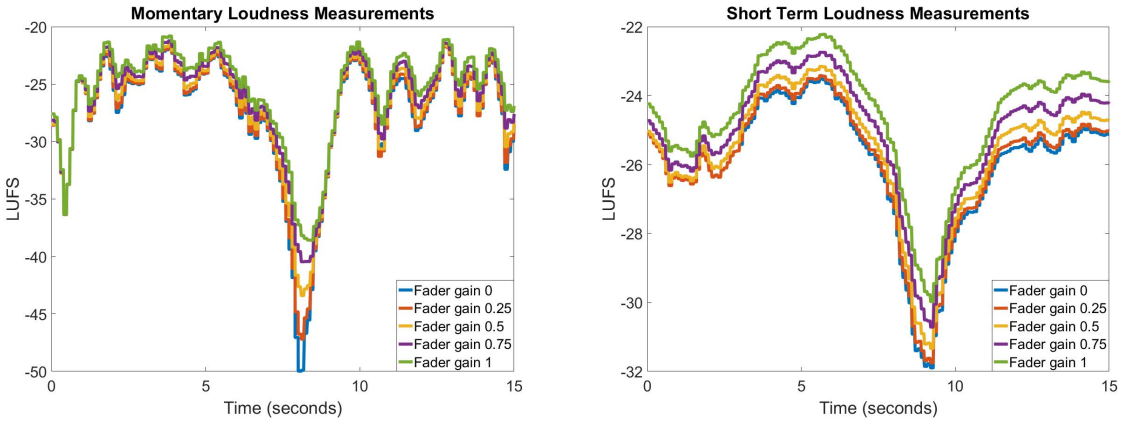


Figure 7.2: Loudness Variation of 15 sec Lead Vocals segment with increasing added Reverb

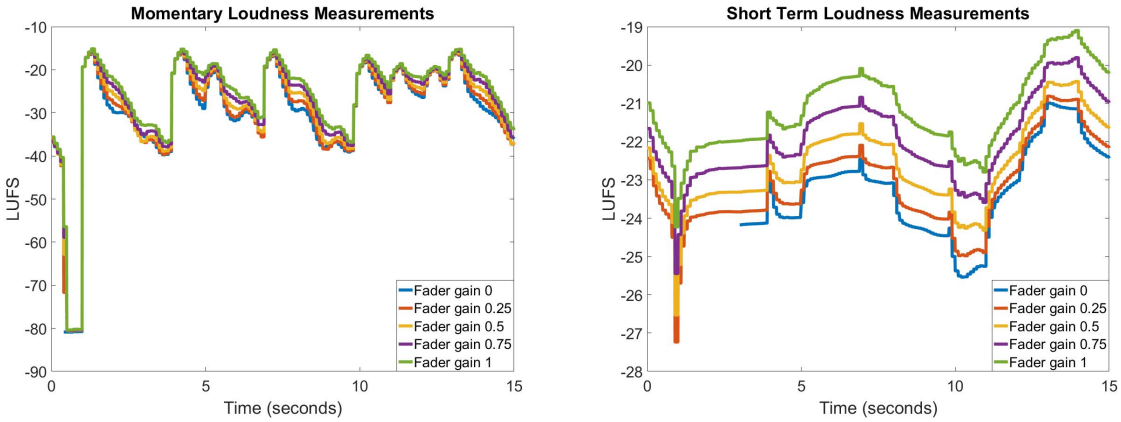


Figure 7.3: Loudness Variation of 15 sec Piano segment with increasing added Reverb

7.2.1. Consequences on Design

Based on the effect of added Reverb on loudness, we can maintain track hierarchy by constraining the loudness balance to be the same as observed before the addition of Reverb. Loudness is bound to increase with addition of Reverb and this has to be factored in by allowing increase in loudness by a set value. This set value has the potential to be varied according to the mixing preferences of the producer or the genre being mixed, but for the purpose of study and analysis, we will fix the value to be 2 LUFS. In engineering parlance, an increase by 6 LUFS doubles the perceived loudness of a track, so setting the value to be 2 LUFS gives us enough headroom for adding a perceptually relevant amount of Reverb, without needing further radical changes in fader gain of individual tracks.

The optimization, with this singular constraint of limited increase in overall loudness can get stuck in local minimums if specific fader gains with the maximum sensitivity to loudness increase are changed, with no change in other fader gain values. This necessitates another constraint in order to encourage an even increase over all the fader gain values, and an effective search of the entire solution space. The sum of all fader gain values must be maximized in order for this to occur.

7.3. Core Changes: Design Choices

The core changes implemented in the adaptation as compared to the paper it is based on, lie in the theoretical groundwork that helps us understand the changes in loudness with increasing added Reverberation. Building upon this groundwork, we add extra constraints and error variables that help us give mathematical credence to the adaptation.

The key contributions implemented include:

1. Changing the mixing set-up from direct fader gain to producer recommended additive Reverb set-up; as displayed in Chapter 6.

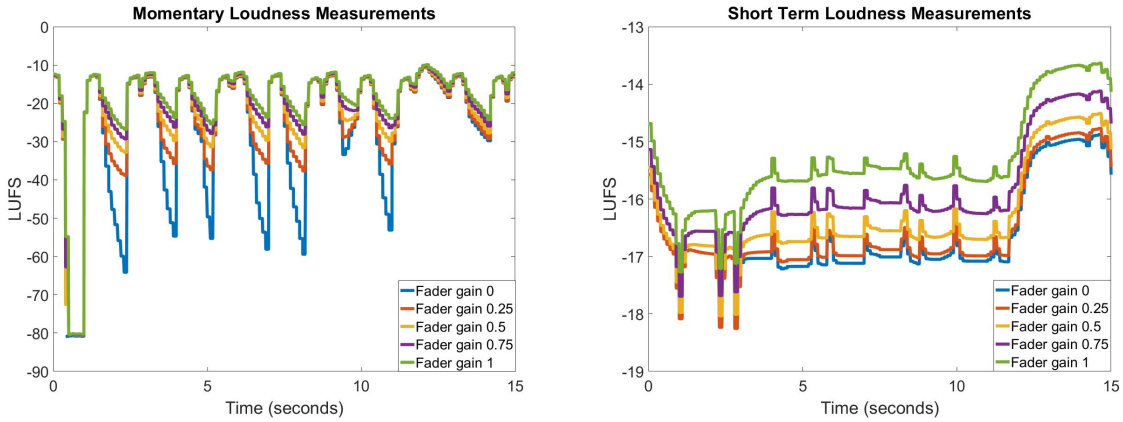


Figure 7.4: Loudness Variation of 15 sec Drums segment with increasing added Reverb

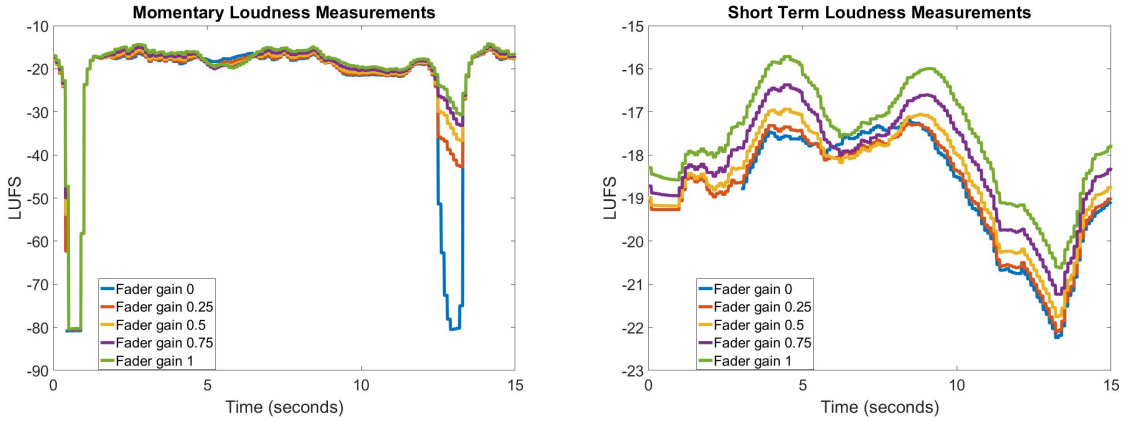


Figure 7.5: Loudness Variation of 15 sec electric guitar segment with increasing added Reverb

2. Moving from an academic loudness standard (Moore-Glasberg) [1] to Industry-standard loudness measurement (ITU-R BS.1770-4) [40]; for compatibility with modern DAWs and faster computation, as explained in section 7.2.

3. Documenting the sensitivity of loudness to Reverb addition on multiple instruments; the observations recorded in section 7.2 help us better understand the behavior of the additive Reverb setup on loudness optimization, aiding us in making critical design choices.

4. Loudness balance no longer based on unrealistic assumption of equal loudness of all tracks; studio practice of set track hierarchy implemented.

5. New constraints that take into account the additive reverberation set-up and the nature of loudness sensitivity to added Reverb.

6. Newer Constraints also make the Jacobian square, non-singular and invertible, ensuring a unique solution exists.

All the constraints mentioned will be explained in detail in the subsequent section.

7.4. Adaptation

Here too, we start with a vector of fader gain values. But, since we have 2 different Reverbs set up with varying parameters (long and short Reverb) and each is fed by 5 tracks through their respective send faders, there will be 2 fader gain vectors, each having five values for each of the 5 tracks that feed it.

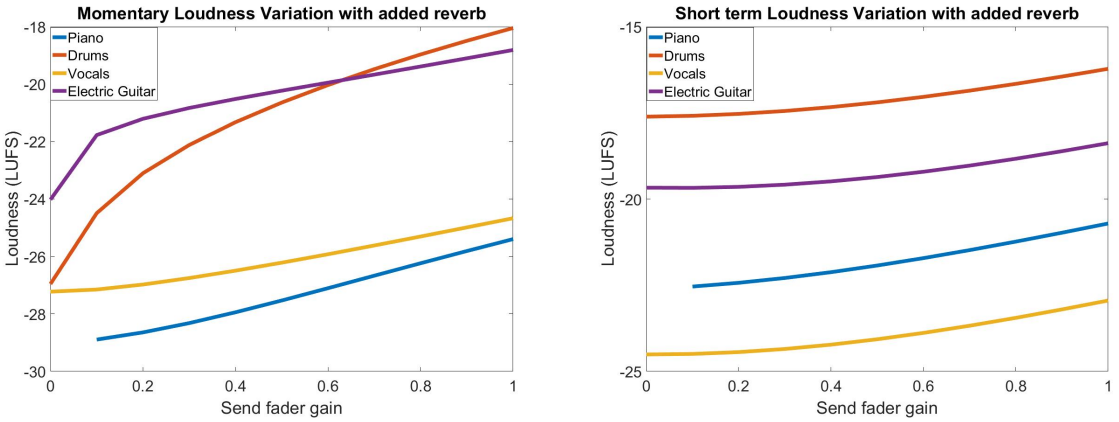


Figure 7.6: Mean Loudness Variation of multiple instruments with increasing added Reverb

$$\mathbf{g}_{long} = \begin{bmatrix} g_{l1} & g_{l2} & g_{l3} & g_{l4} & g_{l5} \end{bmatrix}^T \tag{7.19}$$

$$\mathbf{g}_{short} = \begin{bmatrix} g_{s1} & g_{s2} & g_{s3} & g_{s4} & g_{s5} \end{bmatrix}^T \tag{7.20}$$

The vector **a** represents the audio streams of the different tracks of a particular mix, and $\mathbf{h}_{long}$ and $\mathbf{h}_{short}$ represent the impulse responses of the long and short Reverbs respectively.

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}^T \tag{7.21}$$

Since Reverb is an added effect, each audio stream after the addition of Reverb, denoted by $\mathbf{p}_i$, is as defined in the next equation.

$$\mathbf{p}_i = \mathbf{a}_i + [(g_{li}\mathbf{a}_i) * \mathbf{h}_{long}] + [(g_{si}\mathbf{a}_i) * \mathbf{h}_{short}] \tag{7.22}$$

Performing partial loudness calculation as defined in equation 7.5 (Note that the following equation is now representative of the broadcast standards of calculating loudness rather than the Moore-Glasberg model).

$$\mathbf{s}_i = c\left(\mathbf{E}_i, f\left(\sum_{j=1, j\neq i}^{n} \mathbf{p}_j\right)\right) \tag{7.23}$$

Loudness averaging, consolidation and calculation of loudness balance as defined in equations 7.6 to 7.8 is performed to give us a vector containing loudness balance of all the tracks.

$$\mathbf{b} = \begin{bmatrix} b_1 & b_2 & b_3 & \dots & b_n \end{bmatrix}^T \tag{7.24}$$

The error vector can then be calculated by calculating the difference in the loudness balance at a particular iteration versus the target loudness balance. The target loudness balance in this adaptation will be set to the same balance before the addition of Reverb, with the intent of preserving the track hierarchy. The target total mix loudness however, will be 2 LUFS greater than the pre-reverb total mix loudness. This helps us implement points 4 and 5, mentioned in section 7.3.

$$\mathbf{b}_t = \begin{bmatrix} b_{t1} & b_{t2} & b_{t3} & \dots & b_{tn} \end{bmatrix}^T \tag{7.25}$$

The subscript t identifies the target variables.

$$l_{tm} = l_m + 2 \tag{7.26}$$

The target loudness for the individual tracks and the total loudness gives us the following error variables:

$$e_i = b_i - b_{t_i} \tag{7.27}$$

$$e_m = l_m - l_{t_m} \tag{7.28}$$

where i represents the track index and the subscript m represents the total mix.

As mentioned in the previous section, we still need to maximize the sum of the fader gain values, in order to encourage consistent Reverb addition over all the tracks, this gives us the following 2 error variables:

$$e_{gml} = \mathbb{e}^{-\sum_{i=1}^{5} g_{li}} \tag{7.29}$$

$$e_{gms} = \mathbb{e}^{-\sum_{i=1}^{5} g_{si}} \tag{7.30}$$

where the subscripts gml and gms stand for gain-maximization-long and gain-maximization-short respectively. The exponential function (represented here by $\mathbb{e}$) is chosen so as to not produce undefined values at the start points $g_{li} = 0$ and $g_{si} = 0$.

In order to constrain the values of the fader gains to between 0 and 1, we introduce a couple more error variables.

$$e_{conl} = \sum_{i=1}^{5} |1 - g_{li}| \tag{7.31}$$

$$e_{cons} = \sum_{i=1}^{5} |1 - g_{si}| \tag{7.32}$$

where the subscripts $conl$ and $cons$ stand for constraint-long and constraint-short respectively.

This gives us the error vector **e**, defined as

$$\mathbf{e} = \begin{bmatrix} e_1 & e_2 & e_3 & e_4 & e_5 & e_m & e_{gml} & e_{gms} & e_{conl} & e_{cons} \end{bmatrix}^T \tag{7.33}$$

The optimal values for the gain variables can now be found by minimizing the total error, formed by the sum of squared errors of vector **e**. This is similar to the operations from equation 7.13 onward.

7.4.1. Constraints and their Mathematical Relevance

The constraints applied in this adaptation are:

1. **The loudness balance of the tracks is maintained to pre-reverb addition levels:** This maintains track hierarchy and preserves the front-back sound localization provided by inter-track loudness levels.

2. **The momentary mean total loudness is constrained to increase by 2 LUFS:** This takes into consideration the increase in loudness facilitated by our additive Reverberation set-up, that fills up spaces of low loudness in the temporal field due to the sound smearing effect of Reverb.

3. **The sum of fader gain values is maximized:** The sums of the fader gain values connected to both the long and the short reverbs are maximized individually. This is necessary to avoid local minimums in the search space that can be formed due to excessive reverberation applied to particular loudness-sensitive tracks, that can individually fulfill the loudness increase constraint. Maximizing the sum of the fader gains will lead to a more consistent increase over all the tracks rather than maximization of the fader gains of the most loudness-sensitive tracks.

   The sums of the fader gains connected to the short and the long reverbs are maximized individually since they have distinct impulse responses and hence display different sensitivities to increase in fader gain values.

4. **The fader gains are constrained to values between 0 and 1:** This constraint is self explanatory since the level of the reverberated track should not rise above the level of the dry track (this can cause severe clutter and spatial confusion). Negative values will cause phase cancellation.

   The fader gains for both short and long reverbs are constrained individually as well, owing to their differing mathematical qualities.

Applying all the above constraints gives us 10 different error variables, which can then be squared and minimized in order to optimize the fader gain values. The important constraints added here are the constraints 3 and 4 enumerated above, since they help us equalize the number of error variables and the number of fader gains to be optimized. As mentioned in point 6 in section 7.3, this means that the Jacobian will no longer be rank-deficient and a unique solution is possible.

A rank deficient matrix would be singular and non-invertible, and would cause very large gain changes in any optimization that depends upon finite difference methods to calculate the error gradient.

7.5. Implementation

The adaptation was implemented on MATLAB using the functionality of the Audio System Toolbox. The optimization was performed using the function `fmincon`.

The fader gain values obtained for each genre are implemented in the additive reverberation set-up to create tracks that will be used in the comparative testing phase. The fader gain values obtained by the algorithm for different genres are cataloged in the tables below.

Table 7.1: Fader gain values for genre Classical

Track Name	Pair Front MS	Pair Overhead ORTF	Spot High Strings	Spot Low Strings	Vocals
Long Reverb	0.2018	0.1855	0.2100	0.0898	0.2312
Short Reverb	0.8323	0.5662	0.8053	0.6196	0.7170

Table 7.2: Fader gain values for genre Drum and Bass

Track Name	Bass	Drums	Harp	Piano	Synth
Long Reverb	0.2798	0.3212	0.2593	0.3158	0.3095
Short Reverb	0.8261	0.4952	0.6564	0.9996	0.7969

Table 7.3: Fader gain values for genre Heavy Metal

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.0622	0.5310	0.1752	0.2240	0.5979
Short Reverb	0.1387	0.5575	0.7402	0.9964	0.3196

Table 7.4: Fader gain values for genre Pop

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.0327	0.3291	0.9107	0.2358	0.3669
Short Reverb	0.0968	0.1421	0.3717	0.8105	0.8396

Table 7.5: Fader gain values for genre Singer-Songwriter

Track Name	Bass	Drums	fiddle	Guitar and Mandolin	Vocals
Long Reverb	0.1766	0.0583	0.4328	0.2514	0.3224
Short Reverb	0.5471	0.1882	0.6216	0.6903	0.6835

On a cursory glance, it can be seen that the fader gain values for long reverb are smaller, on average, than the values for short reverb. Since it is known that long reverb has greater potential to cause temporal masking, this suggests the algorithm does inversely correlate temporal masking potential with the final fader gain values.

This chapter dealt with the adaptation and implementation of a mathematical algorithm designed to minimize temporal masking. The next chapter will detail the adaptation of an algorithm designed to reduce the effect of spectral masking.

8

Spectral Masking Minimization

This approach is based on the paper "Improved Control for Selective Minimization of Masking using Inter-Channel Dependency effects" by Gonzalez and Reiss [28]. The authors describe their approach to spectral masking minimization using a specific architecture that logs the spectral characteristics of multiple tracks, and uses this information to attenuate other tracks in the mix that might cause masking of the master channel.

This chapter is demarcated as follows. Section 8.1 describes the control architecture used by the authors, section 8.2 enumerates certain observations concerning the effect of Reverb on the spectral attributes of a track and how they will affect design choices for our adaptation. Section 8.3 enumerates the specific design choices made and their justification, before their use is explained in greater detail in the following section describing the Adaptation. Section 8.5 explains the details of the implementation and the results obtained.

8.1. Description

This section details the algorithm and the control architecture used by the authors to achieve their intended means. Starting out, the authors explain the different control architectures possible for effects implementations: direct user control, auto-adaptive, external-adaptive and cross adaptive effects. Direct user control effects depend upon the user's judgment and experience, auto-adaptive effects derive their control parameters from a feature extracted from an input track, external-adaptive effects derive the control parameters from a different track to the one on which it is applied, and cross adaptive effects derive their control parameters through the analysis of the contents of each track with respect to the other tracks.

Firmly qualifying the control architecture used as being in the category of a cross adaptive effect, its implementation is further detailed. The objective cross track characteristic used for deriving the control parameters is Spectral masking; a sound artifact which causes loss of perception of the spectral characteristics of one or more channels when they are mixed together.

8.1.1. Quantifying Spectral Masking

The authors define spectral masking for a specific track Ch_m against the rest of the mix as:

$$\mathbf{SM}_t = \left[\mathbf{FFT}(CH_m)\right]^2 - \left[\mathbf{FFT}(mix - CH_m)\right]^2 \tag{8.1}$$

where $SM > 0$ defines the track as unmasked and $SM < 0$ defines the track as masked by the mix. **FFT** denotes the Fast-Fourier transform. Note that equation 8.1 defines the **SM** per time frame, denoted by the subscipt t. $\mathbf{SM}_t$ is thus a vector containing values for a particular time frame over multiple frequency bins. The values obtained can change with the FFT resolution used, and no windowing is to be used since spectral masking is an amplitude difference measurement.

Based on the previous equation, an accumulated spectral masking measure or ASM can be obtained

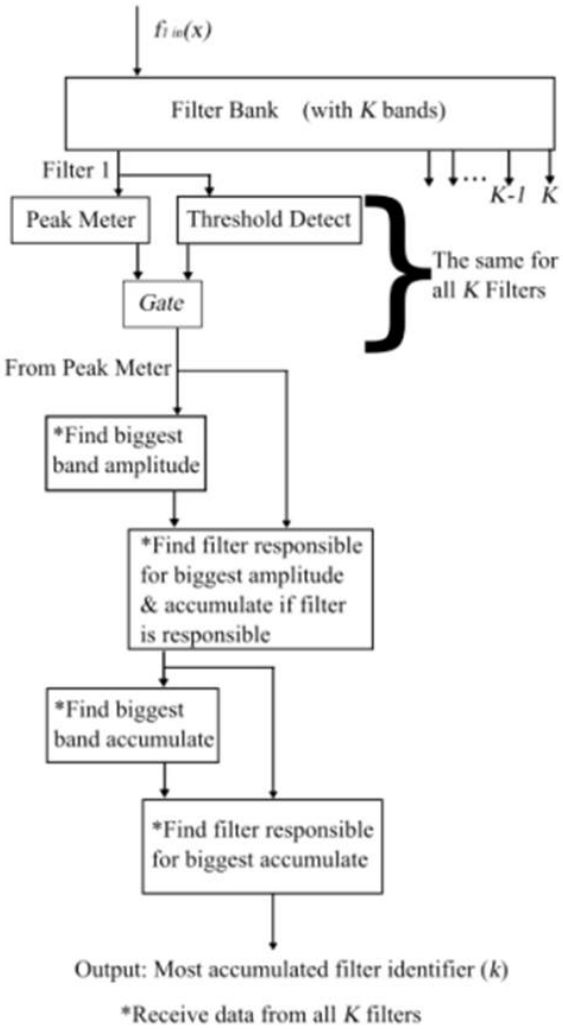


Figure 8.1: Accumulative Spectral Decomposition Architecture, taken from [28]

as follows:

$$\mathbf{ASM} = \sum_{t=0}^{\inf} \mathbf{SM}_t \tag{8.2}$$

This measure, which gives us an accumulated spectral masking for a particular track over multiple frames, can be used as an objective quantity to optimize the control parameters for multiple effects implementation.

The authors apply the basis described above, to create a control architecture that lowers the levels of tracks that have similar spectral content to the track to be enhanced. Control parameters are derived by comparing the spectral characteristics of all the tracks involved in the mix and hence, this is a cross-adaptive effect.

8.1.2. Mathematical Implementation

The implementation starts with the spectral analysis of all the tracks involved. An accumulative spectral decomposition classification method presented by the same authors in [6] is used.

Figure 8.1 details the control architecture used for analysis of the spectral content of the different tracks. The filter bank has K filters, which equals the number of tracks being mixed, say N. Each filter is assigned a k_n value that varies between 1 and N. The filter bank analyses each input track, and a score related to the maximum peak excitation filter is accumulated every 1 ms. Thus, depending on the spectral content of the track, the accumulative spectral decomposition architecture assigns each

track a k_n value which depends on the most accumulated filter at the end of the analysis. The higher the value of n, the greater frequency content the track has in the higher end of the spectrum.

The ASD classification architecture thus, outputs a feature corresponding to a particular filter in the filter bank.

$$k_n = \textbf{ASD}(Ch_n) \tag{8.3}$$

The filters are chosen such that they boost the low frequencies while gently attenuating the gain of the higher frequencies within each filter bandwidth. The authors state that this approach is taken to reduce noise associated with higher frequencies.

The k_n value obtained for each track is then subsequently used to implement a Gaussian dependency that defines the attenuation values for the other tracks, in relation to the track being enhanced.

A unitarily normalized gaussian function is used:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \boldsymbol{e}^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{8.4}$$

where x represents frequency, μ is a constant that represents the mean of the Gaussian function, σ controls the width of the Gaussian and can be user controlled using a variable Q that controls the rate of attenuation of the different tracks.

$f(x)$ is now modified to satisfy the design requirements of the algorithm. First, $f(x)$ is normalized to one, the absolute value of its complement calculated and a user controllable attenuation function A is added. This gives us $G(x)$ which is an inter-track dependency mapping function:

$$G(x) = \left| \left[A\left(\frac{f(x)}{\frac{1}{\sigma\sqrt{2\pi}}} \right) \right] - 1 \right| \tag{8.5}$$

Because the Gaussian function must be centered at k_m, which is the ASD classification of the master track to be enhanced, μ must be related to k_m.

$$\mu = \left[\frac{2}{N-1}(k_m - 1) \right] - 1 \tag{8.6}$$

where N is the total number of tracks and hence the total number of filters in the filterbank and the subscript m identifies the master channel to be enhanced.

The master track gain G_m, is kept unchanged at 1, while the gain value for the other tracks G_n is calculated by evaluating x in equation 8.5 with respect to the k_n spectral classification obtained for the track. This gives us:

$$x = \left[\frac{2}{N-1}(k_n - 1) \right] - 1 \tag{8.7}$$

Thus, the algorithm uses five important variables:

1. **Track number location:** The location of a track asking for a control value.
2. **Total number of tracks:** This is user selected and helps determine the number of filters present in the filterbank.
3. **Master track:** User selected variable determining the track to be enhanced.
4. **Attenuation:** User selected variable that determines the maximum attenuation applied to tracks that have the same spectral classification as the master track.
5. **Q:** User selected variable that controls the smoothness and spread of the Gaussian attenuation curve.

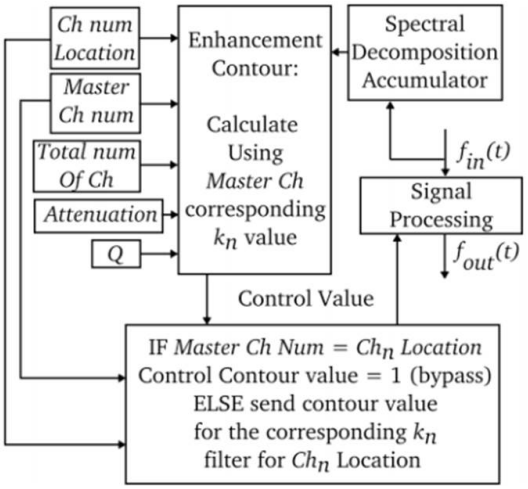


Figure 8.2: Block diagram of the Gaussian dependency algorithm, taken from [28]

Figure 8.2 shows the block diagram of the Gaussian dependency algorithm and its dependency on the Accumulative Spectral Decomposition architecture detailed before. The final mix after processing and attenuation of select tracks is given by:

$$mix_g(t) = \sum_{n=1}^{N} G_n Ch_n(t) \tag{8.8}$$

where N is the total number of tracks, G_n is the gain for channel Ch_n. The authors then go on to explain the algorithm's utility to reduce spectral masking without the use of equalization filters and also its adaptability, by applying the same algorithm to decorrelating the phase information of the right and left channels to reduce directional masking.

8.1.3. Analysis

The authors present an adaptable algorithm and a cross adaptive effects implementation architecture to reduce spectral masking on a master track by suitable attenuation of other tracks with similar spectral content. The strengths of this application are:

1. Cross adaptive architecture ensures that tracks are not treated in isolation.
2. Track priority is implemented by choosing a master channel.
3. The accumulative spectral decomposition algorithm is capable of quick analysis and feature extraction from the input tracks.
4. Gaussian dependency on attenuation ensures a smooth transition for fader values and reduced attenuation of tracks that do not cause spectral masking.
5. The simple and elegant architecture ensures wide adaptability to effects that influence spectral characteristics of a track.
6. A computationally efficient algorithm ensures real time compatibility.

The weaknesses of the application are:

1. **The number of filters are limited to the number of tracks being processed:** This severely limits the spectral resolution of the filterbanks when the number of tracks is low.
2. **Detachment from studio practices:** The implementation of track loudness balance using faders and spectral balance using equalization are very distinct processing steps in the studio. Equalization can help us improve loudness balance of the tracks by eliminating or attenuating

spectral regions of the track that are either unnecessary or may cause spectral imbalance and masking. But, reducing spectral masking merely through attenuation can cause loudness imbalance due to unnecessary spectral regions of tracks still being present and increasing the attenuation values further.

3. **Lack of implementation of perceptual features:** Equal loudness curves dictate that the perceptivity across frequencies is not equal and will change with either loudness increase or decrease. This change in perceptivity is not taken into account with the simplistic spectral characterization of tracks performed.

4. **Focus on optimization of a single track:** Using this algorithm, only a single track can be spectrally unmasked in the mix. Applying the same algorithm further to the other tracks would lead to a change in the optimal parameters derived in the first iteration. Hence, there is no real way to optimize parameters to minimize masking across tracks, rather to just minimize the masking of a singular master track.

8.2. Observations

The nature of Reverb and its effect on the spectral power density of an instrument was clearly discussed in Chapter 3. Spectral imbalance and masking can be a result of excessive reverberation due to the frequency specific nature of sound spreading and room acoustics. On the surface this method seems to be directly adaptable to the application of Reverb, with greater attenuation attributed to tracks more prone to spectral clutter and masking. The challenge lies in attributing track hierarchy, around which attenuation should be based and taking steps to minimize spectral masking and audio artifacts.

8.2.1. Possible Issues

The possible issues to be solved and challenges to be addressed are as follows:

1. **Assigning Track priority**: With loudness, assigning track priority is straightforward as an increase in loudness pushes the track "forward" and makes it more noticeable. Reverb application however can push a track to the background and make it sound farther; excessive reverberation can also cause sound smearing and timing clashes.

2. **Limiting "Booming"**: Due to the disparate effect of reverberation on low and high frequencies, tracks with low frequency content are generally assigned lesser reverberation than other tracks [23].

3. **Short and long Reverbs will show distinctly different spectral effects**: Acoustic rooms of different sizes tend to highlight different parts of the spectrum depending on the eigenfrequencies excited. Implementing the same track hierarchy over both these Reverbs may lead to setting the fader gain values for both reverbs of each track to the same value; this will be both spectrally and sonically suboptimal.

8.3. Design Choices

This sections enumerates and explains important design choices made for our implementation. These are as follows:

1. **Track Priority**: Track priority will be based on the extent of masking calculated with the difference in power spectral densities of completely dry and completely wet (reverb fader gain 1) tracks, as defined in equations 8.1 and 8.2, but for a single track rather than comparing with the entire mix.

$$\mathbf{SM}_t = \left[\mathbf{FFT}(CH_{wet})\right]^2 - \left[\mathbf{FFT}(CH_{dry})\right]^2 \tag{8.9}$$

$$\mathbf{ASM} = \sum_{t=0}^{\inf} \mathbf{SM}_t \tag{8.10}$$

This ASM represents a numerical value which can be used for assigning track priority. The larger the ASM value, the greater the masking potential of the track. This measure is more suited to our implementation against the ASD architecture as shown in Figure 8.1 because:

- The frequency resolution for track analysis is much greater than the ASD architecture used by the authors.
- This allows us to treat long and short reverbs independently and get separate track priority assignments for either.
- This track priority is more attuned to the spectral masking caused due to peculiar characteristics of the reverb (eigenfrequencies excited, standing waves formed) than just the spectral characteristics of the track itself; and hence, is a more accurate reflection of extent of spectral masking due to reverb, while being highly adaptable.

2. A **value of unitarily normalized gaussian**: This represents the maximum fader gain value that will be assigned to the track with the least **ASM** value. Since fader gain values are always between 0 and 1 [23], the maximum value will be set to 0.8. This is an arbitrarily defined value chosen thus:
   - Fader gain values are always between 0 and 1 [23] and in the author's experience, the maximum value of 1 is rarely seen in practice.
   - By definition, an automatic reverb implementation ideally should not require user input.
   - This maximum value could possibly be track and/or genre specific and require further study to optimize. But for an introductory study on the viability of automatic Reverb implementation, this offers a good starting point.

8.4. Adaptation

We start our adaptation with the spectral masking measure as defined by the authors, but adapted for measuring the spectral difference between dry and wet (completely reverberated) tracks, as defined in equations 8.9 and 8.10.

$$\mathbf{SM}_t = \left[\mathbf{FFT}(CH_{wet})\right]^2 - \left[\mathbf{FFT}(CH_{dry})\right]^2 \tag{8.11}$$

$$\mathbf{ASM}_n = \sum_{t=0}^{\inf} \mathbf{SM}_t \tag{8.12}$$

where the subscript n denotes a specific track. This track specific $\mathbf{ASM}_n$ value gives us the measure we use in order to assign track priority separately for both long and short Reverbs. The track with the highest value gets the lowest priority and vice-versa.

Using the unitarily normalized gaussian function as defined in equation 8.4, we can obtain an inter-track dependency mapping function.

$$G(x) = \left|\left[A\left(\frac{f(x)}{\frac{1}{\sigma\sqrt{2\pi}}}\right)\right]\right| \tag{8.13}$$

Note, that in contrast with equation 8.5, we no longer invert the Gaussian by calculating its complement. This gaussian function is centered at k_m, which is the track with the highest priority i.e; the track with the lowest **ASM** value. The μ value of the gaussian is related to k_m thus:

$$\mu = \left[\frac{2}{N-1}(k_m - 1)\right] \tag{8.14}$$

N here refers to the total number of tracks, which in our case is 5.

The fader gain values for all the tracks is then calculated using their priority values k_n to evaluate x which is then input into equation 8.13 to get the final value. x is calculated using the following equation:

$$x = \left[\frac{2}{N-1}(k_n - 1)\right] \tag{8.15}$$

After calculating the fader gain values, the final mix will be given by:

$$mix_g(t) = \sum_{n=1}^{N} \Big(Ch_n(t) + G_{nShort}\mathbf{R_s}\big(Ch_n(t)\big) + G_{nLong}\mathbf{R_l}\big(Ch_n(t)\big) \Big) \tag{8.16}$$

where G_{nShort} and G_{nLong} are track specific fader gain values calculated for both Short and Long reverbs individually. $\mathbf{R_s}$ and $\mathbf{R_l}$ denote the short and long reverberation operations.

The fader gain values obtained by the algorithm for different genres are organized in the tables below.

Table 8.1: Fader gain values for genre Classical

Track Name	Pair Front MS	Pair Overhead ORTF	Spot High Strings	Spot Low Strings	Vocals
Long Reverb	0.7999	0.5809	0.6682	0.7384	0.7841
Short Reverb	0.7999	0.5809	0.7841	0.7384	0.6682

Table 8.2: Fader gain values for genre Drum and Bass

Track Name	Bass	Drums	Harp	Piano	Synth
Long Reverb	0.7999	0.6682	0.5809	0.7384	0.7841
Short Reverb	0.7999	0.5809	0.7841	0.7384	0.6682

Table 8.3: Fader gain values for genre Heavy Metal

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.7999	0.5809	0.7384	0.7841	0.6682
Short Reverb	0.7999	0.5809	0.7841	0.7384	0.6682

Table 8.4: Fader gain values for genre Pop

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.7999	0.5809	0.6682	0.7384	0.7841
Short Reverb	0.7999	0.5809	0.6682	0.7384	0.7841

Table 8.5: Fader gain values for genre Singer-Songwriter

Track Name	Bass	Drums	fiddle	Guitar and Mandolin	Vocals
Long Reverb	0.7999	0.5809	0.6682	0.7384	0.7841
Short Reverb	0.7999	0.5809	0.6216	0.7384	0.7841

8.4.1. Further adaptation: Alternate implementation

The fader gain values are fixed values varying from 0.5809 to 0.7999 allocated to tracks sorted according to their **ASM** value. This gives us rather unrealistic values which suffer from the following issues:

1. No wide variation in values as seen in chapter 7 or in values set by studio professionals.
2. No direct correlation between actual masking potential of the track and final fader gain value.
3. Lack of adaptability to the greater spectral masking potential of long reverb versus short reverb.

It is evident that these issues can be fixed with an adaptation that offers an inverse correlation between the masking potential as defined by the **ASM** value and the final fader gain values. This can be mathematically defined as follows:

$$MP_n = \left(\frac{N_{masked}}{N_{total}} \right) 100 \tag{8.17}$$

where MP_n stands for masking potential of the n^{th} track and N represents the number of frames, the subscripts being self explanatory. Note that, the determination of whether a frame is masked or not depends upon the frame specific value of equation 8.11 being positive.

In order to inversely correlate this percentage value to the fader gain value, we use the following equation.

$$G_{nR} = 1 - \left(\frac{MP_n}{100} \right) \tag{8.18}$$

where the subscripts n and R identify the fader gain value G to be specific to a track and a reverb (long or short).

This simplistic representation is advantageous for a number of reasons:

1. Inversely correlates spectral masking potential to reverb fader gain value.
2. No need for arbitrary setting of Q and A values as required by the previous gaussian function, making this more adaptable and truly automatic as an implementation.
3. Highly adaptable to the varying nature of tracks, specific instruments and even artifical reverb emulators.
4. It is specifically adaptable to the loudness of individual tracks as the masking potential can be recalculated if the track loudness is altered.
5. Gives us a clear way to determine the effect of minimizing spectral masking on subjective perception of mixed multi-track audio, before proceeding with more complicated adaptations.

8.5. Implementation

The designed adaptation was implemented in MATLAB. The fader gain values obtained for each genre are implemented in the additive reverberation set-up to create tracks that will be used in the comparative testing phase. The reverb fader gain values obtained from the adjusted adaptation are as follows:

Table 8.6: Fader gain values for genre Classical(Alternate Implementation)

Track Name	Pair Front MS	Pair Overhead ORTF	Spot High Strings	Spot Low Strings	Vocals
Long Reverb	0.4592	0.5391	0.4467	0.2708	0.5309
Short Reverb	0.6238	0.6723	0.6631	0.5819	0.6625

Table 8.7: Fader gain values for genre Drum and Bass(Alternate Implementation)

Track Name	Bass	Drums	Harp	Piano	Synth
Long Reverb	4648	0.5554	0.3922	0.1548	0.5833
Short Reverb	0.5740	0.6929	0.50621	0.2938	0.7138

Table 8.8: Fader gain values for genre Heavy Metal(Alternate Implementation)

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.5892	0.5356	0.5771	0.4718	0.5920
Short Reverb	0.7178	0.7220	0.7058	0.7074	0.7220

Table 8.9: Fader gain values for genre Pop(Alternate Implementation)

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.5937	0.4614	0.4212	0.5812	0.5927
Short Reverb	0.7245	0.7215	0.5333	0.7200	0.7214

Table 8.10: Fader gain values for genre Singer-Songwriter(Alternate Implementation)

Track Name	Bass	Drums	fiddle	Guitar and Mandolin	Vocals
Long Reverb	0.5968	0.5963	0.5330	0.5828	0.5923
Short Reverb	0.7251	0.7257	0.6966	0.7136	0.7250

We now have our results from both the mathematical optimization and spectral masking minimization algorithms from the last 2 chapters. Including results from a professional producer, this gives us enough preliminary data that we can use to conduct subjective listening tests. The next chapter introduces the layout and the format of these listening tests.

9

Subjective Listening Tests

Music being an inherently subjective experience, introduces multiple challenges in its evaluation. Since the central question of this thesis is whether algorithmic addition of Reverb to multi-track audio can compare to a professional producer's input, this chapter is focused on the selection and implementation of a subjective test method that evaluates tracks processed through the algorithms designed and implemented in this thesis, against a professional producer's analogous results. This chapter is structured as follow. The first section documents some details about the professional producer and the gain values set according to his subjective preferences. The second section introduces the different types of subjective listening tests in use today. Section 3 includes the selection and justification of the MUSHRA listening test for this thesis and section 4 details the process involved in the conduct of the listening tests.

9.1. Professional Producer Input

A professional producer with 8 years of live sound arrangement experience and 1.5 years of audio mixing experience[1] was consulted and asked to set the respective short and long Reverb fader gain settings for the different tracks of the 5 genres. His input is as recorded below.

Table 9.1: Fader gain values for genre Classical(Professional Input)

Track Name	Pair Front MS	Pair Overhead ORTF	Spot High Strings	Spot Low Strings	Vocals
Long Reverb	0	0	0.3308	0	0
Short Reverb	0	0.4994	0.2210	0	0.0118

Table 9.2: Fader gain values for genre Drum and Bass(Professional Input)

Track Name	Bass	Drums	Harp	Piano	Synth
Long Reverb	0	0	0	0	0.4028
Short Reverb	0	0	0.0883	0.	0

Table 9.3: Fader gain values for genre Heavy Metal(Professional Input)

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0	0	0	0	0
Short Reverb	0.3493	0	0	0.4179	0.2951

Table 9.4: Fader gain values for genre Pop(Professional Input)

Track Name	Backing Vocals	Bass	Drums	Electric Guitar	Lead Vocals
Long Reverb	0.3197	0	0	0	0
Short Reverb	0	0	0.3962	0	0.2432

[1] http://www.monohive.com

Table 9.5: Fader gain values for genre Singer-Songwriter(Professional Input)					
Track Name	Bass	Drums	fiddle	Guitar and Mandolin	Vocals
Long Reverb	0	0	0	0	0
Short Reverb	0	0.3498	0.2965	0.1827	0.1116

It is clear that the gain values set are sparse and minimal, which follow the guidelines of producer Mike Senior [24] to err on the side of caution and apply minimal Reverb for least clutter and smearing. In a post interview with the producer, he revealed that his ability to set the optimal reverb gains for any mix was limited by the pre-processing already performed. He would have set different values if he had more control over the filtering, compression and panning tasks of the mixing process. This conforms with our view of mixing being a recursive process, with the multiple effects working in tandem to create a cohesive whole according to the experience and judgment of the producer. Even though the producer was limited in his approach to adding reverberation to the mixes, the set-up used necessary to provide equality of basis and opportunity to both the algorithms and the professional producer. This provides a critical control aspect which is necessary for us to be able to judge the performance of the algorithms developed against the work of a professional.

9.2. Cursory Listening Analysis

A preliminary listening analysis of the tracks after initial compilation, results in the following observations:

- **1. Greater Spatial Impression:** The tracks produced using algorithmic fader gain settings show a greater spatial impression as compared to the professional track. This does not translate to a better overall impression however, as the excessive smearing produced reduces the quality of the production.
- **2. Improper Depth hierarchy:** Tracks that aim to showcase the vocals as the most important component of the mix suffer as a result of increased Reverb, as the vocals seem to be pushed further away from the listener in the sound-stage. The increased loudness balance of the vocals is counteracted by the smearing and interaction of the other components in the mix. Pop and Singer-Songwriter tracks are the most affected, although the Classical genre does not display the same level of quality degradation.
- **3. Increased Clutter:** Increased Reverb levels also cause an increase in clutter due to sound smearing, negatively affecting the distinction between the instruments.
- **4. Effect on Timbre:** Different levels of Reverb on the algorithmically produced tracks as against the professionally set track, also create different textures and timbres for multiple instruments across genres. This is readily apparent with the synth sounds in the Drum & Bass tracks.

9.3. Types of listening tests

This section describes the types of listening tests in use today, their possible uses and requirements for their conduct. Loudness equalization and stimulus randomization is the common feature of the all the tests described below.

9.3.1. ITU-R Recommendation BS.1116

Also called an ABX or ABC reference test, described in the ITU-R Recommendation BS.1116 [41] "Methods for the subjective assessment of small impairments in audio systems"; requires users to rate the subjective quality of tracks 'B' and 'C' against a reference track 'A'. The tracks 'B' and 'C' are rated on a five grade impairment scale depending upon the level of impairments present. One of the tracks 'B' or 'C' is a hidden reference.

9.3.2. ITU-R Recommendation BS.1534

Also called a MUSHRA (MUltiple Stimulus with Hidden Reference and Anchor), described in ITU-R Recommendation BS.1534 "Method for the subjective assessment of intermediate quality level of coding systems" [42] ; requires users to rate multiple tracks according to subjective perceptual quality against

a reference tracks. One of the tracks being rated is the reference itself, hidden and placed randomly among the stimuli. And another track is an anchor, which is a low quality, low pass filtered version of the reference track under test. The stimuli are rated on a continuous quality scale which varies from 0-100, where 100 denotes the best possible score.

9.3.3. Audio Perceptual Evaluation test

Described by Reiss and De Man [43] as a more flexible and versatile test environment; it allows users to rate tracks against each other on multiple scales and based on different perceptual and quality metrics. The choice of including a reference and an anchor are entirely optional.

9.4. Selection and Justification

We will use the MUSHRA listening test to derive subjective evaluation results. The justification for this selection is as follows:

- **1. Use of multiple test samples:** This precludes the use of ABX test which can only acquire subjective ratings for one test sample at a time.
- **2. Necessity of a reference:** A reference for every genre can possibly reduce bias caused due to preferences for a certain kind of music. The central question of this thesis (can algorithms compare to a professional producer?) also require a reference to be used. Although all three tests detailed can include a hidden reference, the APE test does not allow the test subject to play the reference track at will, in order to be able to do a fair comparison. It is assumed that the reference would be rated the highest, if included. For a test that spans multiple genres that can display differing impairments and quality attributes, the ability to play a reference track at will is important, and enables the test subjects to fairly judge each genre individually.
- **3. Tracks rated against a reference, not against each other:** The APE test interface is designed to allow users to implicitly rate tracks against each other. Even if a reference is provided, it is always hidden, not allowing for an open reference or a recursive adjustment of results based on multiple playback and firm knowledge of the reference track. As mentioned in the previous point, the aims of this thesis require firm judgment of the test samples against the reference track and not necessarily against each other.
- **4. Hidden anchor and reference provide greater spread of quality ratings:** The hidden reference is expected to be given the maximum value and even though the users are not aware of the anchor, it is required to be of a perceptually obvious lower quality, allowing for a spread of ratings through the entire scale. This is an advantage as opposed to the ABX test which does not allow for a hidden anchor.
- **6. Rating the entire impression rather than selective subjective parameters:** Production quality audio is prepared for the masses, for large scale consumption across varying cultures, preferences and environments and hence, in general, it is judged as a cohesive whole rather than specifically on individual parameters. The previous chapters have also revealed that Reverberation influences production audio in multiple ways beyond just providing spatial impression, it is also important to create realism, improving timbre of instruments, creating/enhancing moods, filling time gaps etc.

  In order for us to then get a realistic impression of such a judgment, it is important for the test samples to be rated on their overall impression against the reference, rather than on Reverb specific attributes such as spatial impression, depth etc. The APE test works best when samples provided are rated on specific attributes; the lack of an open reference also makes such attribute independent judgment difficult, which makes the MUSHRA test preferable in this case.
- **7. MUSHRA test provides greater grading resolution:** The ABX test only allows ratings on a five grade impairment scale, whereas the MUSHRA test allows test subjects to rate stimuli on a continuous scale from 0-100. The greater resolution provided by this scale allows us to make more robust statistical conclusions.

- **8. The MUSHRA test is designed to allow judgment of wider impairments:** The ABX reference test is poor at discriminating between small differences in quality at the bottom of the scale [42]. Aside from providing greater rating resolution, the MUSHRA test is also capable of allowing the judgment of both small and large scale impairments rather than just small scale ones.

9.5. Suitable Modifications

It is important to note that the guidelines provided for the test are recommendations rather than hard rules. This allows us to incorporate context dependent modifications that are more suited to the requirements of this thesis. These modifications are as follows:

- **1. Requirement of critical listeners unnecessary:** The MUSHRA test recommendation document [42] suggests the use of experienced listeners who are screened based on their discrimination ability and reliability of ratings. This is considered unnecessary in our context, as production audio is not specifically made for critical listening but for mass consumption across varying listening environments, genre preferences, nationality etc. The use of critical listeners also introduces additional complexity and resource intensiveness that can discourage large scale participation in the listening test.
- **2. At least one stimuli required to be given the maximum rating:** The recommendations suggest that no explicit rating guidelines be given during the tests. The hidden reference is expected to be rated above 90. However, in our test we constrain the user to rate at least one of the samples 100, i.e; the maximum rating. This is necessary not just to obtain a greater spread of scores, but also because it underscores the importance of rating the test samples against the reference track. This is especially important since the tracks are implicitly being judged on their overall impression rather than on subjective perceptual parameters. A greater spread of scores allows us to deduce exactly how comparable our algorithmic reverb methods are to a professional producer, but the method also allows a test subject to rate a non-reference track higher based on a favorable impression.

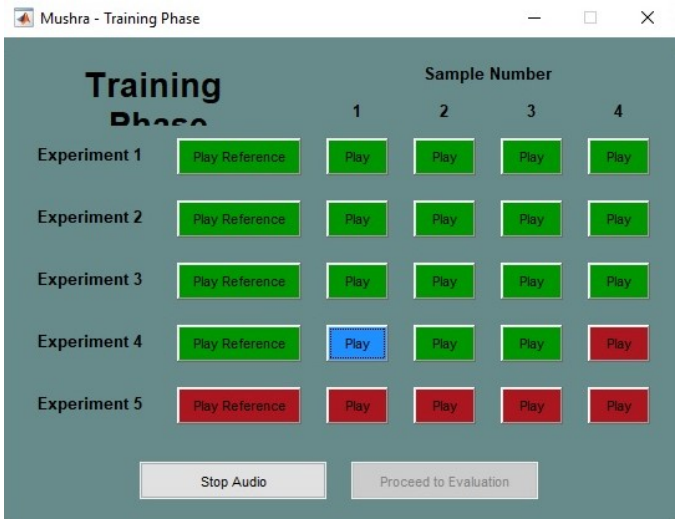
9.6. Listening Tests

MUSHRA listening tests were conducted using the MUSHRA MATLAB toolbox developed by Sean Enderby[2]; suitably modified for compatibilty on MATLAB R2016b. A pre-test survey was included, inquiring about the participant's age, whether they have any audio mixing experience, if they play an instrument, hours music listened to per week and whether they have any preferences for certain genres. 22 participants completed the listening tests. The mean age of the participants was 25.6 years. Only 2 test subjects had experience with mixing audio, and the mean number of hours the participants listened to music was 28.6 hours. None of the participants showed a strong inclination towards or preference for a particular genre.

The test stimuli included the hidden reference (producer set fader gain values), track produced using gain values obtained by mathematical optimization (Chapter 7), track produced using gain values obtained by spectral masking minimization (Chapter 8), an unreverberated track (all fader gain values set to 0) and a low pass filtered version of the unreverberated track as a hidden anchor.

This chapter focused on the selection, customization and set up of the subjective listening test. The next chapter compiles the results obtained, upon which an analysis is conducted.

[2]https://sourceforge.net/projects/matlabmushra/



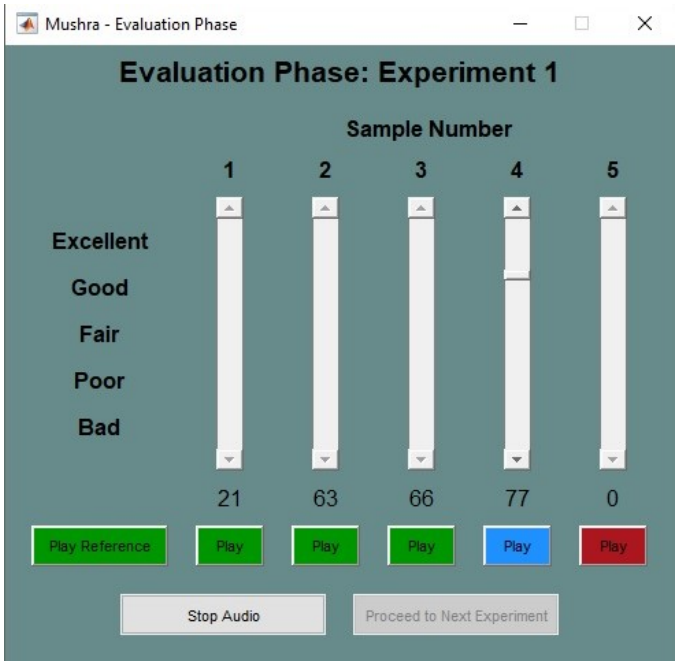


Figure 9.1: Training and Test interfaces for the MUSHRA test implemented on MATLAB. The colors of the buttons represent the play status: a blue button denotes a playing track, a green one denotes a track that has been played at least once and a red button denotes a track that has not been played yet

10

Results, Analysis and Conclusions

This chapter catalogues the results obtained through the subjective listening tests conducted. This chapter is set up as follows, section 1 contains the histograms depicting the normalized grades obtained by the test stimuli for each genre experiment. Section 2 contains an analysis based on the results and Section 3 details the conclusions. The chapter culminates with a discussion of the contributions made to the research field and possible future directions.

10.1. Results

Figures 10.1 to 10.3 display the histograms of the normalized grades obtained for the multiple genres under test. Sample numbers 1-3 represent the tracks produced using the professional producer input, mathematical optimization and spectral masking minimization. Sample numbers 4 and 5 represent the unreverberated track and the low pass filtered anchor respectively. The small red bars denote the 95% confidence intervals.

Table 10.1 tabulates the numerical values for the normalized grades obtained for the multiple genres.

Table 10.1 Normalized grades for MUSHRA Test Stimuli

Genre/Track	Professional	Mathematical	Spectral	unreverberated	Anchor
Classical	92.2401	74.3215	56.0689	65.8725	7.0242
Drum and Bass	93.2885	48.1415	28.1835	83.7368	10.5182
Heavy Metal	96.1172	50.8991	29.8163	79.8380	2.6730
Pop	93.6764	48.9300	31.7471	77.6477	3.6034
Singer-Songwriter	94.9062	40.5594	24.3085	91.9182	3.0426

It can be seen that the anchor (sample 5) is rated consistently lower than all other tracks and the reference (sample 1; professionally produced track) is rated higher than all the other tracks. The numerical grades support the histograms in showing that the reference track (produced using professional input) is always rated above 90 points and the anchor is always rated below 15 points.

Conducting a two-way analysis of variance on the data in table 10.2 yields p-values of 0 and 0.6501 across the columns and rows respectively. This suggests that the test subjects consistently differentiate between the different tracks within a genre, however, the ratings show similar trends across genres, which is to be expected.

10.2. Analysis and Discussion

This section conducts a basic analysis of the obtained results followed by a discussion. For brevity the different tracks under test will be referred to by the method used to obtain the values for the fader gain settings i.e; professional, mathematical, spectral, unreverberated and anchor.

10.2.1. Analysis of Subjective Test Results

The numerical normalized grades obtained for the multiple tracks across genres allow us to make the following observations:

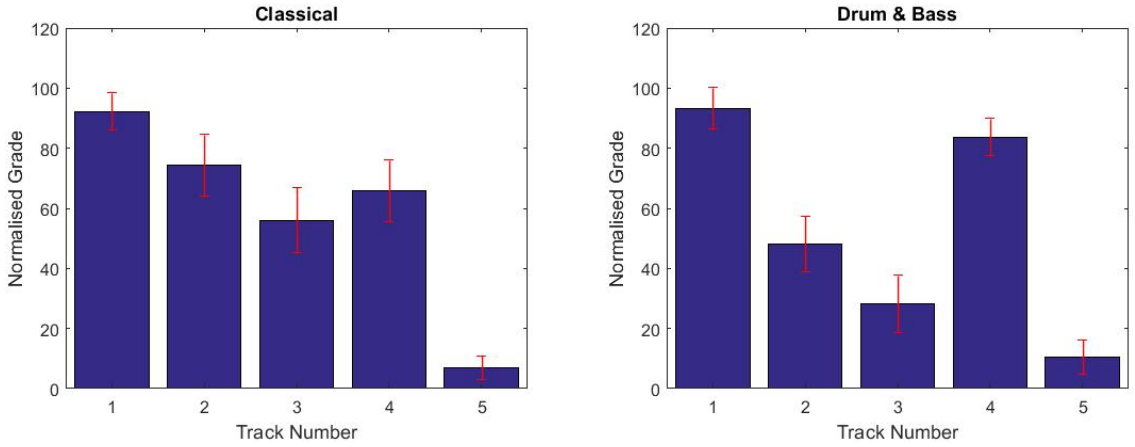


Figure 10.1: Results for Classical and Drum and Bass; tracks 1-5 are Professional, mathematical optimization, spectral masking minimization, unreverberated track and hidden anchor

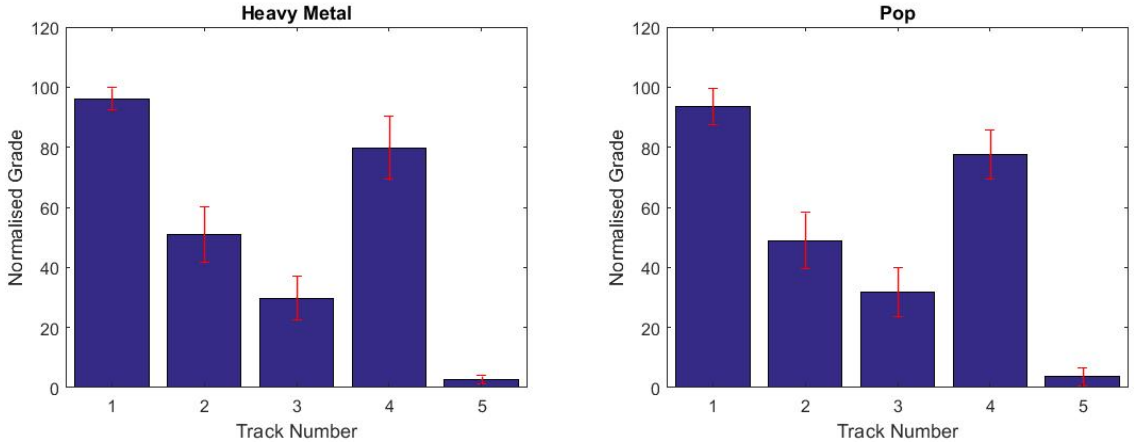


Figure 10.2: Results for Heavy Metal and Pop music; tracks 1-5 are Professional, mathematical optimization, spectral masking minimization, unreverberated track and hidden anchor

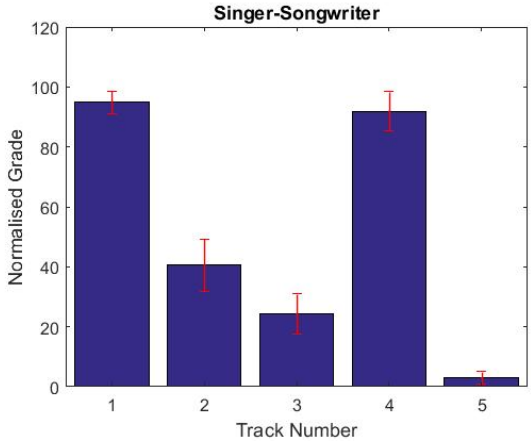


Figure 10.3: Results for singer-songwriter; tracks 1-5 are Professional, mathematical optimization, spectral masking minimization, unreverberated track and hidden anchor

- **1. Professional track is rated highest:** Subjects reliably rated the reference track higher than all other test stimuli. In a post test interview with the test subjects, 2 subjects did explicitly mention the close contest between a couple of test samples, and how the sample eventually

rated higher sounded "fuller" and more "spacious". This validates our theoretical research of Reverb contributing to a more positive overall impression by melding tracks together in a common acoustic space (section 4.3.1, points 1,4 and 5).

- **2. Unreverberated track is rated lower than the reference:** The test subjects are able to differentiate between the professional reference track and the unreverberated tracks in-spite of the sparse Reverb gains set. The ratings for the professional track defer by a mean of 14.2358 against the unreverberated tracks. The lowest difference is observed for the Singer-Songwriter track and the highest for the Classical track.

- **3. Algorithmically produced tracks rated lower than reference:** The algorithmically produced tracks differ by a mean of 41.4754 and 60.0209 rating points for the mathematical and spectral tracks respectively.

- **4. Algorithmically produced tracks rated lower than the unreverberated track:** The algorithmically produced tracks are also outperformed in subjective ratings by the unreverberated track, indicating excessive reveberation. This trend though has an exception in the classical genre, where the mathematically produced track is rated higher than the unreverberated track.

- **5. Mathematical track outscores spectral tracks on subjective ratings:** The ratings for the mathematical track are on average 18.5423 rating points higher than the spectral track.

- **6. Algorithmic tracks for the classical genre show better results:** The mathematical track for the classical genre outscores the unreverberated track by 8.449 rating points. The rating points for the mathematical and spectral ratings obtained for the classical genre are 21.7512 and 22.0441 respectively, above the mean for the method of implementation.

10.3. Conclusions and Discussion

Through the observations made by an analysis of the subjective test results and from a cursory listening, the following conclusions can be drawn:

- **1. Greater Spatial Impression does not translate to a higher quality production:** As reiterated before in the observations on a cursory listening (section 9.2), the intensity of a spatial impression is not directly related to the perceptual quality of a production. This is clearly demonstrated by the ratings for both the mathematical and spectral track ratings across genres being lower than both the unreverberated and the professional reference tracks.

  The spectral masking minimization algorithm does not specifically modify the spatial impression, but the act of maximizing the sum of fader gain values for a set allowable increase in loudness for the mathematical algorithm (section 7.4.1, point 2), encourages a bounded increase in spatial impression. This may be counter-productive in hindsight. The use of a lower bounded value could help reduce the excessive Reverb fader gain values obtained.

- **2. Depth Hierarchy must be maintained:** Loudness balance is often the first step of the recursive mixing process, and arguably the most important, considering the initial loudness balance set is always readjusted recursively after further adjustments to maintain the initial impression obtained. Reverb has the potential to upset this balance not via any explicit change in track loudness, but through the smearing of sound events, causing a spillover of the loudness of track components into quieter sections. This is most often seen with an excessive use of long Reverb as longer delay times cause greater smearing and clutter between closely placed sound events and track components. Short Reverb has a lesser tendency to cause smearing as compared to long Reverb, but can affect the nature of timbre more clearly.

  Although the mathematical algorithm does look to preserve loudness hierarchy, the use of a singular value averaged over the entire song segment (equation 7.5) does not account for the spectral characteristics of the components or the sound event density, and in hindsight is too poor in temporal resolution and information compression to be a reliable and effective loudness-balance preserving measure.

The spectral masking minimization algorithm does not take the loudness influenced depth hierarchy into account at all, which can be another reason for the unrealistic Reverb fader gain values obtained through it.

- **3. Sound Event density needs to be factored in:** It is noticed that the subjective rating results obtained for the classical genre are better than the other genres. This is postulated to be due to the sparseness of the sound event density. The sparse sound event density leaves greater periods of temporal and spectral silence in between sound events and track components that can tolerate higher levels of added Reverb (section 4.3.1, points 6,7 and 9). The sparse sound event density in this case also makes any added Reverb more readily noticeable, which can account for the unreverberated track being rated lower than the mathematical track. Neither of the two Reverb addition algorithms account for the sound event density or sparseness.

- **4. Spectral masking minimization is not a goal of Reverb addition:** The spectral track ratings are the lowest among the test stimuli if we do not consider the anchor track. This confirms our viewpoint of the mixing process as not using Reverb as part of the strategy to minimize spectral masking among components. Equalization and filtering is the distinct step that is specifically used to remove or reduce attributes that help to reduce spectral conflict among track components, while Reverb often adds more sound components to a track. This also explains why producers recommend post Reverb equalization [23].

  This helps us infer that any Reverb addition algorithm based specifically on spectral masking minimization will not be realistic in terms of studio practices.

- **5. Long and Short Reverb used for different purposes:** Through the cursory listening analysis (section 9.2) it is clear that longer delay times are responsible for loss of depth hierarchy and excessive clutter, while short Reverb aids in modifying instrument timbres. The negative effects of excessive Reverb is clearly noticed through the extremely low scores for algorithmically processed tracks obtained for the singer-songwriter genre (figure 10.3). This genre is highly focused on the vocals and dependent on their clarity and intelligibility. Long Reverb times can not only make the vocals sound distant, but the basis for the vocals i.e; the combination of supporting instruments can fade into a cluttered background, severely denting the overall impression.

  This observation makes it extremely important to notice and categorize the disparate effects of long and short Reverb addition on specific components of the track. This will possibly lead to distinct strategies for the application of long and short Reverb. Neither of the two automatic algorithms implemented treated long and short Reverbs independently.

- **6. Different temporal resolutions need to be investigated:** Although only one particular section of a song is used for each genre, analysis of track components at different temporal resolutions can possibly help account for the information concerning the spectral conflicts and sound event density.

  In both our Reverb addition algorithms, the entire song section is considered to optimally benefit from a constant level of added Reverb. Considering the importance of sound event density on the level of tolerable Reverb, it could be possible to add varying Reverb levels on increasingly smaller segments of a song that feature stable sound event density/sound energy.

  Although the spectral algorithm implemented does not feature any adjustable variables that could allow us to investigate multiple temporal resolutions, the use of loudness averaging (equation 7.5) as a singular loudness measure for the entire song section can possibly be amended to account for analysis around varying temporal resolutions and preserve more temporal information.

- **7. The use of Reverb as a minimally invasive melding effect must be given more importance:** It is readily observed that both the spectral and mathematical algorithms add excessive reverberation to the multi-track audio being processed. Although the level of added Reverb is not directly influenced in the spectral algorithm, this level can be adjusted in the mathematical algorithm by modifying the minimum allowable increase in overall loudness (section 7.4.1, point 2). In hindsight, allowing the total loudness to increase by 2 LUFS represented excessive headroom and allowed for an over-generous addition of Reverb. Re-framing the algorithm with

a lower value for total allowable increase in loudness would have set lower values for both long and short Reverb and could have lead to potentially better subjective test results.

Through the results, it is evident that the test subjects were highly sensitive to excessive Reverb and the audio artifacts it generated. The test subjects were also able to identify the sparse and minimal reverberation added by the professional producer in comparison with the unreverberated track. Both these observations suggest that at automatic algorithms aiming for wide applicability must give more priority to the melding effect provided by smaller Reverb values over the other, more creative uses brought about by a more liberal application of Reverb.

- **8. The automatic Reverb algorithms implemented are not comparable to a professional producer:** To answer the central question of our thesis, the algorithms implemented herein are certainly are not perceptually similar or as subjectively appreciable as the tracks produced using the professional producer's input. Although, the detailed and interdisciplinary nature of this thesis makes it a strong reference for further work in this field.

10.3.1. Possible Criticisms

Even though sufficient care has been taken to provide optimal characteristics to conduct a fair experiment and subjective judgment of the experimental results, there can still be certain criticisms made about the procedures that can help guide us better through possible future implementations. This sub-section catalogues some such faults, without claiming to be exhaustive or complete.

- **1. One song is not representative of an entire genre:** The choice of multiple genres was made in order to test the adaptability and flexibility of our designed algorithms. However in an age of music proliferation, experimentation and fusion, a single song can never be representative of an entire genre. For an introductory study into automatic addition of Reverb, this setup can be argued to be sufficient, however any successful algorithm would have to be extensively tested and modified in order to avoid overfitting.
- **2. Pre-processing affects the producer's ability to set optimal fader gain values:** As mentioned in Chapter 9, the producer did mention that his ability to set optimal gain values according to his subjective preferences was affected by the pre-processing operations already performed. This was only discovered late into the completion of the experimental phase, but nonetheless was an important step in providing equality of opportunity to both the algorithms designed and a professional producer. The ideal workflow would involve a professional producer performing all the pre-processing steps himself to prepare the tracks for input to any Reverb addition algorithms being investigated. However, this can be resource and time intensive.
- **3. Producer's input biased towards a live audio setup:** The sparse Reverb values set can possibly be considered to be ideal for a live audio setup, which does not require excessive Reverb, considering the added spatial impression from the performance space. This deduction is further strengthened by the knowledge of the professional producer's background in live audio production. However, the fact that listening test subjects were reliably able to differentiate between the professional track and the unreverberated track proves that even the sparse Reverb fader gain values set were enough to make an important perceptual impression. In any case, the use of multiple professionals for future work would help to obtain more reliable results.
- **4. One producer is not representative of the entire community:** As mentioned in the previous point, the use of multiple producers would lead to more robust subjective ratings. This is due to the subjective nature of audio mixing and perception and the myriad creative ways in which a mix can be envisioned. 1
- **5. Treatment of mixing as a linear rather than a recursive process:** This assumption made was explained in Chapter 6, and was acknowledged to be unrealistic with regards to modern studio practices. This approach was utilized to be able to effectively assess the capabilities of our Reverb addition algorithms.
- **6. The spectral algorithm implemented is not truly cross adaptive:** The spectral algorithm implemented does not set the values for Reverb fader gains after an analysis of all the tracks

concurrently; the analysis is focused upon the track under consideration. A true cross adaptive algorithm would look to minimize spectral masking across tracks rather that within the same track. The fruitfulness of such an approach can only be speculated, however, as established, spectral masking minimization is not an explicit goal of Reverb addition.

- **7. Variance of Reverberation parameters has not been tested:** The short and long Reverbs used had fixed parameter values. Apart from the Wet/Dry mix, which needs to be maintained at 1 for the Reverb addition set-up used, the variance in perceptual impression with the change in the other parameters could have been tested. However, as mentioned in section 6.1, point 3, a collection of 2-3 Reverbs are sufficient for a multi-track mix. This assumption, coupled with the increase in mathematical complexity with variable parameters could have lead to potentially less conclusive results, and a large increase in the computation required. An automatic Reverb addition algorithm with variable parameters would be more prudent after a study of the variance in perceptual impression with change in individual parameters.

10.4. Contributions to the Research Field and Future Directions

This thesis has been an important introductory study into the possibility of adding Reverberation to multi-track production audio. A study that involved multiple disciplines and human interaction. This section details the contributions made to this nascent research field through this thesis.

10.4.1. Contributions from Theoretical Survey

The theoretical survey helped us understand the role of Reverb in the process of music mixing, the contributions it makes to the overall sound and its multi-faceted utility to the modern producer. The major contribution made from the compilation of therotical and practical knowledge was the Reverb addition set-up as designed in Figure 6.3. This is the setup recommended [24] by producers and used in Digital Audio Workstations to add reverberation to multi-track audio. This set-up and its integration with an automatic Reverb addition algorithm acts as the most basic link between modern studio practices and the academic research field.

The focused compendium of information on the practical use of Reverb, in Chapter 4, can also act as an important reference for any researcher looking to design an automatic Reverberation addition algorithm.

10.4.2. Contributions from Experimental Research

The important conclusions drawn and the contributions to the field from the experimental research is enumerated as follows:

- **Reverb gain setting needs to be genre/instrument dependent:** Future Reverb addition algorithms need to account for the different spectral, spatial and temporal characteristics of the track under application to be adaptable and successful.

- **Spectral masking minimization should not be the explicit goal of Reverb application:** Spectral masking is inevitable with Reverb application, producers suggest the use of post Reverb equalization [23] to correct this. Hence, the avoidance or minimization of spectral masking should not be one of the goals of optimal Reverb application.

- **Sparseness of sound events needs to be taken into account:** It has been found that the tolerance of the level of applied Reverb is dependent upon the sparseness of sound activity, and the presence of regions of silence between them. Minimal sound activity leads to minimal clutter due to temporal smearing of sound.

  The Reverb in this case is more noticeable, not just in case of spatial impression but also with its instrument timbre altering effects. The level of applied Reverb is weakly directly related to the sparseness of sound event density; i.e; minimal recordings can tolerate greater level of applied Reverb before clutter and smearing can degrade the overall impression and quality.

- **Long and Short Reverbs need to be applied independently:** Since long and short Reverbs are used for different uses (long Reverbs for melding and spatial impression, short Reverbs for

timbre changes), using the same strategies for calculation and optimization of fader gain values for both Reverbs is unrealistic.

- **The melding effect of Reverb must be given the highest importance:** As explained in the conclusions (section 10.3, point 7), automatic algorithms aiming for wide applicability across genres and instruments must first focus on the minimally invasive melding application of Reverb, that requires minimal and miserly values to be set. This approach is also prescribed by Mike Senior [24] for amateur producers.

10.4.3. Future Directions

The author hopes that this thesis has provided a sense of clarity and direction to this research field. The work conducted and the recommendations made open up further avenues for exploration and algorithm design.

First and foremost, as mentioned in the closing words of Chapter 5, there needs to be a thorough analysis of the change in perceptual impression with variance of certain Reverberation parameters, in order to create an importance hierarchy and also to distill the multiple subjective parameters in use today to a few important and relevant ones.

Although spectral masking minimization has unequivocally been proven to be unrealistic in its underlying assumption and practical utility, the mathematical optimization algorithm still shows growth potential. The choice of constraining the increase of overall loudness by 2 LUFS can be lowered to obtain smaller fader gain values that may reduce masking and clutter. Although, more research would be needed in order to obtain optimum values for this parameter for different genres.

Abandoning the assumption of constant Reverb fader gain values for a particular section of a song may also pay dividends, if future algorithms are adapted to analyze segments of music at varying time scales. This can help identify moments of spectral or temporal silence, however small, which can be leveraged in order to create a more cohesive production out of multiple tracks.

Optimization of Reverb fader gain values should also be done on the basis of perceptual features rather than mathematical ones. Objective quality metrics such as the one defined in [44] can serve as a good preliminary starting point for such an implementation.

Bibliography

[1] B. C. Moore, B. R. Glasberg, and T. Baer, *A model for the prediction of thresholds, loudness, and partial loudness,* Journal of the Audio Engineering Society **45**, 224 (1997).

[2] M. Terrell, A. Simpson, and M. Sandler, *The mathematics of mixing,* Journal of the Audio Engineering Society **62**, 4 (2014).

[3] D. Ward, J. D. Reiss, and C. Athwal, *Multitrack mixing using a model of loudness and partial loudness,* in *Audio Engineering Society Convention 133* (Audio Engineering Society, 2012).

[4] M. J. Terrell and J. D. Reiss, *Automatic monitor mixing for live musical performance,* Journal of the Audio Engineering Society **57**, 927 (2009).

[5] E. Perez-Gonzalez and J. Reiss, *Automatic gain and fader control for live mixing,* in *IEEE Workshop on applications of signal processing to audio and acoustics* (2009) pp. 1–4.

[6] E. P. Gonzalez and J. Reiss, *Automatic mixing: live downmixing stereo panner,* in *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx'07)* (2007) pp. 63–68.

[7] B. De Man and J. D. Reiss, *A semantic approach to autonomous mixing,* in *APR13, 2013. Ma, et al. PARTIAL LOUDNESS IN MULTITRACK MIXING* (Citeseer, 2013).

[8] N. Peters, J. Choi, and H. Lei, *Matching artificial reverb settings to unknown room recordings: a recommendation system for reverb plugins,* in *Audio Engineering Society Convention 133* (Audio Engineering Society, 2012).

[9] S. Heise, M. Hlatky, and J. Loviscach, *Automatic adjustment of off-the-shelf reverberation effects,* in *Audio Engineering Society Convention 126* (Audio Engineering Society, 2009).

[10] Z. Rafii and B. Pardo, *Learning to control a reverberator using subjective perceptual descriptors.* in *ISMIR* (2009) pp. 285–290.

[11] B. A. Blesser, *An interdisciplinary synthesis of reverberation viewpoints,* Journal of the Audio Engineering Society **49**, 867 (2001).

[12] H. Kuttruff, *On the audibility of phase distortions in rooms and its significance for sound reproduction and digital simulation in room acoustics,* Acta Acustica united with Acustica **74**, 3 (1991).

[13] P. M. Morse, *Some aspects of the theory of room acoustics,* The Journal of the Acoustical Society of America **11**, 56 (1939).

[14] W. C. Sabine and M. D. Egan, *Collected papers on acoustics,* The Journal of the Acoustical Society of America **95**, 3679 (1994).

[15] J. A. Moorer, *About this reverberation business,* Computer music journal , 13 (1979).

[16] M. R. Schroeder, *The "schroeder frequency"revisited,* The Journal of the Acoustical Society of America **99**, 3240 (1996).

[17] D. Pressnitzer and S. McAdams, *Two phase effects in roughness perception,* The Journal of the Acoustical Society of America **105**, 2773 (1999).

[18] J. Blauert, *Spatial hearing: the psychophysics of human sound localization* (MIT press, 1997).

[19] R. Izhaki, *Mixing audio: concepts, practices and tools* (Taylor & Francis, 2013).

[20] M. R. Schroeder, *Natural sounding artificial reverberation,* in *Audio Engineering Society Convention 13* (Audio Engineering Society, 1961).

[21] M. Senior, *Use reverb like a pro: 2, Sound on Sound Magazine, August 2008*, (2008).

[22] B. Owsinski, *The music producer's handbook* (Hal Leonard Corporation, 2010).

[23] P. Pestana and J. Reiss, *Intelligent audio production strategies informed by best practices,* in *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio* (Audio Engineering Society, 2014).

[24] M. Senior, *Use reverb like a pro: 1, Sound on Sound Magazine, July 2008*, (2008).

[25] A. M. Sarroff and J. P. Bello, *Toward a computational model of perceived spaciousness in recorded music,* Journal of the Audio Engineering Society **59**, 498 (2011).

[26] M. Terrell and M. Sandler, *An offline, automatic mixing method for live music, incorporating multiple sources, loudspeakers, and room effects,* Computer Music Journal **36**, 37 (2012).

[27] M. J. Terrell, A. J. Simpson, and M. B. Sandler, *A perceptual audio mixing device,* in *Audio Engineering Society Convention 134* (Audio Engineering Society, 2013).

[28] E. P. Gonzalez, J. D. Reiss, *et al.*, *Improved control for selective minimization of masking using inter-channel dependancy effects,* in *11th Int. Conference on Digital Audio Effects (DAFx)* (2008).

[29] J. Scott, M. Prockup, E. M. Schmidt, and Y. E. Kim, *Automatic multi-track mixing using linear dynamical systems,* in *Proceedings of the 8th Sound and Music Computing Conference, Padova, Italy* (2011).

[30] P. D. Pestana, Z. Ma, J. D. Reiss, A. Barbosa, and D. A. Black, *Spectral characteristics of popular commercial recordings 1950-2010,* in *Audio Engineering Society Convention 135* (Audio Engineering Society, 2013).

[31] B. R. Glasberg and B. C. Moore, *A model of loudness applicable to time-varying sounds,* Journal of the Audio Engineering Society **50**, 331 (2002).

[32] A. J. Simpson, M. J. Terrell, and J. D. Reiss, *A practical step-by-step guide to the time-varying loudness model of moore, glasberg, and baer (1997; 2002),* in *Audio Engineering Society Convention 134* (Audio Engineering Society, 2013).

[33] B. De Man, M. Boerum, B. Leonard, R. King, G. Massenburg, and J. D. Reiss, *Perceptual evaluation of music mixing practices,* in *Audio Engineering Society Convention 138* (Audio Engineering Society, 2015).

[34] E. Skovenborg and T. Lund, *Loudness descriptors to characterize programs and music tracks,* in *Audio Engineering Society Convention 125* (Audio Engineering Society, 2008).

[35] J. J. O'Donovan and D. J. Furlong, *Perceptually motivated time-frequency analysis,* The Journal of the Acoustical Society of America **117**, 250 (2005).

[36] D. Furlong and J. O'Donovan, *Quantitative characterisation of perceptually relevant artifacts of synthetic reverberation using the earwig distribution,* Digital Audio Effects **1** (2001).

[37] M. Karjalainen and H. Jarvelainen, *More about this reverberation science: Perceptually good late reverberation,* in *Audio Engineering Society Convention 111* (Audio Engineering Society, 2001).

[38] T. Necciari, P. Balazs, R. Kronland-Martinet, S. Ystad, B. Laback, S. Savel, and S. Meunier, *Perceptual optimization of audio representations based on time-frequency masking data for maximally-compact stimuli,* in *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio* (Audio Engineering Society, 2012).

[39] T. Necciari, P. Balazs, N. Holighaus, and P. L. Søndergaard, *The erblet transform: An auditory-based time-frequency representation with perfect reconstruction,* in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2013) pp. 498–502.

[40] B. Series, *Algorithms to measure audio programme loudness and true-peak audio level,* (2016).

[41] I. Recommendation, *1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,* International Telecommunication Union, Geneva (1997).

[42] I. Recommendation, *1534-1: Method for the subjective assessment of intermediate quality level of coding systems,* International Telecommunication Union (2003).

[43] B. De Man and J. D. Reiss, *Ape: Audio perceptual evaluation toolbox for matlab,* in *Audio Engineering Society Convention 136* (Audio Engineering Society, 2014).

[44] B. Recommendation, *1387: Method for objective measurements of perceived audio quality,* International Telecommunication Union, Geneva, Switzerland (2001).