

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The School of the Arts (Music)

A research dissertation submitted in partial fulfilment of the requirements for the degree of Master of Music in Music Technology

Name:

Michael-John Peter Anderson

Department:

The School of the Arts (Music)

Contact details:

Cell phone: 072 709 1272

Email address: m2audioproductions@gmail.com

Physical address:

176 Craig Street

Grootfontein Country Estate

Pretoria

0056

Supervisor:

Dr Miles Warrington

DECLARATION OF ORIGINALITY

UNIVERSITY OF PRETORIA

The Department of .....Music as part of The School of the Arts..... places great emphasis upon integrity and ethical conduct in the preparation of all written work submitted for academic evaluation.

While academic staff teach you about referencing techniques and how to avoid plagiarism, you too have a responsibility in this regard. If you are at any stage uncertain as to what is required, you should speak to your lecturer before any written work is submitted.

You are guilty of plagiarism if you copy something from another author's work (eg a book, an article or a website) without acknowledging the source and pass it off as your own. In effect you are stealing something that belongs to someone else. This is not only the case when you copy work word-for-word (verbatim), but also when you submit someone else's work in a slightly altered form (paraphrase) or use a line of argument without acknowledging it. You are not allowed to use work previously produced by another student. You are also not allowed to let anybody copy your work with the intention of passing if off as his/her work.

Students who commit plagiarism will not be given any credit for plagiarised work. The matter may also be referred to the Disciplinary Committee (Students) for a ruling. Plagiarism is regarded as a serious contravention of the University's rules and can lead to expulsion from the University.

The declaration which follows must accompany all written work submitted while you are a student of the Department of .....Music as part of The School of the Arts..... No written work will be accepted unless the declaration has been completed and attached.

Full names of student: .....Michael-John Peter Anderson.....

Student number: .....13239912.....

Topic of work: .....The Amalgamation of Acoustic and Digital Audio Techniques for the Creation of Adaptable Sound Output for Musical Theatre.....

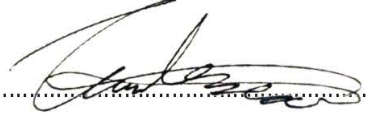
Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.

2. I declare that this .....Masters dissertation..... (eg essay, report, project, assignment, dissertation, thesis, etc) is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE .....  .....

The Amalgamation of Acoustic and Digital Audio Techniques for the Creation of Adaptable Sound Output for Musical Theatre

by

Michael-John Peter Anderson - 13239912

A research dissertation submitted in partial fulfilment of the requirements for the degree of Master of Music in Music Technology

Department of Music
Faculty of Humanities
University of Pretoria

September 2019

Abstract

There are many facets that influence the quality of a musical theatre production. The visual appeal is created from the décor, costumes and lighting, whereas the plot, pace, and relationship a listener develops with the characters are fundamental to the performance quality. However, one often overlooked factor is the impact of sound quality. The perception of sound quality is subjective but is greatly impacted by the environment in which the listener finds themselves. If the projection of the music is underwhelming in depth and expression, or the balance of the dynamics and timbre are badly mixed, this can jeopardise the production's success, regardless of the quality of the composition or the visual aspects.

The production budget for a musical performance can be prohibitive. As a result, pre-recorded music is often used as an alternative substitute to live musicians. However, the subjective authenticity of a musical may be jeopardized by the exclusion of live musicians and create additional challenges and performance limitations. One such challenge is the environment in which music will be played. Recorded music is usually created in a single format such as compact disc or for broadcasting, and the cost of recording be can just as expensive as a live performance, especially on large scale works. Time and budget constraints may impact the sound quality. In addition to this, the varying acoustic properties of potential venues may emphasise sonic gaps and flaws contributing to a listener's negative perception of the sound quality, resulting in a compromised experience of the performance as a whole.

This mixed method dissertation offers a systematic explanation to potentially resolve these challenges and limitations by conceptualising established knowledge of sound, audio and acoustics to formulate a framework for adaptive sound. These concepts are put into practice by creating a specifically designed audio recording that is experimented with in multiple theatre scenarios to successfully achieve optimal adaptation of the sound for the theatre environment.

Keywords: Sound, Audio, DSP, Acoustics, Psychoacoustics, Adaptive Sound, Sound Transmission, Sound Reproduction, Music Technology

Table of Contents

Declaration of originality ii

Abstract iv

Keywords to the study iv

Chapter 1: Introduction 1

1.1 Background to the study 1

1.2 Statement of the research problem and objective 3

1.3 Research questions 3

1.3.1. Main research questions 3

1.3.2. Secondary research questions 3

1.4 Research design 4

1.5 Rationale and value of the study 8

1.6 Dissertation statement and hypothesis 8

1.7 Delineations and limitation 9

Chapter 2: Literature review and Data Collection 10

2.1 Adaptable sound 10

2.1.1. Sound and audio 10

2.1.2. Acoustics 14

2.1.3. Manipulation of acoustic and audio properties 21

2.2 Recording *Welcome to the Show* (2019) 23

2.2.1. The Sound of *Welcome to the Show* (2019) 23

2.2.2. Microphone theory and application 25

Directional response and polar patterns 25

Impedance 26

Frequency response and generating element 27

Stereophonic arrays 27

Stereophonic dimensions and imagery 29

Monaural placement and critical distance 30

2.2.3. Acoustic and digital instrument layering 31

MIDI, digital sequencing and samples 32

2.3 Mixing for adaptable sound 33
2.3.1. Monitoring and phase 34
2.3.2. Mixing practice for adaptable sound 35
Equalisation (EQ) 35
Spatial processing 37
Dynamic Processors 39
Temporal processing 42
2.4 Mastering for adaptable sound 43
2.4.1. High-resolution mastering 44
2.4.2. Tonal balance 45
2.4.3. Loudness range 45
2.5 Analysing audio 47
2.6 Data collection conclusion 50

Chapter 3: Methods, statistics and conceptual framework 51
3.1 Environmental analysis and numerical data 51
3.1.1. Ambience, signal and noise 51
3.1.2. Critical distance 53
3.1.3. Frequency response and modal resonance 53
3.1.4. Venue size and reverberation time 55
3.1.5. Environmental analysis summary 58
3.2 Recording process and data application 59
3.2.1. Recording preparations 59
3.2.2. MIDI: Tempo tracks and digital instruments 62
3.2.3. Recording instruments and samples 63
Stereophonic microphone placement 64
Monaural microphone placement 65
Acoustic samples 69
3.2.4. Recording process summary 70
3.3 Mixing and mastering for adaptive capabilities 71
3.3.1. Mixing adaptive audio 71
Overdubbing, panning and profiling 72
Editing and restoration 75
Inverted 60-phon spectral application 76

Dynamic stability and reinforcement 79
Acoustic and artificial temporal processing 81
Spatial enhancements 84
3.3.2. Mastering adaptive audio 85
Spectrographic analysis and restoration 85
Inverted 60-phon frequency range reinforcement 86
Dynamic signal processing 89
Post-mastering DSP applications 91
3.3.3. Mixing and mastering summary 91

Chapter 4: Adaption testing and discussion 93
4.1 Acoustic and audio testing and comparisons 93
4.1.1. Signal chain: hardware and software 93
4.1.2. Sonic differences between playback environments 94
4.1.3. Adapted benchmark creation and statistics 96
4.2 Adaption: DSP applications 100
4.2.1. First-order adaption: IFM filter and maximizing 100
4.2.2. Second-order adaption: Dynamic and spectral enhancements 101
4.2.3. Third-order adaption: Live adaption and additions 102
4.3 Adaption summary 103

Chapter 5: Results and conclusion 105
5.1 Discussion and interpretation of the main findings 105
5.2 Critical assessment 108
5.3 Production comparisons 110
5.4 Limitations and implications of the research 112
5.5 Conclusion 114
5.6 Recommendations and future research 116

References 118
Appendices 122
List of figures 144
Glossary 146

Chapter 1: Introduction

1.1 Background to the Study

No one can predict with certainty what will happen in the future. Likewise, no sound engineer can precisely determine the transformation of sound in an acoustic environment after a show begins. Sound reacts differently in diverse environments and is affected by circumstantial change such as the size of the venue or the capacity of the audience (Leonard, 2001: 12).

In a live production such as a musical, the sound design is only one aspect of the full production. Set-design, choreography, special effects and décor are equally important components (Everett, 2002: 8). However, if one considers the production of a mega-musical, which is a lavish large-scale stage production with scenes constructed from collections of through-composed music with very little or no dialogue, it would rely heavily on the sound quality (Lamb, 2001: 454). Sound quality is determined by two factors: performance-related considerations, such as the capabilities and technique of singers and instrumentalists, and audio-related elements. Audio-related factors are controlled by several technical processes, such as the placement of microphones, resolving sonic issues such as phasing, mixing during the sound check, and fixing the mix after the performance begins to compensate for these circumstantial changes. Consequently, an audience may perceive the entire production of a musical as mediocre if the sound quality is jeopardised in any way (Leonard, 2001: 141).

Commonly known examples of mega-musicals include *Les Misérables* (1981) by Claude-Michel Schönberg and *The Phantom of the Opera* (1986) by Andrew Lloyd-Webber. My mega-musical composition, *Welcome to the Show* (2019), also uses this construction and is used as the subject material for this study. The logistics involved in a production of a mega-musical, which generally uses a large orchestra and cast, would require many months of rehearsals, accommodating practice and performance venues, multiple sound technicians and a plethora of sound equipment. All this makes for an immensely time-consuming and potentially very costly pre-production process (Novak & Novak, 1996: 4). Considering the expense of a such a production, many producers use pre-recorded tracks to eliminate many of these expenses but, as a consequence, there can be no variation or improvisation (Leonard, 2001: 122). Furthermore, these tracks are not likely created for every potential acoustic environment.

Acoustics, in the architectural sense, is the behaviour of sound in an enclosed space (Brown, 2008: 28). Resonance and reverberation are features of sound that react according to a venue's shape and size (Izhaki, 2012: 401). Bass frequencies have longer wavelengths (20Hz – 250Hz) and are less prominent in large spaces in relation to higher frequencies. This is because the physical distance of a wavelength decreases in length as frequencies increase and result in less reflection (Huber & Runstein, 2014: 49). In addition, bass frequencies are further absorbed by the presence of an audience because pliable material and water, which makes up approximately 90% of the human body (Huber & Runstein, 2014: 103), have physical properties that trap bass frequencies. As a result, higher frequencies (above 2 kHz) are easily accentuated from this loss of bass frequencies and are also prone to reflect off even the smallest of objects (Leonard, 2001: 21).

Considering the above example, the frequencies and reverberation of output are the most affected by the room's acoustic properties, but a recorded track is dependent on the mix of the audio. Izhaki (2012:3) explains the important role of a mix by emphasising that the mix can either improve the music of the performer or completely change it. Mixing involves manipulating and enriching the raw audio in a complimentary and structured process. For example: a mixing engineer should not add reverberation before equalisation (EQ) because it would make the process more difficult and the final mix slower to achieve (Izhaki, 2012: 37). Each stem is processed by adding spectral, dynamic, temporal and spatial audio processing techniques (Izhaki, 2012: 181-401). After completion, the tracks are sent to a mastering engineer for refining and to prepare the tracks for distribution (Katz, 2015: 10).

The above information is the general process in which recorded music is created. However, this standard practice is not commonly utilised over multiple formats or to optimise live sound. It is usually created one-dimensionally for radio, CDs or YouTube to name a few (Katz, 2015: 8). Therefore, if unchanging tracks are exposed to a sporadic environment, the sound quality will likely be inconsistent from one space to another. Hence, this study will investigate the different acoustic and digital audio techniques needed to create adaptable sound, in which its audio properties can be further manipulated after production to conform to any environment for optimum sound output.

1.2 Statement of the Research Problem and Objective

The product created in this study is not intended to be the ultimate process to create any commercial recording but rather a dependant variable that, when affected by external influences, can be manipulated to remain dependant. The subjective perception of the mix or master is not the focus of this study. This product is created objectively to enable adaption according to the criteria highlighted in chapter 3. Nonetheless, the recording has to be of the highest quality. Therefore, the problem that this study aspires to resolve is to ensure that the product, as an outcome of the study, is valid in terms of sound quality and adaption.

In other words, the product is a constant variable and should remain constant under different conditions. For this to materialise, it must be able to acclimatise to its environment and retain a level of 'authenticity', or fidelity to a benchmark of high sound quality. Therefore, the objective of this study is to develop an authentic but economical recording process that recreates large-scale orchestral music for musical theatre use. The product should be able to adapt its sound output adequately to the varying acoustic parameters found in disparate venues. Coincidentally, whatever information is gleaned from this can be used on other products.

1.3 Research Questions

1.3.1 Main Research Question

How are audio techniques used to reproduce authentic full-scaled orchestral music, in a reduced setting, using a combination of acoustic and digital mediums that may be manipulated to adapt to any acoustic environment?

1.3.2 Secondary Research Questions

- Does the combination of digital and acoustic sound affect the authenticity of the product?
- What is the difference between the techniques used in this study compared to producing the same music for a different platform such as radio or compact disc?
- How would the product of this study compare in expense, sound quality, reliability and production time to recording any large-scale orchestral recording project?

1.4 Research Design

This study follows a mixed-method approach for the different stages of production. According to Creswell (2009: 188), mixed method methodologies use the strengths of both qualitative and quantitative methods to expand the understanding of interdisciplinary research, such as what is being presented in this paper. One of the several salient factors in mixed-method papers in the weighting or priority of quantitative and qualitative research (Creswell, 2009: 191). In the case of this study, the quantitative aspects of this study contribute to approximately one third of this paper because, as the hierarchy tables below each method show, each methodological process it built from the quantitative data that is collected.

The second and third approaches are somewhat interchangeable as practice-led research and true experiments tend to be strong in validity because independent variables are manipulated to determine if it influences the dependents (Bryman, 2012: 50). This is particularly true for this study because the parameters can be altered from the practice-led research on the experimental research. This is a field experiment, where the events of real-life implementations affect the outcome of the research, unlike in a controlled laboratory environment (Ibid).

Data collection verifies the descriptive and numerical data to create an adaptable sound product before its physical creation. This is seen in the literature review chapter. The collecting of theoretical information within the reviewed literature will indicate the parametric statistics needed for the pre-production stage of this study before its physical application. This data is analysed and documented according to a nominal scale of measurement whereby the data identifies different processes needed to formulate a conceptual framework to create adaptive sound capabilities. The numerical data is intervallic measurements that identifies the acoustic properties of both the recording and playback environments. These constant variables will be documented in a measurement table as they are inferential statistics which may affect the adaptive sound capabilities of the audio product.

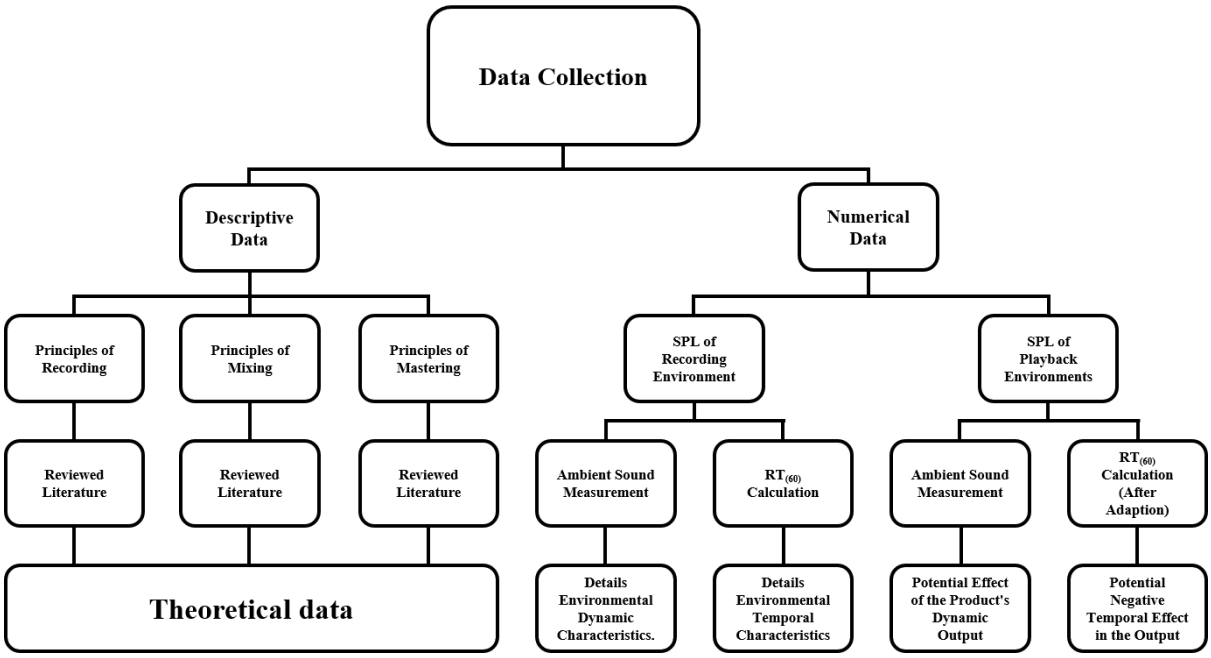


Figure 1.1: Data collection

A practice-led approach is used for the physical application of the obtained descriptive and numerical data. The data will serve as a reference to the physical recording, mixing and mastering process in order to create the audio product. During the creation processes, a comparison between the relationship of the quantitative data and its physical application will be conducted and documented by using a 'Solomon Four-Group' design. This design involves investigating the possible effects of selected processers applied to the audio product where one variable (audio process) influences another variable (output). The four test groups expose the different audio processing procedures. The first group is the audio processes applied in the mix; secondly, the audio processes applied in the mastering; thirdly, the audio processes applied in the playback environment; finally, additional audio processes applied to the product after testing to finish the final product and conclude this study.

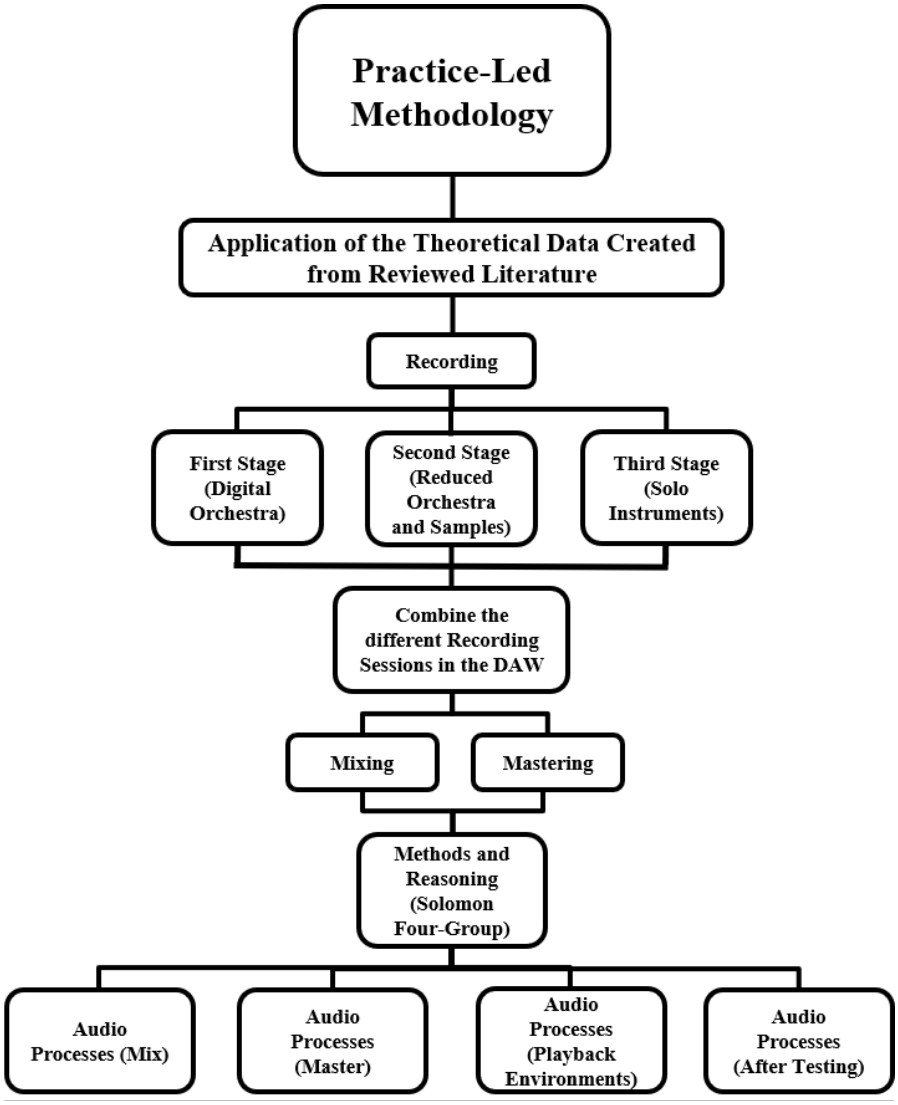


Figure 1.2: Practice-led research design

In correlation with the methods described above, a true-experimental research approach is undertaken after completing the first audio render of the audio product. This is tested in the different acoustic playback environments. It is expected that there will be a difference in sound output from venue to venue. If consistent results are not acquired after the testing of the audio product, adjustments need to be made to either the mix or master depending on the problem present in each venue. This presents a second audio render of the product and the testing is repeated. This is noted using a rubric that contains a list of criteria pertaining to optimal sound quality achieved in a studio environment. This comparison validates this test of sound quality because, if the output in playback corresponds with the output in a studio environment then adaptive sound capability is achieved, regardless of subjective interference.

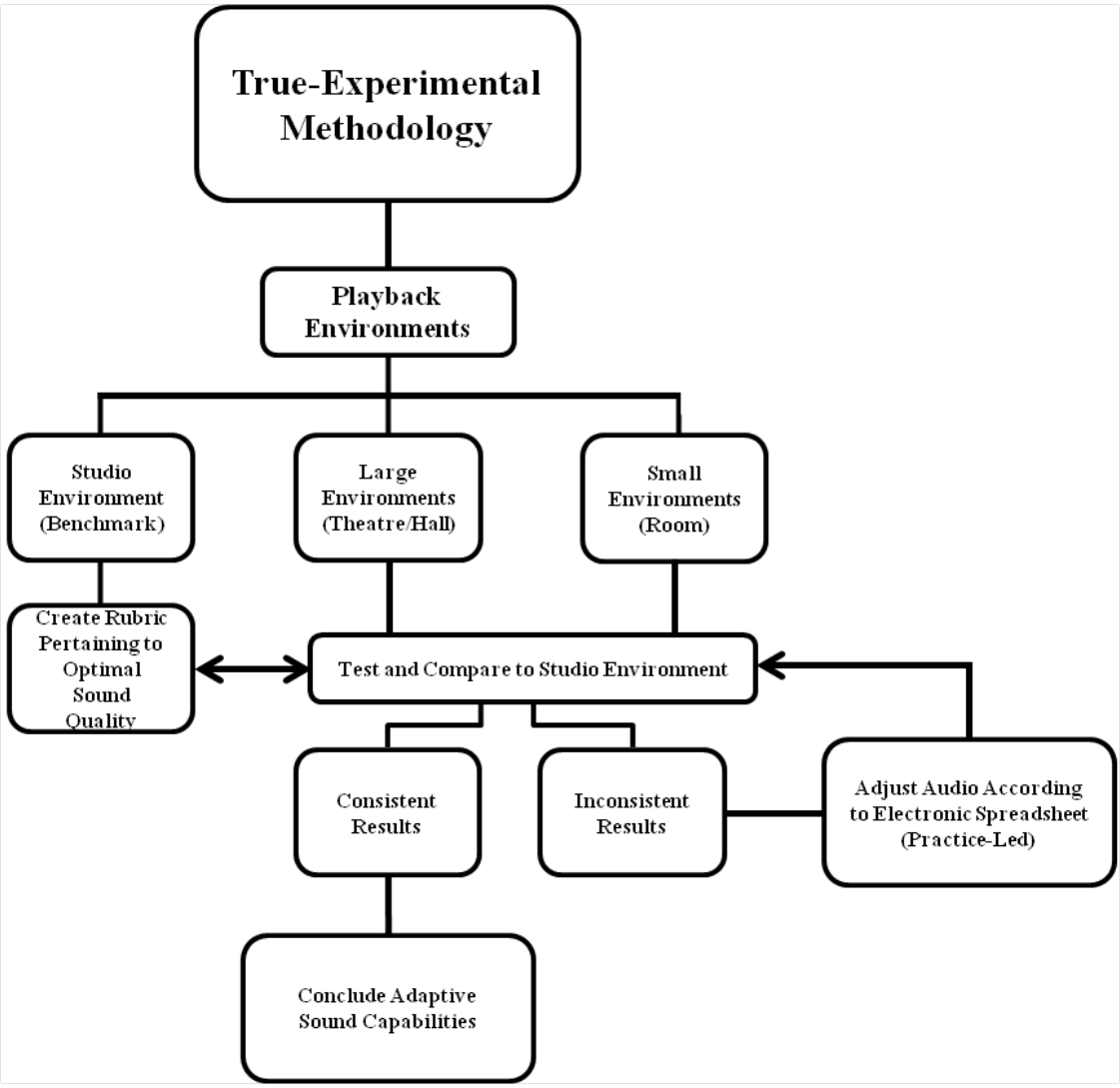


Figure 1.3: True-experimental research design

In summary, the various methods may be viewed as inter-linked. The data collection theorises the principles of recording, mixing and mastering, which is used in the practice-led and experimental research methods of this study. Using the theoretical data alongside the measurement data collected in the different acoustic environments, a conceptual framework is created that is applied in practice to create the audio product. The final audio product (adapted sound) is the outcome from the theoretical principles and is therefore renewed data formulated from the quantitative research.

1.5 Rationale and Value of the Study

There is always controversy between the perception of sound amongst different listeners. These reasons may be physical or phycological in nature depending on the particular situation. From an audio production standpoint, the best possible sound is the most important factor. However, as stated previously, large stage productions such as mega-musicals are expensive to produce, and producers have to weigh comparatively the importance and cost of each aspect of a production versus what they would potentially profit. This production problem is the foundation of this research: to reduce the cost of such a production by reproducing music that can adapt to an environment but retain the quality of the sound and the interest of listeners. However, the importance of this research is not rationalised from the producer's perspective but rather from an audience's perspective. From the first note of music played in a concert, an audience should expect, apart from visuals, good quality sound. The reasoning behind this research is to ensure that this will always be the case. Because this research aims to create a process whereby sound can adapt to an environment's acoustic parameters, this adaption in theory would ensure this.

The product created from this study would eliminate the need for an orchestra in a hypothetical production of *Welcome to the Show* and would be able to be used repeatedly, thus ultimately reducing the cost a full production. Therefore, the value of this study aims to provide an alternative approach in recreating the sound capabilities of a full orchestra using a combination of acoustic and digital media to reduce the cost of recording large-scale orchestral music without sacrificing sound quality. This study hypothetically could be used as a method to reduce the cost of any mega-musical production although it could also be used as an alternative process to create other large-scale orchestral productions such as a soundtrack for a film, a concerto recording or any other large-scale orchestral production.

1.6 Dissertation Statement and Hypothesis

This study will investigate the different acoustic and digital audio techniques needed to create adaptable sound using *Welcome to the Show* (2019) as the subject material. Only the instrumental music is recorded in this study because the intention is to remove the need of an orchestra for a hypothetical production of *Welcome to the Show* (2019), not the performers. With this information, a practice-led application of these audio techniques will be applied to a physical audio product. This product will be created with the intent that it can be manually

adapted by a singular sound engineer in a performance venue to accommodate any potential acoustic environment. The product itself is not the focus of the study but one of the outcomes of the research. In other words, this study will examine whether or not this amalgamation method reduces the human resources and expense of an orchestral production without compromising the sound quality. Such methods could possibly result in the enhancement of the sound quality from the exclusion of human error or external impact, such as noise or instrument quality.

The process of creating an illusion of a full-scale orchestra using the amalgamation between orchestral acoustic and digital sound. is discussed in this study. The effect of the environment on the music determines whether or not this illusion is objectively successful. The outputs from the sound systems are measured to determine the change in sonic properties, after which the adaption process commences. If the output of sound in the performance environments sufficiently adapts to their acoustic parameters, then optimal sound output is achieved. As a result, the process of creating this type of recording could provide producers with a cost-effective method to supply consistent high-quality sound to potential audiences.

1.7 Delineations and Limitations

Although the use of recorded music could never fully replicate the authenticity of live musicians, the processes used in this study should accurately depict the characters and mood of the musical. A logistical problem faced in this study is the size of the musical itself. As a consequence, only one piece of the music from the musical is recorded to reduce the expense of this study, although accurate estimations of the cost of a complete recording can be asserted.

This study explores the audio technique used to produce a recording of *Welcome to the Show* (2019), but the characters and plot is excluded from this study as they are irrelevant its purpose. However, this study does explain how the music is conceptualised, composed and recorded to determine the best possible audio production techniques needed for *Welcome to the Show* (2019), but the official compositions are not included in this study. This study applies to music technology and audio engineering, not musical theatre. A full production and performance of the mega-musical, *Welcome to the Show* (2019), is not the focus of the study.

Chapter 2: Literature Review and Data Collection

2.1 Adaptable Sound

To understand what, why and how particular processes are carried out to create sound that can adapt, one must observe the phenomena of sound, audio and acoustics. For sound to conform to an environment, one of two things must take place: either the physical acoustic properties of an environment must be manipulated, or the audio properties of sound presented in such environments must be modified. Therefore, one or both options are used to generate identical sound output consistently. The later of these options would seem simpler but, even so, every different environment has its own unique characteristics, and some music may not completely conform to the construction of the room. As Harold Geerdes (1975: 40) puts it, "once these halls are completed, they stay the same for viols or Moog synthesizers, for solo recitals or programs of massive choral and orchestral works." In most part, this is true but in chapter 2.1.3, Geerdes (1975) explains how some physical adjustments may be used to amend this. Sound and audio is discussed next and then how acoustics affect timbre.

2.1.1. Sound and Audio

There is a duality of interests within the subject of sound. If the interest lies with the physical disturbance of air pressure within an environment, this is a matter of the physics of sound. Alternatively, if the interest is focused on the human response and perspective of sound, it is considered a matter of psychoacoustics (Everest & Pohlmann, 2015: 1). If sound is conceptually formulated, it is the brain perceiving and interoperating variations of atmospheric pressure to correspond with the physical auditory stimulus of the human brain (Huber & Runstein, 2014: 43). Sound oscillations consist of *compressions*, which are air molecules moving outwards from a resting state, and *rarefactions,* which reverse the compression in an inwards direction due to atmospheric inertia. Therefore, sound in the physical domain is air particles pulsing (vibrating) by compressing and relaxing continuously. The simplest example of sound propagating is known as simple harmonic motion (Everest and Pohlmann, 2015: 2) as seen in figure A and B below, which displays how particles react to atmospheric pressure and how they are graphically represented.

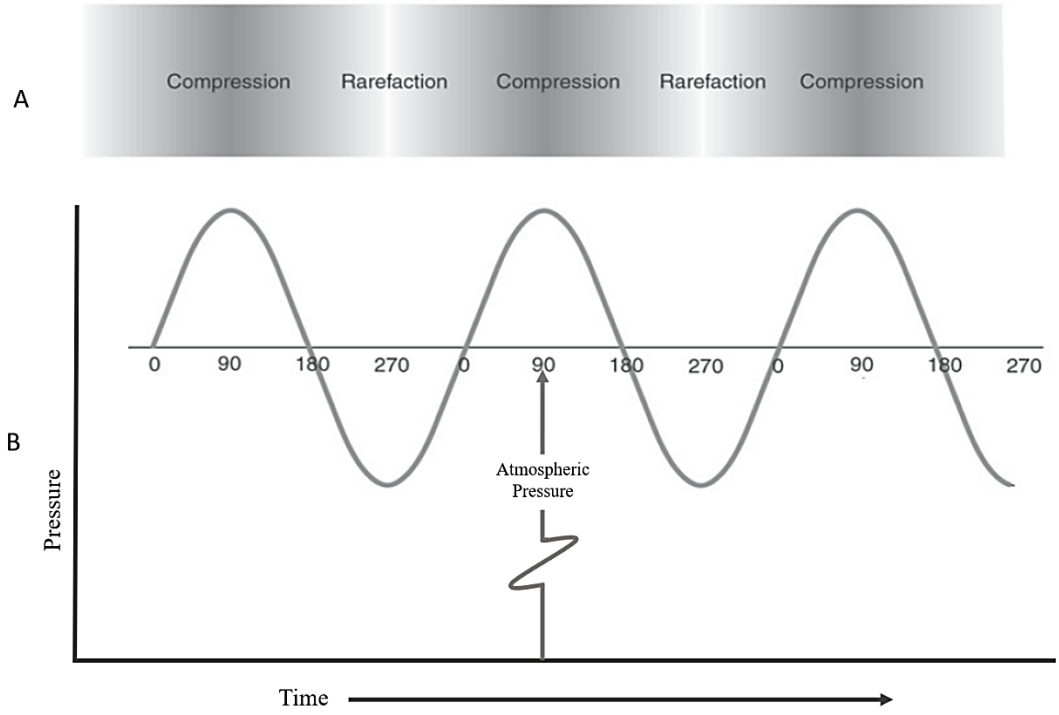


Figure 2.1: Sound propagation (Everest & Pohlmann, 2015: 6)

As an objective phenomenon, sound consists of a collection of frequencies that specifies the number of waveforms within a second of time. This determines the pitch of these frequencies. However, the human ear perceives pitch subjectively depending on the pitch's intensity, or "volume" (Everest & Pohlmann, 2015: 1). It is noticed that the representation of sound is portrayed as a sine wave. This is how the sound-pressure level (L_p) is graphically represented in a waveform as it moves through a medium over time (Huber & Runstein, 2014: 44). The speed of sound is approximately 334m/sec (1,240 km/hr) at normal atmospheric temperature (20°C) and pressure (101.325 kPa) or sea level (Everest & Pohlmann, 2015: 6). The speed increases by approximately 0.6m/sec for every degree Celsius increased whereas humidity and air pressure have very little effect on the speed (Everest & Pohlmann, 2015: 7). The speed in which a sound wave propagates through a medium (air, water, metal, etc.) determines the wave's physical size or wavelength (Brown, 2008c: 27).

A wavelength (λ) is measured either by the succession of peaks or between corresponding points in a cycle, or the distance between two corresponding points on the sine wave. Frequency (f) is the amount of wave cycles to be completed in a second and is measured in Hertz (Hz) (Everest and Pohlmann, 2015: 7). Figure 2.2 displays peridoic sound pulses as sinusoidal waves compared to its radial projection.

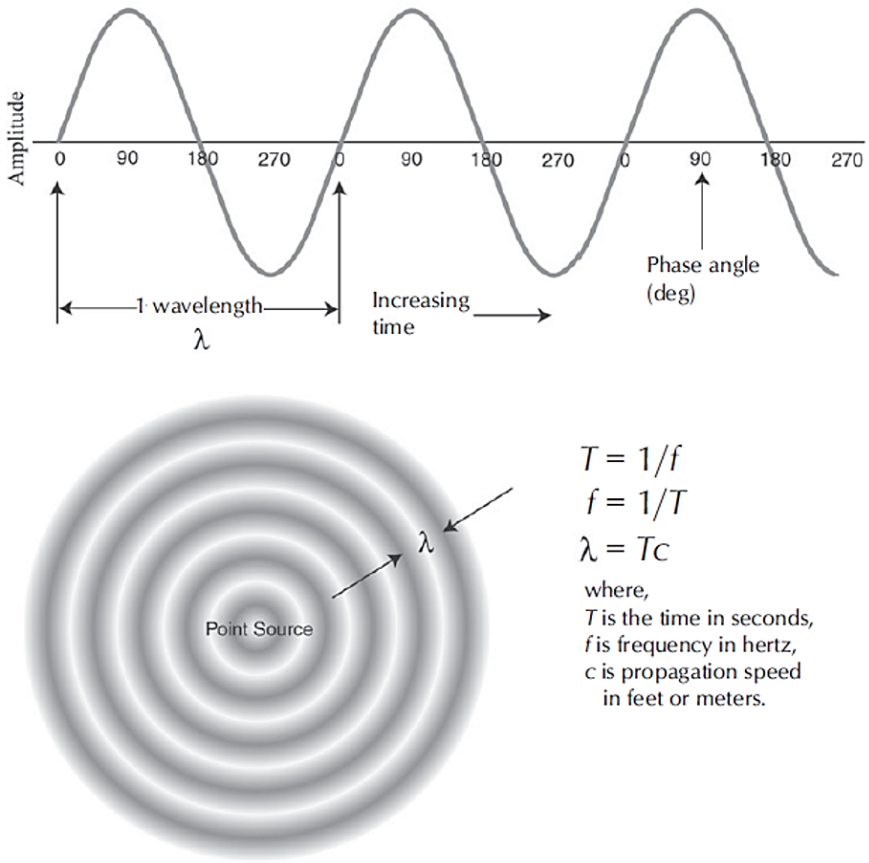


Figure 2.2: Sound pulses (Brown, 2008: 29)

Psychoacoustics refers to the subjective human perception of sound and includes a relationship between psychology, acoustics, biology and physiology to name a few, but there is a clear understanding between certain objective and subjective quantities of sound (Zhang, 2008: 43). For example, pitch and frequency are not linearly related but is the perception of one function that may be dependent on the other (Everest & Pohlmann, 2015: 44). Pitch perception is another of these subjective quantities that can be affected by its sound level or loudness (Zhang, 2008: 43). In this instance, a sound-level metre (SLM) reads the sound level in an environment, but a smaller frequency bandwidth could sound quieter than a wider bandwidth and the L_p reading will remain the same due to equivalent rectangular bandwidth filtering (Zhang, 2008: 49). These quantities can be controlled using audio devices such as mixing desks. These are intended to compensate for the human ear's subjective response towards different frequencies at different L_p. Furthermore, they would be calibrated to vary adaptively the sound pressure relative to the listener (Everest & Pohlmann, 2015: 48). Figure 2.3 shows the average dynamic range of music and speech in relation to the human ear's frequency response.

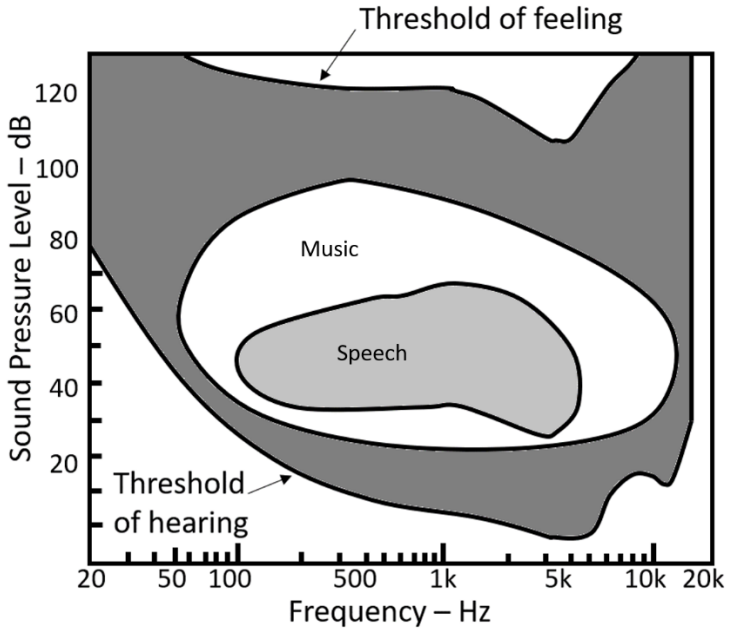


Figure 2.3: Human hearing (Zhang, 2008: 51)

L_p (or SPL) is a measurement of acoustic pressure (dB) in a defined atmospheric area (Huber & Runstein, 2014: 61). An SLM reads sound pressure at 20µPa, however as seen in figure 2.3, the human ear does not have a 'flat' response across the sound spectrum. Frequencies above 10 kHz and below 100 Hz are not as distinguished compared to our 'comfortable' hearing range between these two extremes. To emulate the human ear, SLMs use a variety of weighting (A, B and C) networks to reduce the L_p in high and low frequency bands as seen in figure 2.4 below (Everest & Pohlmann, 2015: 29). The different weightings are used for different applications based on the average sound level, which is discussed in chapter 3.1.1.

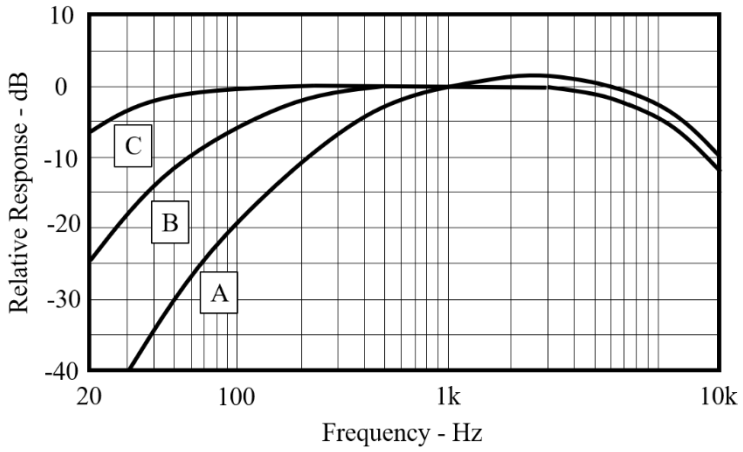


Figure 2.4: SLM weightings (Everest & Pohlmann, 2015: 30)

2.1.2. Acoustics

The acoustic design of an environment is intended to enhance a listener's experience by creating space with reflected sound fields. When sound waves interact with a surface, a multitude of reactions occur such as reflection, absorption or diffraction, to name a few (Brown, 2008: 28). To reproduce sound in environment, the different types of acoustic spaces need to be observed. Leonard (2001: 12) discusses the different types of theatre shapes and their typical sound qualities. One common type of theatre is a 'thrust stage' theatre, which is designed to project sound into the auditorium. This building is based on the ancient Greek amphitheatres, having a protruding stage that is surrounded on three sides by audience seating. If the sound is reinforced by an electric-acoustic sound system, the loudspeakers are usually located on the sides of the stage for maximum coverage.

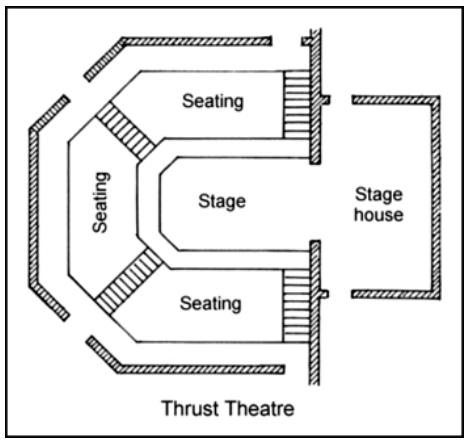


Figure 2.5: The thrust theatre (Chittom, L: 2018)[1]

There are many other types of theatres such as the promenade and open-air theatres (Leonard, 2001: 15) however the proscenium arch is most likely the type of venue where *Welcome to the Show* (2019) would typically be performed. This is because the auditorium of a proscenium arch is separated from the stage, under which there is an orchestra pit. Even without an entire orchestra, these theatres generally have well-treated acoustic environments and are ideal for pre-recorded music with loudspeakers placed on the sides and hung by a floating bar above the stage for the best sound coverage (Leonard, 2001: 11). These theatres, also known as *Singspiele,* are described by Ahnert and Tennardt (2008: 164) as types of concert halls that use electroacoustic sound reinforcement in collaboration with the natural acoustics of the room where one does not outweigh the other. The acoustic parameters of

[1] For more information, see L. Chittom's associated website: https://www.tes.com/lessons

these types of theatres should be treated in such a way as to have a high diffusivity of sound as well as the reverberation time not exceeding a frequency-independent value of approximately 1.4 seconds (Ahnert & Tennardt, 2008b: 164).

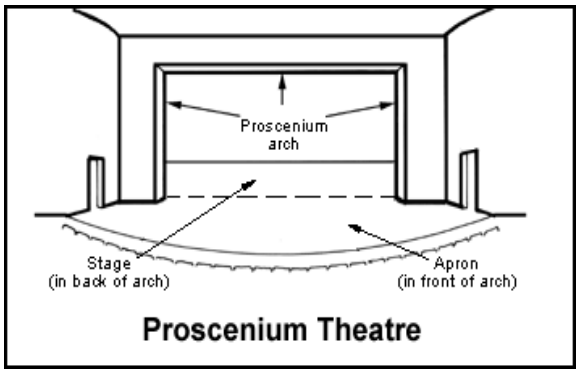


Figure 2.6: *The proscenium theatre* (Chittom, L: 2018)[2]

In a large theatre, like a proscenium theatre, there are generally no structural ornamentations (pillars, wall corners and so on). This is because acoustic treatment is designed to have free circulation. This causes sound to diffuse freely over a longer period of time. Consequently, this creates a magnitude of sound reflection in a small environment and a "dead" sound in large environments (Barbieri, 1998: 281). Sharp corners diffuse sound sporadically and, if there are warmer temperatures, sound absorption increases furthering the "dead" sound (Barbieri, 1998: 283). Furthermore, Barbieri (1998: 287) states that the size and shape of a proscenium arch, for opera, enhanced the female voice. This is because higher frequencies compress and refract more rapidly than lower frequencies (Barbieri, 1998: 294). As a result, lower frequencies suffer because they are lost behind the protruding arch on the physical stage and the orchestra reflecting sound out of the pit into the large auditorium (Barbieri, 1998: 289).

Objective acoustic criteria are factors such as the acoustic treatment in the auditorium and the sitting position of an audience member whereas subjective acoustical criteria would be a performer's musical ability or the mixing (Ahnert & Tennardt, 2008: 147). Theatres are generally designed to be well treated in terms of acoustics but electric-acoustic sound reinforcement systems (sound systems) are usually installed for maximum sound coverage.

[2] For more information, see L. Chittom's associated website: https://www.tes.com/lessons

Considering that this study aims to create adaptive sound across multiple environments, be it a general concert hall or sacral room, the duration of reverberation should be limited as much as possible (below 1.4 seconds). This is because electric-acoustic manipulation, known as variable acoustics, is used to artificially manipulate the acoustic properties of the environment (Ahnert & Tennardt, 2008: 188). This sonic reinforcement is intended to amplify the output but not detract from the sound quality of a recording as is discussed in chapter 3.3.1.

In most types of acoustic environment, the distance sound travels will affect the L_p at different points. This is known as the *inverse square law for sound intensity* and indicates the intensity of a sound source in a free field in inverse proportion to the distance squared whereby sound intensity (sound power/unit) decreases at the square of the radius (Everest & Pohlmann, 2015: 34). This equates the loss of sound pressure over a distance. In other words; for each unit of distance sound travels, the sound intensity halves (-6 dB).

$$I = \frac{W}{4\pi r^2}.$$

Where:

I = The sound intensity per area unit.

W = The power source (watts).

$4\pi r^2$ = The area of the sound sphere (radius).

(Everest & Pohlmann, 2015: 35)

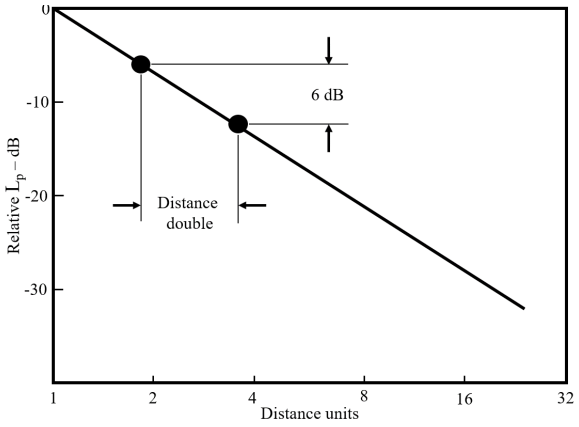


Figure 2.7: The inverse square law for sound intensity (Everest &Pohlmann, 2015: 36)

Theoretically, sound in open space, or a free field, travels in linear motion and is unimpeded until interacting with a surface (Everest & Pohlmann, 2015: 33). Furthermore, sound in free space also propagates uniformly from a source spherically as seen in figure 2.2. Here, the sound intensity decreases as the distance increases due to geometric spreading of energy (Everest & Pohlmann, 2015: 34). The difference between sound-pressure ($L1$) in the point-source of the innermost part of the sound sphere ($r1$) and sound-pressure ($L2$) at the outermost part of the sound sphere ($r1$) is calculated:

$$L2 = L1 - 20\log\frac{r2}{r1}\text{decibels}$$

Therefore, the difference in L_p between two points can be calculated:

$$L2 - L1 = 20\log\frac{r2}{r1}\text{decibels}$$

(Everest & Pohlmann, 2015: 35)

However, free fields exist only in enclosed anechoic spaces and are not cosmological. This means that the sound requires a medium to move and is unhindered by overlapping resistance from other sound waves. This intension of anechoic space is to distribute **direct** sound uniformly from a source enabling the truest calculation of the inverse square law for sound intensity (Everest & Pohlmann, 2015: 33). Opposed to a free field, a **reverberant** sound field has equal sound level throughout the space no matter the distance from the source. Nearly limitless reflections in these enclosed spaces disrupt the natural decay of sound resulting in a contradiction of the inverse square/distance law. The reality is that normal rooms use the extremes of both anechoic and reverberant sound fields, combining both direct and reflected sound especially if the sound source doesn't originate from a single point (Everest & Pohlmann, 2015: 37). Both of these acoustic conditions directly relate to the laws of energy regarding sound transference. An immense amount of absorption is needed to create an anechoic environment whereas reflection is needed for reverberant environments. This means that, when measuring and analysing sound, this distance factor is considered in the scope of average sound output in the environment. This is discussed in chapter 3.1.1.

The distance sound travels in the space of an environment is one of two physical factors that affects sound in an acoustic environment. The other is the construction of the environment around the travel space. In physics, the law of the conservation of energy states that energy cannot be created nor destroyed[3], but it can be transformed or change state. In the subject of acoustics, disproportionate sound energy in an environment is treated with sound-disrupting materials.

As discussed previously, sound is vibratory (kinetic) energy. In some situations, there may be an influx of a particular frequency. This could be from the mix, the frequency response of loudspeakers or digital noise created from the PA system. One important thing to note is the concept of standing waves. The resonant conditions of the room may conflict with the output from the PA system due to the interaction between sound reflections in the environment resulting in a poorer perception of sound (Everest & Pohlmann, 2015: 103). Acoustically reducing an excess of frequencies is achieved by disrupting this energy. Absorption materials are used for this purpose by converting it into heat, thus slowing the rate in which the energy vibrates, and ultimately reducing the sound intensity (Everest & Pohlmann, 2015: 183).

Absorption is one method to dampen certain frequencies from an environment, but sometimes acoustic problems are not caused by the type of frequencies but rather their distribution. If sound in being reproduced from a sound system, the sound source is a localised from only the loudspeakers instead of multiple points such as an orchestra. Diffusion is acoustic treatment technique used to provide a more even frequency response in an environment while maintaining the sound's level of directionality. Diffusion's foremost objective is to achieve a uniformity in sound energy across the audible frequency spectrum. This is not completely possible in the physical domain, but it can be used to significantly aid in the effort (Everest & Pohlmann, 2015: 127).

Reverberation is the result of multiple reflections. When sound moves though a theatre, the first wave of sound that a listener hears is called the *direct sound*. This is because the distance between the source and the listener is the shortest. Sound reflection is simply sound changing direction after interacting with the boundary of the room (Everest & Pohlmann, 2015: 97). The first reflections that a listener hears are called *early reflections* which determine the size

[3] For more information, see: https://isaacphysics.org/concepts/cc_laws_of_thermodynamics

and nature of the acoustic space (Huber & Runstein, 2014: 72). Most acoustic materials feature reflection, absorption and diffusion to some degree, but these are once-off events whereas reverberation is a reoccurring event as seen in figure 2.9.

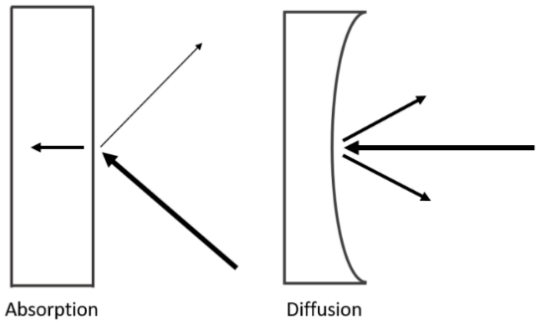


Figure 2.8: Absorption and diffusion.

After the early reflections, sound does not stop but repetitively reflects off the boundaries of the room. This creates a set of randomly decaying sounds over an amount of time, and this is known as reverberation. Figure 2.9 below is a representation of how reflections are heard by a listener. Reverberation, be it natural in room or recreated in a studio, can be used to deceive the brain into perceiving the sound as broader than it actually is (Huber & Runstein, 2014: 73). However, it can also degrade the sound quality by saturating the initial attack frequencies of music which effects the clarity of instruments (Everest & Pohlmann, 2015: 97).

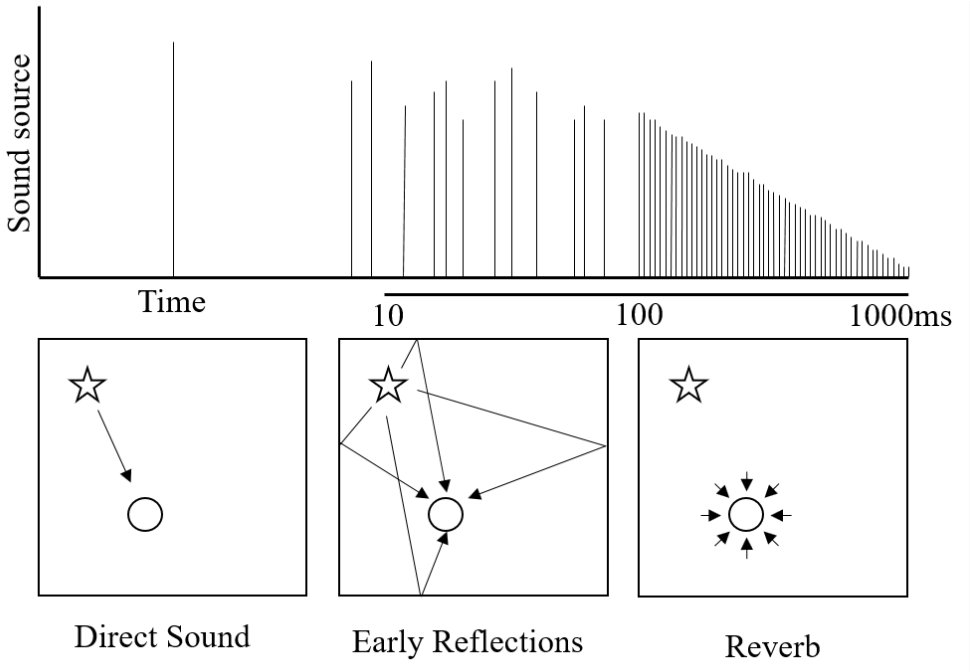


Figure 2.9: Reverberation (Huber & Runstein, 2014: 72)

In an enclosed space, if a loudspeaker is suddenly turned off or an orchestra stops playing there is a moment that the room is filled with reverberant sound. The sound energy will begin to decay from the lack of continued energy input (Everest & Pohlmann, 2015: 155). The rate at which sound decays over time is called *reverberation time* (RT). This rate is calculated to determine the time sound takes to revert to the room's steady-state of sound energy. The rate at which this decay occurs is determined by how much time it takes for the sound energy to revert to 1/1,000,000 of the initial value of sound energy density (-60dB). This defined loss of L_p corresponds with the dynamic range of a large orchestra and is known as the RT_{60}. However, RT_{60} is a subjective quantity because the perception of sound decay is relative to a listener's perspective. This is due to multiple psychoacoustic quantities such as presbycusis (see chapter 2.3.2.) or listening location, thus the decay process, or $RT_{duration}$, is followed until the noise level in the environment becomes perceptible and is dependent on the levels of excitation and noise (Ahnert & Tennhardt, 2008: 148).

Mathematically, RT_{60} can be calculated using the Sabine Equation, developed by Wallace Clement Sabine in the 1890s at Harvard University, which calculates RT_{60} by quantifying the average distances sound travels between two successive reflections (Everest & Pohlmann, 2015: 157):

$$RT_{60} = \frac{V}{A}$$

Where:
RT_{60} = reverberation time (sec/ms)
V = Volume of the room (m^3)
A = Total absorption of the room (metric sabins – 0.093)

Absorption (A) is calculated from contributing absorption surfaces. Absorption is mathematically estimated by combining the absorption surfaces and multiplying the squared-metre area of each surface (S_1) by its respective coefficient (α_1), which is the absorption properties of the material:

$$A_{average} = \Sigma\, S_1\, \alpha_1$$

Each result is then summed to obtain the total absorption:

$$A_{total} = S_1\ \alpha_1 + S_2\ \alpha_2 + \ldots, \text{ etc.}$$

(Everest & Pohlmann, 2015: 158)

A similar equation called the "Eyring-Norris Equation" is an alternative method to calculate RT_{60} (Everest & Pohlmann, 2015: 159). However, considering that both these equations do not consider factors such as potential audience capacity or psychoacoustic quantities, it becomes quite obvious that this equation is not the most accurate method. The $RT_{duration}$ needs to be accurately determined in both the recording and playback environments of this study in order for the product to be created and tested according to the reverberation criteria discussed by Ahnert and Tennhardt (2008) previously. This is discussed in chapter 3.2.3.

2.1.3. Manipulation of Acoustic and Audio Properties

Composers have understood for centuries how the acoustics of an environment affect their music, and many auditoriums are constructed to manage multiple mediums of music or performances. However, these 'general-purpose' rooms lack acoustical conditioning and usually need electroacoustic sound reinforcement (Ahnert & Tennhardt, 2008: 162). A probable reason for the construction of these rooms is that highly specialized venues are not sufficiently versatile for the owner (Geerdes, 1975: 40). For example, you would rarely have rock bands perform in concert halls designed for classical music as there would be an oversaturation of reverberation due to the environment being constructed for acoustic instruments, and not electric instruments and drum kit through a PA system.

For these general-purpose rooms, some methods can be used to adjust their acoustic properties. Depending on the cubic meterage of air space, volume (L_p) can be altered with the use of a ceiling with movable panels which could be raised or lowered; or using motorized trap doors built into the ceiling which would permit sound to pass into a reverberant chamber and decrease the rate in which sound decays if too dry. Alternatively, movable diffusion panels may be placed adjacent to walls to effectively direct sound energy into an auditorium. However, if dealing with an environment with an influx of extensive reverberation (wet environment), reverberation may be removed using acoustical draperies made from heavy velour lined with flannel, or movable absorption panels made from porous material such as rockwool (Geerdes, 1975: 41).

These examples are physical methods that changes the acoustics of an environment, but this is not very feasible if treating a very large environment. An alternative is not to change the room, but to manipulate the audio properties of sound in the room. Both live mixing and pre-recording music have their own challenges. Obviously live mixes incorporate both the raw audio directly from the instrument (drum kit, piano, etc.) and the output from the loudspeakers that that has to be balanced. In contrast, pre-recorded music is static where boosting or attenuating a singular frequency affects everything in the mix.

The frequency spectrum is divided into four basic bands: lows, low-mids, high-mids and highs. The bands are not standardised from where they begin or end, but rather crossfade into each other at point on the spectrum. These points are around 250 Hz, 2 kHz and 6 kHz (Izhaki, 2012: 57). During a live or studio mix, different points on the spectrum are equalised (filtered) by attenuating or boosting different frequencies on the frequency spectrum. By adjusting these four different bands in different positions, numerous effects occur in different situations which changes the perception of the sound output as seen in figure 2.10 (Izhaki, 2012: 210).

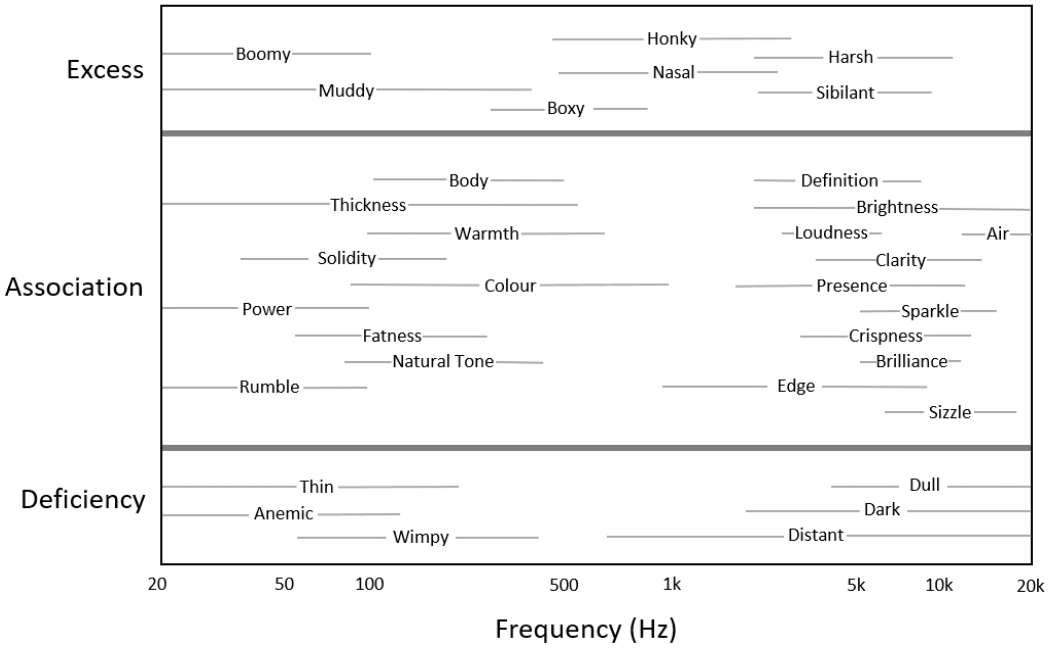


Figure 2.10: Frequency perception (Izhaki, 2012: 211)

These manipulations are made to align the sound output from the loudspeakers into the ‘association’ zone by using post-equalisation. Every environment where sound is present changes the overall tonality of sound output. For example: if sound is played in an

environment constructed from only concrete, high frequencies will be accentuated due to immense reflection off these non-absorbent surfaces which is why they are equalised to compensate for this (Toole, 2018: 60). Another example is when pre-recorded music is played on commercially produced output devices. On many occasions, commercial loudspeakers or headphones are built to over-accentuate bass frequencies, which can ultimately muffle the sound (Toole, 2018: 48). The sonic intent of the music is discussed next and how microphones are used to achieve the intended tonality and impact of the sound.

2.2 Recording *Welcome to the Show* (2019)

2.2.1. The Sound of *Welcome to the Show* (2019)

In the previous section, an explanation of how to manipulate the sound is given, but to what degree should the sound be manipulated? In other words, what parameters should the outcome of the study adhere to? This has to do with the subject material of this study, and its intended tonality and impact.

A musical genre, as described by Beard and Gloag (2016: 116), is a set of constructs and expressions within a musical work that categorizes music according to its structure, meaning and arrangement. Some genre examples are rock 'n roll, jazz, musical theatre and classical music to name a few. Musical theatre is believed to have originated from mid-17th and early 18th century Italian opera when the compositions were deeply rooted in the western art style of music (Burkholder, 2010: 860). A style of music, according to Pascall (2001: 638), is also a mode of expression but is exclusive to a specific work or composer. This is determined by observing the main phenomena of form, texture, harmony, melody and rhythm. Musicals could be considered a bridge between commercial music (pop) and western art music (classical) because they either feature popular songs or full orchestration, and sometimes both. The pieces of music in a musical are framed by a loose plot, and the structures of these plots resemble Italian opera (Burkholder, 2010: 860).

Welcome to the Show's (2019) plot results in the protagonists dying, thus the genre is further defined as a musical tragedy (Novak & Novak, 1996: 2). The composition of this 'dark-themed' musical is stylized in multiple sub-genres featuring collections of late 19th century to 21st century classical and commercial music techniques. These are techniques such as extended harmonies, varying textures, jazz and heavy metal riffs, and sections of 12-tone

systems, to name a few. Even though these techniques sound as if they confuse the identity of the main genre, there are stylistically incorporated using repeating thematic idioms in variation to link the different sub-genres such as the Argentinean tango or rock ballad (Anderson, 2019: v).

Besides some non-traditional orchestral instruments such as the classical guitar, drum kit and synthesizer, it uses fairly traditional orchestration also derived from the 19th/20th century. Although not as large as the symphonic works by composers such as Berlioz (1803-1869) or Prokofiev (1891-1953), which may consist of 100 players (Adler, 2001: 6), its full arrangement is written for a 43-piece orchestra (reducible to 25) and 17 actors/singers. The orchestra is significantly larger than a standard Broadway orchestra which according to Suskin (2009) rarely has more than 26 players (Writing the Score section, para. 1). The reasoning for *Welcome to the Show's* size is due to timbral balance that is intended to closer resemble an opera rather than a typical Broadway musical. The orchestration table in appendix B1 displays the different sections of the orchestra, the instrumentation and the number of performers in each section.

The content of the story defines the overall mood of the music and should be expressed in the recorded product. Considering this, the music and mix should emphasize the "darkness" of the musical and evoke the emotions of the characters onto the audience. This is expressed in multiple ways but the main contributing factor to this is a musical term called *timbre*. Timbre is a difficult term to describe as it is multidimensional, attributing technical impressions such as spectral characteristics and decay time. However, it has been found that spectral centroid (brightness) and temporal attack time are the two fundamental dimensions of timbre that justifies emotional response and is essential to genre recognition and discrimination (Wu, Horner & Lee, 2014: 928). Wu et al (2014: 930) categorised the emotional responses of subjects into eight categories when exposed to changes in musical tone: happy, sad, heroic, scary, comic, shy, joyful and depressed. These different emotional responses were chosen as they are considered to be the most salient (Ibid).

Their results show that instruments with a slower attack time and an accentuated spectral centroid become more depressing while at lower levels are happier. In contrast, brass instruments generally oppose this having a more heroic emotional response at louder levels and boldness (joyful) when lower. Fast attack instruments such as the string instruments are

neutral (Wu et al, 2014: 931). This likely because this family of instruments have multiple strings with different thicknesses that can produce the same pitch in various positions creating a softer attack (Adler, 1989: 57). Although these results are subjective in nature, it shows that the use of temporal and spectral audio changes on the instruments are the main components in this study that are manipulated to form the timbre of the musical during creation and testing. However, the majority of timbre and emotion is conveyed in the orchestration of composition itself. Thus, the way in which the music is performed and recorded are the first steps in capturing the intentions of the composer.

2.2.2. Microphone Theory and Application

The microphone is a transducer used to convert kinetic energy to electrical energy (signal) and is usually the first device of many in the recording chain (Huber & Runstein, 2014: 109). There are external and design variables of microphones that determine the quality of a recording. Design variables are the microphone's operation/directional response, design characteristics or frequency response, and quality (construction). The external variables include the placement of the microphone with regard to distance and height, the acoustic environment the microphone is in, and the instruments that are being recorded (Ibid). Every instrument has particular characteristics such as range, build material and resonance. Because of this, varying types of microphones and configurations are used to extract the best possible sound for a particular recording (Ibid).

Directional Response and Polar Patterns:

Different microphones contain several variations of directional configurations. When used in an ensemble recording setting, some configurations can be used to isolate specific instruments or sections of instruments to reduce the amount of sound leakage from other instruments. Alternatively, other configurations can be used to capture the holistic sound within the recording environment. These configurations refer to a microphone's directional response which can be graphically represented as polar patterns (Owsinski, 2017b: 32). All recording microphones are classified as either *pressure* or *velocity* type microphones. Pressure type microphones are zero-order gradient and have a diaphragm of which a singular surface is exposed to sound. This causes it to flex back-and-forth converting sound pressure (atmospheric pressure) into signal (Ballou, Ciaudelli & Schmitt, 2008: 493).

Velocity or *first and second order directional* microphones form the majority of the cardioid family of microphones (Rayburn, 2012: 81). Microphones can be constructed with multiple or singular polar patterns and directional responses however their placement and reactance are determined by several factors (Ballou et al, 2008: 493). *Random Efficiency (RE)* is a ratio between the microphone's on-axis directivity in relation to its omnidirectional (reverberant) response. A microphone with a RE of 0.25 (hyper-cardioid), for example, captures the reverberant acoustic power with one-quarter sensitivity compared to its acoustic power along its major axis. The *distance factor (DF)* compares the *reach* of a microphone's pattern relative to an omnidirectional pattern. Reach minimizes feedback, and in combination with *the forward acceptance angle* (Rayburn, 2012: 85).

Impedance:

In electrical circuits, the electrical potential between two points (potential difference) causes electrons to flow from the higher potential point to the point of lower potential. This is called *electromotive force* (EMF) which is measured in volts (V) (Brown, 2008: 32). The rate in which electrons move from one point to the other is called current, which is measured in *ampere* (A), and the ratio between the voltage and current in measured in *Ohms* (Ω) and is called resistance. Current, resistance and potential difference are all directly proportional to each other. In other words, if one changes, so do the others. This relationship is known as *Ohm's Law* (Ibid).

Alternating current (AC) is the frequency of a waveform alternating, changing its polarity back and forth over time. This causes potential difference and current to be out of phase which reduces power flow. This shift in phase causes reactance which reduces power transference by storing the energy and reflecting it back towards the source. This opposing current flow, which is cause by reactance and resistance, is called impedance (Brown, 2008: 33). A microphone with high impedance is created by the resistance of current through components resulting in loud electrical noise (Edwards, 1980: 5). The RØDE NT2-A[4] pressure gradient condenser microphone is considered a low-noise microphone with an impedance of approximately 200Ω. Likewise, the Lewitt MTP 440[5] dynamic microphone has a rated impedance <500Ω is also a low-noise microphone. Considering this, impedance can

[4] RØDE NT2-A instruction manual, specifications, pg3: (www.rode.com)
[5] Lewitt MTP 440 DM instruction manual, tech data, pg16. (www.lewitt-audio.com)

be low-noise depending on microphone construction. Most professional microphones are designed to work up to a load of 2000Ω before creating high electrical noise (Ballou et al, 2008: 522).

Frequency Response and Generating Element:

One of the most important design variables of a microphone is its frequency response. This is the measured audible range (20 Hz to 20 kHz) of output frequencies when driven by a constant input signal. This generally indicates how a microphone reacts to frequencies found in a recording or playback environment (Huber & Runstein, 2014: 121).

Despite this, there are external variables that affect the frequency response. Different types of microphones generate different frequency responses. Most dynamic microphones typically are designed to have emphasized resonant frequencies (1 kHz – 4kHz) but frequencies around 10 kHz are less prominent due their moving coil construction (Owsinski, 2017b: 24). Whereas, the capacitor, or condenser, microphone generates signal using an electrostatic principle. They are constructed with a movable diaphragm and a fixed backplate which, in combination, stores an electrical charge (Huber & Runstein, 2014: 113) provided by an external power source generating 48V called phantom power (Huber & Runstein, 2014: 114). Some condenser microphones contain vacuum tubes which amplify the impedance, resulting in slight harmonic distortion. These are favoured among recording engineers as they give a unique character to a recording (Huber & Runstein, 2014: 116).

There are other microphone types such as the ribbon or Piezoelectric microphones, but they are not used in this study. The frequency response of the microphones used in this study can be observed in appendix A1. The use of these microphones is discussed in chapter 3.2.3.

Stereophonic Arrays

The polar patterns mentioned above are observed in singular/monaural (mono) microphones, but very often combinations of various polar patters are used to customize the directionality of a multi-microphone array. Several microphones can be arranged in very close proximity so that the sound from the source arrives to each microphone's diaphragm simultaneously allowing multiple variations (or combinations) to be possible. This array of microphones placed closely together is called a *coincidental* microphone technique (Senior, 2015: 217). However, if this is done incorrectly the signal to the microphones can be out of phase (see

chapter 2.3.1.) creating a series of peaks and dips that result in various frequency cancellations. This is referred to as the *comb-filtering effect* which effects the tone quality of recorded instruments (Ballou et al, 2008: 523). When recording a section of instruments, a primary mono microphone can be used to capture the optimum representation of the sound source. However, in some instances a singular microphone is not sufficient for this purpose, so a second or third ('spot') microphone is used to supplement the primary microphone. This is referred to as the *dominant array* (Senior, 2015: 269). Another method to record an ensemble is to use separate mono microphones for each instrument or section and is balanced via mixing. This is called a *peer array* and is a method that prevents any dominant sound source. This method captures a clearer signal from the instruments but requires more equipment and, because of this, is more susceptible to comb-filtering (Senior, 2015: 270).

There are multiple variations of coincidental peer arrays. The most commonly used is the XY stereo microphone technique which is a pair of matched cardioid or super-cardioid microphones standardly configured at a 90° (Owsinski, 2017b: 112). Often this angle increased up to 135°, as the standard 90° can overaccentuate mid-stage balance. Even so, the stereo image may still appear narrow (Rayburn, 2012: 221). Although considered "near-coincidental" arrays, there are variations of the XY technique that are also commonly used to enhance a narrow stereophonic spread. The Nederlandshe Omroep Stichting (NOS), which translates to the Netherlands Broadcasting System, also uses a matched pair of cardioid microphones at a 90° angle, but the capsules are 30cm (11.8 inches) apart (Owsinski, 2017b: 118). The other is the ORTF (Office de Radio-Television Diffusion Française) that uses a match pair of cardioid microphones 17cm (7 inches) at an angle of 110° (Rayburn, 2012: 230). Others do exist, but these are the most common.

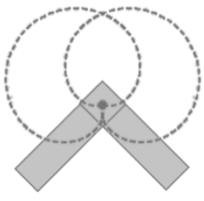
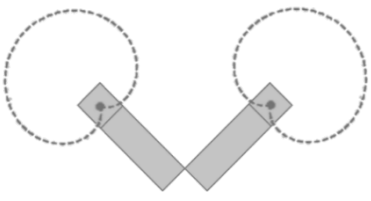
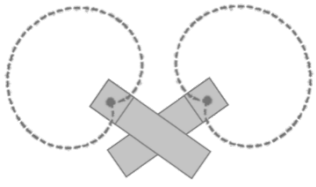
XY	NOS	ORTF
		
90°-135° inwards	90° outwards	110° outwards
0cm apart	30cm apart	17cm apart

Figure 2.11: Stereo arrays

The spaced stereo microphone technique is a *non-coincidental* peer array that uses a pair of matched cardioid or omnidirectional microphones that are generally used to record a singular instrument or a small chamber ensemble (Rayburn, 2012: 224). The main sonic characteristic of this configuration is defining a prominent left and right stereophonic image but is prone to phase discrepancies in large ensembles due to sectional sound arriving at each microphone at different times (Huber & Runstein, 2014: 139). The spaced pair also captures the room's ambient sound which includes a greater amount of reverberation in the recording with natural reverberation.

Stereophonic Dimensions and Imagery

Camilleri (2010: 200) explains that playback tools on recording platforms have drastically changed the way music is created by producers and perceived by listeners. During the time of the Beatles' (1960s) later albums, music would be created in studio but could not be performed live; rather, it could be played as a backtrack/supporting track on top of a live performance (Camilleri, 2010: 200). Even though the music of *Welcome to the Show* (2019) can be performed live without a backtrack, this practice-led experiment uses this idea to blend the different recorded mediums. This is based on Camilleri's (2010) definition of stereophonic (stereo) space in three dimensions.

The first dimension is *localised space,* which is the directional sound in a mix such as depth and width that is created with the use of panning and faders (output) control. *Spectral space* is the second dimension which has no direction, but it is experienced as saturation or emptiness. In audio processing, this would include signal processes such as distortion and EQ. *Morphological space* is the third dimension which is similar to localised space but is subject to change. In other words, it can be seen as a middle ground that blends the spectral and localised planes together. In digital processing, this would mean processes such as reverberation and stereo delays (Camilleri, 2010: 202).

Stereo microphone arrays are used to capture the stereoscopic (L/R) imagery of the orchestra. Above, Camilleri (2010) discusses recordings in three dimensions; however, a less obvious dimension needs to be addressed – height. In most situations, orchestral instruments have upward and forward directionality. At low frequencies, the instruments resonate sound with no defined direction. As sound pressure increases with brass instruments, the sound propagates more along the axis of the bell while the sound from woodwinds and strings

disperse through the sound holes and bells randomly (Rayburn, 2012: 248). When recording an ensemble of various instruments, high frequencies are accentuated on the x-axis from the sound source as high frequencies are more directional than low frequencies (Ballou et al, 2008: 520).

A combination of a central pair of coincidental (XY) or near-coincidental (NOS or ORTF) microphones and omnidirectional flanking (spaced pair) microphones are usually used to record ensembles. This provides flexibility for stereophonic perception (Rayburn, 2012: 232). These are positioned in front of and above the orchestra to roll-off much of the high frequencies. Flanking microphones are usually placed to what is called the *three-to-one* ratio where the distance from the microphone to the target sound source is no greater than a third of the distance from the microphone to the nearest interfering source. In other words, if the distance (see monaural placement and critical distance below) is 1 metre away from the orchestra, the microphones are placed at least 3 metres apart which reduces the chance of comb-filtering (Rayburn, 2012: 300). All things considered, "The best mic position cannot be predicted, it must be found" (Owsinski, 2017b: 102).

Monaural Placement and Critical Distance

Dyar (1961: 50) explains that there are fundamental practices for mono microphone placement such as testing the recording environment and reducing as much external noise as possible. Every sound that is present in a recording environment is converted to electrical energy. This makes groups are harder to record than individuals because, in the situation of having multiple sound sources, such as brass, woodwind and string sections in an orchestra, there will be significant sound leakage into an opposing group's microphones. There are ways to reduce this leakage such as placing the microphones closer to an instrument group or individual, to emphasise clarity and loudness, or placing an acoustically treated barricade between sections, or placing instruments and microphones further apart (Ibid).

The distance mono microphones are placed from the sound source cannot be predetermined because it is established by identifying the *critical distance* of the room. Microphones placed at a distance where the level of direct sound matches the room reflections is called the critical distance. When placing mono (spot) microphones for individual instruments or sections, the distance of the microphone from the sound source should be less than the critical distance of the room. This is because mono microphones are used as accentuators of the stereography of

the mix. Directional microphones, such as cardioid or bi-directional microphones, interpret low frequencies differently depending on the distance they are placed from the source. This is known as the *proximity effect*. The measurement data on any microphone product should display the distance from the sound source to the microphone to assess the degree of proximity saturation, which in most cases is one metre (Rayburn, 2012: 130).

Depending on the distance factor, placing a spot microphone within the critical distance increases the effect of proximity and can be used in situations to add character and fullness in the recording (Owsinski, 2017b: 35). However, the true distance is determined by the tone that the orchestra needs to achieve according the repertoire (Rayburn, 2012: 260). Initially, the timbre of the various subgenres found in *Welcome to the Show* (2019) would be dependent on the players' technical abilities. The microphone placement and mixing of the music should be used as the tool to reinforce the desired tone produced by the orchestra. This tone should incorporate aggressive bass resonance, sharp high frequencies and a colourful middle spectrum to emphasize the extended harmonies (Anderson, 2019: v). Placing a directional microphone in the area of proximity creates a tight and present sound quality, and also reduces acoustic resonance and sound leakage (Huber & Runstein, 2014: 133).

2.2.3. Acoustic and Digital Instrument Layering

On the subject of production cost, digital instruments are one of the most effective ways to limit a production's expense (Huber, 2008: 1114). To refer back to objective of this study, the idea of amalgamating acoustic and digital instruments, from a producer's perspective, revolves around creating a high-quality economic product. Unlike traditional multi-track recordings, where layering and overdubbing are used to create an audio product over a span of time, live recordings are created on location (Huber & Runstein, 2014: 17). This can be a relatively simple task because very little post-production processing is needed. However, in the case of *Welcome to the Show* (2019) and this study, this will prove to be one of the more challenging aspects of the recording. This is because the recordings of acoustic instruments are supplemented by sequenced digital instruments. This combination of acoustic and digital sound needs to conjoin precisely in order to amalgamate the different stereoscopic properties and produce natural-sounding music. In other words, if any section of the music is incorrect in either recording stages with regards to tuning, timbre or timing, the music will appear mechanical or unrealistic which disproves this hypothesis of combining the multiple media for symphonic music.

MIDI, Digital Sequencing and Samples

Huber (2008: 1099) discusses the properties of Musical Instrument Digital Interface (MIDI) controlling and sampling. MIDI does not communicate sound but triggers signal from a hardware or software device to produce it: digital instruments for the orchestra in this case. MIDI data may be sequenced (recorded) into software where it can be edited to generate a sound or control a set of parameters (Huber, 2008: 1101). The editing of any digital instrument is intended to match an acoustic instrument's timbral equivalent.

Zhang, Granzow, Ren, and Bocko (2014: 2) state that the prevalent mode of experimental instruments is solo performances which lack timbre compared to an ensemble setting. Their research was an attempt to eliminate this lack of timbre to allow a merger between experimental sound and conventional ensemble music. They characterized the timbre characteristics of a violin and a daxophone[6] by observing the magnitude of the spectrums of these instruments (Ibid). The method used two cognitive concepts of imitation and adaption whereby they first simulated the timbre of the violin and implemented it onto the daxophone. Secondly, they applied the performance techniques of the violin to the daxophone Both of these where done using parametric spectral analysis and manipulation based on a short-time Fourier transform (Zhang et al, 2014: 4). In short, a short-time Fourier transform (STFT) takes non-stationary signals and analyses frequency in free-time by segmenting the signal and analysing each segment separately. This is then summed for the total frequency analysis (Blandford & Parr, 2013: 606).

The results from the experiment by Zhang et al (2014: 7) show that the timbre of one instrument can be superimposed onto another using spectral processing in order to merge the different instruments in terms of frequency distribution. A variation of this experiment to conform the timbre of one instrument to another is used in this study, as the digitally controlled instruments should accurately depict the acoustic orchestral sound. This is implemented by shaping the timbre of the digital instruments to their acoustic counterparts (see *overdubbing, panning and profiling* in chapter 3.3.1).

[6] For more information and demonstrations, see: www.daxo.de

Welcome to the Show (2019) was composed in Avid Sibelius Notation Software.[7] To create the digital orchestra, the composition's MIDI data is exported from the notation software and imported into a digital audio workstation (DAW). A DAW is a software-based program that can be used to communicate MIDI in a digital instrument plugin (Huber & Runstein, 2014: 233), allowing a variety of digital instruments to be sequenced. The majority of digital instruments used in this study are from the Vienna Symphonic Library[8] (VSL) plugin. VSL, like many other digital instrument plugins, are created from acoustically recorded instrument samples, but there are other hardware or software samplers that can be used. These devices can be thought of as wavetable synthesizers that allow a user to record, load and edit musical samples into Random Access Memory (RAM) and then recall them back into the DAW output (Huber, 2008: 1115). However, some plug-in software or samplers don't accurately emulate the sonic capabilities of these acoustic instruments. To overcome this, some samples are manually created as the individual instruments, as they need to match the acoustic properties of the other instruments and reduce the amount of artificial temporal audio processing to match the rest of the orchestra (see chapter 3.3.1.).

The next subchapters discuss the role of mixing and mastering and the effect these have on the final audio product.

2.3 Mixing for Adaptable Sound

The importance of a mix is to convey the moods and expressions of musical pieces. The foundation of the mix is built from the quality of the recording, and the role of a mixing engineer is to build on this foundation by using the available tools (Izhaki, 2012: 4). A mix is created by combining, balancing and treating recorded material into a multichannel format (Izhaki, 2012: 5). Preparation for a mix is the most essential element when mixing. Creative mixing revolves around what is wanted in the mix, what actions are taken, and then evaluating the results from these actions. This is repeated until the envisioned sound is achieved (Izhaki, 2012: 20).

[7] For more information, see: https://www.avid.com/sibelius-ultimate
[8] For more information, see: https://www.vsl.co.at/en

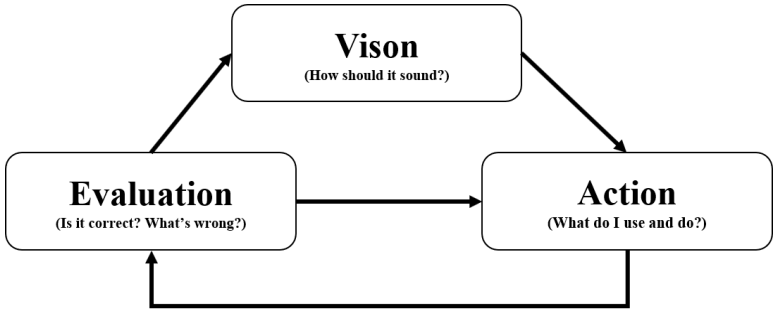


Figure 2.12: Mixing actions (Izhaki, 2012: 20)

There are both technical and creative tasks when mixing audio. Technical processes usually do not affect the sound but are essential for efficiency and clarity. These are tasks such as naming the track files, grouping instrument sections, editing performance corrections and track restoration, such as removing noise, clicks and buzzes. Within the recording, mixing and mastering processes, the ability to assess and adjust sound is done primarily through a monitoring system (Huber & Runstein, 2014: 523).

2.3.1. Monitoring and Phase

Loudspeakers used in a studio environment are called *monitors* and are one of the most vital, yet weakest, pieces of equipment in the recording chain (Huber & Runstein, 2014: 523). This is because they are affected by nonlinear qualities present in the environment such as the monitor's frequency response in contrast to an environment's acoustic properties, or the acoustic treatment of the environment and the position of the monitors (Izhaki, 2012: 73). In a small studio environment, a pair of nearfield monitors usually provide the most defined stereo image when placed equilaterally in distance to the listener (Senior, 2013: 15). As discussed in section 2.1.2, the acoustic design of the environment is intended to enhance a listener's experience. This is not only exclusive to a performance environment, but also for a studio environment. Monitoring is equally important as the environment's acoustics, because output from the monitors may become out of phase due to reflection, which causes comb-filtering ripples, and frequency responses become unbalanced from sporadic diffusion (Senior, 2013: 17).

One of the more significant technical processes is the checking of phase issues (Izhaki, 2012: 37). *Phase* is a degree of time related between multiple waveforms. When identical waveforms begin at different stages within 360°, the waveforms are said to be out of phase. The degree of a phase shift is dependent on the space of time between periodic waves and

their frequency (Izhaki, 2012: 161). These phase issues appear when combining waveforms in the mix which may shift or invert them. This causes comb-filtering when the time between waves is less than 35ms; and if waveforms are inverted or shifted to 180°, sound cancelations occur (Izhaki, 2012: 163). There are some phase issues that cannot be fixed if created in the recording stages, but some may be corrected. For example: top/bottom or front/back stereo microphones (usually on snare or guitar cabinet recording) can be fixed by inverting the phase of one of the microphones, or a peer array with spot microphones can be shifted or inverted over instrument sections (Izhaki, 2012: 164). Once optimal monitoring is established and technical interferences are addressed, the creative processes of the mix begins (Izhaki, 2012: 37).

2.3.2. Mixing Practice for Adaptable Sound

There are similarities and distinct differences in the mixing techniques used to create music for adaptable sound capabilities and for standard practice. The standard mixing objective is to establish the mood, balance, definition and interest in the music (Izhaki, 2012: 54). Similarly, adaptive sound is mixed primarily for mood and interest, but the true definition and balance of the music is established after the music has adapted to the playback environment.

From a holistic view, the multiple sub-genres in *Welcome to the Show* (2019) would require each piece to have different approaches when mixing in standard practice (Izhaki, 2012: 36). To some extent, this would also be true for adaptive sound. The orchestration of the music is meant to highlight certain instruments in different pieces, putting them "in the front" of the mix, but it must be kept in mind that this is only the instrumental part of the musical. Singers should hypothetically be able to sing on top this recording without competing with these instruments. Additionally, many of the pieces share similar thematic characterises such as comedy, drama or tension that will likely conflict with opposing emotive subjects. Besides from putting certain sections in front of the mix, adaptive sound in theory would not have different mixes for different pieces as it focuses on the relative response of the human ear, and not the true definition or balance before adaption.

Equalisation (EQ)

The human ear is said to have an approximate frequency range between 20 Hz and 20 kHz but, as discussed in chapter 2.1.1, humans do not hear all frequencies equally. Additionally, as we get older, our high frequency range reduces due to what is called presbycusis (Howard

& Angus, 2009: 38). In reality, the average usable dynamic range of music spans approximately between 10dB and 110dB of sound pressure between the frequency range of 50 Hz to 15 000 Hz (Howard & Angus, 2009: 39). The louder music is played, the more apparent low and high frequencies become compared to mid frequency range (Izhaki, 2012: 12). This makes mid frequencies the key to a balanced mix at varying levels, as good sound at low levels is likely to sound good at high levels (Izhaki, 2012: 13). This is because of the human ear's perception of loudness, where the ear perceives variations of sound level pressure with different frequency variations.

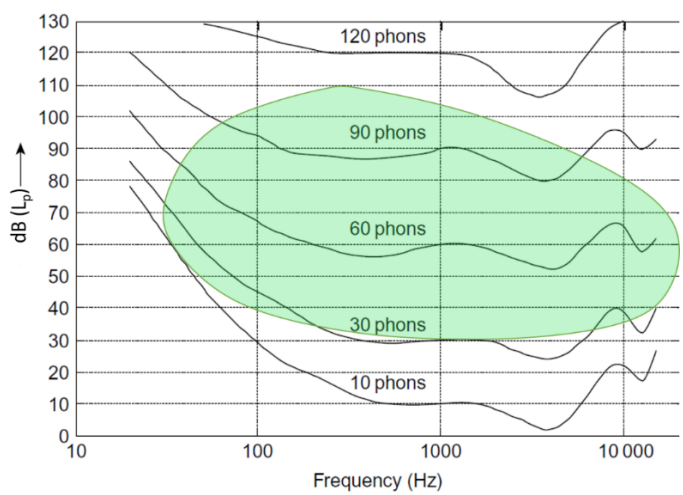


Figure 2.13: Fletcher-Munson and music frequency range (Howard & Angus, 2009: 41)

Figure 2.13 displays a Fletcher-Munson (1933) graph indicating the contours of the human ear's relationship between perceived loudness and measured sound pressure at different levels (Howard & Angus, 2009: 41). The highlighted section is the average frequency range of music discussed in chapter 2.1.1. These contours can be compared to the microphone frequency response discussed in chapter 2.2.2. (see appendix A1), where the contours represent a human ear's 'frequency response', and the L_p variations represent the 'proximity' of loudness. The phon scale is based on subjective loudness and is determined by comparing the 1 kHz frequency L_p in direct proportionality to the change in L_p along the frequency spectrum. This identifies a human's 'flat' frequency response in relation to frequency L_p. The resonant frequencies of our ears (3.4 kHz and 13 kHz) increase the closer sound moves towards our ears or the louder it becomes (Ibid). As can be seen in the figure 2.13, the contour of 60-phon curve has the majority of the frequency spectrum encased by the average sonic range of music.

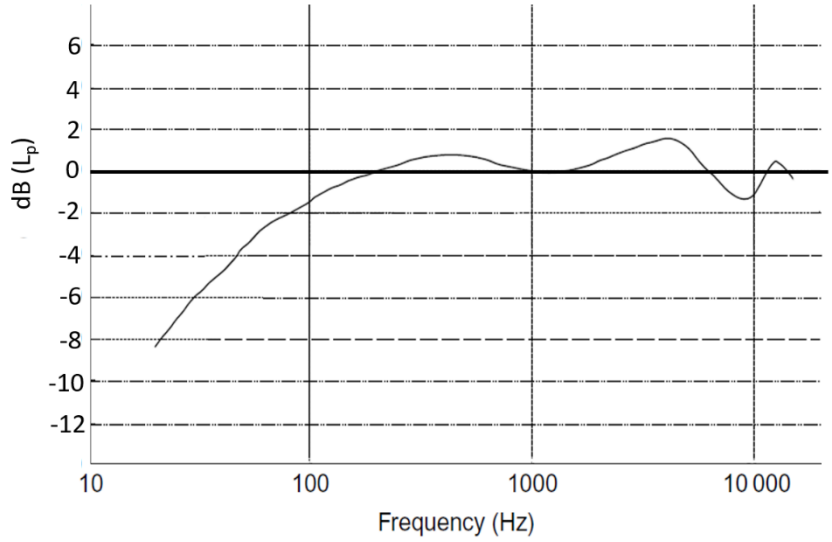


Figure 2.14: Inverted 60 Phon Fletcher-Munson contour (EQ Curve)

When processing audio for adaptive sound, the 60 phon contour is inverted and is represented on an EQ graph as seen figure 2.14 above. This is because this inversion represents the EQ filtering contours that enhance definition according to the human ear's frequency response (A-weighting). Additionally, the gradient of the Fletcher-Munson (1933) contour represent the L_p intensity of the frequencies compared to the benchmark L_p at 1 kHz. In other words, the higher the contours on the Fletcher-Munson (1933) graph, the more sound intensity (volume) is needed to match the 1 kHz benchmark L_p. Therefore, the average range of frequencies and L_p around the inverted 60-phon curve is where the music is equalised for adaptive capability. This is because, if the music is equalised to conform to this range, the degree of change in frequency response in the playback environments can be measured and adjusted to this contour (see chapter 3.3.1). If the output after adjustment of EQ has no detrimental effect on the sound quality, the theory of adaptable sound is concluded with regard to EQ.

Spatial processing

As discussed previously, both mono and stereo microphone techniques are used to record the acoustic instruments. The required stereo image is observed in the original composed arrangement of the orchestra (see appendix A2) because like a live performance the objective of stereo imaging is to identify where each instrument is positioned in the mix (Izhaki, 2012: 196). Both the acoustic and digital instruments are panned to a position in the stereo panorama, but it was found that the use of dozens of processed audio stems put immense

strain on the RAM and CPU. This is resolved by bouncing the digital instruments out as mono or stereo tracks (WAV). As a consequence, no more adjustments can be made in the plug-in software unless returning to the digital instruments and re-bouncing the audio files (Senior, 2013: 81). Additional processing can be made to the bounced WAV files from other plugins (see overdubbing, panning and profiling in chapter 3.3.1).

A new problem is created from this exporting. The version of VSL that was used in this recording could only bounce the audio as stereo tracks. Mono tracks can be panned in any direction in the stereo panorama as they have no spatial cue (Izhaki, 2012: 191) whereas stereo tracks result in spatial reproduction as explained in chapter 2.2.2, but this doesn't mean they can't be used outside of a stereo pair. The VSL tracks are exported as a coincidental pair thus can be summed internally without causing comb-filtering. This is because this group of stereo pairs provides the best mono-compatibility as phase discrepancies are non-existent due to close proximity. This allows a mixing engineer to narrow or widen the stereo image as desired as seen in figure 2.15 below (Izhaki, 2012: 192).

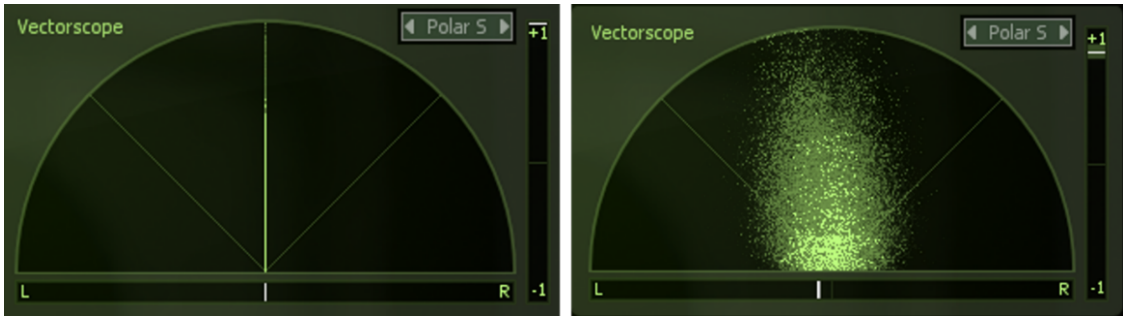


Figure 2.15: Stereo track of suspended cymbal in absolute mono and enhanced stereo

In comparison to Camilleri's (2010) theory of stereoscopic space, mono tracks have directionality that creates width and is therefore localised space. Stereo tracks are the opposite, creating depth. In the recording process, the stereo microphones are placed to cause amplitude and time differences between the different signals captured by each of the stereo microphones in the arrays. Spatial reproduction is created by panning the stereo microphones to the opposite extremes (Izhaki, 2012: 192). The exported stereo tracks or recorded stereo microphones generally provide the image of the orchestra but have no immediate direction and the image is considered spectral space, creating saturation (Camilleri, 2010: 203). The individual instruments in the stereo panorama may be localised by over-laying the stereo

tracks with the mono instruments panned to the same pseudo-position found in the stereo mix which gives the illusion for spatial depth (Izhaki, 2012: 194).

There are some exceptions to placing instruments in exact positions. Traditionally, the drum kit is placed in the middle of the orchestra or band but uses multiple microphones to record, including its own stereo pairs (overheads and room mics). This allows the drum kit to be spread across the entire panorama instead of being localised in one area (Izhaki, 2012: 195). In the case of this study, XLN Audio – Addictive Drums[9] software, is used to replace authentic drums. This plug-in allows micro-manipulation of the drum kit very similarly to the real thing. Additionally, the piano and guitar are generally considered the "main" instruments in the music and other instruments recorded in the orchestra have solo sections that should be highlighted in the mix. To bring these instruments out in the mix, we refer back to, what Camilleri (2010: 202) describes as, morphological space: sound that is subject to change. For example: in the *Entr'acte from* Act 2 of *Welcome to the Show* (2019), the piano performs ascending and descending runs which should stand out in the mix. By using panning automation, these runs stand out as the piano part changes position which catches the listener's attention (see appendix A3). Although this is not what would be heard in a live production of the musical, the use of automated panning in a pre-recorded track could be considered an example of how this process can enhance the experience of the sound.

Dynamic Processors

There is a correlation between an environment's L_p, reverberation time and equivalent sound absorption area (Ahnert & Tennhardt, 2008: 165). Suffice it to say, the volume in a room should not fall short of a minimum L_p if the desired reverberation time is to be established regardless of its capacity. In other words: the desired natural reverberation time cannot be established if the L_p of output falls short of the room's volume index. (Ahnert & Tennhardt, 2008: 166). There are multiple ways to calculate this regarding indexes and metric volume but, in this case, the recorded music is mixed to the 60-phon curve. Therefore, the average output L_p of the room becomes irrelevant as long as the inverted 60-phon curve can be re-established at different levels of sound intensity and excess reverberation is not created. However, there is a serious acoustic problem connected to dynamics within a confined space especially in the classical orchestral sense. The dynamic range of an orchestra in many

[9] For more information, see: https://www.xlnaudio.com/products/addictive_drums_2

circumstances can generate between 10dB at its quietest moments to 100dB at its loudest in addition to the ambient sound of the environment. The difference between the L_p at these two points is known as dynamic range (Izhaki, 2012: 258). The problem is the relationship between the distance from the sound source and the position of the listener. This is according to the inverse square law discussed in section 2.1.2.

As determined in free space and acoustic properties, when sound radiates from a loudspeaker into an acoustic environment, the sound energy spreads until interacting with a surface. The energy undergoes absorption and reflection until reaching a steady state of energy, where the rate of the decay is equal to the rate of creation. The total sound energy of an environment is the sum of the total direct energy and reverberant energy (Davis & Davis, 2009: 508). During the testing of adaption, the music is played in the different playback environments. The average sound intensity of the music should establish the desired reverberation time. As discussed in chapter 2.1.2, the product is mixed to have a reverberation time below 1.4 second but playing it back in an environment causes additional reverb to be added to the music due to an increase in space. After adaption, the music must be loud enough to hear clearly, and the additional reverberation created by the space in the room should be in proportion to the loudness for every appropriate listening position. For example, opera houses are designed to have greater spaciousness as a musical requirement, having an average reverberation time up to 1.8 seconds (Ahnert & Tennhardt, 2008: 163). Loudness is achieved by controlling the dynamics using digital signal processing (DSP).

Dynamics in terms of mixing refers to the variation of sound level. In the case of orchestral music, there are vastly noticeable variations between different sections of music. These vibrant dynamics are the differences in the dynamic ranges in the music, but there are several audio processes that can be used to control these levels in the mix. These are known as *macro-dynamic* controls, where variations of sound level are manipulated in large sections, and *micro-dynamic* controls, which are used to control the dynamic envelope of sound. The dynamic or audio envelope refers to the attack, decay, sustain and release (ADSR) time of frequencies (Izhaki, 2012: 260).

Different compression and expansion processors are used to manipulate the dynamics of a mix, be it in the larger scheme of the entire work or in micro-management (Izhaki, 2012: 262). Input signal that peaks over a particular sound pressure limit can be treated by adjusting

a parameter called the **threshold**. The threshold level bi-divides the input signal and is positioned below the input signal limit as seen in the figure 2.16.

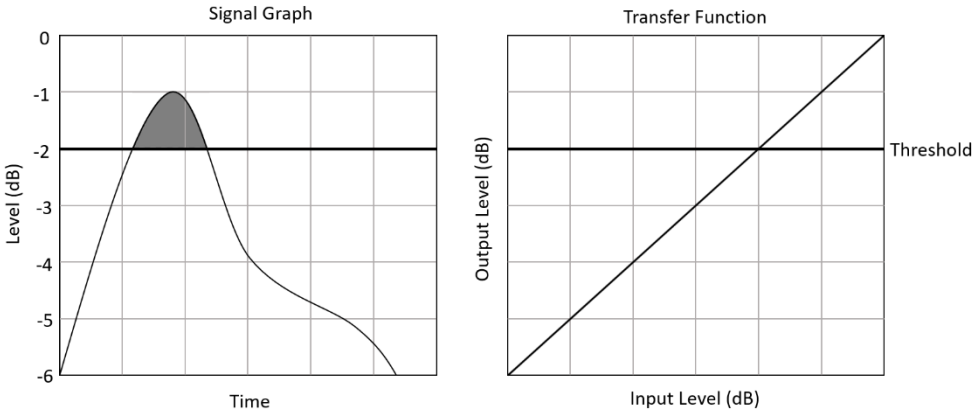


Figure 2.16: Uncompressed input signal peak (-1 dB) with a set threshold (-2 dB)

If a **compressor** is used, the signal level will be reduced below the threshold (Izhaki, 2012: 262). Other dynamic processors can also be used depending on the circumstances. **Limiters** refuse signal past the threshold, **upward compressors** boost the signal below the threshold, **expanders** reduce signal below the threshold, and **upward expanders** boost signal above the set threshold (Izhaki, 2012: 264):

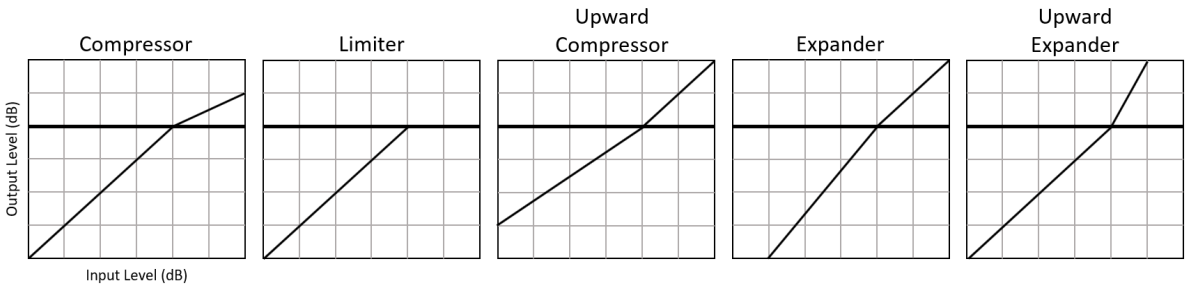


Figure 2.17: Dynamic processors using a ratio of 1:2 or 2:1 (Izhaki, 2012: 263)

There are other variations of dynamic processors such as gates and duckers. However, considering that most of the recorded music is recorded digitally, this becomes less prevalent. However, this dynamic processing for pre-recorded adaptable sound is identical to standard procedure, whereby the output needs to be optimized without being detrimental to the sound quality.

Temporal Processing

Time-based processors are used to alter or augment signals to create perceived depth or dimensional space. There is a large variety of time-based effects such as flange, chorus and regenerated echo, but they all revolve around the use of signal **delay**. Digital delay is a common DSP tool that alters the parameters of time by storing sampled audio directly into Random Access Memory (RAM), then reading it out after a set time, creating various lengths of reflections in a mix (Huber & Runstein, 2014: 506). As discussed in chapter 2.1.2, acoustic reverberation is multiple reflections that randomly decay over an amount of time (Huber & Runstein, 2014: 73). In the digital domain, simple delays are more controllable echoic patterns as they usually use less stereophonic space, and are rather used as reverberation enhancers (Senior, 2013: 255). These 'echoes' can be copied and repeated over the original audio stem for a set time period or continuously to simulate large spaces or for creative effects. This is known as **feedback** (Izhaki, 2012: 378).

Even though time-based processors are not used in this study to create such effects, as the objective is to recreate and improve the sound of an authentic orchestra with digital means, there are another uses for delay. The doubling of instruments in an orchestral score is widely used to create timbral balance and stability as is the case of *Welcome to the Show* (2019) (Adler, 1989: 5), but many digital instrument plug-ins do not offer the choice of multiple instrument configurations. The use of closely spaced delays, between 15 and 35 milliseconds (Haas effect), can be used to create a doubling effect that can fool the listener into believing there is a large ensemble that what has been recorded (Huber & Runstein, 2014: 507).

The most important use of delay and reverberation in the scope of amalgamating the digital and acoustic recordings is to add a sense of space and depth (Izhaki, 2012: 389). During the recording stages, the natural distance that musicians are placed from the stereo microphones allows each instrument's sound to reach the microphones at different intervals. The digital orchestra has no such natural delay because it has direct output. Therefore, the digital orchestra depth is spatially enhanced by introducing an equivalent digital delay and reverberation time (less than 1.4 seconds). This artificially recreates a digital copy of the environment the acoustic instruments are recorded in, which can be merged.

2.4 Mastering for Adaptable Sound

Mastering is a highly specialized process that takes place after the mix is finalized. The objective of mastering audio is to fine-tune the small details of the mix and to prepare the final product for distribution, or adaption in this case (Owsinski, 2017a: 17). One of the most valuable mastering tools is a dedicated mastering DAW. Unlike many mixing workstations, dedicated mastering workstations facilitate different features and conveniences that makes mastering more efficient, such as the software that can embed metadata and CD text (Katz, 2015: 13). In this case, iZotope Ozone 5[10] is the dedicated mastering software used that also features high resolution and data integrity, powerful crossfade functions and conversion software (Katz, 2012: 14).

The main difference between a mixing and mastering engineer is the approach of listening to music. A mixing engineer specialises in improving the sonic integrity between the individual elements of the recorded audio and stereo output, while a mastering engineer works on the complete mixes in their entirety to create optimal output correspondence in the stereo tracks and sonic resemblance between corresponding pieces/songs (Katz, 2015: 25).

Any audio processors applied to the master tracks will affect everything, not just individual audio stems, as in mixing. For example, if attenuating the low frequencies in the master, the perception of high frequencies will be intensified (Katz, 2015: 55). Furthermore, the cutting and boosting of the frequency spectrum in mastering will affect all the instruments in the frequency spectrum which intensifies the use of equalisation. Hence, it is important not to over-emphasize EQ when mixing. For similar reasons, over-compression must also be avoided when mixing because the waveforms hyper-compress if additional dynamic changes are used in the mastering process (Owsinski, 2017a: 31).

The final responsibility of a mastering engineer is to embed the metadata and the International Standard Recording Codes (ISRC) to protect the intellectual property of the artist (Kats, 2015: 16). Shaw (2010: 137) explains that the libretto (lyrics), choreography, costumes and plot all form part of one entity. If *Welcome to the Show* (2019) had to be performed, grand rights have to be obtained through the Dramatic, Artistic and Literary Rights Organization (DALRO). DALRO is a division of the South African Music Rights

[10] For more information, see: https://www.izotope.com/en/products/master-and-deliver/ozone.html

Organization (SAMRO), which is where the music will be separately registered for copyrights. The actual recording will need to be submitted to the Composers, Authors and Publishers Association (CAPASSO) for the mechanical copyrights of the recording. However, it is one small administrative stage of the post-production and does not feature extensively in this study.

2.4.1. High-resolution Mastering

There are similarities in the mastering techniques used in standard practice and for adaptive capabilities. One such similarity is the choice of *sample rate* and *bit depth*. Throughout this discussion of audio processing, there has been mention of the word "signal". Analogue signal refers to sound being stored (recorded) and reproduced as voltage that changes level over a period of time. However, digital signal cannot be stored in such a fashion. Digital signal is recorded periodically by capturing samples of the analogue waveform over time and converting the voltage level into binary code of numerical equivalent at any point in time. During this time, the analogue signal is sampled at precisely timed intervals (sample rate) where the analogue signal is held (Huber & Runstein, 2014: 201). Sample rate and bitstream (samples per second) determine the resolution in which digitally captured sound is recorded, and the *bit depth* is a relation between the amount of quantization steps between samples (Huber & Runstein, 2014: 256). For example, the standard CD has a sample rate of 44,100 samples per second with a bit depth of 16 bits (Owsinski, 2017b: 94). Figure 2.18 displays the conversion of analogue to digital signal.

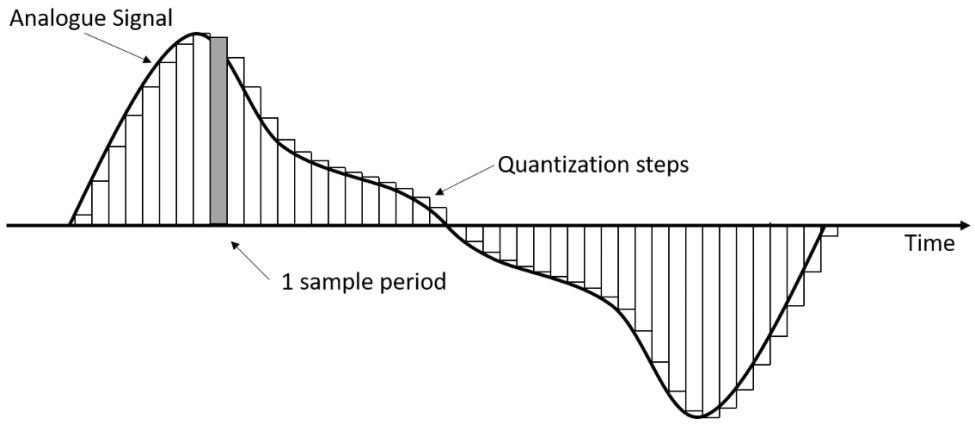


Figure 2.18: Analogue signal curve and digital signal steps

The mastering similarities of the different procedures is in regard to using sample rates and bit depths in high resolution. High resolution mastering varies depending on the situation, but

the use of 96,000 samples per second (96 kHz) at 24 bits is the minimum requirement (Owsinski, 2017a: 129). The difference between standard mastering and adaptive sound mastering is that commercial audio products such as CDs (44.1 kHz/16 bit) and music mastered for iTunes (44.1 kHz/256 Kbit) are down-sampled and have a dithered bit rate to conform to these formats, whereas this does not apply in the case of adaptable sound. This recording is not a commercial product, which is why it is created in high resolution (96 kHz/24 bit) to enable more accurate and effective adaptive capabilities.

2.4.2. Tonal Balance

The mastered EQ balance from recording and mixing a symphonic orchestra should always exhibit a gradual roll-off of high frequencies on a spectrum analyser. The task of creating tonal balance in standard mastering practice is done to provide tonal colour for a genre-specific product with the ability to translate the final master through a variety of different playback systems (Katz, 2015: 56). The process of mastering EQ for adaptable sound is derived from this procedure but is not specific to a genre. Instead, the tonal balance is created after adaption. Incidentally, the EQ parameters are observed in appendix A4. This is to ensure optimum frequency coverage over the entire sound spectrum. The mastering process reinforces the inverted 60-phon EQ curve by utilising different parametric band and filter adjustments. When music is adapted in the environment, post-production EQ is used to re-establish the 60 phon curve in the playback environments. Band EQ compensates for the resonant frequencies and standing waves while filters attend to noise.

2.4.3. Loudness Range

There are two international audio standards for loudness. The first is the International Telecommunication Union (ITU) BS.177-3 that measures both the *programme's loudness* (PL), or integrated loudness, and its true-peak level (Katz, 2015, 73). The second is a recommendation by the European Broadcasting Union (EBU) called the R-128 which defines the PL dynamic range from 0 dB to 128 dB of maximum intensity. The latter standard is used when defining loudness in this study (Katz, 2015, 74). Regardless whether using the ITU or EBU standards, PL is measured in LUFS, loudness units below full scale, which is the level of encoding in a digital format (0 dB) for total output. Essentially, the digital sound clips above 0 decibels. The negative decibel readings (e.g.-6 dB) are known as LUFS (Katz, 2015: 219).

As mentioned before, the dynamic range of this recording can range between 10 and 100 dB of L_p. This is measured on a dynamic metre and identifies the true dynamic range of the music (Owsinski, 2017a: 60). This range can be manipulated using dynamic processors as mentioned in chapter 2.3.2. The difference between dynamic processors in mixing and mastering is that the ratios are set lower (1.5:1 to 3:1) as the sound becomes unnatural. This may be desired in a mix to acquire a certain effect, but mastering is intended to keep the original sound of the mix as intact as possible (Owsinski, 2017a: 46). One of the most commonly used dynamic controllers used in mastering is the multi-band compressor. Unlike a standard compressor, a multi-band compressor has the ability to split the input signal into different frequency bands (Owsinski, 2017a: 47). Using bandpass filters, the multi-band compressor isolates different frequency ranges to have different compression settings (Huber & Runstein, 2014: 501).

This audio process is used to change the dynamic range of the music to a more appropriate playback level. On average, mastered classical music usually measures a dynamic range of approximately 16.63 dB between its loudest and quietest moments, and music for stage or screen is around 14.29 dB (Owsinski, 2017a: 61). Although *Welcome to the Show* (2019) is fundamentally a classical composition, in the sense of dynamic range, the mastered tracks should not exceed the digital output of 0 dB or be less than -14.29 LUFS for the entire work. This is because *Welcome to the Show* (2019) is intended for stage, therefore the dynamic range should be moulded to these limits. There are some deviations to this due to the different sub-genres found in the work, but the maximum limit still applies (See chapter 3.3.1.)

The mixing and mastering processes are important aspects of this study, but according to objectives laid out in this research, then environment plays an important role in the global timbre. Therefore, a discussion of analysing audio in the acoustic environments is discussed next.

2.5 Analysing Audio

Considering all the aforementioned information, accurately creating this particular audio product requires constant procedural analysis of the audio pertaining to the sound pressure and the frequency and time domains. These audio properties are what will likely change in the environments during playback and what are compensated for during adaption. To measure and analyse the audio properties, specialized microphones and software are used.

Measurement microphones differ from the those used in recording. They are all pressure gradient and omnidirectional, but their main defining characteristic is that they are designed to have an as-close-to-flat frequency response as possible. In the case of a theatre acoustics, the 6mm (1/4 inch) measurement microphones are the most commonly used for analysis because they provide very accurate data of most of the frequency range (Toole, 2015: 395). However, as can be seen on the Behringer ECM 8000 (see appendix A1), the upper hemisphere of its frequency response is jaggedly accentuated. This is because this measurement microphone compensates for misorientation which causes dips in the upper response. Therefore, this sound level calibration is unlikely to be affected by the accentuation (Ibid). Analytical software such as iZotope Insight[11] offers, among other uses, the ability to analyse multiple facets of audio data including mono-to-stereo changes, spectral analysis or LUFS metering. It is a method to 'see' output and input signals for comparison.

When analysing output from loudspeakers in an environment, the first major change that affects the sound is the environment itself. Rooms such as theatres have low-frequency resonance and have minimal phase discrepancies, having no prominent peaks above the average spectrum level (Toole, 2015: 89). In a non-anechoic environment that has two sound fields (stereo) such as a theatre, we find that analysed audio consists of irregular troughs and peaks (see figure 2.19).

[11] For more information, see: https://www.izotope.com/en/products/mix/insight.html

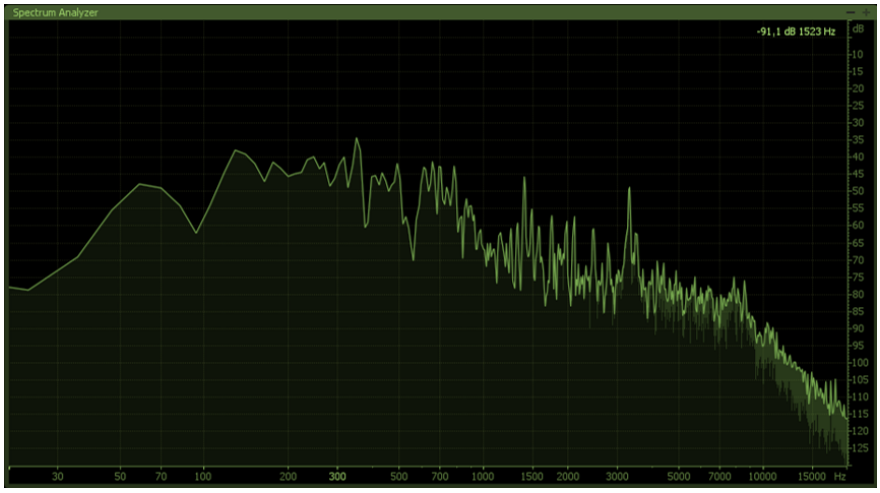


Figure 2.19: Audio spectrum analysis example

This is a result from the dominance of relatively isolated room modes and associated standing waves of low frequencies and overlapping modes from reflection of high frequencies (Toole, 2015: 148). The middle resonant zone between this occurrence in auditoriums is defined as cross-over or Schroeder (1954) frequencies. Due to this, some frequencies may be accentuated or lost during playback (Toole, 2015: 149).

The space between the loudspeakers and the microphone makes sound susceptible to unwanted noise. This is likely to occur from external sources, be it the digital noise from cables and preamps or acoustical noise from air-conditioning or simply footsteps on the stage. The human brain has the ability to filter out much of the noise, but when analysing audio, a computer does not have this cognitive function. Therefore, noise filtering using DSP is used to separate particular frequencies or bands of frequencies, or to remove interference in order to make the signal more prominent (Blandford & Parr, 2013: 1). The other use of a filter is to change signal from one form to another, as seen in the Fourier transform-based spectrum analyser where frequency can be observed in the context of time (Blandford & Parr, 2013: 2).

In filter design and analysis, an audio engineer has to decide whether to use an *infinite impulse response* (IIR) filter or a *finite impulse response* (FIR) filter, and there are different variants of these filters. The type of frequency analysis usually determines what kind of filter is implemented (Blandford & Parr, 2013: 205). FIR filters (e.g. pass-filters) are zero-centred as the poles are set to have uniformity across all frequencies whereas IIR (e.g. comb-filters)

filters have poles in locations other than zero, creating gain feedback and an extent of gain increase depending on the distance between bands. Although the narrower interchangeable bands of IIR filters seem more efficient compared to FIR filters, because of the feedback, noise can be circulated in a loop which makes IIR unstable in terms of linear phase. This is not possible in FIR filters (Blandford & Parr, 2013: 287). Considering the use of filters in this study is to eliminate noise, the focus for this section of analysis is on FIR filters.

FIR filters are designed by using the *Fourier series* and then *window functions* are applied (Blandford & Parr, 2013: 206). In brief, the Fourier series in applied mathematics allows input signal to be periodically represented as a sum of functions. Each function is selected by means of superposition so that the system response of each function can be found. In other words, the output signal is the sum of the function responses to equal the approximate input signal (Blandford & Parr, 2013: 82). However, when using the Fourier series, a set of coefficients are created ranging from negative to positive infinity, which can't be implemented as this just removes all sound. Therefore, the coefficient limit is truncated to a certain range. This is done using what is called a window function (Blandford & Parr, 2013: 221).

There are several variations of windows, the two most significant being the Hamming window and the von Hann window (see figure 2.20). These are both gradually descending window functions that result in narrow bandwidth causing less distortion in a discreate Fourier Transform (Blandford & Parr, 2013: 122). Other windows also exist such as the Kaiser, Blackman and Triangular but these have either broader or narrower bandwidths.[12] Considering the environments that are being tested are acoustically treated, these will likely filter too much or not enough. Therefore, either the von Hann or Hamming window will be used during analysis depending on the size and acoustic criteria of the testing environment.

[12] For more information about FIT filters, see Blandford and Parr, 2013: 225-237 and 718-719

$$w[n] = 0.54 - 0.46\ \cos\left(\frac{2\pi n}{N-1}\right), 0\ \leq n\ \leq N-1$$

$$w[n] = 0.50 - 0.50\ \cos\left(\frac{2\pi n}{N-1}\right), 0\ \leq n\ \leq N-1$$

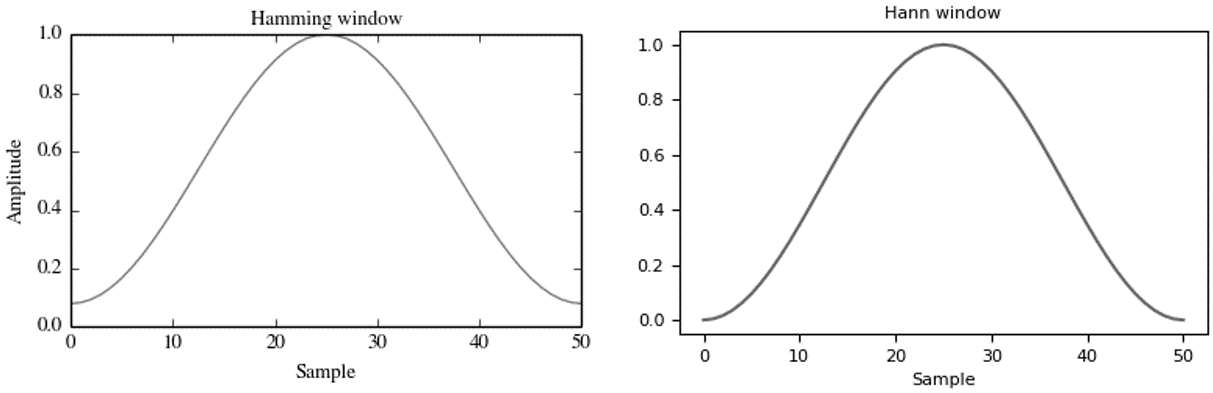


Figure 2.18: Filter windows

2.6 Data Collection Conclusion

This concludes the theoretical preparation for the product's creation. All the numerical and descriptive data discussed in the above sections is considered to create the audio product. All this data and further data is summarised and displayed in the appendices and is used as the reference notes in order to apply the correct production techniques to create this product in the practice-led and experimental sections of this study.

In chapter 3, the methods, statistics and conceptual framework on how to achieve the results are discussed.

Chapter 3: Methods, Statistics and Conceptual Framework

3.1 Environmental Analysis and Numerical Data

The ability to understand and control sound in a recording or playback environment is the key component in identifying what is to be expected during all stages of production. The multiple reactions of sound in an environment ultimately change how sound is perceived by a listener, but the task of creating uniformity in the distribution of frequencies across the audio spectrum does not. By analysing the environments in which sound is introduced, the properties of the environments serve as reference to what DSP adjustments are needed before recording or adaption. These properties are the ambient sound pressure level, critical distance, reverberation time and frequency response.

3.1.1. Ambience, Signal and Noise

The amount of ambient L_p in an environment is not a constant statistic because its controlling factors involve information, or a lack of information, that stems from two entities – signal and noise. 'Signal' refers to the information that is intended, whereas noise, in the general sense of the word, is undesirable and occurs at different degrees of L_p. However, noise can also be a carrier of information that can be shaped into a desirable form, or it could be used as a measurement tool to identify the acoustic quality of an environment (Everest & Pohlmann, 2015: 82).

Considering that the majority of environments are not anechoic in nature, various types of noise are generated within different acoustic conditions. As sound travels, it becomes susceptible to unwanted interference. This is classified as undesirable or unwanted noise which is usually created from external sound sources. In quieter spaces, undesirable noise is more prone to sound transference (under a door, through cracks in walls, air-conditioning hum, etc.) because, similarly to liquids, sound travels the path of least resistance. However, with an increase in airspace (room size), there will be an increase in sound isolation and loss in sound transference.[13] Furthermore, the level of noise increases with an increase in reverberation because of the increase of decay time causing saturation. As reverberation time increases, a decrease in sound isolation and an increase in sound transference occurs (Ahnert & Tennhardt, 2008: 148).

[13] For more information, see: http://www.stcratings.com

The environments in which acoustic instruments are recorded and playback occurs are tested using a Sound Level Metre (SLM) and pink noise. An SLM measures the acoustic response and noise level in an environment including its ambiences (Rayburn, 2012: 48). Both white and pink noise are often used to test signal although white noise exhibits a uniform distribution of energy across the frequency spectrum, doubling in energy in each higher octave and creating a loud hissing noise as it contains twice as many 1-Hz bands at each octave increase. Thus, pink noise is used as it contains the same average power in each octave with a more prominent low-frequency response that better corresponds to acoustic measurements (Everest & Pohlmann, 2015: 85).

The L_p readings from each environment are considered variables that are continuous. Due to this, they are statistically inferential in nature but provide sufficient data to make reasonable estimations on how playback environments may infer the adaption properties of the audio product. The readings taken from the recording environment determine the amount of manipulation an engineer would apply to the mix in order to adapt the sound to the playback environment. In the playback environments, the music is performed through a PA system and re-recorded using measurement microphones to essentially replace the human ears. Due to this, the measurements taken must be weighted. There are a variety of different weightings such as Z, B and C weighting but they rarely correspond to human perceptions (Everest & Pohlmann, 2015: 49). A-weighted (dBA) L_p is used to determine the effect of the environment on the frequency spectrum as it is the it compares the frequency response in relation to the L_p of the environment. The objective is to have no audible difference of frequency sound level within the different environments after adaption.

Regardless of the sound pressure at which the music is being outputted, the recorded data taken by the measurements is *normalised* to insure even levels when analysing and adapting the sound. Normalising is an audio process that maximises the gain signal to the desired level[14] in this case, the same peak-level of the mastered track which is -0.1 LUFS. (see appendix B2).

[14] For more information, see: https://steinberg.help/cubase_pro_artist/v9.5/en/cubase_nuendo/topics/direct_offline_processing/direct_offline_processing_normalize_r.html

3.1.2. Critical Distance

As previously discussed, the critical distance is the distance at which the direct L_p of the sound sources and the L_p of room reverberation are of equal quantity. To identify this, a sound source is placed at the end of the room, outputting pink noise, and the distance from one end of the room to the other is measured. The SLM is A weighted, set to a 90-dBA range (0 dB to 90 dB), and placed 1 metre away from the source. When the pink noise L_p decreases, from moving the SLM a further 1 metre away from the source, the SLM will also decrease, displaying a decrease in L_p between 4 dB and 6 dB due to the inverse square law of sound intensity. This process is repeated multiple times after resetting the SLM to the same set range of 90 dBA and beginning the measurement at the doubled distance point. At the point where the SLM no longer recognises a change in L_p, micro-variations of this experiment are conducted, which identifying the critical distance.[15] This data is displayed in the spreadsheet in the appendix B2.

The critical distance is used to determine the extent of proximity (distance) that microphones are placed from a sound source. Microphones placed further away from the sound source are less prone to feedback in all digital systems but also reduce the microphone-to-source variations, creating a product with sonic gaps in the stereophonics (Ballou et al, 2008: 519). However, critical distance is not a priority during the playback experiment because the microphones are certainly placed beyond where the measurements take place. In the auditoriums where the measurement microphones are placed in the playback environments, the distance between the microphones and the loudspeakers will differ, creating different variations of audio signals. The most significant changes will be to the reverberation time and the frequency response of the measured sound.

3.1.3. Frequency Response and Modal Resonance

Most rooms exhibit an acoustic feature called modal resonance which is when certain frequencies are concentrated in different areas resulting in an increase in sound pressure. As sound travels outwards from a source in an enclosed space, modal resonances are established at the intersections between direct sound and reflections. This causes sound to vary in EQ and pressure in different areas in the space. As a consequence, other frequencies decrease in energy elsewhere. Both these frequency attenuations and intensifications distribute in the

[15] For more information, see: ProSoundWeb article, referred to by M, Senior (2015, 273) at www.prosoundweb.com

three-dimensional environment (Everest & Pohlmann, 2015: 229). Simple harmonic motion (see figure 2.1) is the representation a singular frequency, but music comprises of complex waveforms and a large range of wavelengths. When a low frequency (large wavelength) radiates from the source, it is omnidirectional in comparison to the dimensions of the source, but a high frequency will have more directionality due to its shorter wavelength (Toole, 2018: 99). This creates distribution issues with total frequency distribution as higher frequencies reflect on-axis while low frequencies reflect off-axis. Therefore, a room's reflected sound does not have the same spectral balance as the direct sound (Toole, 2018: 100).

The adaption experiment in this study involves playing music through loudspeakers. Consequently, the different modes of the environments need to be considered. The objective of loudspeaker placement is to achieve a neutral environment with even sound distribution in different locations, although this is not completely possible to execute in a room containing reverberation due to sporadic diffusion and the comb-filtering effect. The use of three classes of room modes are the fundamental features that aid in frequency neutrality. The *axial modes* contain the most sound energy as they are created by the direct sound. *Tangential modes* are the early reflections which have the second-most amount of sound energy. *Oblique modes* (reverberation) are rarely problematic in an environment as they have the least energy (Toole, 2018: 216):

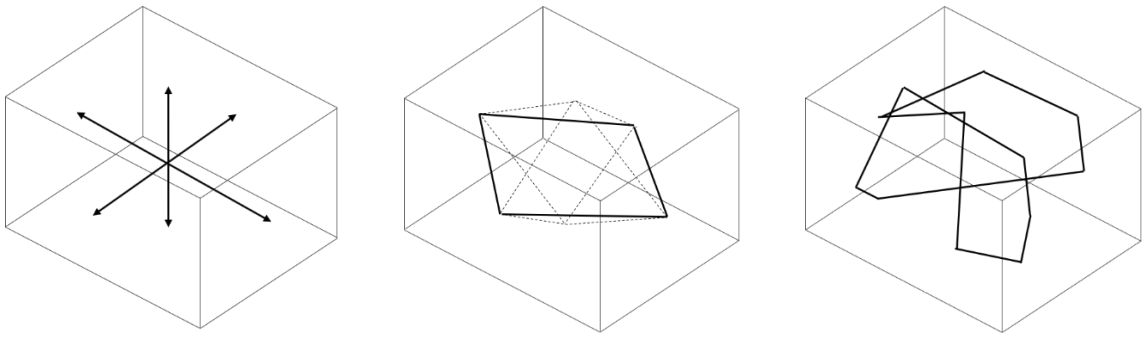


Figure 3.1: Room modes (Toole, 2018: 217)

The first method to achieve neutrality is to deliver the same quantity of bass frequencies to all listening positions in the room, because bass frequencies resonate sporadically past the

critical distance in contrast to high frequencies, which are more linear (Toole, 2018: 215). However, this is easier said than done. Excitation of noise from loudspeakers creates energy in a room. As this energy excites one mode, it excites the other modes in a chain reaction (Everest & Pohlmann, 2015: 164). As a result, when the sound eventually decays from a lack of sonic support, there are fluctuations in the decay time at different frequencies (Everest & Pohlmann, 2015: 165). This variation in frequency decay is also affected by the capacity of the audience in the room. With a full auditorium, spectators could absorb between 55 and 60 percent of low frequencies (Barbieri, 1998: 290).

Therefore, the loudspeakers should be placed to optimize the resonant modes, creating a close evenness in high-frequency distribution along the axial mode while computer software pertaining to L_p metering (A-weighted) is used to monitor the sound intensity of the room. This can indicate what can be altered to achieve neutrality. In addition, a spectrum analyser is also used to identify the sporadic frequencies produced in each room. Coincidentally, this method could also be used as a tool to identify flaws in the acoustic properties of an environment if physical modifications were to be made to adjust its acoustics, although physical adjustments to the loudspeakers are not always possible (see chapter 4.1.1).

3.1.4. Venue Size and Reverberation Time

The Sabine or Eyring-Norris Equation is used to determine mathematically the reverberation time of an environment. However, a more accurate and a practical method is used to acquire the RT_{60} of the recording and playback environments in which the decay of L_p is observed over time. Firstly, the ambience of the room is needed to determine its steady-state of L_p. This is identified by using measurement microphones placed in various positions in the environments. The empty space of the room is recorded, and the average of these different recordings establishes the ambient sound. This ambient sound serves two purposes; the first being the aforementioned and the second is to fill silent gaps in the music during playback (see chapter 3.3.2).

Once each environment's steady-state of L_p is determined, a wideband pink-noise signal is applied to the stereo sound source (loudspeaker system), which is positioned to maximize the resonant modes of the rooms, which are the corners. Depending on the size of the environment, a singular or several measurement microphones are placed in various positions beyond the critical distance in each environment to record the incoming signal from the

sound source and additional environmental noise (Brown, 2009: 29). If several microphones are used, they are evenly placed according to the inverse-square law due to direct sound attenuating over larger distances (Brown, 2009: 30).

The test for RT_{60} involves outputting four bursts of gated pink noise in four-second intervals with a four-second gap between bursts. This gated pink noise has no reverberation in the recording, as can be seen in the figure below.

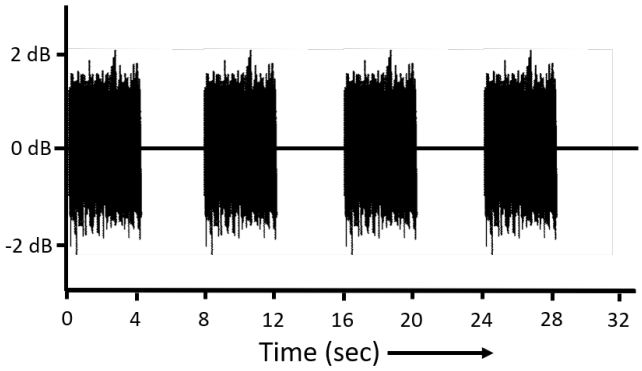


Figure 3.2: Gated pink noise sound waves

This is recorded with the measurement microphones, sent through a preamplifier and AD converter to convert the analogue signal into digital[16]. The total RT_{60} is determined from the average time the sound takes to revert back to the steady state of sound energy in the room at a normalised level of sound intensity. The different spaces will display variants of RT_{60} that identify the nature of the decaying field (Brown, 2009: 13). The possible variants are shown in the figures below.

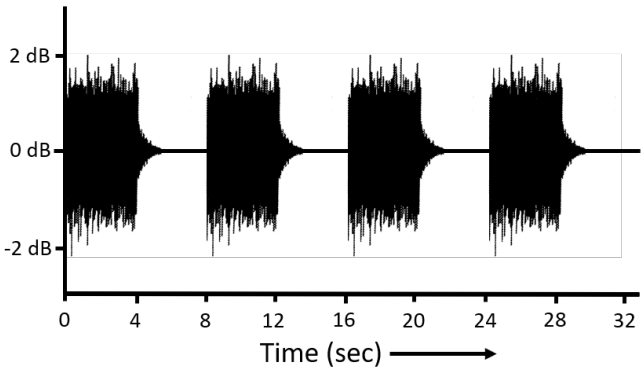


Figure 3.3: RT_{60} example with a decay time of less than 4 seconds

[16] For more information, see: https://www.analog.com/en/products/audio-video/audio-ad-converters.html

The figure above illustrates the signal from the loudspeaker stopping after exactly 4 seconds. The decay of sound energy indicates that a RT_{60} of less than 2 seconds has occurred. This reaction occurs under different circumstances such as in large acoustically treated environments with high absorption properties, or a small room with little sound transmission. Either way, this amount of decay is considered to have a relatively low reverberant field (dry environment).

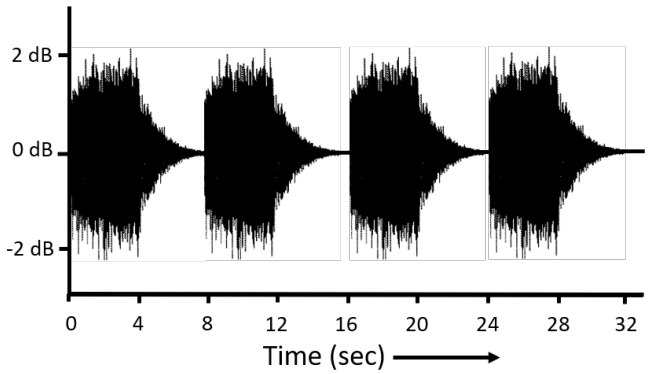


Figure 3.4: RT_{60} example with a decay time of exactly 4 seconds

An environment that takes between 2 and 4 seconds for the sound energy to decay to a steady state of sound energy is a highly reverberant field (wet environment). Although this is relatively uncommon for a multipurpose theatre, this reaction is likely to occur in environments with a larger sound intensity or in rooms constructed with highly reflective or resonant materials seen in sacral rooms, such as a cathedral, or classical concert halls (Ahnert & Tennhardt, 2008: 163).

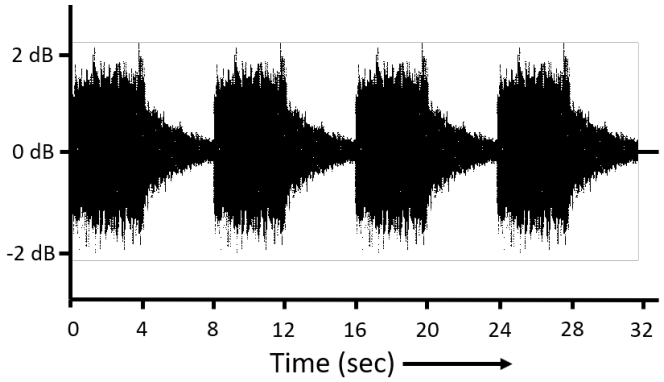


Figure 3.5: RT_{60} example with a decay time of more than 4 seconds

A decay time exceeding 4 seconds is a phenomenon that usually occurs in specifically constructed environments that are designed have an overly reflective sound field under immense L_p. This type of reverberant response is not likely to be relevant in the scope of this study because it is not a usual occurrence but, in a situation such as this, very little can be done using DSP to adapt. However, in this unlikely event, physical changes are the only possible solution to adapt the sound, such as adding absorption materials to the environment or changing the modal position of the loudspeakers to weaken the resonant modes.

3.1.5. Environmental Analysis Summary

In a live performance that uses no sonic reinforcement, the decay of sound energy lessens in an environment that is equipped with sound absorption materials such as rockwool panels. However, if playing pre-recorded tracks through loudspeakers, the reverberation of the room compounds the reverberation embedded into the tracks and cannot be easily removed. Artificial reverberation can be digitally introduced to both pre-recorded and live sound by use of external components if there is a lack of natural reverberation (dry environment). In theory, the artificial reverberation is a dependent controlling factor, but the additional reverberation caused by the environment is an independent statistic. Therefore, less than normal reverberation is used when creating the audio product to avoid oversaturation and to allow the sound to adapt to the environment with regards to reverberation. Less than the established 1.4 seconds, discussed in chapter 2.1.2, is the target for the mix as this is considered a fairly dry reverberation which is supplemented by the natural reverberation in the environment or by artificial reverberation to fulfil the needs of a particular space (see chapter 4.3).

The link between a room's acoustic qualities and the loudspeakers placed in all spaces influences the quality of the sound as it is reproduced (Toole, 2018: 16). It does not matter if the room is well acoustically treated; if the loudspeakers are inappropriate for the environment in which the sound must be performed, adaptive capabilities with regards to EQ and RT_{60} become unattainable (Toole, 2018: 17). Extreme examples of inadequate loudspeakers would be using a pair of 10-inch studio monitors in a theatre or using 24-inch subwoofers in a small room.

The completed tracks are tested in various environments but before testing commences, several regulations are exercised to standardize the different acoustic variables in each setting. The metric data of the environments is gathered that describes the characteristics of the listening spaces. The output of the tracks is played from a sound system, but sound systems are greatly affected by the sound energy persistence in an enclosed space. The main contributing factors to this are sound intensity and RT_{60} (Brown, 2009: 12). This experiment is repeated using the finished audio product in a controlled studio environment, to determine the benchmark of sound quality, then in the playback environments with an equivalent intensity-to-size ratio to determine the difference in sound.

3.2 Recording process and data application

This adaptable sound technology can be used for both live and pre-recorded music. However, this study focuses on the pre-recorded instrumental music of *Welcome to the Show* (2019) because this music is specifically created for adaption whereas music for commercial use is created one-dimensionally, usually for radio, CDs and YouTube (Katz, 2015: 8). The music recorded in this study can be adapted for this use too but is rather aimed at enhancing the sonic experience in a theatre environment.

3.2.1. Recording Preparations

Several procedures differ in this recording process compared to a regular commercial recording of symphonic music. Firstly, the process of recording the orchestra is different because the orchestra has been reduced for the intention of digital overlaying. As a result, a sonic imbalance is created in terms of the total frequency distribution. The figure A2 in the appendix displays the layout of the full orchestra for a holistic performance of *Welcome to the Show* (2019). The colour coding used in the figure displays the general area in which the instruments fall on the frequency spectrum.[17] If this is compared to the reduced orchestra (see appendix A5), it is noticed that the ratio between the different frequency responses of instruments becomes more in favour of bass and low-mid instruments.

[17] According to Katz's (2015) musical pitch relation chart.
For more information, see: https://www.soundonsound.com/reviews/bob-katz-mastering-audio

As observed in appendix A2, the orchestra's layout is designed to have an even response of frequencies in different layers. Most of the high-frequency instruments (ignoring the percussion section) are focused in the front of the orchestra. The low-frequency instruments are predominantly placed on the outskirts, and the mid-frequency instruments mainly reside within the middle. This relatively traditional layout is used to evenly distribute the body of sound and create an immersive experience for the listeners (Anderson, 2019: v).

The purpose of recording acoustic instruments is to supplement the digital orchestra, providing a sense of authenticity. *Welcome to the Show's* instrumentation is divided into seven sections as seen in the orchestration table in appendix B1. As in typical orchestration, the string section is subdivided into first violins and second violins, violas, celli and double basses, with an amplified classical (nylon) guitar as a non-traditional addition. The woodwind and brass instruments are also divided into first and second parts. The ten percussion instruments are performed by three players. There is also one dedicated drummer who would also have a full set of chimes included on the kit setup. The last instrument is a digital keyboard which changes to different digital instruments, as indicated in the score.

The fundamental instruments are selected by necessity. In many cases, many digital instrument plugins do not completely imitate an instrument's physiognomy, such as a flute player's breathing or realistic glissandi played by trombones. This provides listeners with a greater sense of authenticity although, when re-creating this authentic experience, most of the acoustically recorded instruments do not have to be completely utilised. Instead, musicians have to focus on selected sections of the music where these particular circumstances transpire, and when these instruments must become more prominent. This prominence is in solo sections, unaccompanied underscore, or phrases where instruments must be emphasised (*en dehors*). This allows the editing process to be more efficient.

In addition, each of these selected instruments requires a maximum of two microphones each and is recorded in a group or separately. The acoustic properties of the room are analysed with regard to L_p and critical distance, which allows for a better control of the microphone placement, reverberation, diffusion and sound leakage from microphone to microphone. This is documented in appendix B3, which is used as reference in order to add this data artificially to the digital orchestra in the mixing stage. This is how the acoustics properties of the room are used to merge the different recording mediums together, producing the illusion of a

completely authentic orchestra. Furthermore, this also allows for easier spatial and dynamic audio processing during mixing.

Sections	Full Orchestra (min)	Fundamental Orchestra (max)
Strings	24	3
Woodwinds	8	4
Brass	6	3
Percussion	3	1
Drum Kit	1	0
Keyboard	1	0

Figure 3.6: Reduction of the acoustic instruments

The music sections and players are selected after analysing the musical score. Figure 3.6 displays the required players in the full orchestra compared to the mandatory players. Recording an orchestra is one thing but having all the participants available at the same time is a logistical issue (Senior, 2015: 271). In this case, some instruments are recorded separately, and samples can be sourced or created to replace digital instruments. The individual instruments that are recorded are the classical guitar, cello and oboe parts. The guitar was not recorded as part of the orchestra because it is a solo instrument which is easily overpowered by the other instruments. Adding to this, the guitar also does not have as much dynamic projection compared to other orchestral instruments. Hence, in a live setting, it would require amplification.

The cello and oboe were a result of the logistical issue mentioned above and were also recorded separately. Besides requiring more recording time, several positive attributes become apparent. Multiple microphones can be used to record these instruments for greater variation without resulting in immense sound leakage from other instruments, allowing easier micro DSP without affecting the stereo arrays (see chapter 2.4.2). The ability to micro-manipulate these instruments in the mix also provides the ability to change the stereo position of the instruments for solo sections, particularly for the guitar and cello as they have important solo sections in the music. In addition, these three instruments use the same stereo microphone configuration as the orchestra and can be merged with its stereo array.

3.2.2. MIDI: Tempo Tracks and Digital Instruments

The first practical task of recording is acquiring a digital recording of the bulk of the orchestra to formulate the guide track and tempo indication (click track) for the acoustic orchestra to perform over. This is a simple task that is performed by exporting the MIDI data from the notation software and integrating it into recording software. There are two standard types of MIDI files. Type 0 merges all data into a single MIDI file and includes all instruments, tempo changes and time signatures. Type 1 MIDI retains the original track information by exporting the individual tracks separately, allowing digital instruments to be allocated to each track (Huber & Runstein, 2014: 384).

Once all the MIDI is successfully exported, it is reimported into the DAW – Steinberg Cubase[18] in this case. After a few minor adjustments to the tempo track, a separate click track is created. This is done by associating percussive taps (clicks) to each beat. This click track is used as a metronome for the performing musicians and is an effective way to record music with efficiency and accuracy. Traditionally, symphonic music is recorded with an entire orchestra in one sitting. If mistakes are made by the instrumentalist, the recording would have to restart from the beginning or from a particular section of the music. By using a click track, mistakes can be addressed from any position in the music and retain a consistent tempo.

The type 1 MIDI file is imported into the DAW and the corresponding digital instruments are assigned to each track, thus creating the raw digital orchestra. Unfortunately, the MIDI data does not differentiate between the doubling of instruments (divisi or a2) or solo instruments. This creates less dynamic and timbre variations intended by the composer. However, instead of manually separating the instruments further, which would require more mixing and editing, the dynamic variations are controlled using the faders in collaboration with an editing tool called *velocity control*. The only exceptions to this are the keyboard, as the grand piano setting changes to a cathedral organ setting, and drum kit which makes use of different sticks (see *acoustic samples* in chapter 3.2.3).

Controlling the velocity of the digital instruments involves compressing expanding the dynamic range of a digital instrument between 0 and 300%. Velocity below 100% produces a smooth ensemble-like sound when set to a lower fader output, whereas a velocity above

[18] For more information, see: https://new.steinberg.net/cubase/

100% with less fader output creates more attack to resemble solo instruments.[19] The variations of the velocity and output faders are applied to each instrument to form the guide track. The guide track is the simplest unmixed form of the music. Further manipulations to the digital instruments are needed to create a more authentic-sounding orchestra, which is discussed in chapter 3.4.1. The guide and click tracks are exported as WAV files, ready to be used for the acoustic instrumental recordings. The updated tempo track and digital instruments with their edited velocities is also exported, as type 0 and type 1 respectively, to be re-imported at the recording venue.

3.2.3. Recording Instruments and Samples

The second stage of recording process is to record the acoustic instruments. This requires a combination of stereo arrays and mono microphones (Rayburn, 2012: 260). There are no precise rules but there are recommended methods that are followed (Ballou et al, 2008: 519).

As seen in appendix A5, the individual instruments in this reduced setting retain their approximate positions from the full arrangement. Even though this is done on a smaller scale, the reason for this is to capture the stereo image when combining the different recording media during the mixing process. During the mixing stage, mono microphones are panned to the respective positions of each instrument to correspond with a precise physical position in the stereo panorama (Rayburn, 2012: 261). However, this differs from room to room due to the critical distance and ambience found in the room (Senior, 2015: 273). It was found that the critical distance of the recording environment is approximately 1.57m, with an ambient noise of 37.95 dB and a reverberation time of 0.76 secs. Additionally, during the testing of RT_{60}, the time between the pink noise outputting and the sound being measured was approximately 0.35 sec. (attack time). These statistics are used to determine the placement of microphones, particularly mono microphones because the degree of proximity would also be dependent on the distance between the microphones and the instruments.

[19] For more information, see:
https://steinberg.help/cubase_pro_artist/v9/en/cubase_nuendo/topics/midi_processing/midi_processing_velocity_r.html

Stereophonic Microphone Placement

It will be noticed in figure 3.7 that the orchestra points towards a central position. This is done for two correlating reasons: first, that this is the position of the conductor; secondly, the position of the conductor is precisely measured as the central critical distance (1.57 m). Just behind this is where the central stereo microphone pair is placed, beyond the closest central sound source of the orchestra. With the established placement of the central stereo pair, the decision of what stereo microphone configuration to use has to be made. Due to the critical distance being strangely short and the spread of the orchestra wide, a cardioid ORTF (110°) configuration is selected instead of an XY or NOS. This is because the acceptance angle of both hyper-cardioid and cardioid XY and NOS configurations would hypothetically provide a relatively dominant forward directionality and a narrow stereo image compared to the ORTF (see *stereophonic arrays* in chapter 2.2.2.). To provide more panoramic flexibility during mixing, a flanking pair of microphones is used to alter the perspective of space (Rayburn, 2012: 232). These are placed at a 1:3 ratio from the ORTF to eliminate the possibility of phasing. However, it was found that when testing this configuration, there was still minor phasing, likely because of the omnidirectional response of the flanking pair. This was corrected by moving the flanking pair slightly behind the ORTF. This stereo configuration can be observed in figure x bellow.

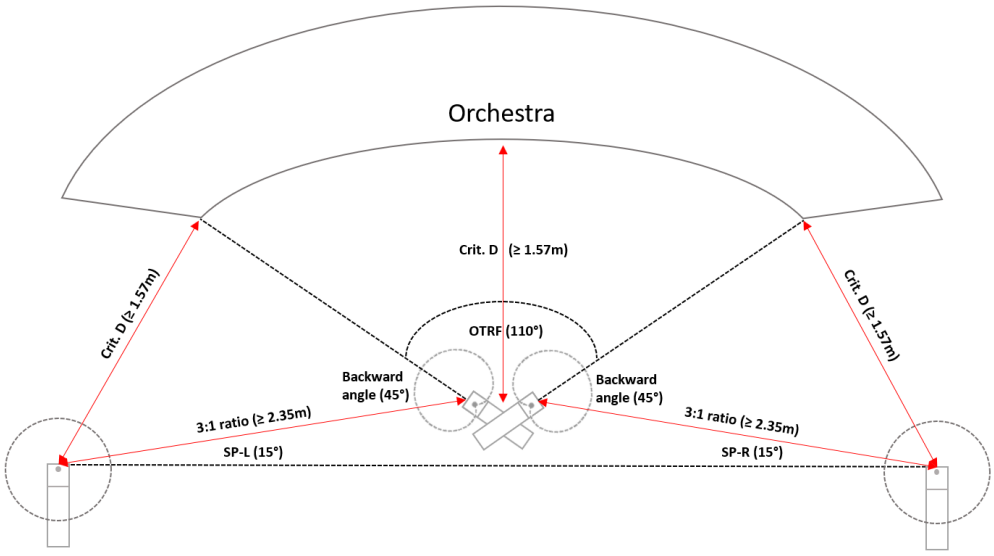


Figure x: Stereo Microphone placement

The stereo microphones are raised approximately 1 m above the head of the conductor to reduce much of the on-axis high frequencies created by the linear directionality of the

orchestra. The centre ORTF pair of microphones is aimed downwards usually at a point two-thirds the depth of the orchestra to avoid recording the majority of the front of the orchestra. According to Rayburn (2012: 264), this off-axis pattern attenuates the front of the orchestra by about 2.5 dB. Additionally, this creates a distance effect that exhibits a more natural sound (Rayburn, 2012: 265).

With the objective of this study in mind, the choice of microphones is important particularly for the stereo image. The environment in which the acoustic instrument is recorded has a very low reverberation time (0.76 sec.), and because of this, a pair of RØDE NT2-A microphones is used as the flanking pair. Although not theoretically matched, these microphones practically share fairly identical qualities as they were built at the same time (serial numbers being S/N0122717 and S/N0122718). These are set to an omnidirectional polar pattern, to capture as much of the natural reverberation as possible and to emphasise more low frequencies (Senior, 2015: 291). A matched pair of RØDE NT6s with the NT45 cardioid capsules are selected as the ORTF microphones. These are selected because they are reasonably flat with a slight accentuation around 10 kHz, although much of the high frequencies are reduced because of their downward positioning as mentioned above.

Monaural Microphone Placement

Each acoustic instrument uses one or two mono microphones placed within close proximity for this recording. This is not a normal operation for orchestral recording as it usually leads to immense sound leakage and phase discrepancies but, in the case of this recording, phasing is not an issue as the distance between instruments is much further apart, due to the retention of the instrument positions from the full arrangement of the orchestra. It could be said that each instrument is a representation of an instrument section of the orchestra. Considering that the bulk of the orchestra is digital, these microphones are used to capture the timbre of each instrument in order to duplicate their characteristics on the digital instruments, thus formulating the amalgamation between digital and acoustic instruments with regards to timbre. Furthermore, another reason for this is to allow solo sections to be authentically expressed, such as hearing the breath from a flute player, the valve noise from a horn or the plucking of guitar strings. This is a matter of subjectivity, but it is also a method to give the illusion of natural sound as digital instruments tend to sound mechanical and can't easily be masked in quieter sections.

Instrument	Microphones
Flute	Sennheiser E914
Oboe (solo)	Sennheiser MK40-P48, AKG C414 B
Clarinet – Bb and A	Neumann TLM103
Bassoon	RØDE NT2-A
Horn – F	Sennheiser E914
Trumpet – Bb and C (excluded)	Shure SM57
Trombone	Neumann TLM103
Guitar (solo)	Sennheiser MK40-P48, AKG C414 B
Viola	Lewitt MTP 440 DM
Cello (solo)	Sennheiser MK40-P48, AKG C414 B

Figure 3.7: Microphone selection

There are many facets behind the choice of microphone such as using a wide or small diaphragm microphone, or to use a dynamic or condenser microphone. The choice of microphone can only be selected from what is available and no microphone is completely appropriate for every instrument. The design, polar pattern, environmental field and proximity are all equally valid elements that are considered when selecting a microphone for a specific instrument (Owsinski, 2017b: 101). Figure 3.7 shows the intended microphones that would have been used for each instrument. However, one discrepancy is noted in this table, which is the exclusion of the trumpet. This is because the trumpeter had to be dismissed (see chapter 5.2) and is digitally replaced as a result (see chapter 3.4.1).

Besides the excluded trumpet that was initially intended to be recorded with a Shure SM57[20] dynamic microphone, the only other instrument that uses a dynamic microphone is the viola. The viola is the only string instrument that is recorded in the ensemble setting. Usually multiple violas are recorded in an orchestra but, for the intention of this study, it is recorded as an individual. As a consequence, the dynamic range of the viola is overpowered by the rest of the ensemble. However, this is resolved by placing the Lewitt MTP 440 DM dynamic microphone in close proximity to the instrument above the strings, approximately 15cm (≈6

[20] For more information, see: https://www.shure.com/en-US/products/microphones/sm57

inches). This is because bass/low-mid frequencies are accentuated within this proximity (Owsinski, 2017b: 101). Considering that the viola is a mid-range instrument and the frequency range of Lewitt is less pronounced at frequencies above 14 kHz (see chapter 2.2.2), it is the microphone that is chosen. In addition, it is a small-diaphragm cardioid microphone that, in close proximity, limits the amount of sound leakage from other instruments.

The horn and flute are both recorded using Sennheiser E914s for a similar reason. The flute's tone is easily captured as it usually generates very little transience, producing a very pure tone. The microphone is placed around 60cm (≈2 feet) above the flute pointing downwards towards the middle of the instrument to avoid key clicking (Owsinski, 2017b: 189). Brass instruments require distance to develop natural tone to reduce tongue noise and excessive air (Owsinski, 2017b: 177). The horn microphone is placed around 90cm (≈3 feet) off-axis at the top edge of the bell but pointed towards the centre of the bell (Owsinski, 2017b: 179). The similarity between the flute and horn microphone placement is that they are placed away from any other direct sound sources, with the flute microphone pointed downwards and the horn mic placed behind the player. The E914 is a small-diaphragm cardioid condenser microphone with a narrow acceptance angle and emphasised upper frequencies and flat mid and low frequencies, resulting in accentuation of the flute's brightness and allowing easy manipulation of the horn's lower range during equalisation. Much like the horn, the trombone microphone is placed in a similar fashion except the bell is wider than the horn. Thus, the wider-diaphragm Neumann TLM103 is used for its clarity and easier equalisation manipulation due to its flat low-frequency response (see appendix A1).

The bassoon and clarinet share similar characteristics in the scope of microphone placement as the sound emanates along the body of both instruments and rarely along the bell. The bassoon microphone is placed approximately 90cm (≈3 feet) away from the top half of the body (Owsinski, 2017b: 175) and, because of the length of the instrument, a wide-diaphragm RØDE NT2-A set to cardioid is used to capture its sound, which slightly accentuates the low-mid and high-mid frequencies (see appendix 1.5). The clarinet is more complex as it does sometimes resonate along the bell if overblown and usually requires two microphones to be placed; however, this is more prevalent in jazz or solo works. Therefore, one microphone is placed 60cm (≈2 feet) at the top of the instrument, pointed downwards towards the first key or "A" key (Owsinski, 2017b: 181). The cardioid Neumann TLM103 is used as it is a fairly

wide-diaphragm microphone that is fairly flat, with slight emphasis above 4 kHz that rarely affects its attack frequencies.[21]

Two mono microphones are used to record the oboe, cello and guitar. As these instruments don't form part of the ensemble, they are recorded to bring out as much of their timbres as possible. It is noticed that all three instruments are recorded using the same mono microphones, the AKG C414 B and the Sennheiser MK40-P48. The reasoning is that the wide-diaphragm C414 B has a slight emphasis on the low-mid frequencies, around 350 Hz, with a light dip around 1.5 kHz. After this dip, it has a sharply accentuated hi-mid to high frequency response. This very contoured microphone is used to capture the majority of all the instruments' resonance and attack frequencies and is a great accompaniment to MK40, which has a very flat frequency response. In addition to the stereo arrays, these microphones capture the fullest potential timbre of each instrument, with appropriate placement, allowing complete control over the sound when mixing.

The oboe has related qualities from all the other woodwinds. Like the flute, it produces a very pure tone, but it has a double reed like the bassoon, and it can also project along the bell similarly to a clarinet. Therefore, it is treated very much like the clarinet where a MK40-P48 is placed 60cm (≈2 feet) away from the instrument and pointed downwards towards the first key to capture better attack response. However, the C414 B is placed just off the bell to capture the bell's resonant response (Owsinski, 2017b: 181). The guitar uses the same microphones in a configuration similar to a spaced pair in cardioid or off-axis XY. The C414B in placed below the instrument pointed between the sound hole and bridge for the majority of the sound, while the MK40 is place 30cm away from the neck, pointed towards the 12th (octave) fret for more control over attack frequencies and fret noise (Owsinski, 2017b: 191).The cello is more of a challenge. Like a double bass, microphones at the F-holes create a booming bass without definition (Owsinski, 2017b: 171), but strangely, it is microphoned more like a guitar, having the C414 B placed close to the bridge for its resonance and the MK40 positioned at the top of the body (lower fingerboard).

[21] The clarinet range usually peaks below 2 kHz according to Katz's (2015) musical pitch relation chart. For more information, see: https://www.soundonsound.com/reviews/bob-katz-mastering-audio

Acoustic Samples

Digital instruments and samples are discussed in chapter 2.3.3. but traditionally samples are very rarely, if ever, used in classical music. Samples are either individual sounds or a collection of categorized sounds found in sound libraries, such as seen on Freesound.org[22] or SoundSnap[23], that are used to insert effects or instruments into a recording session. In the case of this recording, the majority of the orchestra is built from digital instruments and, for the most part, serves as a fairly effective replacement for authentic instruments. However, in one circumstance in this recording, a digital instrument completely fails in replicating the authenticity of an instrument.

The drum kit is a multi-phonic instrument that incorporates a wide variety of drums, cymbals, and additional-effects instruments such as windchimes or cowbells. Huber (2008: 1117) states that multiple software plug-ins and hardware drum-machines are available for a plethora of playing styles. Although this is true in most instances, it was found that the drum sampler used in this study, XLN Audio – Addictive Drums, does not sufficiently replicate the cymbal techniques used in this music. The reason is that the drumkit, besides being used in the traditional sense, is also used to create different effects (cymbal rolls and scrapes), with the additional requirement of alternative sticking options. A MIDI track in a DAW indicates the range of attributes and is represented in block form. These blocks can be manipulated to change the duration of a sustained note or to manipulate its volume with velocity control among other features (Huber, 2008: 1120). To create a cymbal roll, multiple blocks are placed in close proximity and each of these blocks uses different velocities to develop a crescendo, sustain or diminuendo (Huber, 2008: 1121). This does not only apply to cymbals, but also to tambourine, triangle and castanet tremolo, among others. To the ear, this sounds unnatural and difficult to control. A solution is to use samples to replace what the software cannot fully replicate.

Unfortunately, because the drumkit cymbals are used for the aforementioned effects, the timbre and environmental characteristics of the cymbals would have to match the cymbals provided by the software plugin and recording space. However, no sample library provides the ability to do so, therefore acoustic samples are created. Much like recording the individual

[22] For more information, see: https://freesound.org/help/faq/#what-is-this-site-anyway
[23] For more information, see: https://www.soundsnap.com/?gclid=Cj0KCQjw1MXpBRDjARIsAHtdN-0HE9DkpUFvlIDpC6UmdMY7VfEcTPy8Bdkntnx7YkbEuBg6thGN_p0aAt83EALw_wcB

instruments, the cymbals are recorded using the ORTF stereo microphone configuration as before, but no flanking microphone pair or spot microphones are used because cymbals generate immense sound pressure variations (Owsinski, 2017b: 155). The microphones are placed above and behind the cymbal player to capture the room reverberation and to avoid stick noise on the cymbals. Considering that cymbals are also high-pitched instruments, positioning the stereo microphones in this arrangement allows an attenuation of their extremely bright timbre. The player hits the cymbals with a variety of sticks (wood, wire brushes and mallets) to create the required effects. Once this recording is complete, these effects can be placed, edited and merged with the drumkit plugin (see *overdubbing, panning and profiling* in chapter 3.3.1).

3.2.4. Recording Process Summary

The above information covers several steps that are taken before the music is recorded but there are several other practices that take place before, during and after these recordings. After the preparations are complete, the recording does not immediately begin. Rehearsals and other logistics have to be organised, cables have to be laid, interfaces and DAW sessions have to be set up, and so on. All of which are small yet vital tasks for the recording process.

The microphones are plugged into the recording interfaces (MOTU 896HD[24] and Focusrite Saffire Pro 40[25]). The preamps in these interfaces and microphone cables also affect the quality of the recording among many other things such as placing reflector or absorption panels between instruments. Although the interfaces and cables do contribute to the sound quality due to latency and gain staging, they are not considered as primary elements that affect the signal in this study. In a situation where live instruments were to be used, this would be considered more thoroughly. During the recording process, sound-checks take place before recording to identify and resolve phase discrepancies and microphone adjustments, hence the change of position of the spaced microphone pair. Furthermore, the guide and click tracks need to be outputted to the performers through headphone amplifiers simultaneously but not excessively so not to be identified though the multiple microphones. Sample recordings are taken before the final takes. These are analysed in the control room to further optimise every aspect of the recording process.

[24] For more information, see: https://cdn-data.motu.com/manuals/firewire-usb-audio/896HD_Manual_Win.pdf
[25] For more information, see: https://www.soundonsound.com/reviews/focusrite-saffire-pro-40

Once the recording stages are complete, the next process begins – mixing. A major portion of the mixing stage is to be able to manipulate the audio tracks of the recorded sound. If the quality of the recording is lacking in some way, the quality of the entire project becomes susceptible to failure. This is why all these checks and safeguards are conducted during the entire recording process. (Listen to *Unmixed recording_No Drumkit* track found in appendix D)

3.3 Mixing and Mastering for Adaptive Capabilities

The process of mixing and mastering is usually done to manipulate the audio properties of sound in order to create a commercially viable product designed for a specific medium, such as compact disc. A mixing/mastering engineer has to conform to a regulated system and process the audio according to what is required of that particular medium (Owsinski, 2017a: 79). In the case of this experiment, the output of product must be able to compensate for as many acoustic environments as possible for musical theatre use. Usually, mixing and mastering are two separate entities used to create a product for a specific market and are done by different people but, in this case, mixing and mastering are both done in association to conform to the criteria set in chapter 2. Although they are still performed separately, DSP done in the mix has a massive effect on the mastering process. For example, the objective is to EQ the master to the inverted 60-phon frequency spectrum. If bass frequencies on a particular instrument are boosted beyond the 60-phon bass criterion, this would have to be amended on another instrument's frequency spectrum or in the master to create balance.

3.3.1. Mixing Adaptive Audio

A mix is a sonic presentation of emotion, creative ideas and performance but the task of mixing audio is a subjective operation, due to engineers perceiving and processing sound in their own way. However, the fundamental applications of processing audio remain the same (Izhaki, 2012: 5). The mixing carried out in this study discusses the processes in which acoustic and digital mediums conjoin to form an authentic sounding orchestra, although simply combining the different recordings is not sufficient. With a plethora of DSP tools and plugins available in today's market, there are no excuses for a semi-decent mixing engineer to develop a bad mix. However, the mix of this particular product will sound different compared to a normal orchestral mix because it is intended that the mix will reach its fullest potential after adaption takes place. There is very little in this chapter that has not been

covered in chapter 2.3 on the scope of mixing practices. This section discusses the practical implementation of the collected data and explains why certain processes are used and how they influence the music.

Overdubbing, Panning and Profiling

After the recordings are complete, the next process is to combine the different recording sessions and mix the raw audio. The first part of this process is to establish the foundation of the final digital orchestra. The guide and click tracks that were used are now null and void as they have served their purpose in the recording session. The exported type 0 and type 1 MIDI files from the Cubase session are re-imported into a new Avid Pro Tools[26] session for the implementation of the VSL digital instrument with the correct tempos and time signatures. Each MIDI track, save the drumkit, is once again assigned its equivalent digital instrument but without the addition of artificial reverberation or any other effect. The DSP for the digital and acoustic instrument will take place after the editing is complete, with the exception of stereo panning. The reason for early panning is to achieve a sense of localization (Izhaki, 2012: 182). The level of signal may rise, drop or remain constant as adjustments are made to the panning position and, as mentioned before, the VSL version used in this mix allows only for stereo tracks to be created, and the panning of stereo and mono tracks behave differently (Izhaki, 2012: 185). Due to the changes in the signal levels, some noises or unanticipated effects, such as comb-filtering, may become apparent that are not noticeable when all the signals are in a central position.

According to the panning laws, described by Izhaki (2012: 185), when two channels (stereo) are summed to mono, there is a 6 dB attenuation across both channels because half of each level is sent to its opposite (L + R = 0.5L + 0.5R), but the perceived loudness of each speaker will have a 3 dB attenuation because the 6 dB cut is divided between the two monitors. This results in level being exactly the same in each speaker – mono. The centre image still appears the same but there is a drop of 3 dB at the extremes of each speaker (Izhaki, 2012: 186):

[26] For more information, see: https://www.avid.com/pro-tools

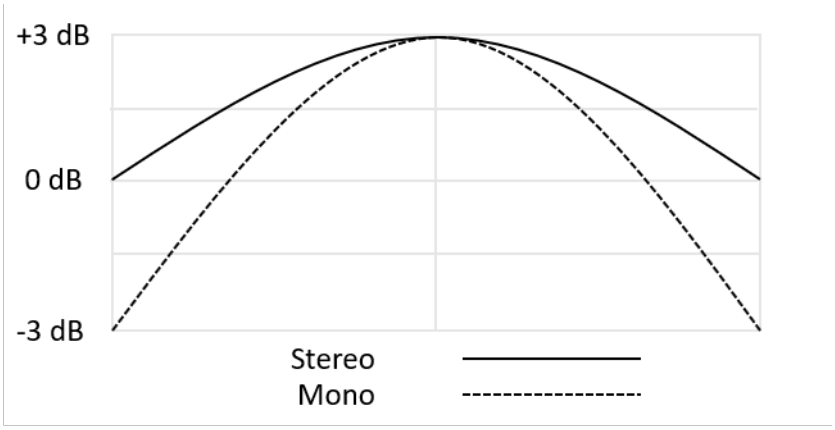


Figure 3.8: 0 dB panning law (Izhaki, 2012: 187)

Therefore, a boost of 3 dB is apparent in the centre of the stereo mix. To compensate for this 3-dB centre boost that occurs, two options can be exploited to create even levels in the stereo panorama. If the channel is panned to either of the extremes, the level becomes stabilized at 0 dB, but this is only relevant for the ORTF and spaced pairs used for the stereo microphones. The majority of individual instruments are not usually panned to either extreme nor are exactly in the middle unless the instrument in question is the focal point (e.g. voice for a song or solo instrument in a concerto). Considering that singers would hypothetically sing over this recording, there are no focal points that consistently have a centre position. Thus, the second option is to attenuate the tracks by -3 dB using a gain reduction tool, creating an even level across the stereo panorama (Izhaki, 2012: 187):

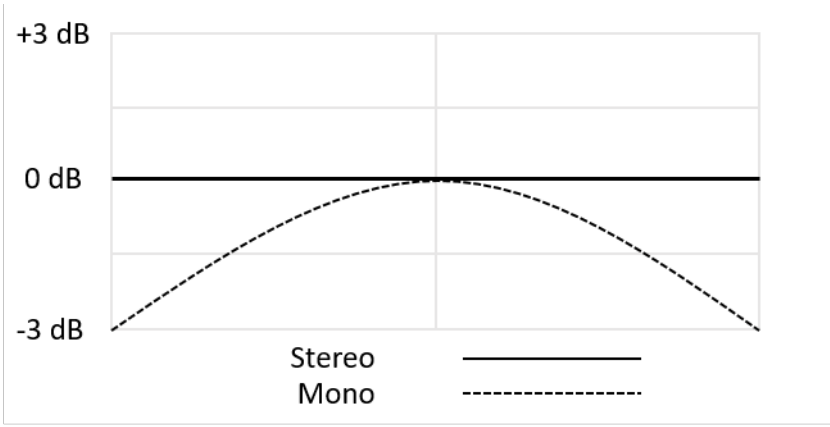


Figure 3.9: -3 dB panning law (Izhaki, 2012: 187)

Once this is applied to all the VSL tracks, they are panned to the approximate position of the orchestra observed in appendix A2. Panning left or right in the stereo with these tracks will

keep the same level of loudness in any position and does not require any subsequent fader adjustments until mixing the total sum of the digital and acoustic instruments (Izhaki, 2012: 187). Thus, the foundation of the digital orchestra is finalized, allowing for the overdubbing of the different acoustically recorded instruments. Overdubbing is a commonly used audio technique in the production of popular music. This is the layering of multiple recordings to create a track, but this is very rarely implemented in a large symphonic setting (Owsinski, 2017b: 239). The difference in timbre between the acoustic and digital instruments is noticeably different but there are steps to integrate them, which is what this study refers to as **profiling**.

Profiling, in terms of mixing, could be referred to as a type of *referencing* which is a common operation among mixing engineers in which a particular instrument or timbre from a different track (reference track) is analysed in terms its sound. This sound is then recreated by the engineer in order to be used on an instrument in their mix. In other words, the sound of the instrument is copied (Senior, 2013: 64). The difference between profiling and referencing is that profiling does not use a reference track for recreating the timbre of an instrument, but rather the raw sound of a recorded acoustic instrument. In the case of this study, the timbre of each of the digital instruments is profiled to imitate closely the timbre of its equivalent acoustic instrument. To do this, the mono channels of the acoustic instruments have to be added to the session and panned to the same position as their digital counterpart, then each sub-section (viola, horns, flute, etc.) is profiled individually by soloing them and making adjustments. Unfortunately, the plugin does not account for different styles of playing in one pre-set, such as brass playing sforzando or strings preforming tremolo. As a solution, each MIDI file is divided into multiple sections and assigned the correct performance pre-sets, and all the tracks are sent through a bus.

The main adjustment made to the digital instruments is the EQ, specifically manipulating the attack and sustain frequencies in the VSL plugin. For example: attack frequencies could be attenuated if the sound is too pronounced, or the sustain frequencies could be accentuated to create a larger resonant sound. The reason EQ is the main contributing factor for integration is because it is the vital DSP task to achieving a sonic balance (Senior, 2013: 171). Oddly enough, the purpose of profiling in this fashion is to mask the frequencies of the acoustic instrument behind the digital ensemble in highly textured sections, and vice-verse in thinly textured sections (see dynamic stability and reinforcement section below). *Frequency*

masking is a subjective perception of desensitizing a frequency region (Senior, 2013: 172). Masking is generally avoided as it affects the clarity of certain instruments in similar frequency ranges, but this is only in the entirety of a mix. Masking is used as a profiling tool to integrate each acoustic and digital instrument section in order to "blur" the individual identities of either media, creating a unison in timbre. This sectional masking does not have an effect on the overall mix when mixing one instrument section and another.

There are well over 50 VSL and gain-staging plugins running simultaneously at this stage and it is clearly noticed that the CPU and RAM are taking a lot of strain. Considering that the use of these plugin tools is complete, each digital instrument bus, acoustic instrument and stereo track is bounced out of Pro Tools as WAV files to be reimported into a new session. However, before further editing in the new session can take place, the drumkit has to be added. The Addictive Drums plugin is copied from the first Cubase session and placed in the new session. The plugin itself contains its own separate panning, microphone placement and additional effects settings. Panning the drumkit is unique compared to other instruments as different parts of the drumkit can be localized in different locations of the stereo panorama, but this is standard practice. What makes this different is the addition of acoustic cymbal samples. The acoustic cymbal (17" Sabian Vault Series[27]) is summed to mono, panned and profiled to integrate its positioning and sound to the stage-right digital cymbal (17" Sabian AAX Studio Series[28]). The overhead microphone sound of the digital cymbal is removed but remains in the room microphone (see *editing and restoration* section below). Once the profiling of the audio is complete, the next stage of mixing, or rather editing, can commence.

Editing and Restoration

There are two types of editing: selective and corrective. Selective editing is the task of selecting the best sections within the different recording takes to create a composite master take (Izhaki, 2012: 33). In this case, the orchestra and individual instruments were recorded three separate times and each take, including the stereo arrays, have their own qualities where certain sections in one take sound better than in another. The best takes from each section are trimmed and combined creating the master take. The output level of the ORTF and spaced pair stereo arrays of the orchestra and individual instruments are adjusted to sound as if

[27] For more information, see: https://sabian.com/en/home
[28] For more information, see: https://sabian.com/en/cymbal/21706x-17-inch-aax-studio-crash

recorded together and merged to form one stereo array. Additionally, the drumkit's room microphones are artificially recreated to imitate the acoustic criteria of the room (appendix B3), thus blending it into the stereo mix. By leaving the digital cymbal's room microphone sound in the mix of the plugin, the summed effect from the more pronounced acoustic cymbal creates the illusion of spatial authenticity without any sonic gaps in the stereo panorama. Once the best takes are selected, corrective editing takes place. This is done to repair a bad performance, but this is limited to only the individually recorded instruments. In the orchestra recording, sound leakage from other instruments is also affected by corrective DSP, and sounds very mechanical. Over-correcting has a quality penalty and should be used sparingly if at all (Izhaki, 2012: 34).

Both selective and corrective editing are technical editing tasks, but additional technical editing processes also take place to improve the raw audio before creative editing takes place. These tasks include deleting unused audio, such as the takes that are not used, and grouping instrument sections together (brass, string, woodwinds, etc.), but the main jobs are removing unwanted noise and audio restoration. The former of these two involves removing undesirable noise, such as page-turning noise, by gating, trimming and cross-fading sections. Restoration involves using plugins such as de-noising or de-clickers, all of which are types of IIR filters (see chapter 2.5) that are used to eliminate as much external noise as possible. However, restoration uses filters to remove sound which also means some fundamental frequencies could be lost as well and, like correction DSP, are also be used tentatively (Izhaki, 2012: 37).

Inverted 60-phon Spectral Application

Mastering the ability to distinguish frequencies in a recording and manipulating them for the betterment of sound quality is one of the greatest skills a mixing engineer should aspire to achieve (Izhaki, 2012: 202). The equaliser tool is not merely used for practicality but is the primary instrument that creates interest and definition in a mix. The different frequencies in the spectrum have an effect on the human perception of depth. Low frequencies are not as linear compared to high frequencies and are more susceptible to sound transference. The brain perceives this leakage of low frequency as farther away, and this phenomenon is what is employed to create spectral depth after adaption has taken place (Izhaki, 2012: 204).

Considering that the frequency transference is used to create depth, in an acoustic environment, this transference is where the noise floor (ambient sound) develops. This can be observed when comparing the different EQ charts in appendix C. The low frequencies of the master track have a gradual decline in sound pressure to 0 dB whereas these frequencies in all the acoustic environments display a significant increase in L_p. The reason for the master track having this roll-off of low frequencies is because of the equalisation methods used. As previously discussed, an inverted 60-phon Fletcher-Munson curve is used to EQ the different tracks to achieve a flat frequency response over the entire spectrum. However, this EQ response is finalised during the mastering process, but mixing each instrument is done in retrospect to the final outcome. This is because a mix consists of individual elements but each element's sound in contrast to another is what contributes to the overall mix (Izhaki, 2012: 24). Simply put, different instruments have their own sonic qualities; a flute has a very pure tone whereas a triangle does not and cannot be mixed in the same way. Thus, applying the linear inverted Fletcher-Munson EQ curve to each instrument in counterproductive in terms of sound quality. As a solution, each instrument's frequency response is mixed according to its octave band relation.

As discussed in chapter 2.1.3, there are four main octave bands (lows, low-mids, high-mids and highs), but these can be further broken down into smaller categories that contain their own general characteristics. Subsonic (0 Hz – 20 Hz) and low bass frequencies (20 Hz – 60 Hz) are considered as bands that are felt rather that heard. Subsonic frequencies are inaudible to humans and, to some extent, so are the low basses but low bass frequencies do possess some of the fundamental frequencies (lowest sound) of the double basses, kick drums and the lower range of the piano that provides power rather than pitch definition. However, mid-bass frequencies (60 Hz – 120 Hz) are also associated with these instruments and also provide this sense of power but are more perceivable in tonality. The upper bass frequency range (120 Hz – 250 Hz) is probably the most important because it is the fundamental range of most instruments, where the natural timbre of these instruments lies. The low-mid frequencies (250 Hz – 2 kHz) contain most of the low-order harmonics that affect the colour (pitch integrity) and broadness of various instruments that showcase their resonant timbres. Likewise, the high-mids (2 kHz – 6 kHz) contain complex harmonics, integral to the definition, loudness and presence of human perception because our ears are very sensitive to this range of frequencies. The last range is the high frequencies (6 kHz – 20 kHz). Although, this range

contains very little sound energy, it is responsible for bringing brilliance, sparkle and air into the mix (Izhaki, 2012: 209).

Each instrument's relative frequency response (see appendix A2) is an indication to where the majority of the frequencies and overtones are found in the octave bands. The attack and sustain frequencies fall within these bands and are adjusted using different filter options. There are several common varieties of EQ filters, such as high and low pass shelving filters, brick-wall filters, parametric filters, and so on, all of which are used to attenuate or intensify different sections of the audio spectrum. This is common knowledge among mixing engineers, but there is one secondary DSP tool besides the equaliser that can significantly alter the subjective timbre of an instrument. A distortion tool, or harmonic exciter (iZotope Ozone[29]), is a DSP plugin that produces a different type of energy to a band region. This can create new frequency components that do not initially exist or can emphasise a spectral region that does exist by increasing the harmonic density (Senior, 2013: 191). Harmonics are integer multiples of the fundamental and, by increasing the intensity of harmonics, so too we enhance the definition of sound (Izhaki, 2012: 447). In the case of this study, the digital strings and trumpets make use of subtle distortion to accentuate the attack power of these instruments as they don't appear to conform to the desired tone with the use of EQ.

The figure in appendix A4 graphically represents what the desired EQ levels are in the spectrum. However; with the loudness range of the orchestra being very wide, these levels are impossible to consistently generate at any point of time as the spectral processing in mixing affects the spectrum of individual instruments or instrument sections, and each instrument has its own frequency response and does not always perform at the same time. Therefore, to conform as closely as possible to the inverted 60-phon Fletcher-Munson curve, the *average signal level* is analysed over time and equalised to the curve. Thus, spectral processing can also be considered a dynamic process because, as seen in appendix A4, the average signal level is the amount of intensity of a frequency band. Furthermore, the spectral processing is further enhanced for its conformation when fine tuning the entire mix during mastering.

[29] For more information, see: http://help.izotope.com/docs/ozone/pages/mod_harmonic_exciter.htm

Dynamic Stability and Reinforcement

The average signal level is the overall loudness of each instrument, or instrument section, residing in the mix. Logically, if the signal for any instrument is too low, it gets lost in the mix, and if an instrument is too loud, it stands out. Both extremes cause an imbalance in the mix. The average signal level should be balanced in order to have each instrument or instrument section clearly audible in the mix but restrained enough to retain an authentic dynamic range. This is because, even in music with a wide dynamic range such as orchestral music, the quietest moments can be lost and the loudest moments can be unbearable (Huber & Runstein, 2014: 494).

The initial tools used to create dynamic stability are the output faders. The digital fader in a DAW multiplies sample intensity by a coefficient value. For example, doubling (+6 dB) the L_p of a signal will be a coefficient of 2, and halving (-6 dB) the L_p is a 0.5 coefficient value (Izhaki, 2012: 176). However, when levels are set to a particular dB level, they do not hold the same dynamic intensity (gain) throughout an entire performance, especially in orchestral music because dynamic movement is what creates interest and development (Izhaki, 2012: 470). It might be found during the recording stage that some elements are overly pronounced or others too quiet. This is corrected with fader automation which adjusts the output levels at certain points, such as a flute solo or acapella section, to allow parts to stand out or to create evenness between sections when singers perform with the tracks. This is very prevalent in this case because, even though the digital instruments have velocity control, the dynamic range of one section has an effect on another. Some sections would require more brass and fewer woodwinds, and in others less brass and more strings. This is addressed in the mix by automating each collective section bus (brass, strings, woodwinds, etc.). In addition, automation can also be applied to other areas of the mix such as the EQ, reverberation or panning positions of instrument (see *spatial enhancements* section below).

In every piece of music, there are emotional peaks and troughs that build to the next climax which should be supported by the mix. The simplest structure of a pop song usually follows the following construction (Senior, 2013: 119):

Verse 1 → Chorus 1 → Verse 2 → Chorus 2 → Midsection/Bridge → Chorus 3 → Chorus 4

The choruses would be where the dynamics would be at their peak, especially the last two choruses having the most power. The subject material that is recorded in this study is far more complex than this, having three movements (three parts), each with its own mood and expression. Each movement flows into the next, creating an entwined 11-minute piece However, the construction of the sub structures follows the same principles as a typical pop song but with the addition of linking sections (introductions, transitions and codas), subdivisions (sub) and climaxes.

1st Movement; *Prelude*:

Intro → 1st Sub → 2nd Sub → Transition/Bridge → 1st Sub (variation) → Climax/Bridge →

2nd Movement; *An Image Defined by Face*:

2nd Intro → Verse 1 → Chorus 1 → Verse 2 → Climax (Chorus 2) → Bridge →

3rd Movement; *Epilogue*:

1st Sub → Chorus 1 → Transition/Bridge → 2nd Sub → Climax → Coda (Chorus 2)

Compared to a regular pop song, there are more peaks and troughs in the dynamic range in this music, but the idea of dynamic changes remains the same where verses and subdivisions are not as loud as choruses and climaxes. Transitions and bridges can have varying degrees of dynamics, depending on their contents.

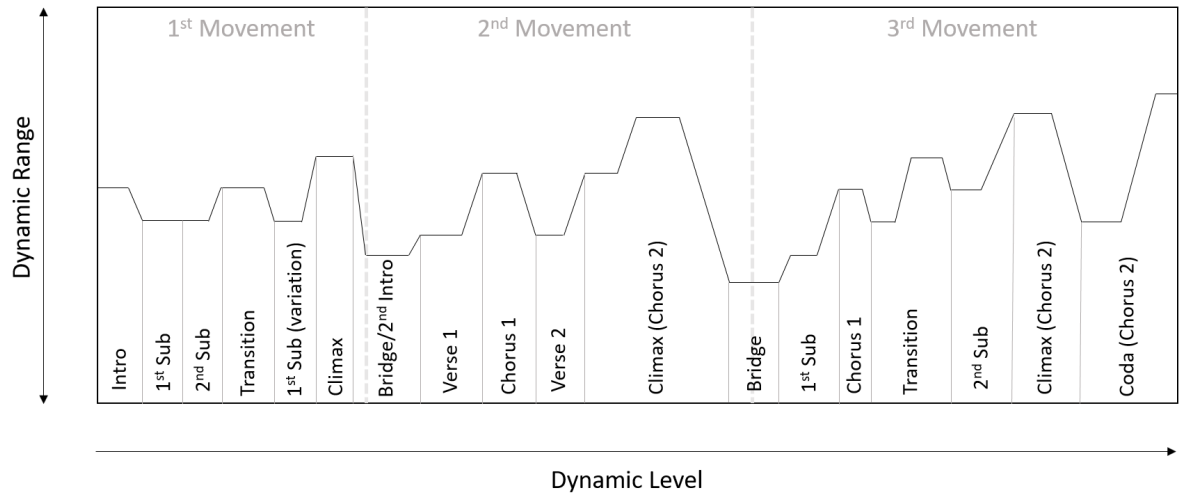


Figure 3.10: Dynamic schematic of subject material

Figure 3.10 displays a rough dynamic contour of the music. EQ, input gain, fader control and automation are key components that make up the dynamic stability of the mix but to some extent other dynamic controls must be used. Compressors and expanders (see chapter 2.3.2) are closely related to automated faders in terms of what they try to achieve. The only difference is that these DSP tools perform balancing changes automatically (Senior, 2013: 143). In spite of this, compressors and expanders should distinguish tightly controlled dynamics yet keep the authentic sound of the instruments (Izhaki, 2012: 266). Each compressor, be it analogue or digital, has different sound qualities. Some add power, others are transparent, some are better for drums, while others are better for vocals. The compressor used for this mix is the iZotope Ozone Dynamics control which is a precise compressor/expander/limiter digital plugin that has a consistent starting point in relation to its threshold, uniformity in signals and automatic crossover frequency points.[30] All of which is what is needed from a precision dynamic tool (Izhaki, 2012: 269).

Dynamic tools do not need to be used on everything, but no musician can maintain perfect dynamic balance (Senior, 2013: 147). Considering that the mix in this study makes use of digital instruments, dynamic tools are used to profile them to the natural dynamic range of the acoustic instruments. At the loudest moments in the music, there should not be any over-compressed sound limiting. The objective is to use compression transparently in these loud sections and to reinforce the quieter sections by expansion. This allows the dynamic range to narrow but still maintain the effect of the changing levels in each section of the music. The acoustic instruments also have their own automated dynamic processes which "activate" during solo sections or in thinly textured sections where they become highlighted (*en dehors*). An example of this thinly textured section is in the introduction of the second movement where the first cello part (*en dehors*) is an acoustic instrument that plays the melody while the second cello (divisi) is digital and plays a low drone with the acoustic oboe.

Acoustic and Artificial Temporal Processing

Thus far, reverberation time has been a large topic of discussion. This is because it is most likely the most important factor for adaption capabilities. Reverberation could be considered a digital balancing act; it does not matter if the EQ or dynamic range of the music is of the

[30] For more information, see: https://www.izotope.com/en/products/master-and-deliver/ozone/features-and-comparison/dynamics.html

highest standard, if the level of reverberation is underwhelming or oversaturated, the product will not adapt to an environment because reverberation is the primary tool that changes the depth and mood of the music (Izhaki, 2012: 403). Furthermore, time gaps in the music and timbre changes on instruments take place when altering the temporal time of a recording that, if used correctly, fills the stereo panorama (Izhaki, 2012: 404). Both acoustic and artificial reverberation exist in this mix. The acoustically recorded instruments already contain natural reverb from the environment in which they were recorded. The digital instruments had all their effects removed when they were implemented including the reverberation, but now that the prior DSP tasks have been completed, it can be artificially reinstated into the mix.

Before adjusting the temporal properties of the music, the potential playback environments are considered. Even though the final mastered track is not subjected to a very large pool of playback environments, the temporal expectations of the playback environments allows for temporal preparation of the mix and master. For example, if a large environment has minimal L_p and sound transference, it will likely have minimal reverberation and is considered "dry" which would indicate to the engineer that additional reverberation should be added. An environment is considered "wet" if it has considerable reverberation and is likely prone to extensive sound transference, indicating that there was too much reverberation added to the mix or master. Four environments are tested (see appendix B2). The first is the studio control room, which is an acoustically treated environment in which sound can be easily controlled as it is designed for mixing. The other three environments are theatres with different size variations and constructions:

- **The Masker** (small theatre): approximately 80 front seats, spoken-drama theatre – proscenium open stage design with a built-in sound reinforcement.
- **The Musaion** (medium theatre): approximately 400 seats, classical concert hall – open auditorium design with no built-in sound reinforcement (portable PA system).
- **The AULA** (large theatre): approximately 1000 seats, general concert hall – proscenium arch design with orchestra pit and built-in sound reinforcement.

(Acoustic statistics of these environments can be observed in appendix B2)

Each environment has its own reverberant field, be it dry or wet, and acoustic characteristics that add to a signal subjected to the space, including ambient frequencies. As discussed in chapter 2.3.2, mid frequencies are key components in balancing a mix because they contain the majority of the lower-order and complex harmonics (Izhaki, 2012: 209). Artificial reverberation is used to compensate for this influx of additional sound by reconstructing the mid-range of the instruments during the mix. This is most prevalent in the stereo array (OTRF, space pair or drumkit room microphones) because the mid frequencies become over-exaggerated due to sound leakage and compression. These frequencies, around the 350-500 Hz range, are addressed by attenuating (filtering) them and reconstructed using artificial reverberation (Izhaki, 2012: 405).

The natural reverberation time in the recording environment is approximately 0.75 seconds, and it was found that attenuating the L_p of this range by approximately 4 dB was the most effective reduction level for this mid-range exaggeration. An artificial hall reverberation is added to the stereo mix using the iZotope Ozone Reverb plugin at the same early reflection/attack time (0.35 seconds) and decay time (0.75 seconds) as the room but with more intensity (100% dry : 36.8% wet). The dry/wet function controls how much reverberation (wet) is added in comparison to the input signal (dry). The dry signal is boosted by 36.8% wet signal, approximately 6 dB, with hall-type reverberation which imitates a large acoustic space (relevant to theatre acoustics), having darker and warmer reflections.[31] Due to this, the reverberation time is slightly longer (0.95 seconds) and appears to have a more dense sound, but is still less than 1.4 seconds.

The different artificial reverb settings are saved as a pre-set in the iZotope plugin and applied to the rest of the instrument busses with varying degrees of dry intensity because the dry-to-wet reverberation ratio allows the different instruments to appear as a natural part of the mix, merging them together (Izhaki, 2012: 402). However, some instrument sections are less dense than others, mainly the strings. This is because of the lack of doubling in the mix compared to the original orchestration. The strings are originally orchestrated for 12 violins and 3 basses. Due to the reduction and digital implementation, this doubling effect is lost. To recreate the balance, 0.25 seconds of mono delays are added to these instruments and, like reverberation, with a 36.8% wet signal and a 0.95 second feedback. The final reverberation time is achieved

[31] For more information, see: http://help.izotope.com/docs/ozone/pages/modules_reverb.htm

during the adaption process because the playback environments also contain reverberation that compounds the reverberation from the track when played into the environment to form the collective reverberation time.

Spatial Enhancements

The last topic of discussion is spatial enhancement. Although the majority of this has been discussed in the overdubbing, panning and profiling section above and spatial processing in chapter 2.3.2, some smaller tweaks to this mix's stereophonics are also made.

Panning the tracks to their designated position is an effective way to create a wide stereo image (Senior, 2013: 262) but, generally speaking, orchestral music tends to remain in these static positions in a live recording to remain authentic. In the case of *Welcome to the Show* (2019), the instruments do not have to stay in these positions. This is an attempt to create interest in the mix. Because most of the instruments are digital and others recorded separately, they change their position in the mix by use of automation. The guitar, being an important solo instrument in the entirety of *Welcome to the Show* (2019), has many sections in the piece that require it to stand out, particularly in the second movement. Besides becoming louder in these sections, it changes its position from a slightly off-centred right position to a centre position. Likewise, the solo cello changes its position from a severe right position to a centre position and back again when less prevalent.

Other instruments follow this idea such as the piano and xylophone chromatic scales moving in contrary motion from both extremes to the other only to return to their original positions close to the end of movement 1 and after the first verse in movement 3. These are all enhancements that do not regularly occur in recorded orchestral music, but this creates subtle interest in the mix without doing a disservice to authenticity. There is also the traditional spatial enhancement to further widen the stereoscopics, such as the M&S stereo enhancer (Senior, 2013: 262), but considering that there is already a very wide stereo spread, it is unnecessary. Although, some harmonic and dynamic enhancements are made during mastering that affect the spatial perception due to an increase or decrease in frequency or dynamic L_p.

3.3.2. Mastering adaptive audio

Mastering audio is the last creative step of an audio production where the final mix is processed to complete the final product (Katz, 2015: 1). The process of mastering audio is a chain of comprehensive procedures that requires precision and objective attentiveness because the master tracks are the final tracks that a potential audience will hear. Any minuscule adjustment in the mastering chain may be detrimental or supportive to the final product depending on the extent of audio techniques applied to the master tracks (Katz, 2015: 12). It was mentioned in the introduction of this chapter that mastering and mixing go hand-in-hand in this particular production. In this final stage of production, the mix is fine-tuned in dynamic levels and EQ balance along the inverted 60-phon frequency curve. The task of mastering follows a similar protocol to that of mixing in this study where the practical implementation of DSP is drawn from the theoretical principles discussed in chapter 2.4.

The main artistic and technical tasks associated with mastering are dynamic DSP, global equalisation, noise reduction and steady levelling (Katz, 2015: 10). In the case of this study, the audio is analysed in order to explain the DSP with reasoning in the scope of adaptive sound mastering. Although not used for the purpose of this study, all the tracks, from the raw recording to the final master and adapted tracks, are converted to an MP3 format and embedded with ISRC and metadata in preparation for distribution (Owsinski, 2017a: 17). The ISRC and metadata do not affect the sound, but they are added nonetheless as it is a recording which should be protected with copyrights, but this is one small part of mastering. The link to snippets of these audio tracks can be found in appendix D. These are used for comparison and to listen to their development.

Spectrographic analysis and restoration

The reasoning behind audio analysis is that it identifies what adjustments are needed in the mix. The mix was done in preparation for mastering, thus there was no extensive equalisation done to the tracks to avoid signal spikes in the frequency spectrum. In addition, very little compression was implemented in the mix and, as a whole, it was recorded at a low level of gain input to emphasise the natural dynamic range and to leave dynamic headroom available for the mastering process (Owsinski, 2017a: 31). Analogue signal is continuous, but music found in the digital domain is not. Dither is a DSP tool that allows digital music to be quantized to a fixed bit rate value (Katz, 2015: 199). A data compressed file (MP3) with a lower bit rate is a smaller file and can be downloaded faster, but at a slightly lower quality as

a consequence (Owsinski, 2017a: 113). Although these tracks are only for progression listening and not for the adaption experiment, all the mastering processes are done in high-resolution (96 kHz/24 bit) as dither is not a requirement in this particular production. However, the track progression tracks are dithered (48 kHz/24 bit) for easy online uploading and downloading.

The audio statistics and EQ graph seen in appendix C display the data of the raw mix. With regards to the stereo levels, it can be seen that there is a slight difference of 0.5 LUFS between the average outputs of the left speaker and right speaker. Although it is not uncommon to have such occurrences in orchestral music, the level difference may have occurred from different L/R gain-staging levels of the stereo microphones or from emphasising the level of instruments in the stereo panorama. Nonetheless, the average level of each speaker is adjusted to have a 0 LUFS difference to balance the average power level. By creating balance between each speaker, the average output can now be adjusted to the average integrated dynamic range for stage music which is approximately -14.29 LUFS (Owsinski, 2017a: 61). However, it was found that the change in the left/right dynamic level had a slight effect on frequency balance during playback because of the particular EQ applied to the master track. Therefore, the average dynamic range is fractionally increased to -14.1 LUFS to amend this issue.

As far as amending other sound issues in concerned, very little restoration DSP is applied to the master track because most of it took place in the pre-mix editing. However, there is one discrepancy that is noticed in appendix A6. The collective signal from all the microphones emphasize noise in a silent section of the music but this is easily reduced by reducing the input gain and crossfading the sections together. This is further corrected when implementing filtering EQ.

Inverted 60-phon frequency range reinforcement

The EQ graphs in appendix C display the average level of frequencies between different octave bands, but it is noticed that the EQ in every case does not resemble the inverted 60-phon Fletcher-Munson curve. This is because this equaliser, iZotope Ozone EQ[32], is a digital

[32] For more information, see: https://www.izotope.com/en/products/master-and-deliver/ozone/features-and-comparison/eq.html

parametric equaliser, and these types of equalisers are active and have several second-order sections (McManus, 2008: 801). Considering this equaliser is digital, its filter coefficients are stored and recalled from memory and modelled off an analogue equivalent (McManus, 2008: 802). Active equalisers compared to their passive counterparts have the ability to boost frequencies. Passive equalisers only cut frequencies and require post-amplification for another device to level the sound (McManus, 2008: 794). The main reason the EQ curve does not resemble the inverted curve is that it the iZotope equalizer imitates a second-order L-type network in its algorithm which rolls-off 12 dB per octave resulting in what resembles a low-pass (weighted) FIR filter creating an arch-like image (McManus, 2008: 788):

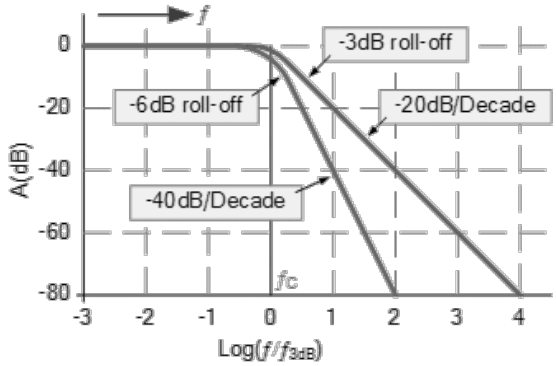


Figure 3.11: second-order filter[33]

The foundation of the inverted curve is first established when mixing. To reinforce the curve, it is implemented in the mastering EQ with some degree of variation. It is obvious that this curve utilizes a high-pass filter. Apart from reducing the booming frequencies in this range, it also acts as a filter that removes much of the undesirable noise and creates a smoother arch.

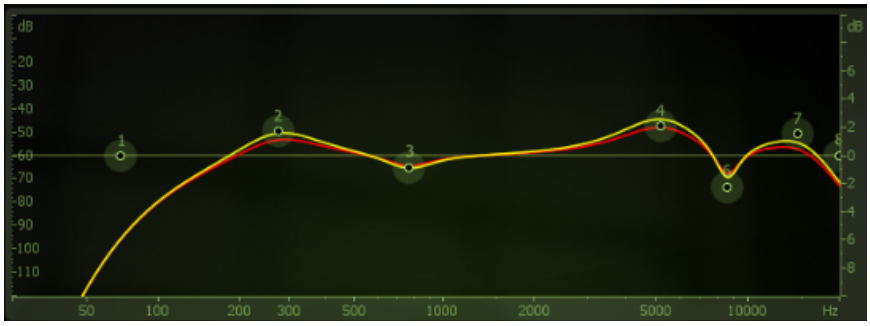


Figure 3.12: Implemented inverted 60-phon Fletcher-Munson Curve

[33] For more information, see: https://www.electronics-tutorials.ws/filter/second-order-filters.html

Unfortunately, the downside to a digital equaliser is that it creates a "sterile" sound, having very little tone colour or character (Katz, 2015: 66). A solution to this is to add subtle distortion to the master file. The iZotope Harmonic Exciter allows for different distortion modes over different octave bands which are the same as those used in multiband compression (see dynamic signal processing section below). This particular distortion tool gives shine to the upper frequencies, boosts the presence of low frequencies and adds warmth to varying areas across the four bands. The different modes have different sonic qualities, but the point of this study is to maintain a sense of authenticity. Therefore, the Tube mode is selected because its modelling is characterized by clear tonal distortion that emphasizes dynamics and transient attack frequencies found mainly in the low-mid and high-mid frequency range.

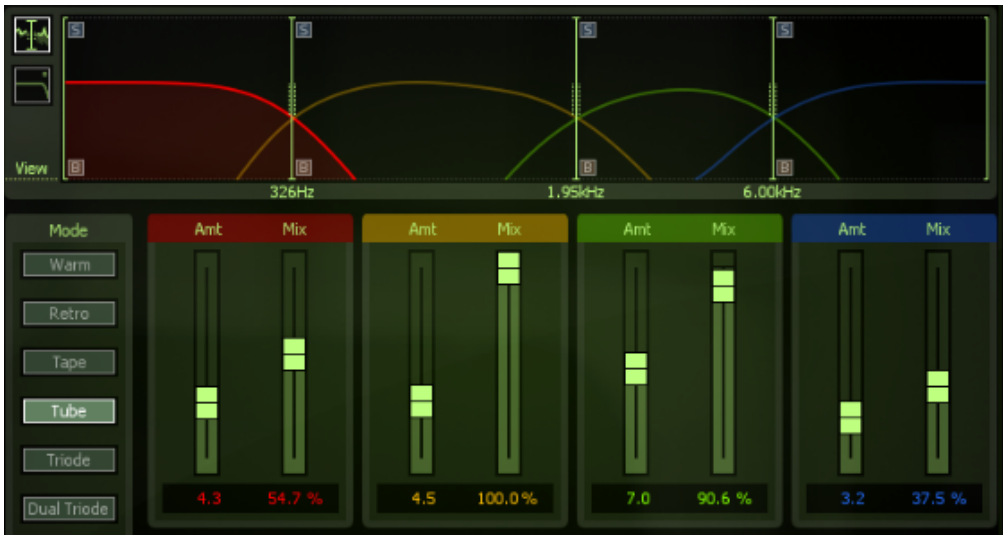


Figure 3.13: Harmonic excitement of the frequency bands

The "amount" scale shown in figure 3.13 controls the intensity of harmonic distortion in the selected band whereas the "mix" scale controls how much of the distortion is added to the original signal.[34] The ratio between the amount of transparent tube distortion and the percentage of distortion added to the mix emphasises the fundamental and complex harmonics in these bands resulting in stronger attack but warmer sustain frequencies (compare track 03 to track 04 in appendix D). To finalize the levels of the EQ, each of these aforementioned bands are processed in terms of dynamics.

[34] For more information, see: http://help.izotope.com/docs/ozone/pages/mod_harmonic_exciter.htm

Dynamic Signal Processing

The two main DSP processing tools for dynamic control are the multiband compressor and a maximizer/limiter. As mentioned in chapter 2.4.3, a multi-band compressor is similar to a standard compressor except it subdivides the entire audio spectrum into four bands (Owsinski, 2017a: 47). In this case, the crossover points are 326 Hz, 1.95 kHz and 6 kHz. These correlate with the harmonic exciter crossover bands discussed previously. These crossover positions are selected because it was found that they complement the EQ curve and the level of harmonic excitement allowing better control over these dynamic ranges.



Figure 3.14: Multi-band compression on master track

The iZotope Ozone Dynamics[35] is used for multi-band compression. The different controls can be observed above with its graphic representation in figure x. It can be seen that each frequency range is compressed, gated and limited to some degree, but the high-mid frequencies (green) appear different from the others. This is because it uses a "gated-expander-compressor", or as Izhaki (2012: 361) puts it, a *compander* that is used essentially to undo a compressor by re-expanding it and has a different tonal effect compared to the others. It is apparent that the mix, especially the digital instruments, does not contain an effective high-mid response related to the inverted 60-phon curve even with harmonic excitement. However, the use of the expander in this range also emphasised many of the

[35] For more information, see: http://help.izotope.com/docs/ozone/pages/modules_multiband_dynamics.htm

complex harmonics, which is why a gate is used to reduce much of the acoustic and digital noise created from this expansion (Izhaki, 2012: 334). Furthermore, the compressor is used on top of the expander to regulate the output of this timbre effect.

Each band is subjectively manipulated to smooth the arch-like appearance of the inverted 60-phon curve. It can be seen in appendix C2 that the use of the multi-band compressor also effects the EQ intensity. The graph shows the average level (lines) have a smoother transition between the different subdivided octave bands. However, the dynamic range has not been adjusted to the ideal dynamic range, but it can be assumed that increasing the level of intensity will also affect the EQ balance. Therefore, the final dynamic range has to be established first before the final EQ, and this is done using a maximizer/limiter DSP tool.

For all intents and purposes, a limiter and maximizer are the same thing. The maximizer/limiter is the principle tool used to achieve high dynamic levels. It functions like a compressor but with a very high ratio with an extremely fast attack (Owsinski, 2017a: 48). The iZotope Maximizer, like many other maximizer/limiters, prevents signal from peaking by implementing a ratio of 10:1 above the threshold (Huber & Runstein, 2014: 502). The threshold determines the point at which sound is limited, and the further the threshold is reduced, the more signal is maximized. The "margin" in the iZotope Maximizer determines how much signal is boosted after passing the threshold. The lower the margin, the less signal is allowed past, almost like a second limiter, hence maximizer/limiter.[36]

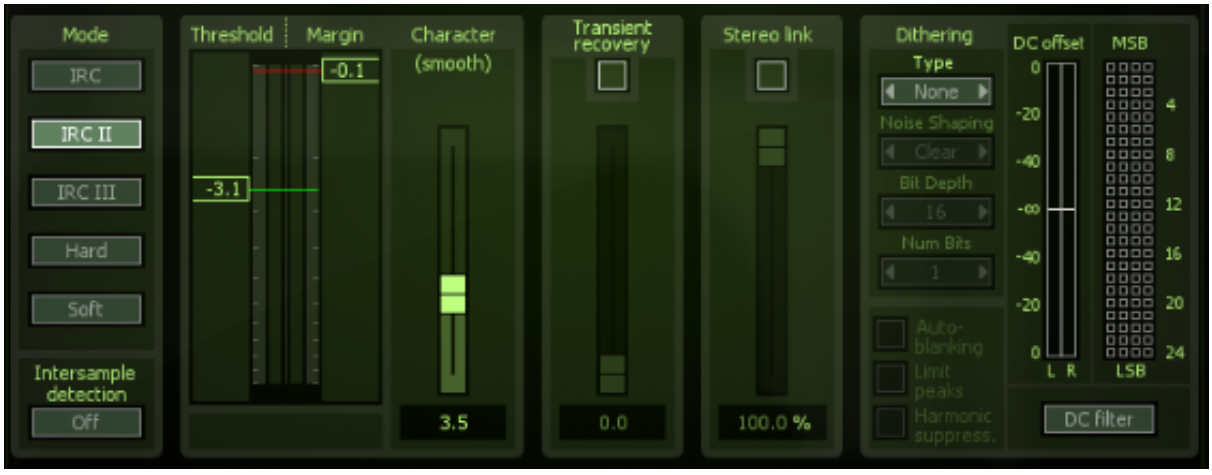


Figure 3.15: Maximizer/limiter used on the master track

[36] For more information, see: http://help.izotope.com/docs/ozone/pages/modules_loudness_maximizer.htm

Figure 3.15 shows the threshold limit and margin that is used to maximize the level of the master track to an integrated dynamic range of -14.1 LUFS, which is observed in iZotope Insight. The way in which the maximizer/limiter is used is to have a natural wave form with no over-limiting that cuts peak levels which can be observed in appendix A6 with the peaks reaching -0.1 LUFS in both the left and right stereo outputs. It can also be seen that there are different modes that can be selected. iZotope Ozone's Intelligent Release controls (IRC) are controlling loudness maximizing options that allow for different degrees of limiting. IRC III is not available in high-definition mastering and the IRC I and soft mode were found to be too transparent while the hard mode oversaturated the lower dynamics. Thus, the IRC II mode is used as it preserves transience, providing a clearer and sharper output signal even with aggressive limiting. Furthermore, there is a "character" scale provided with the IRC control that controls the speed in which the limiter releases.[37]

Post-Mastering DSP Applications

The master track is nearly complete but, due to limiting, the dynamic range has been affected. The quieter sections in the music have become too loud. Furthermore, the arch-like appearance of the inverted Fletch-Munson curve has also been affected with a slight boost around the 700 Hz area as well as a cut at the 6 kHz. A post-EQ is available in iZotope Ozone and these frequencies are cut and boosted accordingly to revert the affected frequencies. The dynamic range is also amended by using fader automation, after which an adjustment of the maximizer/limiter is done to revert the integrated dynamic level back to -14.1 LUFS (Appendix A7 is the final level of threshold and margin).

3.3.3. Mixing and Mastering Summary

Through the entire mastering process, it has probably been noticed that there are no temporal adjustments made to the master track. This is because it is unnecessary to add additional reverberation considering that, when these tracks are performed in the playback environments, the room reverberation will compound the reverberation of the track. Artificial reverberation can always be added during adaption but not taken out. As Katz (2015: 1) puts it, "Sometimes the only process we need to do is – nothing!"

[37] For more information, see: http://help.izotope.com/docs/ozone/pages/modules_loudness_maximizer.htm

This concludes the creation of the physical audio product. The link to the tracks in appendix D may serve as reference to a listener to understand the progressive changes from the raw recordings to the complete adapted sound. In the next chapter, the audio product is exposed to the playback environments to prove this theory of adaptive sound, where the music is tested in the multiple theatres and adjusted to re-conform to the inverted 60-phon Fletcher-Munson curve (arch).

Chapter 4: Adaption Testing and Discussion

4.1 Acoustic and Audio Testing and Comparisons

The experiment of adaptive sound now takes place but, similar to the discussion in chapter 3.2.4, the audio statistics of the master track are not the only variables that need to be considered. In the three different theatre environments, measurement microphones are placed in various positions to "recapture" the master track in order to be analysed and adapted. Besides the microphones, this recapturing requires various hardware and software components, all of which have an effect on the sound in the signal. Furthermore, the acoustic characteristics of the environment also form part of this signal chain whereby the master track's audio characteristics are most affected by the acoustic space. All these factors play a vital role in adaption capabilities.

4.1.1. Signal Chain: Hardware and Software

In the small theatre (henceforth referred to as the Masker), a stereo set of Mackie[38] loudspeakers is used. The medium-sized theatre (henceforth referred to as the Musaion) uses a set of JBL[39] PRX615M loudspeakers with corresponding PRX618S-XLF sub-woofers. The sound system in the large theatre (henceforth referred to as the Aula) cannot be referenced as there is no data available but it uses a stereo line-array system, which characteristically has flat frequency responses in most designs.[40] The Masker and Aula loudspeakers cannot be calibrated as they are permanent fixtures in the venues, whereas the Musaion's loudspeakers cannot be calibrated as they are permanent fixtures in the venues, whereas the Musaion's loudspeakers can be as they are portable. Calibration is done the same way as studio monitors are calibrated, using pink noise and a dBC weighted SLM to level the left and right speakers and sub-woofers to the same level of loudness respectively. In addition, the different loudspeakers in every situation have their own features and problems, such as frequency response and quality, that influence the output (see chapter 5.6). However, considering that the sound is meant to adapt, this is compensated for during adaption.

The same Behringer ECM 8000 measurement microphones used to measure the recording environment are also used for measuring the playback environments (see appendix A1).

[38] For more information, see: https://mackie.com/products/thump-powered-loudspeakers
[39] For more information, see: https://www.jblpro.com/www/products/portable-market/prx800-series
[40] For more information, see: https://www.soundonsound.com/techniques/line-arrays-explained

These are connected to a Focusrite Scarlett 18i20 (gen 1) USB interface. This interface has a 105 dB AD (analogue-to-digital) dynamic range with a very slight boost of 0.5 dB at 20 Hz and 0.1 dB at 20 kHz at minimum gain.[41] This may increase with additional gain and the use of phantom power on the microphones. The figure below shows how the signal moves from the laptop through the Focusrite, is outputted though loudspeaker systems, then re-recorded by the measurement microphones, only to return as new tracks in the DAW. Henceforth, this will be known as the **cyclic process**.

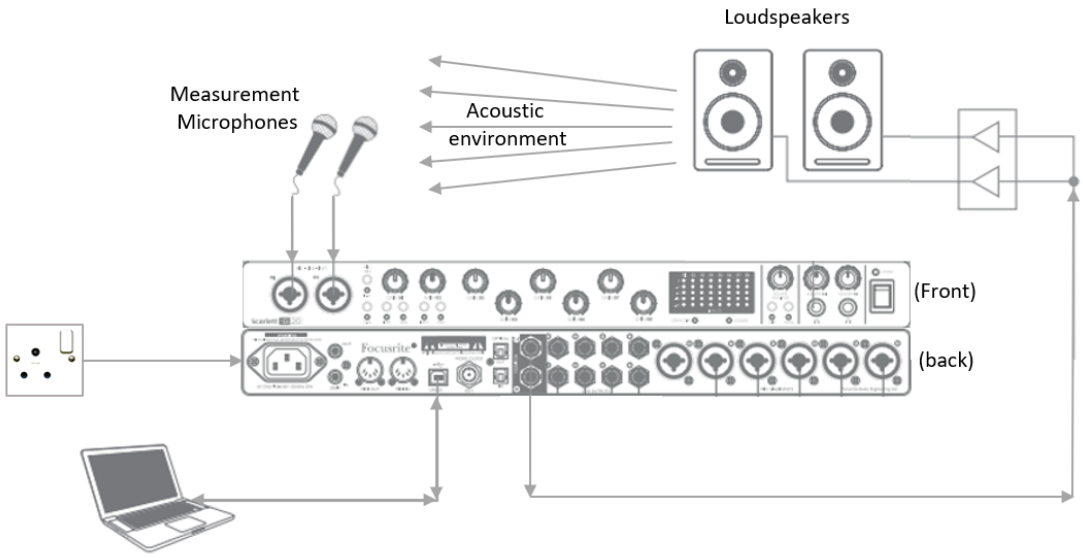


Figure 4.1: Signal chain of the cyclic process

The software used to re-record the music is once again Cubase and the iZotope Ozone mastering suite (including Insight) for analysis and post-DSP for adaption. The adapted sound is then sent once again through the signal chain to be re-analysed to observe the results. This is true for all three theatre environments and the adapted benchmark studio environment (see chapter 4.1.3).

4.1.2. Sonic Differences Between Playback Environments

The EQ data in appendix C shows the difference between the normalised level statistics of the selected takes in each environment. The iZotope Insight spectrum analyser is set to display an infinite time and peak holding of the music in the theatres. The spectrum analyser

[41] For more information, see the specifications (pg35-36) of the Focusrite 18i20 in user manual [accessed on 15 August 2019]: https://customer.focusrite.com/sites/customer/files/focusrite/downloads/8424/scarlett18i20-user-guideeng.pdf

uses a Hamming filter, which allows the linear average frequency spectrum to resemble the spectrum curve of the master track, as seen in the figure 4.2.

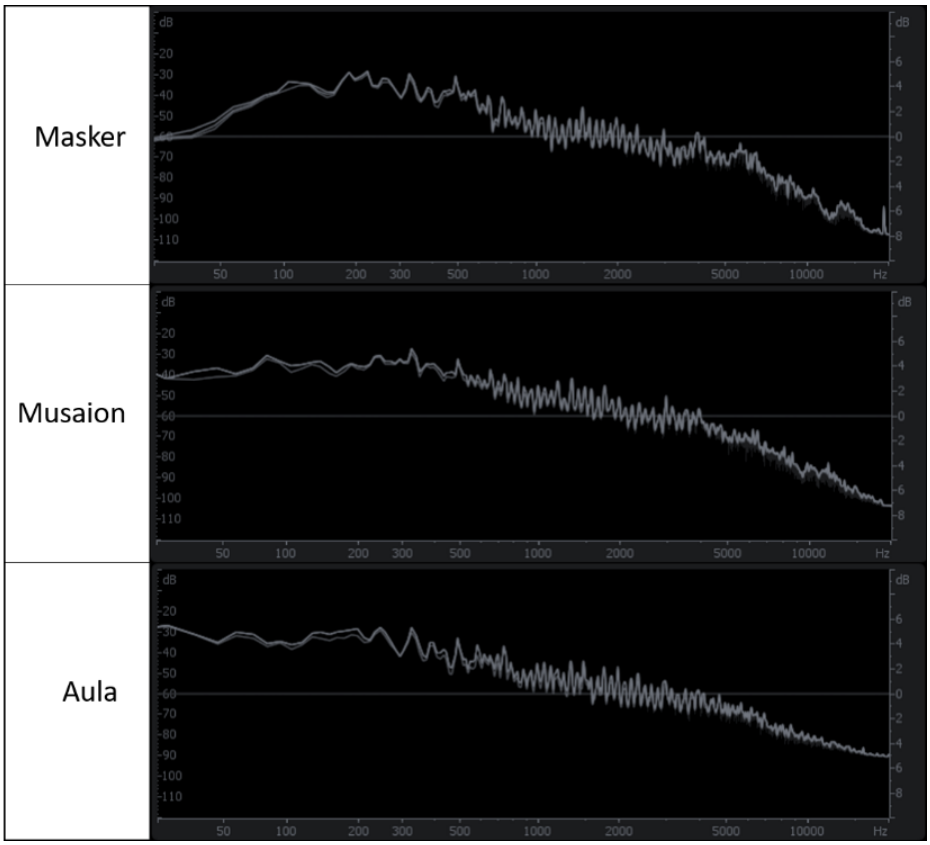


Figure 4.2: Linear spectrum analysis of theatre environments

It is fairly obvious from the above figure that the Masker has a greater roll-off of the low frequencies compared to the Musaion and Aula. This is most likely due to its construction, as it is vastly different in acoustic treatment and shape compared to the other two. Both the Musaion and Aula are constructed as with resonant material (wood) on the walls, both have no sharp corners, and both have curved ceilings so as not to impede the free circulation of sound, as described by Barbieri (1998: 283) in chapter 2.1.2. However, instead of resonant materials, the Masker is lined with heavy drapes that, as Geerdes (1975: 41) explains in chapter 2.1.3, lessen acoustic reverberation and therefore also lessen noise intensity. The ceiling is also constructed in a triangular shape, which affects the diffusion of sound.

Another obvious observation is that the Masker also contains less stability in the high-frequency response compared to the other two theatres. This is likely to be for the same reasons, but additionally due to the axial modes of the speakers. Considering that the speakers

could not be adjusted, cross-over frequencies could have been created at the single measurement microphone's position, accentuating some points while attenuating others in various parts of the spectrum even with different microphone positions, as discussed by Toole (2015: 149) in chapter 2.5. Less obvious observations can be seen in appendix B2. It is generally assumed that environments constructed with resonant and reflective materials would have greater reverberation and an emphasis on fundamental and complex harmonics. This may be true for the Musaion, but it can be seen that the Aula has a smaller reverberation time compared to the Masker and significantly less than the Musaion. This can be for multiple acoustic reasons, but the most probable contributing factor to this loss of energy is the sheer size of the theatre; as discussed in chapter 3.1.1, larger air space contributes to a loss in sound transference.

Nonetheless, it is a far drier environment compared to the others and would require temporal enhancement, especially if it were to be occupied by an audience. In addition, it can be seen in appendix B2 that the ambience (dBA - average RMS) is also varied in terms of loudness. There are several external contributing factors to this but nothing that seriously effects the adaption properties (see chapter 5.6). Furthermore, the L/R peaks of all the environments, including the studio, are mono. This is simply because the measurement microphones cannot be configured to stereo and thus the stereo spread of the master track is lost in the cyclic process. However, this is only limited to the measurement microphones. The stereo panorama in a live scenario is clearly audible and does not have an effect on adaption.

4.1.3. Adapted Benchmark Creation and Statistics

The graphs in figure 4.2 are also observed in appendix C amongst the other spectrum graphs. Although they are exactly the same, these graphs in this appendix serve a different purpose. They are subdivided into the different major frequency bands. The horizontal dashed lines across the bands represent the average loudness of a particular band. This is determined by bisecting the solid lines, which represent the waveforms as an average linear line between one band and another. These average levels are the fundamental equalization properties that are manipulated to adapt to the space in terms of EQ. However, before the audio properties of the theatres can be processed in order to adapt, a benchmark has to be established.

The benchmark environment where the first cyclic process takes place is in the same control room where the music was recorded. This studio's control room is used because it is a small relatively well-treated acoustic space that allows for a greater level of control compared to a somewhat larger and sporadic theatre environment. A single measurement microphone is placed in the listener's position at 'ear height' as if in a sitting position. Three cyclic processes take place at varying degrees of input gain and output intensity, but it was found that louder levels of gain with lower levels of output create intense digital interference (clipping and signal limiting) and are far too susceptible to undesirable noise, whereas lower levels of gain with a louder output are more acceptable. The tracks are then analysed using iZotope Insight and compared to the master track. It can be observed in appendix B2 that there are some subtle differences between the data of the master track's input signal and the normalized recording ('before'), the most important being the integrated output sound pressure and the changes in all the average frequency bands.

If we consider the expanded statistics of these results in appendix B4, a basic calculation is made to find the difference between the two tracks. To create the benchmark used for adaption, this cycled studio track is remastered to resemble as closely as possible the master track in terms of dynamic and spectral properties. Temporal processors are not as relevant as the dynamic and spectral properties of the music because are created naturally during the cyclic process but can be added at a later stage if the environment is considerably dry (such as the Aula), but not removed, as discussed in chapter 3.1.5. The creation of the benchmark track is similar to the processes of adaption (see chapter 4.2). However, this benchmark could make use of more DSP than adapting sound to the theatres. The reason for this is that DSP used in the adaption of the theatres can only be implemented as pre-fader inserts. In other words, DSP will affect the audio in the cyclic process before being outputted into the theatre environments, creating a different sound but not necessarily adapted sound as the music will still be subjected to the environment. However, the final statistics (see appendix B2, 'after') of the benchmark audio are used as a comparative audio track that is not subjected to an environment. Therefore, theatre environments are put through the cyclic process and are compared to the benchmark, not the master track. The remastering of the benchmark track is done as follows.

- Normalising:
    - The cycled track is first normalised to optimize the peaking levels (0.0) as seen in appendix B2 (‘before’).
- Pre-EQ:
    - The equaliser in iZotope Ozone is used to re-establish the inverted 60-phon curve but to a lesser extreme (see appendix A7).
    - Considering that the EQ level differences in appendix C2b shows how the cyclic process has affected the EQ, this data is used in the EQ to compensate for the change. For example, the average low-mid frequency (LMF) band differs by -0.7 dB. Therefore, a boost of +0.7 dB is applied (see appendix B4).
- Dynamics (gate and compression):
    - The objective behind compression is to have as little effect as possible on the timbre of the track.
    - Threshold is set to -20.7 and has a compression ratio of 1.5:1 which is minute, as the intention is not to over-compress. This affects the dynamic range of the music. To further this, a gate with a 1.3:1 ratio is used to reduce some of the noise that the high-pass filter did not remove from the noise floor (see appendix A7).
    - Additionally, a very fast release is used on the compressor because faster release times sound more natural and are less obstructive (Izhaki, 2012: 279). Furthermore, a soft knee at a 3.3 level is introduced to make the transition more transparent (Izhaki, 2012: 285).
- Post-EQ and makeup gain:
    - It was found that the cyclic process had revealed a 4 kHz boost in frequency spectrum due to the addition of the environment (compare appendix C1c to C1a). This is further accentuated when applying dynamic DSP. Therefore, a 1 dB Q cut is implemented; no more could be cut as it would affect the complex harmonics of some instruments.
    - The normalizing process optimized the peak levels but, after analysing the sound after the prior processes, it has been affected. The makeup gain of the compressor is adjusted until establishing a peak at -0.1.
- Maximizer and margin:
    - The last process in to re-establish the integrated (average) output dynamic level. Although the prior DSP allowed a close resemblance to the master track

(-15.1), a maximizer is used. Unfortunately, the maximizer (-1.0 threshold) also clipped the L/R peak at +0.1 dB, thus a -0.2 limiter is put in place, using the IRC mode for transparency (see dynamic signal processing in chapter 3.3.2).

It can be seen in appendix B2 that 'after' the process of adaption is complete the statistics very closely resemble the master track statistics; however, there are some slight variations. The two major differences are the dynamic range and excess frequencies above 2 kHz (HMF and HF). It was found that these frequencies could not be increased any further without completely changing the timbre of the music. The likely reason is because the axial and reflection properties of the environment and loudspeakers react differently in various environments. It can be seen that 'before' adaption takes place in the other environments these frequency bands are the most affected during the cyclic process. Therefore, it can be confirmed that higher frequencies produce more energy over time due to shorter wavelengths, which means they require less spectral change to adapt (dBA weighting). The difference in dynamic range is not unexpected as additional ambient noise and natural reverberation is added to the music, resulting in a smaller dynamic range in every case. This is in fact a beneficial property as it creates desirable noise and a sense of authenticity, as discussed in chapter 3.1.1, and does not hinder the outcome.

If one listens to the master track and the adapted studio track (compare *Final Master* to *Studio Benchmark_Adapted* in appendix D), very little difference can be heard between them, except that the studio track is mono. However, the music does sound subjectively broader in terms of depth and timbre, and this is statistically true as the music now contains compounded reverberation and a longer decay time. Furthermore, it can be observed in appendix C that the master track and the adapted cycled remaster of the studio are very closely related in their spectrums. Thus, the statistics of this benchmark track can be adequately used as comparative material for theatre adaption.

4.2 Adaption: DSP Applications

The DSP used to adapt the cyclic processes of the three theatres takes place in three stages. These are referred to as first, second and third order adaptions. It has been mentioned that DSP for adaption is done on a smaller scale than when creating the benchmark track. Considering that the adaption takes place in less controllable environments compared to a studio, micro-manipulations to the cycled tracks have a profound effect on the tone of not only the music but also the sound of the environment. **First-order adaption** applies to equalization and loudness where the inverted 60-phon Fletcher-Munson (IFM) EQ graph is reapplied to the cycled tracks and the drop in the average output compensated for by use of maximizing. **Second-order adaption** takes place after analysing the outcome of the first-order adaption, and addresses overcompensated inaccuracies in terms of dynamic and spectral influx from the first-order adaption. **Third-order adaption** is where the first and second order adaption data is applied to the master track. This adapted master is then put through the cyclic process for a second time, then micro-manipulations to the fundamental and complex harmonics are made to complete the final adaption. This is also the stage where timbral enhancements are made to the music if necessary.

4.2.1. First-Order Adaption: IFM Filter and Maximizing

Once all three theatre environments are put through the first cyclic process, they can be analysed to determine what adaptions have to be made. A constructed IFM filter is applied to the iZotope equalizer, as seen in figure 3.13, and is maximized. Theoretically, applying an IFM filter (EQ) will not severely affect the timbre of the music as it was mastered to the same curve, but only affect the additional frequencies created by the environment. The maximizer does not use a margin, or limiter, at this time as it is imperative to observe the outcome of dynamic change without additional limiting. Essentially, this maximizer works as a gain-staging tool allowing the signal to be boosted closer to the desired -14.1 LUFS.

The statistics in appendix B2 and the spectrum analyses in appendix C compare the difference in dynamic and spectral levels between the benchmark track, the unprocessed basic cycled tracks, and the first-order adaptions (FOA). It must be noted that the high-pass filtering integrated into the IFM filter does not make a difference when the music goes through the second cyclic process (third-order adaptions) as the majority of this sound is

created by the environment. It is applied only to show how the low frequencies have changed during the first cyclic process.

The data in appendix B5 indicates that the majority of dynamic and spectral sound closely resembles the benchmark, but the use of the maximizer/limiter has had the greatest effect on the integrated output level. This is not an unexpected occurrence, but it is far closer to the final result. The data in red indicates that there has been regression (further away from the final result) whereas the data in green shows that it is closer than the basic cycled track.

4.2.2. Second-Order Adaption: Dynamic and Spectral Enhancements

The second-order adaption protocols are to observe the data and to make micro-changes to the spectral and dynamic levels of the first-order adaptions. It is noticed that in all three cases (see appendix A8) there is a significant drop in the high-mid and high-frequency spectrums again after the cyclic processes have taken place. This has occurred from the measurement microphones placed beyond the critical distance during the process, which diminishes the axial properties of the high frequencies from the loudspeakers, as sound has no proximity. Considering that the Aula and Musaion are concert halls designed to enhance resonant frequencies of instruments (low-mid and high-mid), this loss is more substantial in these environments (Ahnert & Tennhardt, 2008: 147).

If we consider Izhaki's (2012: 211) frequency perceptions in figure 2.9, the lack of these frequencies in this range causes this 'distant' effect. Therefore, the solution is first to modify the intensity of the IFM filter by subjectively attenuating or boosting frequencies to conform to the timbre of the master track and, secondly, to incorporate 'air', or a high-frequency shelf, in the post-EQ to increase these frequencies globally on the IFM EQ. The change in EQ and post-EQ intensity may be a subjective operation but can be objectively analysed and viewed on the spectrum. The high-frequency shelf is applied again during the third-order adaption because, like the high-pass filter, these frequencies contain environmental sound. Additionally, due to the directional axial properties of high frequencies and lack of proximity, these frequencies will have to be compensated.

After the EQ is analysed again, a slight margin (limiter) is introduced to prevent the signal from exceeding the -0.1 LUFS peak. However, it can be seen in appendix B6 that the integrated outputs of each track are closer in value compared to the basic cycled tracks, but

all surpass the desired -14.1 LUFS level and the -0.1 peak in some cases. Again, no compression or expansion is used, as can be seen in the dynamic range, but rather incremental changes between the limiting and maximizing thresholds. Manually controlling these thresholds is the most time-consuming aspect in second-order adaption as it is not an automated process but produces very accurate results as can be seen in appendix B6. An important point to take note of is that the second-order adaption is not the final result. The idea behind this stage is to imitate the shape of EQ arch of the benchmark track. The third-stage adaption is where complete adaption takes place.

4.2.3. Third-Order Adaption: Live Adaption and Additions

The statistics in appendix B6 and the second-order frequency spectrums seen in appendix C are clear indications that the cycled tracks mutate the output to closely imitate the benchmark's audio properties. However, thus far, the statistics do not show how these changes impact the adaption in a live scenario, nor do they provide evidence whether they enhance the perception of music in the performance environment. As discussed in chapter 3.3.1, the mix is intended to reach its fullest potential after adaption takes, and the final adapted tracks are intended to conform fully to the benchmark track's EQ and dynamic level. This is the purpose of the third-order adaption. The master track is once again put through the cyclic process in each environment but, this time, the spectral and dynamic changes are incorporated.

As previously mentioned, these changes can be incorporated only as pre-fader inserts and will still be affected by the environment, hence why neither compression nor expansion can be used for dynamic adaption. However, due to the spectral and dynamic changes, the readings of each environment should closely resemble the benchmark track because there will be the addition of ambient sound that intensifies the lacking frequency levels. This is done by applying the same EQ changes to the master track and increasing the output level according to the thresholds of the maximizer to produce the same level dynamics in the environment. A significant difference between the third-order adaption stage and the previous two is that during this second cyclic process the loudness and frequency spectrums are monitored in real time in iZotope insight. By analysing the audio in real time, micro-changes in terms of EQ and gain can be made to the audio properties to conform completely to the benchmark track. The amount of input signal is used to control the integrated output intensity and a limiter is set at -0.1 in each take to prevent digital clipping as a precaution, but is proven to be

unnecessary because the input signal (gain) and output levels achieve the same level of intensity as the second-order adaption dynamic enhancements.

4.3 Adaption Summary

The statistical results of the final adapted sound can be seen in appendix B2 and the spectrum analysis of each track can be observed in appendix C. To compare the different stages of adaption, samples of each stage can be found in appendix D. As a general note, a high-pass filter is incorporated to the same degree as the IFM filter used in the first-order adaption on each track, after completing the third-order adaption. Although this is not used during playback in the playback environments, it is incorporated in the sound samples and the spectrum analysis to compare the adapted sound in each environment to the benchmark track.

Neither the Masker nor the Musaion required temporal reinforcement. On the contrary, the Musaion could be considered too reverberant, having a compounded decay time exceeding 2 seconds. Nonetheless, the Aula is a considerably dry environment (≈ 1,65 sec), that required additional artificial reverberation for depth enhancement to create a more desirable sound. The degree of additional reverberation is subjective but could be compared to the other environments, including the benchmark studio environment. The compounded decay time of the Masker (≈ 1,97 sec) is more transparent than the Musaion (≈ 2,43 sec), but the benchmark (≈ 1,07 sec) could be considered too dry as it is fractionally more reverberant than the original mastered track. As discussed in chapter 3.1.4 with regard to RT_{60}, a reverberation time exceeding 2 seconds is considered highly reverberant, therefore the best course of action was to incorporate a transparent reverberation to achieve a middle ground. Therefore, a 5.5ms theatre-type reverberation is added during the third-order adaption of the Aula. This barely affects the spectral properties of the music and provides sufficient saturation and depth (≈ 2,00 sec), but does slightly affect the integrated output level and L/R peaking.

During the final adaption stage, moderating the input gain and the output level is a time-consuming yet vital role in the dynamic control of the music. With the addition of artificial reverberation in the Aula and every change made to the EQ in each scenario, the dynamic properties of the output tend to fluctuate. The solution to stabilise the sporadic dynamics is to control the level of gain before these additions are made to the master track, decreasing the L/R peaking points while controlling the output fader into the PA system in order to increase

them again to the correct level. The reason this is time-consuming is because the ratio between input and output needs to correspond with the integrated output level. In other words, the total average output of the music needs to precisely level -14.1 LUFS. This is achieved by applying a lesser percentage of gain to the master track and incrementally experimenting with different degrees on the output faders.

This concludes this chapter. The following chapter will discuss the results gained from the methods applied in the physical domain and the conclusions made from these results.

Chapter 5: Results and Conclusion

5.1 Discussion and Interpretation of the Main Findings

To summarise the results of this research and to answer the main research question of this study, the question is divided into three different components. Each component is used to compare its theoretical and conceptual framework with its nominal results. These components consist of the accuracy of profiling the digital instruments to their acoustic counterparts, the audio techniques used to create the adaptable sound, and the alterations of the audio properties used during playback in the various environments to formulate the final results. The summation of these results from this study are systematically reviewed in reverse from the main findings to secondary findings discovered during the production process. The overall relationship between the adapted sound and the initial recording is partly subjective but is objectively substantiated from the gathered nominal and visual data. In other words, the final results of the data observed in appendix B2 and C6 display a very close sonic relationship in terms of dynamic and spectral properties, but not necessarily a conclusive psychoacoustic perception due to subjective hearing from person to person in terms of timbre, temporal change and frequency range.

In terms of the spectral results, the benchmark track and the adapted tracks in the playback environments accurately share similar properties with regard to frequency distribution. In comparison, the adapted tracks generally demonstrate an average difference of less than 0.1 dB between each octave band. However, it can be noted that the Masker does deliver a lesser output of high frequencies, likely due to reasons explained in chapter 4.1.2. regarding diffusion, cross-over frequencies and the use of only one measurement microphone. Furthermore, it is noted in appendix B2 that the low frequencies' statistics show a "+" symbol. This represents an excess of low frequencies in this range, but this statistic is a result from the IFM high-pass filter applied to the EQ post-adaption to portray the adapted spectrum without noise created from the environment, as discussed in chapter 4.3.

The dynamic properties of the adapted tracks are less subjective than the spectral properties. Although loudness does affect the human perception of pitch (Zhang, 2008: 43), the dynamic range and the overall loudness of the sound within the environments is not affiliated with pitch in these results, as pitch is more closely related to timbre and the temporal effects as discussed in chapter 2.1.1. Furthermore, the spectral data in this study, observed in appendix

C, is gathered from the average sound pressure within each octave band. However, these octave bands contain more variations in data. Although only represented as an average in this study, these are variations such as wavelength, amplitude, sample size and bit depth, whereas the dynamic data is more simplistic – how loud is the music and how loud does it get? Thus, the dynamic data in appendix B2 is more linear in terms of statistics and therefore more objective.

The results indicate that there is less than a 0.1 dB difference between the average integrated output level, and the L/R peaking levels of each adapted track do not peak above the -0.1 LUFS level. These accurately equate the loudness levels of the benchmark track. However, the dynamic range of each track is far more varied from environment to environment. As discussed in chapter 2.4.3, the dynamic range between the loudest and quietest point in stage and screen is usually 14.29 dB. The adapted music in each environment does not surpass this level, therefore the tracks still remain within this limit; but this will likely increase to a wider range with the hypothetical addition of an audience that gives rise to an increase in absorption. Furthermore, the addition of an audience would also theoretically vary the temporal effects and frequency response, but this is discussed in chapter 5.6 with regard to further research.

What can be summarised in terms of temporal adaptions is the effect of the acoustic and artificial reverberation of sound in the environments. The desired effect of the temporal DSP is entirely subjective but certain conclusions can be drawn from the compounded decay time data. The desired outcome of temporal DSP is to create depth and a sense of authenticity in the music but not to sacrifice sound quality. Considering this, the music should not be inundated with excessive reverberation to deter from sharper attack frequencies or a lack depth in the music as discussed in chapter 3.3.1. In other words, the compounded decay time should not exceed 2 seconds as this is considered to have a fairly equal dry-to-wet ratio as discussed in chapter 3.1.4. The Masker and Aula do achieve this but, due to its construction, the Musaion possesses approximately 0.5 seconds of decay time above 2 seconds. However, the compounded decay time variations of all three theatre environments are obtained uninhibitedly and would significantly reduce if occupied as mentioned above. Whether or not this marginal additional decay time in the Musaion is oversaturating can only be established subjectively, but, considering that the Musaion is constructed for classical music, it could be assumed that the adaption is suited to its environment.

The mixing and mastering techniques used to create the audio product follow fairly standard procedures. The difference is that the mastered track achieves its fullest potential only after adaption takes place. By subjecting the music to an environment, it is the addition of the acoustic reverberation and ambient noise that characterizes the adapted music for an authentic sound. Authenticity is mainly a subjective matter but, with the added environmental effects, the music is irrefutably broader in terms of reverberation. Unfortunately, the spatial dimensions of the adapted tracks cannot be objectively justified due to the monoaural limitation of the measurement microphones but, once again, this is only a limitation from the cyclic process. From a listener's point of view in a live scenario, the spatial effects created during mixing are clearly audible during playback in the stereo loudspeaker systems. There are, however, two problems that have not been completely noted up to this point.

The loudspeaker systems themselves had some negative effects on the sound within the environment. It can be heard in the *Final Adaption_Aula* sound sample (see appendix D) that the Aula produces a fairly distinctive hissing noise between 12 kHz and 16 kHz. This noise could not be avoided as it is created from digital interference. It was concluded that this digital noise was not directly linked to the equipment used for adaption (laptop, cables, adaptors or sound desk) but the loudspeakers themselves, as several tests were made in an attempt to remove the noise. A similar circumstance occurred in the playback within the Musaion. Although these speakers could be calibrated, as they are not permanent fixtures in the venue, this portability came with its own issues. Considering that the prior use of these loudspeakers is unknown, the stage-left speaker and the stage-right sub-woofer did contain slight distortion which is audible around the 120 Hz range in *Final Adaption_Musaion* sound sample in appendix D. It is fairly safe to assume that because these portable loudspeakers are quite old and have been moved around for potentially dozens of other events, they may have sustained some internal damage that causes this noise. In both instances, these discrepancies were discovered during the RT_{60} pink noise tests. The analytical software used to identify the decay time and frequency response of the environments indicated these mechanical flaws. Both these minor mechanical issues pose a problem in positively concluding this study, but this is limited only to these particular cases. Holistically, this indicates only a slight variation in sound quality and does not affect the final results.

To conclude this summary and to answer whether acoustic and digital instruments can be merged to reproduce *Welcome to the Show's* orchestral size in a reduced setting, the success of profiling needs to be addressed. The *Unmixed Recording_No Drumkit* sound sample in appendix D can be used to compare the unprofiled digital instruments to their unmixed acoustic counterparts. Very obvious distinctions can be made between the different media, mainly timbre and pitch. Although these are subjective qualities, there are smaller sections in the music, such as the flute solo in the introduction of the second movement, where neither medium accurately portrayed an authentic sound due to the mechanical sound of the digital instruments and the performer's limitations.

Recommendations to amend issues such as this are discussed in chapter 5.6 but the reasoning behind the lack of authenticity mentioned above can be described thus: during the mixing process, the levels of the digital instruments in some sections are more pronounced than the acoustic instruments; in other sections, such as solos and thinly textured sections, the acoustic instruments are accentuated more to create a more authentic experience. The problem lies in the quieter sections where the acoustic instruments become vital for authenticity. Due to the relatively unpractised musicians, some sections had to be digitally replaced, causing these sections to sound more mechanical. Furthermore, some alterations to the mix should be made in future such as the timpani being more forward than was at first desired, but the progressive improvement from the raw recording to the final adapted sound does show that both media amalgamate with significant success.

5.2 Critical Assessment

This study diverges into a multifaceted field of sound, audio, acoustics and psychoacoustics. Each part of the production process is critically assessed separately in order to identify the positive and negative aspects of the audio product as a whole and the degree of success with regard to the acoustic and digital amalgamation.

Regarding the practice-led recording stage of this study, there are several contributing factors that positively and negatively reflect in the audio product. The only major negative critique is that of the skill and level of practice of the performers used to record the acoustic instruments, as discussed previously. It can be heard in *Unmixed Recording_No Drumkit* sound sample that some instrumentalists performed better than others but, unfortunately in

the ensemble setting, the leakage of sound affects all the instruments. This affected accurate pitching negatively in the recording but did not deter from successful profiling, thus it is relatively irrelevant towards the amalgamation of the two different media.

A positive outcome from the recording was clear signal from the microphones with very little discrepancy such as phase issues, especially when recording the individual instruments. Having clear signal provided an easier task when profiling and mixing, and the false notes produced by the instrumentalists could be edited more easily and restored to some extent without having an effect on the sound quality, although some of the recorded music had to be completely replaced digitally, such as the trumpet as explained in chapter 3.2.1. Considering the lack of rehearsal time, it was not surprising that there were some problems. In future, it would be better to use more experienced musicians and have more time for practice and rehearsals.

It is mentioned above that some sections of the music had to be completely digitally recorded. In appendix A5, it can be seen that the violins, double basses and the majority of the percussion instruments were not recorded in the reduced setting. This was done deliberately to observe whether their exclusion would affect the outcome of the audio product. For all intents and purposes, these excluded instruments did not negatively affect the outcome of the audio product; however, there is no evidence that their inclusion could have positively affected the outcome either (see chapter 5.6). The exclusion of the trumpet however is a different case. Percussion and string instruments in the VSL plugin are subjectively easier to synthesize and profile than the brass instruments, which is why all three brass instruments are included in the reduction. As a matter of opinion, the “valve” timbre created from their construction and complex harmonics was found to be more challenging to profile than the other instruments. Thus, the ’authenticity’ of these instruments is not as established compared to the others. This could be amended with better instrument samples and more specialized plugins for brass instruments, or to use more experienced musicians and have more rehearsals as mentioned before.

Mixing and mastering are subjective operations but the outcome of the audio product as an entirety, proven from the results, is positive in what was proposed in this study. Suffice to say there are some changes that could improve the product in future. As a matter of opinion, more strings could be emphasised and the timbre of the brass and the timpani (as mentioned above)

can be improved. These are minor mixing flaws and have no significant effect on the outcome of the study but were noticed after the adaption experiment took place, although the adapted tracks did have a positive effect on the insufficient string and brass timbre due to the increase in high frequencies.

The final adaption on the whole has a positive outcome, as can be observed in the EQ graphs and statistical results in the appendices. However, it can be seen in the EQ graphs that the final adapted tracks do not absolutely correspond to the benchmark track. This is most likely because of the manual adaption processes. Nonetheless, a negative consequence of manual adaption is that it requires a lot of time and concentration, and the human ear experiences a loss of activity over extended lengths of time (Howard & Angus, 2009: 46). This is further discussed in chapter 5.6 for a possible automated future solution.

The answer to the first secondary question of this study - the authenticity of a full-scale orchestra using the amalgamation of digital and acoustic sound - is complicated because it is completely dependent on a listener, 'authenticity' being a psychological perception of sound. Therefore, there is no simple answer as this would require more qualitative research and wide-scale case studies, but both media are used very successfully in commercial and experimental music. Classical composers, especially contemporary composers, also experiment with digital sound, but *Welcome to the Show* is not as experimental compared to works by avant-garde composers such as Stockhausen (1925 – 2007) or Lachenmann (1935 –) for example. *Welcome to the Show* (2019) follows more traditional methods in composition influenced by composers such as Piazzolla (1921-1992), Stravinsky (1882-1971) and Don Davis (1957-), and is also an independent variable based on the subjective interpretation of a listener (Anderson, 2019: v).

5.3 Production Comparisons

The way in which the audio product is created in this study is fairly similar to the standard procedures used to create commercial recordings, with the exception of exact spectral and dynamic DSP. To answer the second and third questions of this study, regarding the logistics and differences between the production techniques used in this study compared to producing the same music for a different platform, comparisons are made between the regulations of these platforms and the outcome of this study.

The difference is that the music that doesn't adapt does not consider the change in dynamics and frequencies when exposed to an environment. As the results indicate in this study, this causes problems when dealing with a musical theatre composition such as *Welcome to the Show* (2019) because each piece of music incorporates different varieties of sub-genres. The adaptive capacities of the dynamic range in this study ensure better dynamics and stability. This is also true for the equalisation of the music because the music adapts to and retains the original equalisation of the music, whereas sound that doesn't adapt changes its spectral properties when exposed to an environment as discovered from the first cyclic process.

One significate difference is that the audio product in this study is not limited by a specific sample rate, bit depth or audio format. Although the bit depth of this audio product in a 24-bit file which is suitable for internet streaming, noise level issues must be considered when downloading and uploading (Katz, 2015: 15). The platforms and many others prefer MP3 format compared to WAV or Advanced Audio Coding (AAC). The data compression of MP3 losses certain audio information that results in a slightly different result (Owsinski, 2017a: 111), but the audio product in this study is not affected by this data compression. The only stage in which the sample rate is affected is when it is down sampled to 48 kHz from 96 kHz for upload and download purposes.

There are many logistical elements involved in creating any audio product and the user is limited by what funding and equipment are available. Because this studio was also subject to these limitations, any comparison of the quality and expense of this study's audio product with another large-scale orchestral work on a similar scale would be entirely subjective. The audio product in this production was entirely self-funded and was limited to what equipment was available. It can be said that more funding could provide better results because better equipment could be purchased, and more experienced musicians hired. However, considering the relatively low cost of this production, the results from this study prove that the adaptive capabilities are achievable without having the most sophisticated equipment or professional musicians. Thus, an audio product created using this method is less expensive and produces a reasonable level of sound quality over a similar sized product using traditional methods with the same funding.

In addition, because this method allows the music to adapt to the environment in which it is played, a pre-recorded track is more reliable than musicians in a live setting, and is even more reliable in a studio environment as the sound quality relies less on the technical skills of performers and more on the profiling skill of the mixing engineer. However, it must be stated that it is still better to use more skilled musicians than those that were available in this study. Less-skilled musicians impact on the production time needed to edit and restore bad sections of music during mixing. Furthermore, the production time, i.e. time needed to create this audio product, is significantly more than producing the same track for a different platform as it is continuously monitored and analysed to achieve the exact requirement discussed in chapter 3.

5.4 Limitations and Implications of the Research

The results of the adaptable sound in this context are limited to the audio product created for the purpose of this study. This is because the audio product is specifically designed for this experiment. However, the principles of applying a spectral IFM filter and altering the average dynamic level is not limited to this study only as it can be implemented on any other track to some degree. It could be said that the objective of the IFM filter is to manipulate the frequency response of a track in order to create even distribution of frequencies across the spectrum. In theory, this can be a method to essentially 'reset' the spectral distribution of sound in pre-recorded music in order to be shaped into a more desirable form.

There is no reason that this adaption process is exclusive to theatre environments. The only limitation is that the dynamic DSP used on the mastered tracks in this study creates a wider dynamic range and specific integrated dynamic output designed primarily for stage and screen. The amalgamation between digital and acoustic instruments is not uncommon but it has been found that some instruments are more suited for this task than others. Therefore, there are some limitations with regard to timbre reconstruction as discussed in chapter 5.2. This study has only used adaption methods to alter pre-recorded music, but real-time live mixing with the use of supervised artificial intelligence is becoming more common in audio engineering as seen in the Sonible Smart:EQ2 plugin.[42]

[42] For more information, see: https://www.sonible.com/smarteq2/ [Accessed on 29 August 2019]

This research is not only aimed at developing optimal sound for a musical theatre performance but could be used in a vast array of commercial applications. The adaptive processes in this research aim to create balance and clarity in sound to improve the sonic balance in terms of frequencies and loudness. Regarding the performance of music, creating that balance and clarity is particularly useful to optimise the sound of a live or studio mix or to identify flaws in an acoustic space. The processes discussed have the potential for uses beyond music and acoustics. With more development and research, the methods used in this study theoretically can improve sound in telecommunications and digital media, such as television or radio.

From personal experience, the input and output of devices such as mobile phones or video cameras are good examples of where adaption capabilities could be beneficial, such as the ability to hear clear signal from one user to another on a mobile phone. This is affected by several factors such as the signal, the carbon microphone construction and the physical size of the speaker. Although these are some factors contributing to sub-optimal sound, adaptive capabilities could improve them by analysing the output of frequencies and adapting the sound to cancel out external noise and improve the sound of a voice. An example of dynamic adaptions can be seen in television broadcasting. There are many instances in which unproportionate sound can be heard, especially between British and American broadcasting channels, as they use different systems. Televisions could incorporate adaptive sound capabilities to bring uniformity in dynamics and spectral enhancements. These are two examples of how the sound in digital media could be improved; chapter 5.6 discusses what future research could be conducted to address these and other possible uses.

5.5 Conclusion

The adaptable sound recording in this study was developed using a variety of acoustic and digital audio techniques. By manipulating the audio properties of the recording in a playback environment, the sound in the space is moulded to generate a consistent output for a theatrical performance of *Welcome to the Show* in venues with differing acoustics. The stated objective of this research is to develop an authentic but economical recording process that recreates large-scale orchestral music for musical theatre. The product should be able to adapt its sound output adequately to the acoustic parameters of a wide range of performance venues to achieve consistent musical authenticity and sound quality.

A sound engineer with an understanding of the fundamental principles of sound and acoustics, as described in chapter 2, and using the conceptual methods in chapter 3 to compare with the theoretical data, can balance the timbre and sound pressure level of pre-recorded music using DSP applications. The engineer can also compensate for any differences that may manifest themselves when exposed to different environments. There are certainly subjective interpretations that can be made from the results in this study, but the gathered data can be assessed according to the objective strengths and shortfalls of these results.

The subjective nature of recording, mixing and mastering can be considered weaker contributing factors in the production chain, even though these tasks formed the larger portion of the practice-led research. The audio product itself, as explained in the statement of the research problem and objective in chapter 1.2, is an independent variable, whereas the adapted tracks are a product of the environment in which they are cycled. Users of this data must further understand that the manual manipulations made to the audio product to conform the cycled tracks to the benchmark, as discussed in chapter 4.2, can also be considered insufficient for attaining absolute adaption. In summary, the human perception of sound is subjective, but the stronger contributing objective data and the impact of this research outweigh auditory subjectivity.

The rationale supports the objective of this research to provide an inexpensive and reliable product that producers can use to reduce the cost of a production, while providing listeners with an authentic sonic experience. The cost of producing a product at full scale with the

same sound quality would be significantly increased, as more resources would be required. Furthermore, the inclusion of additional people inevitably results in an increase of both logistical and technical challenges, as discussed in chapter 3.2. Regardless of the size of the orchestra recorded, the quality of the product is still dependent on the recording, mixing and mastering skills of the engineers.

Substituting the bulk of the orchestra with digital instruments and samples provides additional mixing advantages and is an effective way of controlling the financial implications and time constraints associated with an audio production. In chapter 3.3.1, profiling the digital instruments to their acoustic counterparts is determined to be an effective way to create authenticity in the mix. More significantly, the ability to control the separation of these profiled instrument sections provides more flexibility in DSP when mixing, as altering the audio properties of one section does not directly influence another section. Although the digital instruments may lack the tonal versatility of acoustic instruments, the profiling process makes up for this shortfall fairly accurately, confirming that large orchestras can be digitally recreated in a smaller setting.

Sound engineers should be highly critical of the quality of their products and respect the dramatic impact that sound quality can have on an audience's experience of a musical performance. On the other hand, every production manager must weigh the cost of each element of a production with the potential for profit. Budget constraints can be perceived as detrimental to the recording engineer's ability to produce authentic, large-scale sound quality in a production. However, data in this study suggests that, by reducing human resources and utilizing the adaption methods indicated in chapter 4, a high-quality adaptable product can be engineered on a smaller budget, tailored for a particular environment and still produce optimal sound quality.

The statistical data and implications of this study show objectively that sound under very specific conditions can be adapted to multiple acoustic environments, but that these findings are not limited to a physical space; they suggest the potential for the adaption methods in numerous technological fields.

5.6 Recommendations and Future Research

Given that some sections of the music contain discrepancies in terms of authenticity as discussed in chapter 5.1, some recommendations can be made to improve the contrast between the digital replacement of acoustic instruments. The most apparent recommendation would be to use the most skilled musicians available for a project such as this. However, if this is not possible, more rehearsals should take place over a longer period of time. Secondly, if neither of these options is possible, more specialized plugin software should be used. For example, the Spitfire Symphonic Brass[43] is subjectively more realistic and could be more suitable for profiling than the VSL version that was available for this study.

Given that there were some issues with the loudspeakers used for playback in this study (see chapter 5.1), it is recommended that properly maintained loudspeakers should be used for any playback scenario since, as stated in chapter 3.1.5, inappropriate loudspeakers will yield less accurate results. In addition to loudspeakers, measurement microphones are used to recapture the sound during the cyclic prosses. Although the microphones used in this study were adequate for the purpose of this study, they were limited to a monaural configuration. The use of binaural measurement microphones, such as the Neumann KU 100[44], is recommended as they, in combination with appropriate loudspeakers, will eliminate this limitation and provide higher quality stereo cycled tracks.

The external factors that contributed to the variations of ambient sound in the different environments were due to noisy air conditioning systems, digital noise in the PA systems and people present inside and outside the venue. Although attempts were made to limit these external influences, these interferences should be avoided as much as possible by making sure the equipment in venues is properly maintained, and that the staff vacate the venue when experiments are being performed.

[43] For more information, see: https://www.spitfireaudio.com/shop/a-z/spitfire-symphonic-brass/
[44] For more information, see: https://en-de.neumann.com/ku-100

The analysis of the audio in this study was done using iZotope Insight software. Although this software fulfilled the needs of the practice-led research and the experiments, it is recommended that more specialized software, such as MATLAB,[45] should be used to produce more thorough data.

It was mentioned in chapter 5.1 that some instruments were deliberately excluded from the recording stages, and that no evidence can be provided whether their exclusion had a positive or negative effect on adaption capabilities. Future research may provide the answer to this but would have to be done with a larger budget. With this in mind, it would be interesting to establish whether the ratio of instruments could be reversed. Considering that the bulk of the orchestra in this study is digital, having acoustic instruments supplemented by digital means for adaption capabilities could provide a more authentic experience. In addition, the experiments conducted in this study made use of uninhibited environments and the data indicates that the spectral and dynamic properties of the music adequately adapted to the environments. However, future research could be conducted to further solidify this theory of adaption by having the venues occupied, and by experimenting with a wider variety of performance environments.

Adaption is a time-consuming process and the above-mentioned ideas would require more development and research. Any future research would benefit from a level of automation by introducing a supervised artificial intelligence (AI) program based on this study's adaption protocols. It would have to incorporate a music retrieval system moulded on the frequency response of the human ear and be able to accurately adapt the audio properties on a live mixing console or in a DAW. An AI-driven program would partially address the subjectivity and loss of listening activity in the human ear over time. Furthermore, in theory it could improve the signal from mobile devices and address the broadcasting dynamic and spectral problems found in television broadcasting, radio or any other form of digital media by using the same adaptive DSP protocols in this study in real time. Therefore, the format would likely have to come in several forms, such as software integrated into firmware or a software program (application) for devices, an advanced VST plugin for a DAW or an independent hardware component that integrates with other hardware devices, such as digital mixing consoles and MIDI devices.

[45] For more information, see: https://uk.mathworks.com/products/matlab.html

References and Appendices

References

Adler, S. 1989. *The Study of Orchestration*. 2nd Edition. Pp 5-6. Composed by Vail-Ballou in Meloir. New York: WW Norton & Company.

Ahnert, W & Tennhardt, H. 2008. Acoustics for Auditoriums and Concert Halls. In Ballou (Ed.). *Handbook for Sound Engineers*. 4th edition. Chapter 7. Oxford: Focal Press. 147-197.

Anderson, M. 2019. *Welcome to the Show*. Score edited by M. Sheehan. Script and lyrics edited by K. Chale and P. Rae. Licensing through DALRO.

Ballou, G., Ciaudelli, J., Schmitt, V. 2008. Microphones. In G. Ballou (Ed.). *Handbook for Sound Engineers*. 4th edition. Chapter 16. Oxford: Focal Press. 493-593.

Barbieri, P. 1998. *The Acoustics of Italian Opera Houses and Auditoriums.* (Ca. 1450-1900)." Recercare, vol. 10, pp. 281–300. Retrieved from: www.jstor.org/stable/41692748.

Beard, D & Gloag, K. 2016. *Musicology: The Key Concepts*. London: Routledge.

Blandford, D & Parr, J. 2013. *Introduction to Signal Processing*. Boston: Pearson

Brown, P. 2009. Test and Measurement. In G. Ballou (Ed.). *A Sound Engineer's Guide to Audio Test and Measurement*. Section 1. Oxford: Elsevier. 12-30.

Brown, P. 2008. Fundamentals of Audio and Acoustics. In G. Ballou (Ed.). *Handbook for Sound Engineers*. 4th edition. Chapter 2. Oxford: Focal Press. 32-33.

Bryman, A. 2012. *Social Research Methods*. 4th edition. New York: Oxford University Press.

Burkholder, J and Grout, D and Palisca, C. 2010. *The History of Western Music*. 8th Edition. Pp 860. Edited by M, Payne. New York: WW Norton & Company.

Camilleri, L. 2010. Shaping sounds, shaping spaces. *Popular Music*, Vol. 29 (2). 199-211. Retrieved from http://www.jstor.org/stable/40926918

Chittom, L. 2018. Stage Direction & Types of Stages. Retrieved from: https://www.tes.com/lessons/JU6Tkw3vEp17mw/stage-directions-types-of-stages [Accessed on 18 February 2018] Collaborative arm of: TesTeach.com.

Creswell, J. 2009. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. 3rd edition. California: Sage

Davis, D and Davis, C. 2009. Large Room Acoustics. In D. Self (Ed.). *Audio engineering explained: professional audio recording.* Chapter 17. Oxford: Focal Press.

Dyar, T. 1961. Techniques and Devices - Microphone Placement. *Ethnomusicology*, Vol. 5 (1). 49-50. Retrieved from http://www.jstor.org/stable/924308

Edwards, J. 1980. Choosing the Right Microphone. *The Choral Journal*, Vol. 2 (3), 5-8. Retrieved from http://www.jstor.org/stable/23545594

Everett, W & Laird, P. 2002. *The Cambridge Companion to the Musical*. Edinburgh: Cambridge University.

Everest, F and Pohlmann, K. 2015. *Master Handbook of Acoustics*. 6th edition. New York: McGraw-Hill Education.

Geerdes, H. 1975. Adjustable Acoustics in Music Performance. *Music Educators Journal*, Vol 61 (8), 40-44. Retrieved from http://www.jstor.org/stable/3394777

Howard, D & Angus, J. 2009. Introduction to Hearing. In D. Self (Ed.). *Audio engineering explained: professional audio recording.* Chapter 2. Oxford: Focal Press.

Huber, D. 2008. MIDI. In G. Ballou (Ed.). *Handbook for Sound Engineers*. 4th edition. Chapter 29. Oxford: Focal Press. 1101-1128.

Huber, D & Runstein, R. 2014. *Modern Recording Techniques*. 8th edition. London: Focal Press.

Izhaki, R. 2012. *Mixing Audio: Concepts, Practices and Tools*. 2nd edition. Burlington: Focal Press.

Katz, R. 2015. *Mastering Audio: The Art and Science*. 3rd Edition. Canada: Focal Press.

Lamb. A. 2001. Musical 5(i): History. In the *New Grove Dictionary of Music and Musicians* (2nd edition, Vol. 17 pp. 453-455) Edited by S. Sadie. London: Macmillan.

Lloyd-Webber, A. 1986. The *Phantom of the Opera*. Lyrics by: C. Hart, R. Stilgoe and M. Batt. Story by R. Stilgoe. Copy write to: The Really Useful Group. London: Hal Leonard.

Lenard, J. 2001. *"Theatre Sound"*. Pp 11 – 179. London: A & C Black.

McManus, S. 2008. Filters and Equalizers. In G. Ballou (Ed.). *Handbook for Sound Engineers*. 4th edition. Chapter 23. Oxford: Focal Press. 788-802.

Novak, E & Novak, D. 1996. *Staging Musical Theatre: A Complete Guide for Directors, Choreographers and Producers*. Edited by A. Manolis. Ohio: Betterway Books (F&W Publications).

Owsinski, B. 2017a. *The Mastering Engineer's Handbook*. 4th edition. Burbank: BOMG.

Owsinski, B. 2017b. *The Recording Engineer's Handbook*. 4th edition. Burbank: BOMG

Pascall, R. 2001. Style. In the *New Grove Dictionary of Music and Musicians*. (Vol. 24, pp. 638-639) Edited by S. Sadie. London: Macmillan.

Rayburn, R. 2012. *Eargle's Microphone Book*. 3rd Edition. Burlington: Focal Press.

Schönberg, C. 1980. *Les Misérables*. Story by V. Hugo (1862). Lyrics by H. Kretzmer and J. Fenton. London: Cameron Mackintosh.

Senior, M. 2013. *Mixing Secrets for the small studio.* 2nd edition. Burlington: Focal Press.

Senior, M. 2015. *Recording Secrets for the small studio.* Burlington: Focal Press.

Shaw, J. 2010. *The South African Music Business*. 2nd Edition. Johannesburg: ada enup cc.

STCratings. 2004. *Sound Transmission*. Retrieved from: http://www.stcratings.com/ [Accessed on 11 October 2017]. Collaborative arm of: Acoustics.com

Suskin, S. 2009. *The Sound of Broadway Music*: A Book of Orchestrators and Orchestrations. New York: Oxford University Press.

Toole, F. 2018. *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*. 3rd edition. New York: Routledge.

Wu, B, Horner, A and Lee, C, 2014. *Musical timbre and emotion: The identification of salient timbral features in sustained musical instrument tones equalized in attack time and spectral centroid.* 928-931. In *ICMC*.

Zhang, M, Granzow, J, Ren, G and Bocko, M. 2014. *Timbre Imitation and Adaption for Experimental Music Instruments: An Interactive Approach Using Real-Time Digital Signal Processing Framework*. (Audio Engineering Society [AES] Convention Paper 9196). Affiliated to the University of Rochester [NY] and Stanford University [CA]. Retrieved from: http://www.aes.org/e-lib/browse.cfm?elib=17519

Zhang, P. 2008. Psychoacoustics. In G. Ballou (Ed.). *Handbook for Sound Engineers*. 4th edition. Chapter 3. Oxford: Focal Press. 43-60.

Appendix A

Sound, Audio and Acoustic Figures

Figure A1

Microphone frequency response charts.

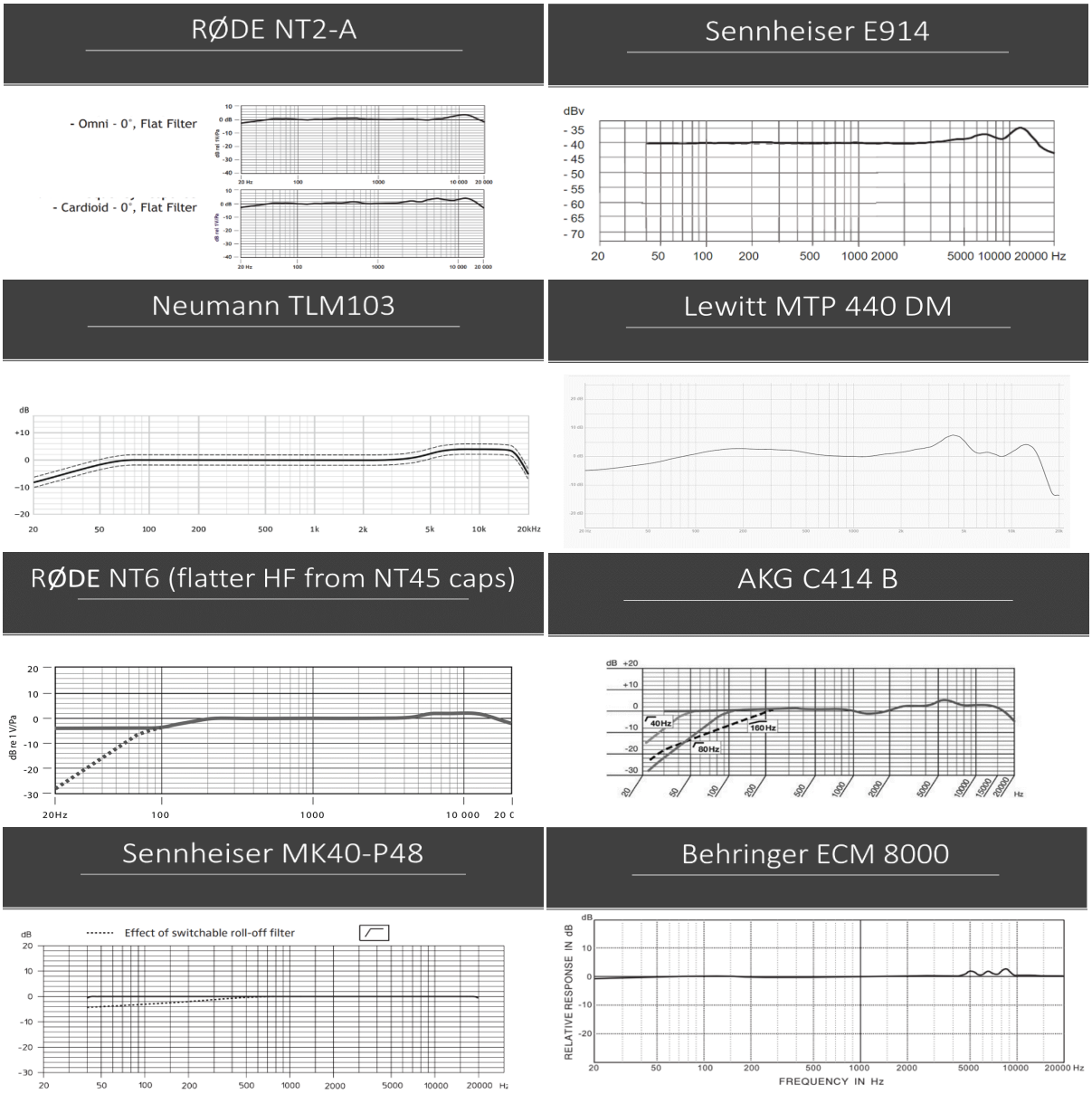


Figure A2

Full orchestra arrangement and frequency response of Welcome to the Show instruments.

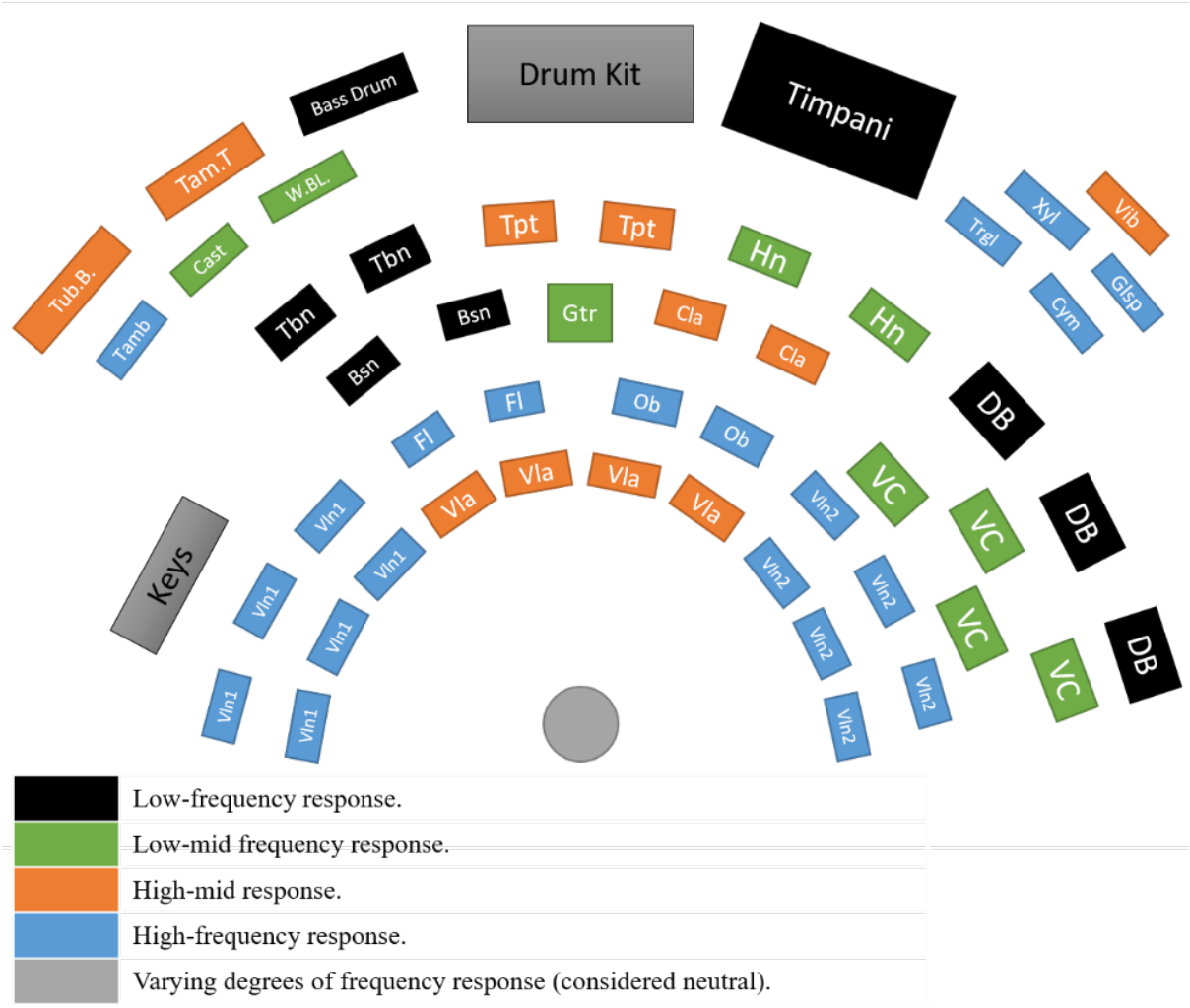
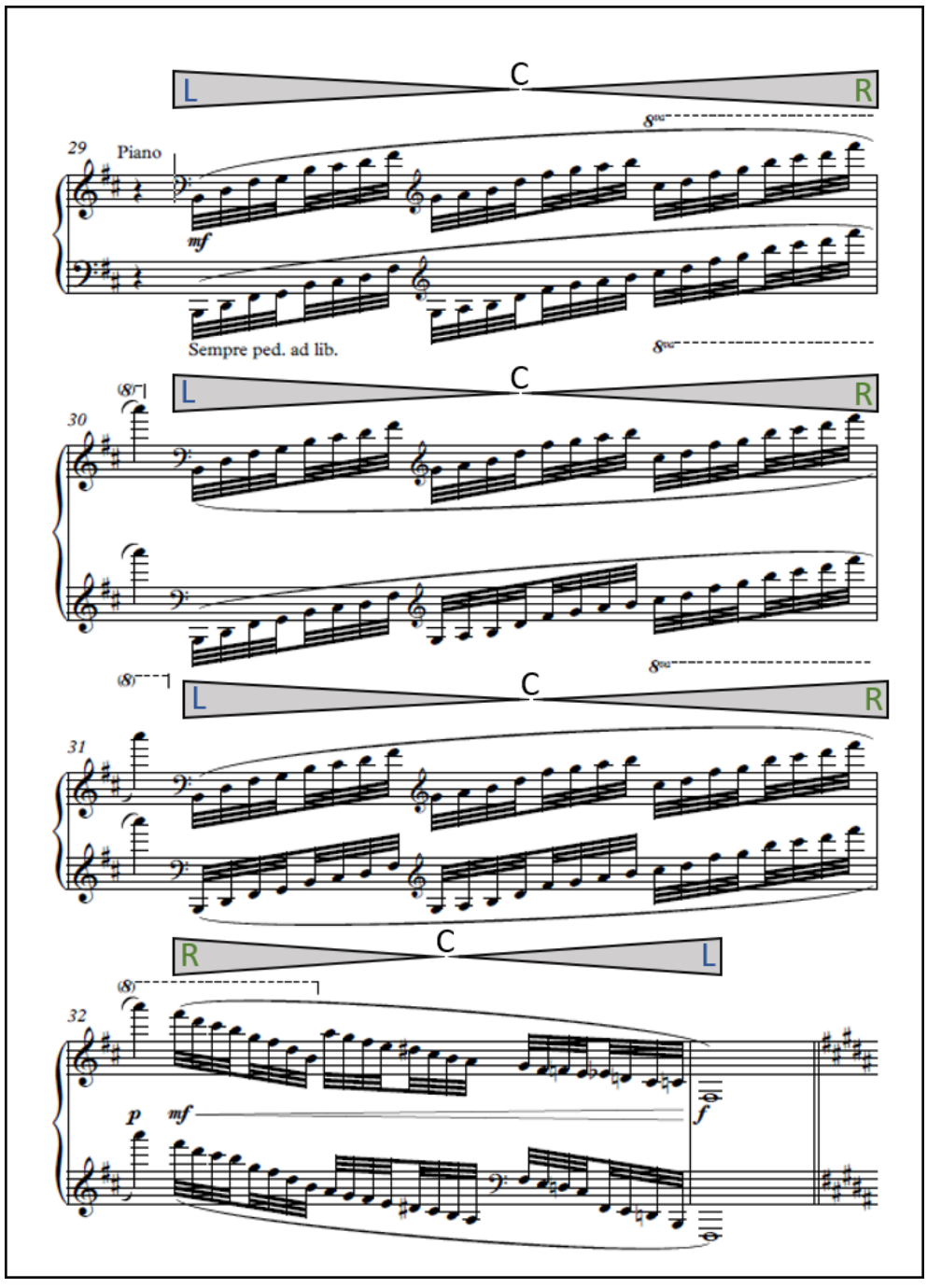


Figure A3

Panning example.



M.P. Anderson, *Entr'acte, Welcome to the Show*. Piano part. Adagio con moto, bars 29 – 33 (act 2, scene 1)

Figure A4

Equalising towards the inverted 60-phon Fletcher-Munson curve

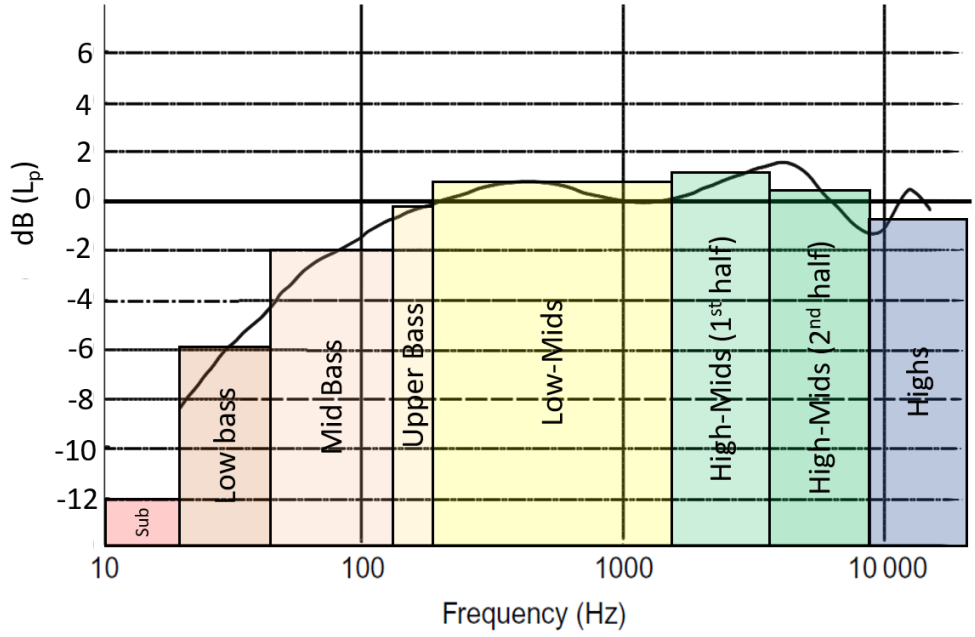


Figure A5

Reduced orchestra wit stereo array.

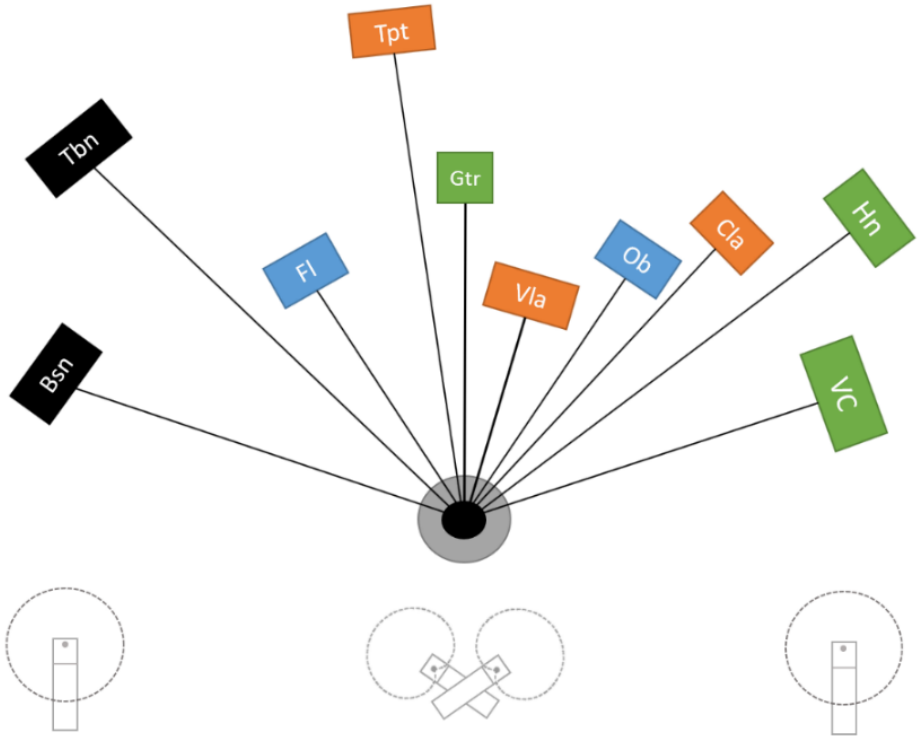
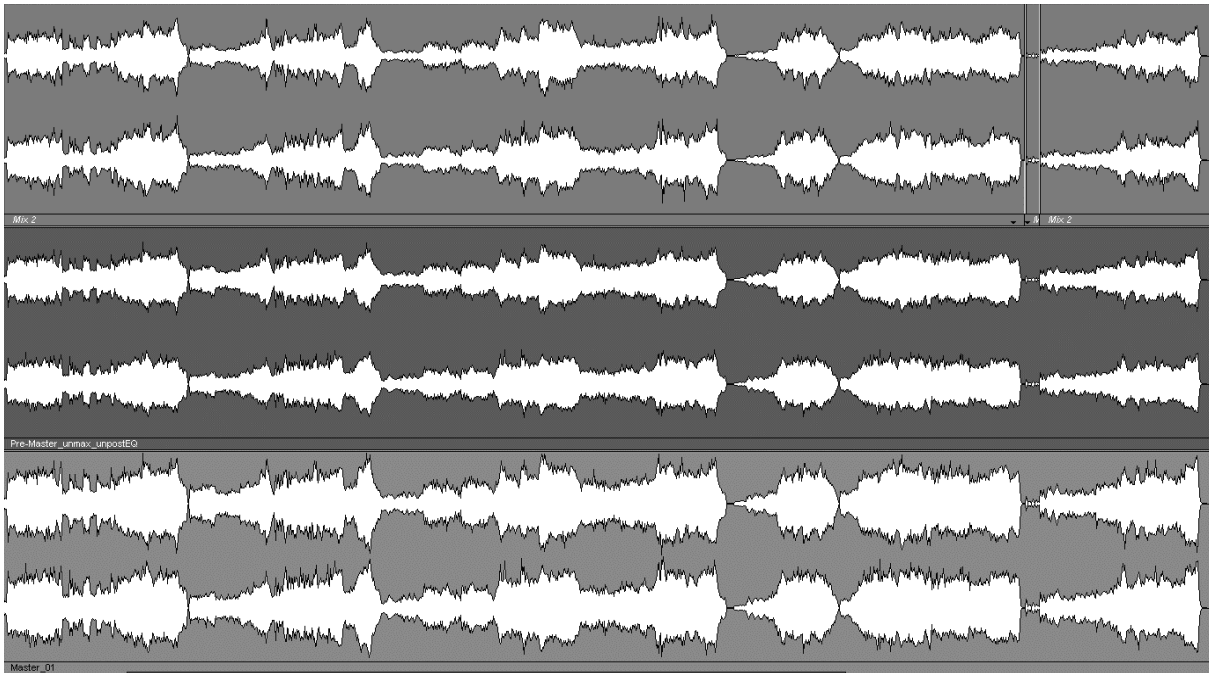


Figure A6

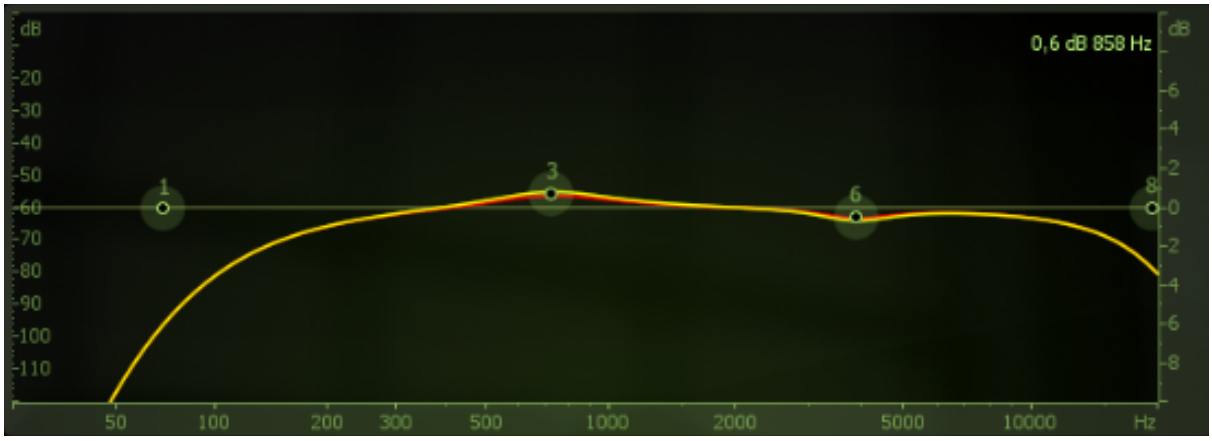
Dynamic changes in mastering.

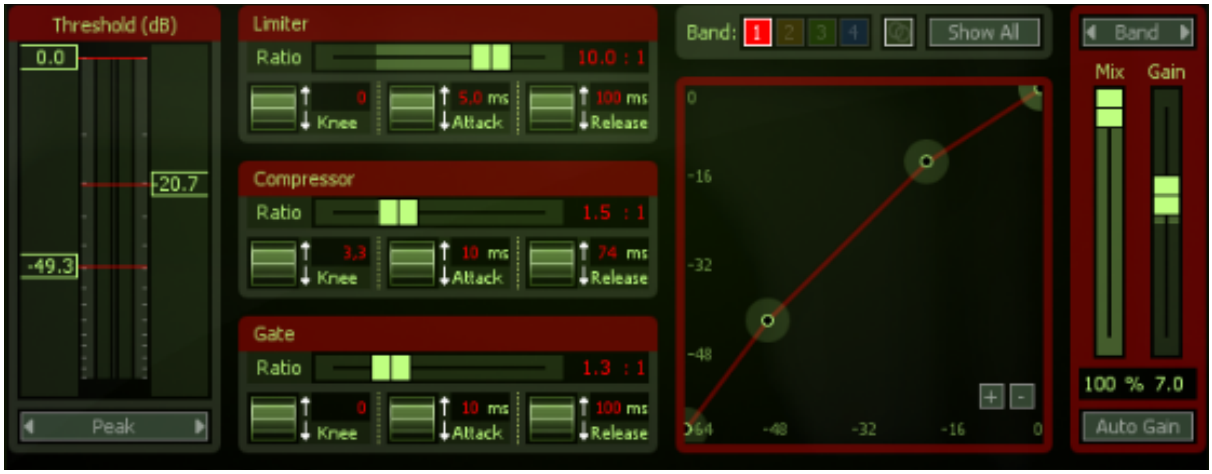


Mix
Master Multiband Compressor – No Automation, No post-EQ, No Maximizer
Final Master

Figure A7

Benchmark adaption tools





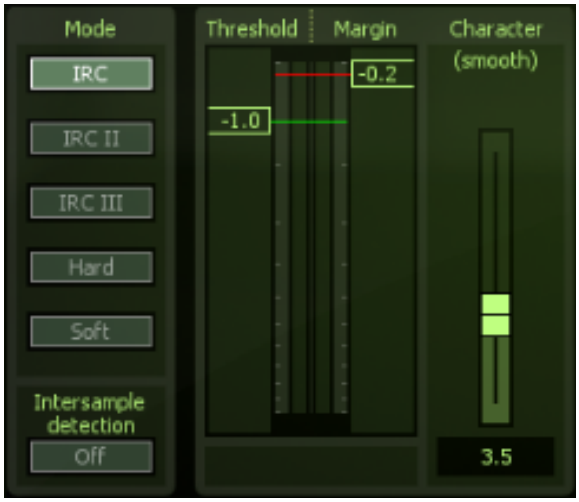
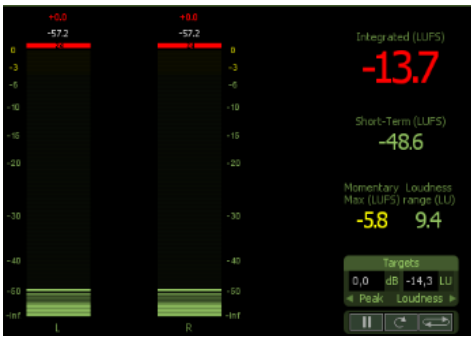


Figure A8

Graphical dynamic and spectral results from the first-order adaption

Figure A8a: Masker



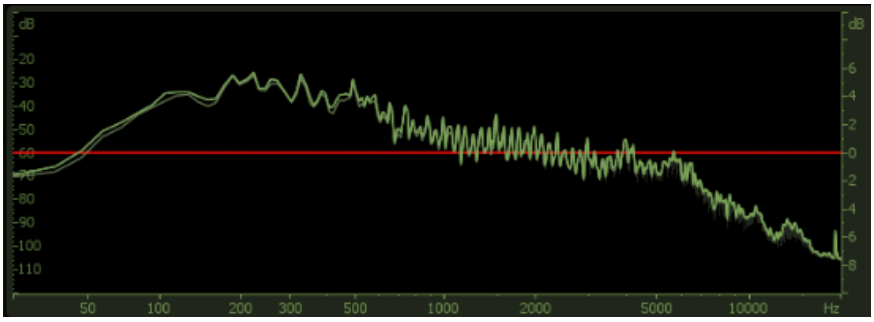
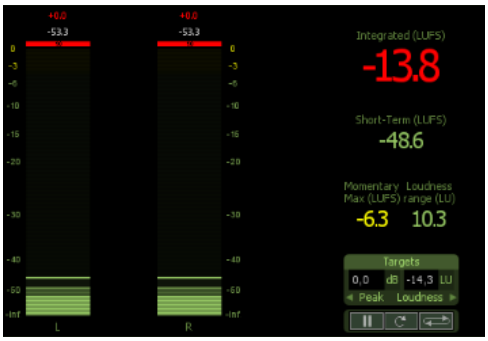


Figure A8b: Musaion



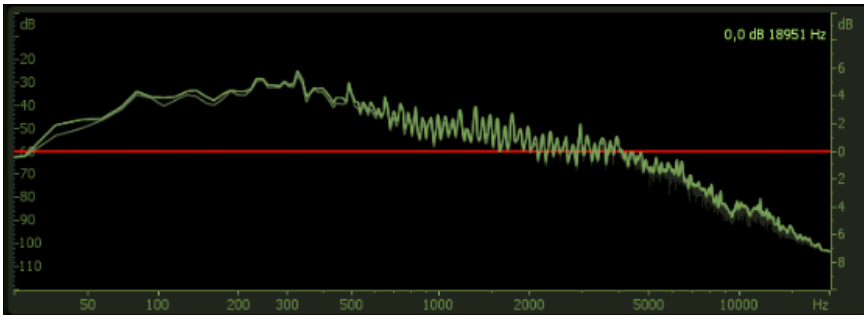
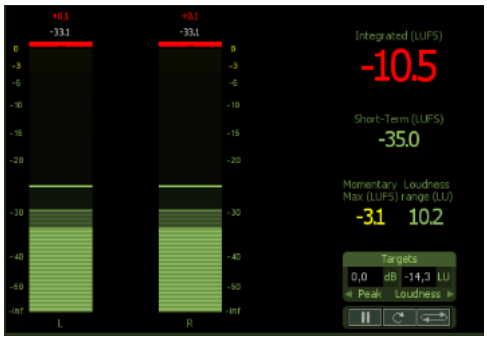
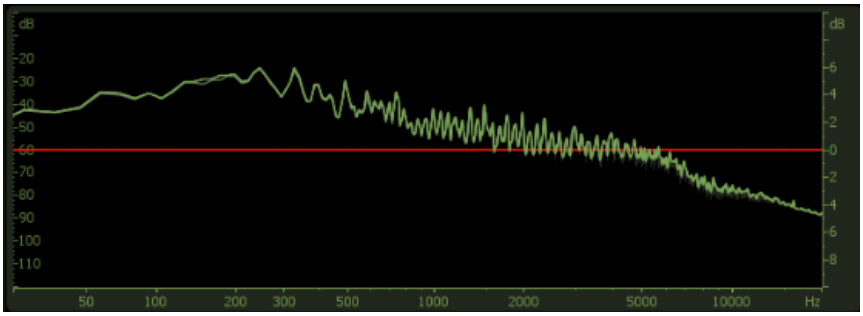


Figure A8c: Aula





Appendix B

Tables

Table B1:

Instrumentation of the full arrangement of Welcome to the Show.

Section	Instrumentation	Number of performers (43)
Strings (24)	First violin	6
	Second violin	6
	Viola	4
	Cello	4
	Double bass	3
	Classical guitar (amplified)	1
Woodwinds (8)	Flute	2
	Clarinet (Bb and A)	2
	Oboe	2
	Bassoon	2
Brass (6)	Trumpet (Bb and C)	2
	Trombone	2
	French horn (F)	2
Percussion (3)	4x Timpani (32”, 30”, 25”, 23”) Concert bass drum Xylophone Glockenspiel Vibraphone Tambourine Tam-tam (at least 36”) Wood block Tubular bells Castanets	3 (Split instrumentation)
Drum kit (1)	5-piece minimum: - Full set of chimes	1
Keyboard and synthesizer (1)	Piano Cathedral organ GS wavetable synth (FX)	1

Sound effects (FX)	Stage FX: - Gun shot - Spotlight - Footsteps - Heartbeat - Metal cage rattling - Beastly growls - Mouse squeak - Bones crushing Musical FX: - Tom-tom (field drum) - Brooms - Boxes - Tables - Chairs	Controlled by engineer, conductor, actor or keyboardist. (Imbedded into the score) Performed by cast (Written in the score)

Table B2

Adaption Statistics

Statistics	Raw Mix Track		Master Track		Studio (Benchmark)		Masker (Small)		Musaion (Medium)		Aula (Large)	
	Input Signal		**Input Signal**		**Before**	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**
Room RT_{60} (sec)	/		/		≈ 0,12	/	≈ 1,02	/	≈ 1,48	/	≈ 0.80	/
Ambient Noise (dBA)	/		/		34,32	/	-40.80	/	36,89	/	42,52	/
L/R Peak Output (LUFS)	-1,9	-1,4	0,1	0,1	0.0 (mono)	-0.1 (mono)	-0.1 (mono)	-0.1 (mono)	-0,1 (mono)	-0.1 (mono)	-0.1 (mono)	-0.1 (mono)
Input Gain (%)	/		/		40%	/	35%	28%	50%	34%	35%	23%
Integrated Output L_p (LUFS)	-17,7		-14,1		-17,5	-14,1	-19,1	-14,1	-15,5	-14,1	-19,1	-14,1
Compound Decay Time (sec)	≈ 0,95		≈ 0,95		≈ 1,07	≈ 1,07	≈ 1,97	≈ 1,97	≈ 2.43	≈ 2,43	≈ 1.65	≈ 2,00
Dynamic Range (dB) (minimum - maximum)	13,1		10,6		11,0	7,7	9,3	10,1	10,5	10,8	9,3	11,5
HF band level (dB)	-4,1		-3,9		-3,1	-3,1	-5,3	(-3,5)	-4,3	(-3,2)	-3,8	(-3,1)
HMF band level (dB)	-0,9		0,4		0,1	0,6	-1,1	(0,6)	-0,3	(0,5)	-0,6	(0,5)
LMF band Level (dB)	2,1		3,1		2,5	3,1	2,1	(3,0)	2,8	(3,0)	2,7	(3,0)
LF band Level (dB)	3,7		3,3		3,8	3,3	2,5	(3,2+)	4,0	(3,2+)	4,2	(3,2+)

Table B3

Expanded statistics of master and cycled studio tracks.

Recording Environment Acoustic Statistics	
Critical Distance (meters)	1.57
RT60 (secs)	0.76
Ambient Noise (dBA)	-37.95
Attack delay (secs)	0.35

Table B4

Expanded statistics of master and cycled studio tracks

Statistic	**Master Track**	**Cycled Track**	**Difference**
Integrated Output level (LUFS)	-14.1	-17.5	-3.4
L/R Peak (LUFS)	-0.1	0.0	+0.1
Momentary Max Level (LUFS)	-8.1	-11.2	-3.1
Maximum RMS Power (LUFS)	-9.54	-7.99	+1.55
Average RMS Power (LUFS)	-17.04	-17.42	+0.38
Dynamic Range (dB)	10.6	11.00	-0.4
Average HF band level (dB)	-3.9	-3.1	+0.8
Average HMF band level (dB)	0.4	-0.1	+0.5
Average LMF band level (dB)	3.1	2.6	-0.7
Average LF band level (dB)	3.3	3.8	-0.5

Table B5

Statistics of first-order adaption.

Statistic	Benchmark	Masker		Musaion		Aula	
		Basic	FOA	Basic	FOA	Basic	FOA
Integrated Output level (LUFS)	-14,1	-19,1	-13.7	-15,5	-13.8	-19,1	-10.5
L/R Peak (LUFS)	-0.1	-0.1	+0.0	-0.1	+0.0	-0.1	+0.1
Dynamic Range (dB)	7,7	9,3	9.4	10,5	10.3	9,3	10.2
Average HF band level (dB)	-3,1	-5,3	-4.2	-4,3	-4.2	-3,8	-3.2
Average HMF band level (dB)	0,6	-1,1	-0.2	-0,3	-0.1	-0,6	0.1
Average LMF band level (dB)	3,1	2,1	2.6	2,8	2.9	2,7	2.4
Average LF band level (dB)	3,3	2,5	1.8	4,0	3.0	4,2	3.3

Figure 4.4: Statistics of first-order adaption

Table B6

Statistics of second-order adaption

Statistic	Benchmark	Masker		Musaion		Aula	
		FOA	SOA	FOA	SOA	FOA	SOA
Integrated Output level (LUFS)	-14,1	-13.7	-14.1	-13.8	-14.1	-10.5	-14.1
L/R Peak (LUFS)	-0.1	+0.0	-0.1	+0.0	-0.1	+0.1	-0.1
Dynamic Range (dB)	7,7	9.4	9.4	10.3	10.7	10.2	10.7
Average HF band level (dB)	-3,1	-4.5	-3.8	-4.2	-3.2	-3.2	-3.3
Average HMF band level (dB)	0,6	-0.2	0.2	-0.1	0.5	0.1	0.2
Average LMF band level (dB)	3,1	2.6	2.7	2.9	3.0	2.4	3.0
Average LF band level (dB)	3,3	1.8	3.1	3.0	3.3	3.3	3.3

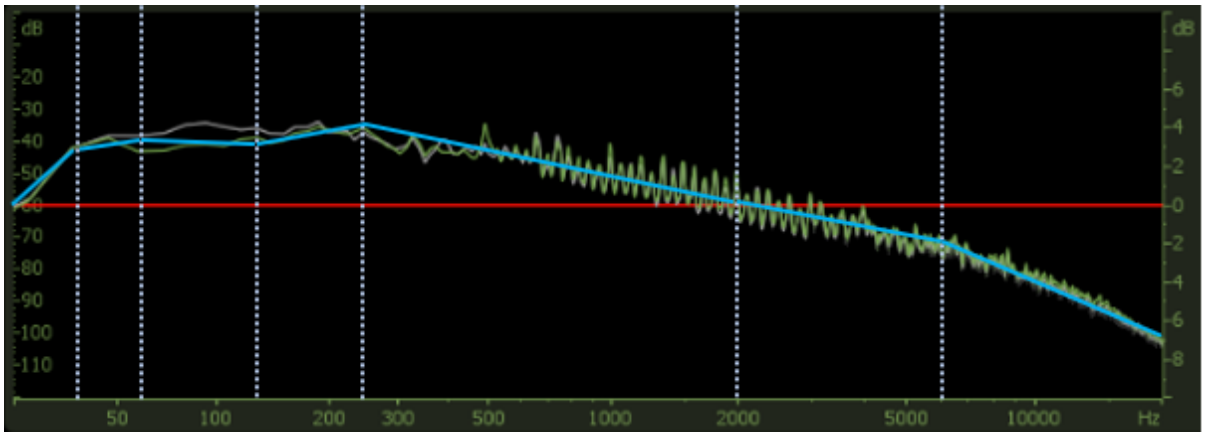
Appendix C
Spectrum Analysis Graphs

The spectrum graphs in this appendix are graphical representations of the adaptive process. The vertical dotted lines represent the cross-over points of the frequency octave bands. The horizontal dashed lines represent the average loudness in each band, and the colour-coded solid lines show the average variations of intensity of the waveforms.

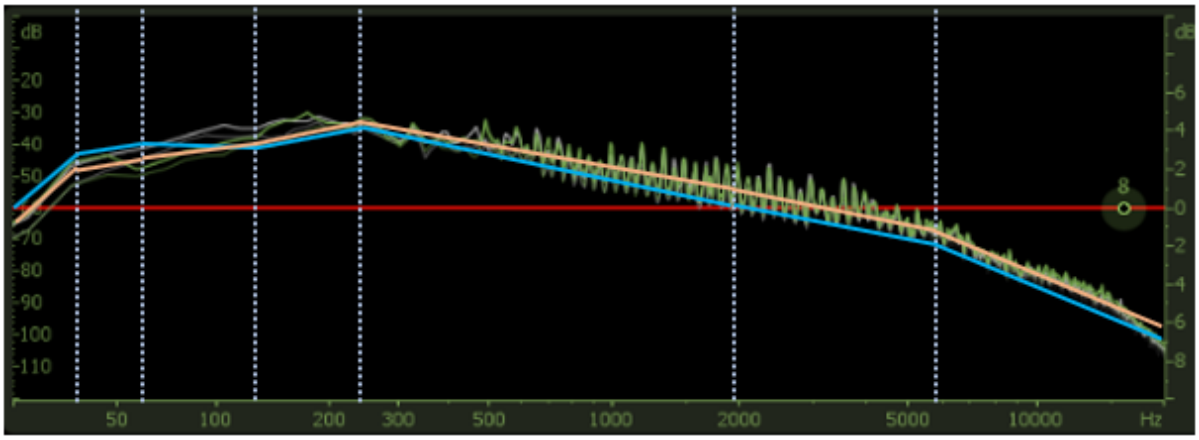


Graphs C1: Mastering Process

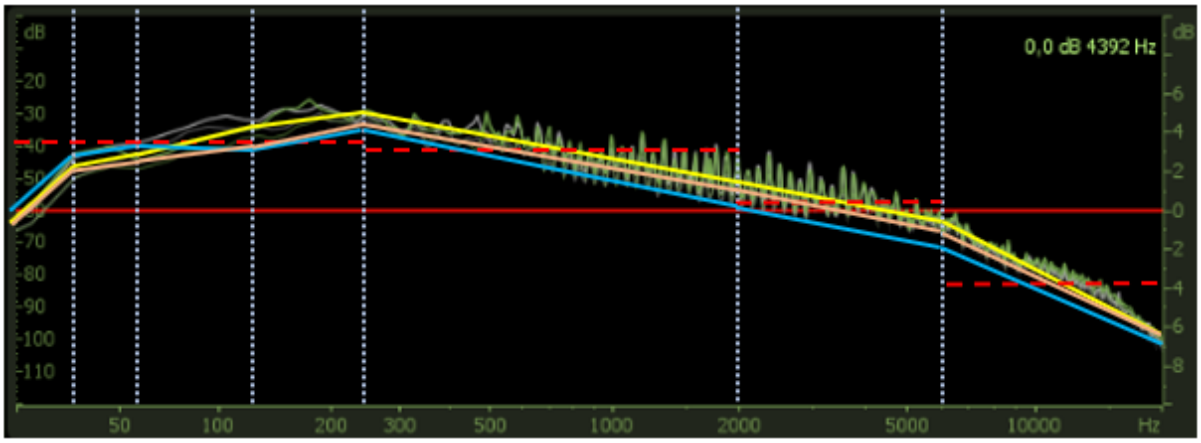
C1a: Final Mix



C1b: Master Multiband Compression (Automation, post-EQ and Maximizing excluded)

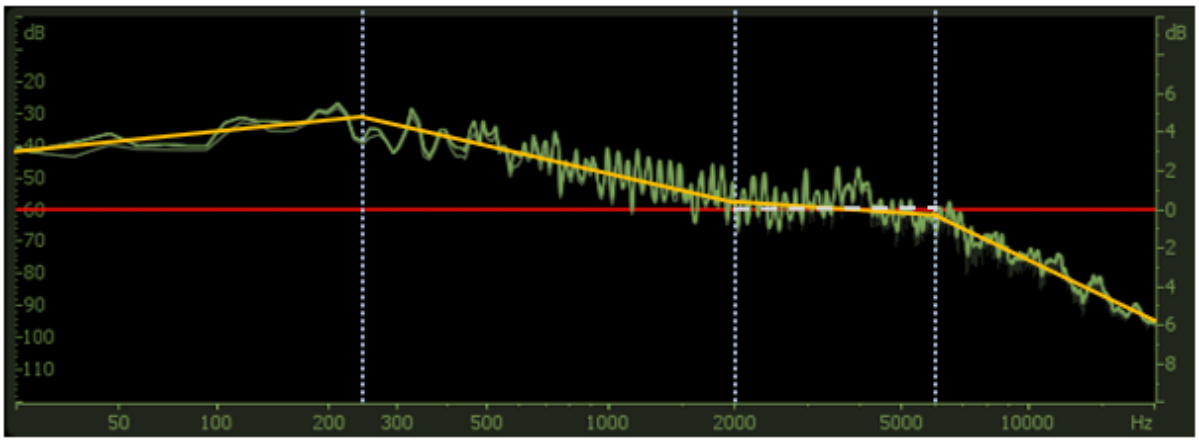


C1c: Final Master

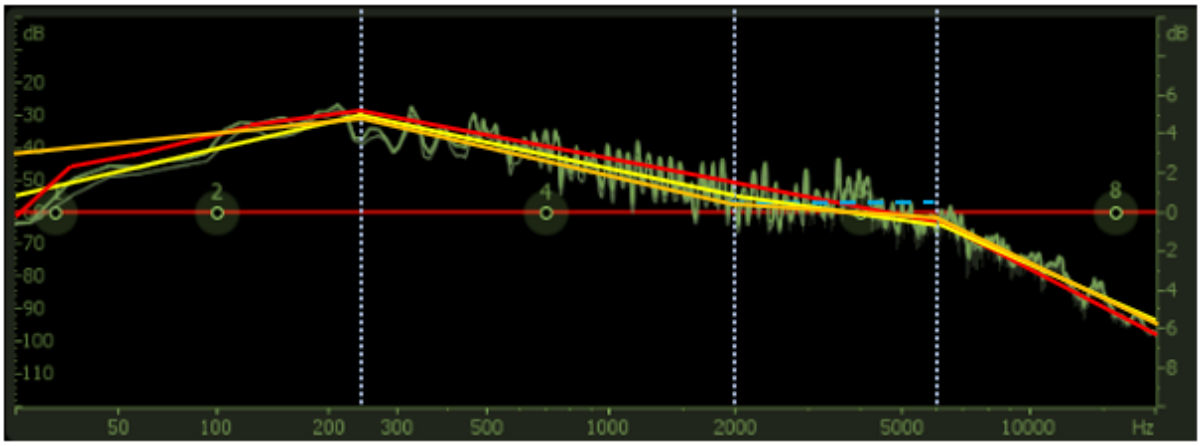


Graphs C2: Benchmark Creation

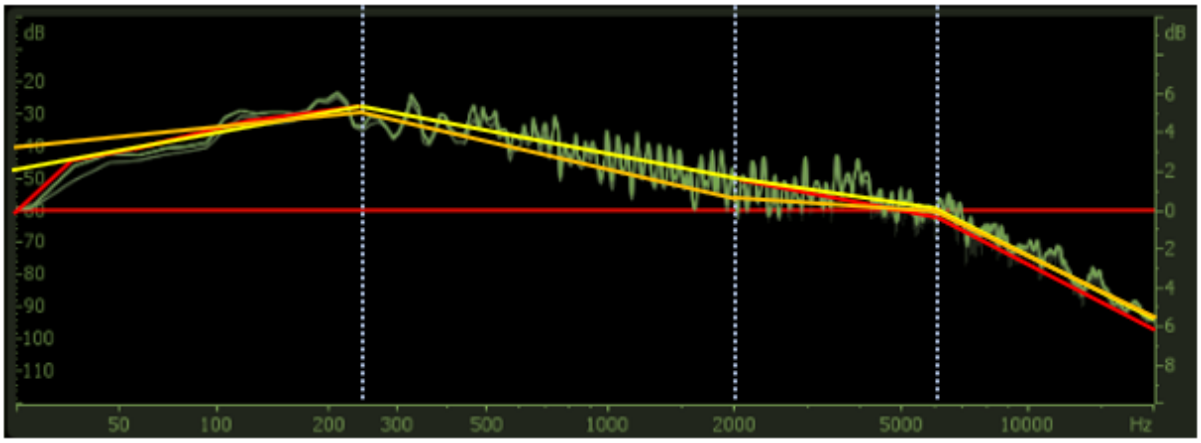
C2a: Unprocessed Cycled Track of Studio



C2b: Primary adaptions

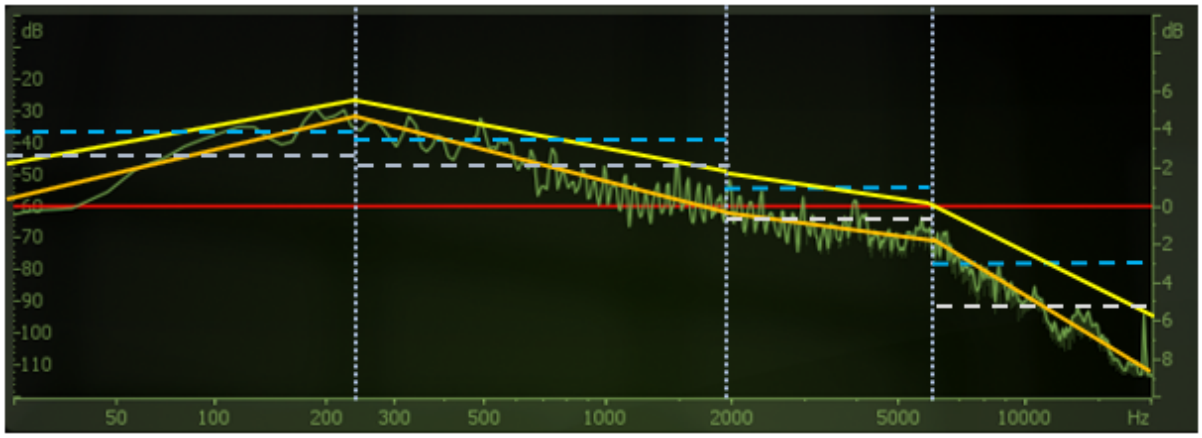


C2c: Final Benchmark Track

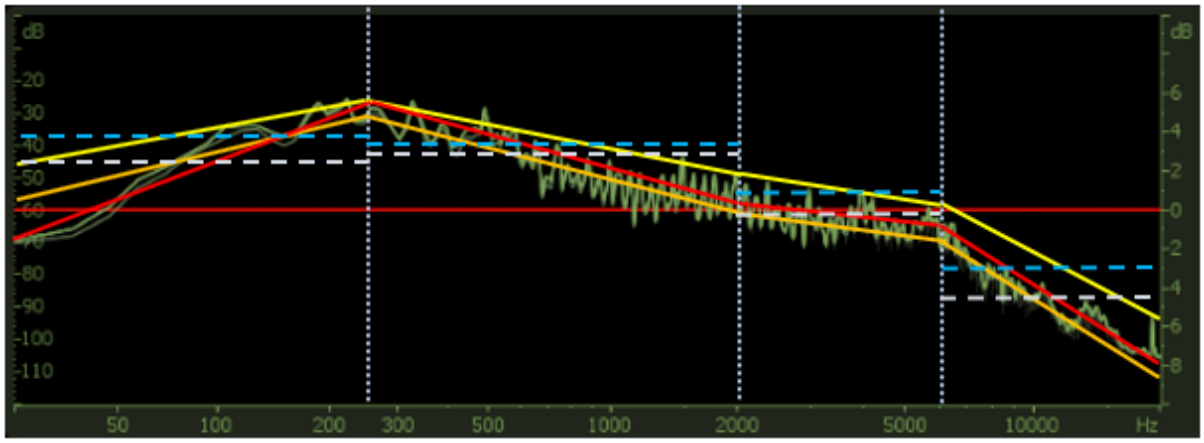


Graphs C3: Adaptions of the Masker

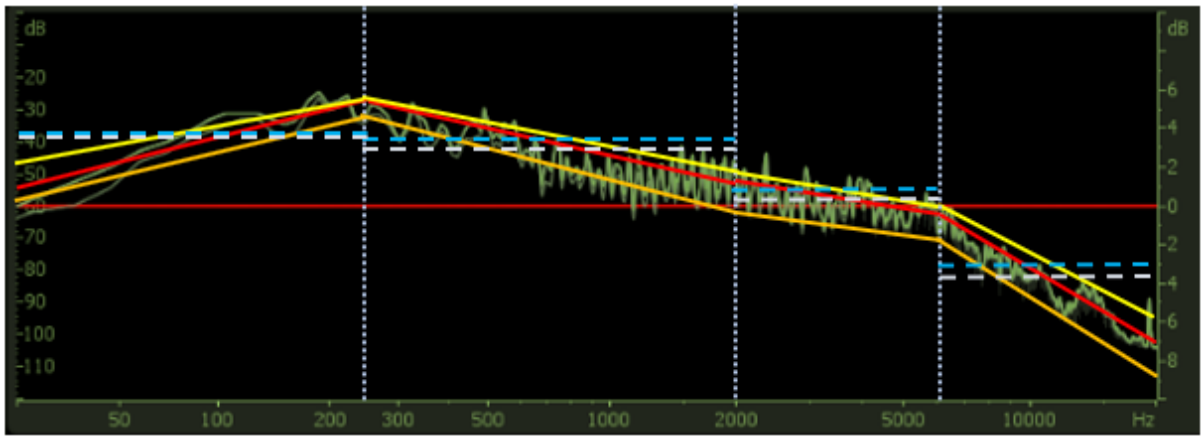
C3a: Unprocessed Cycled Track of Masker



C3b: First-Order Adaptions of Maker (IFM Filter and Maximized)

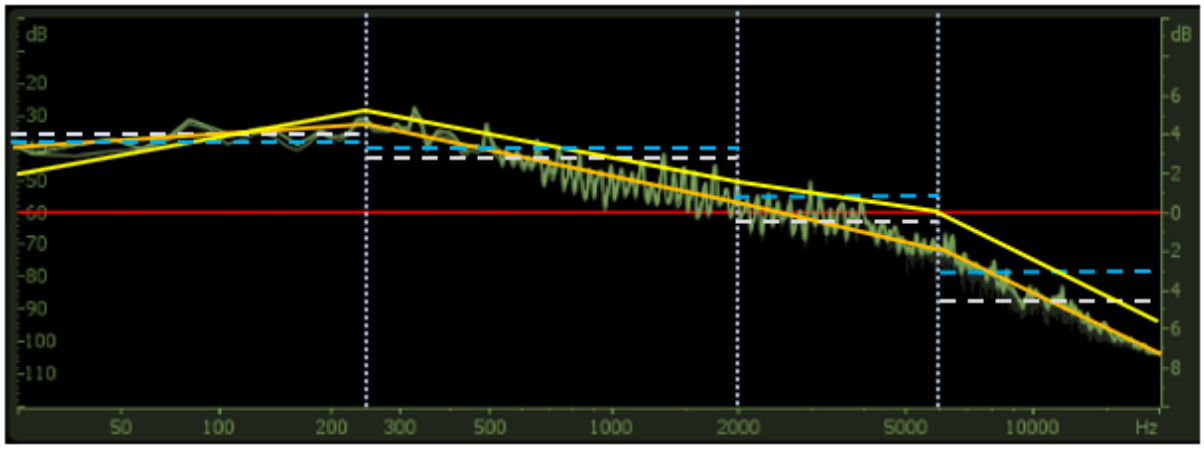


C3c: Second-Order Adaptions of Masker (Adjusted IFM Filter and Margin)

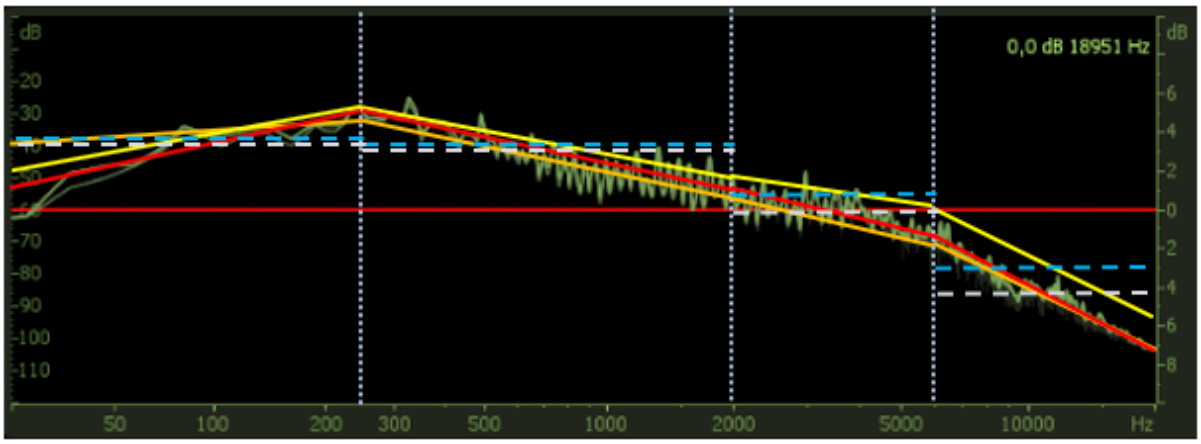


Graphs C4: Adaptions of the Musaion

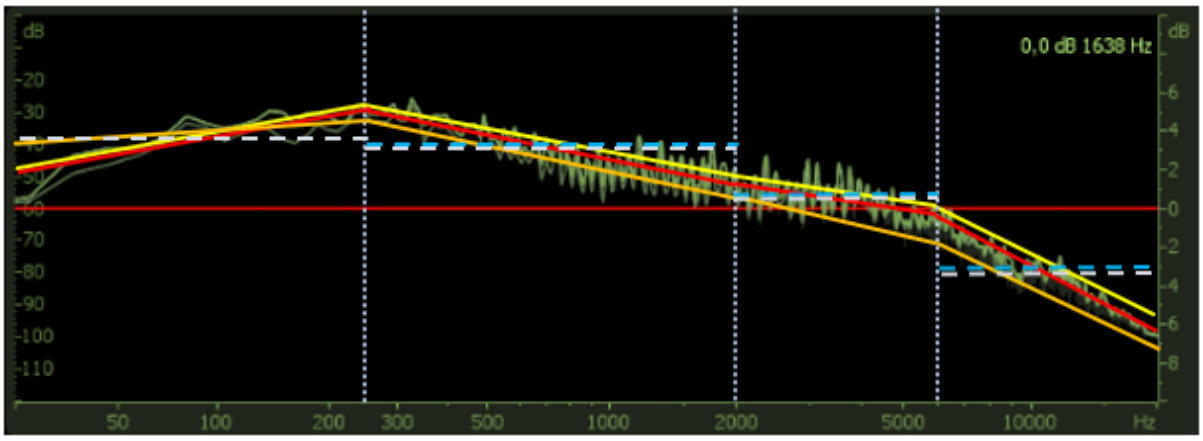
C4a: Unprocessed Cycled Track of Musaion



C4b: First-Order Adaptions of Musaion (IFM Filter and Maximized)

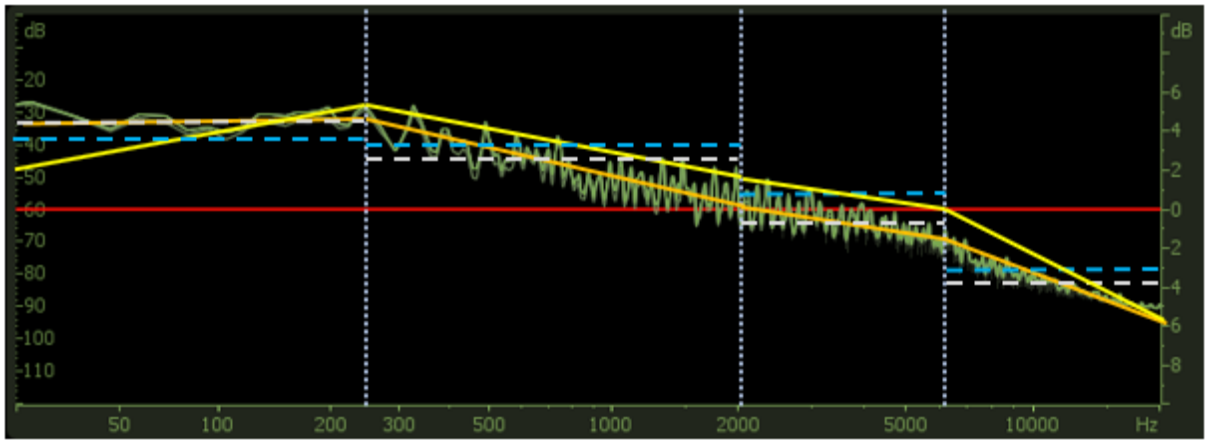


C4c: Second-Order Adaptions of Musaion (Adjusted IFM Filter and Margin)

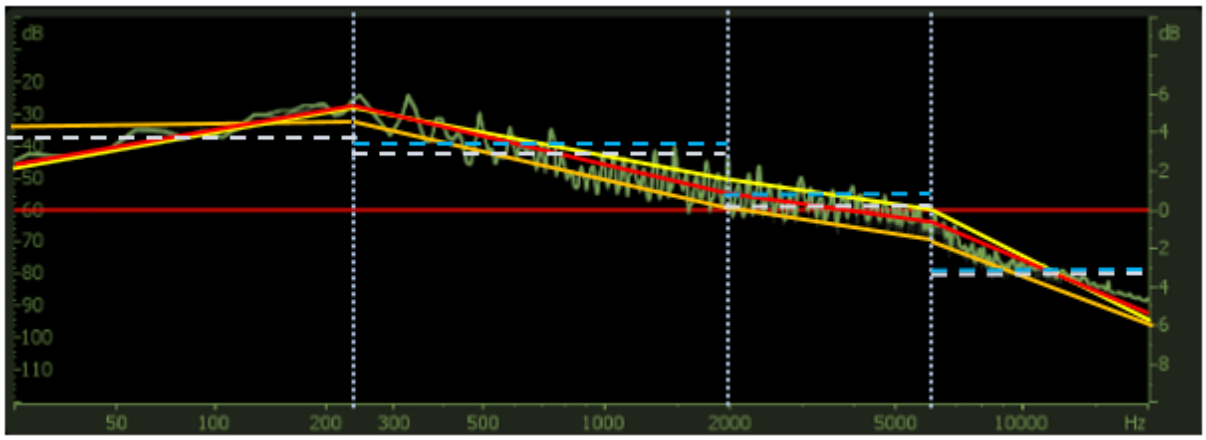


Graphs C5: Adaptions of the Aula

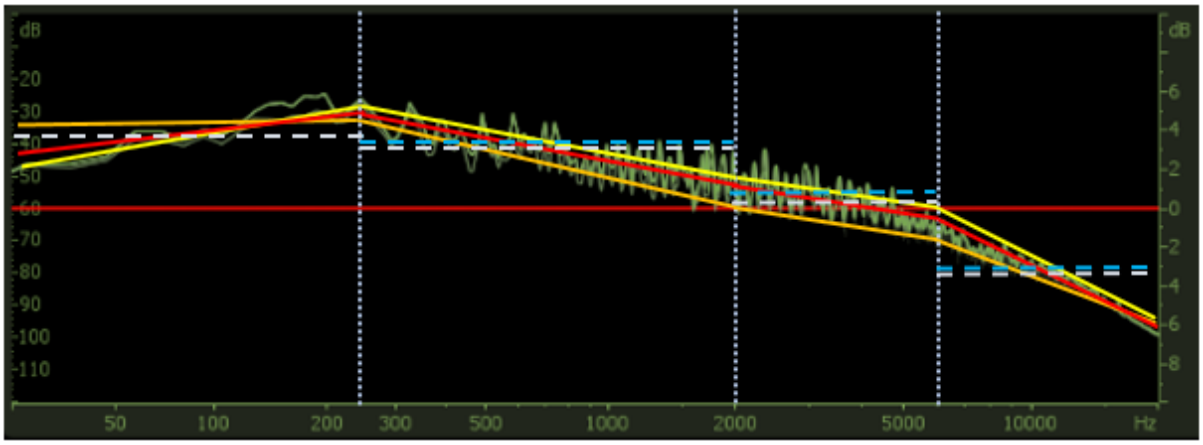
C5a: Unprocessed Cycled Track of Aula



C5b: First-Order Adaptions of Aula (IFM Filter and Maximized)

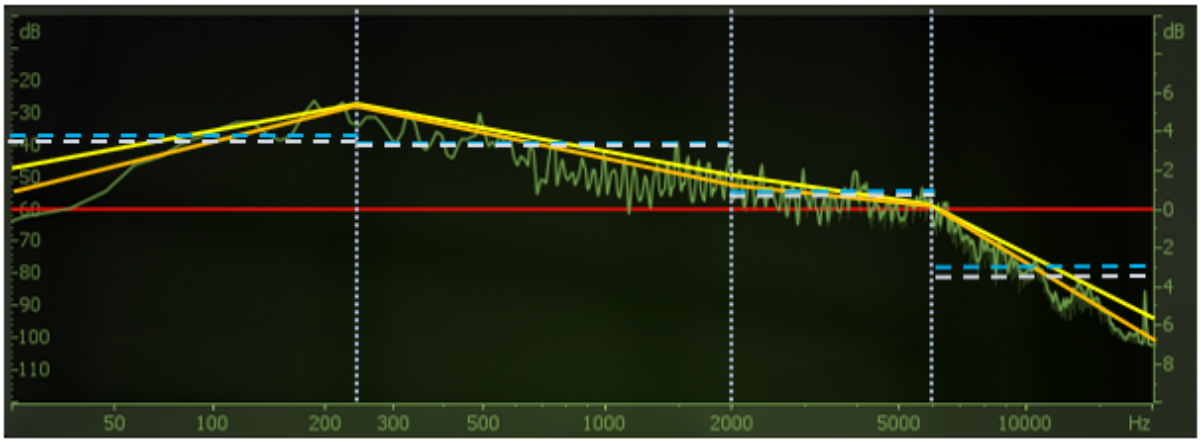


C5c: Second-Order Adaptions of Aula (Adjusted IFM Filter and Margin)

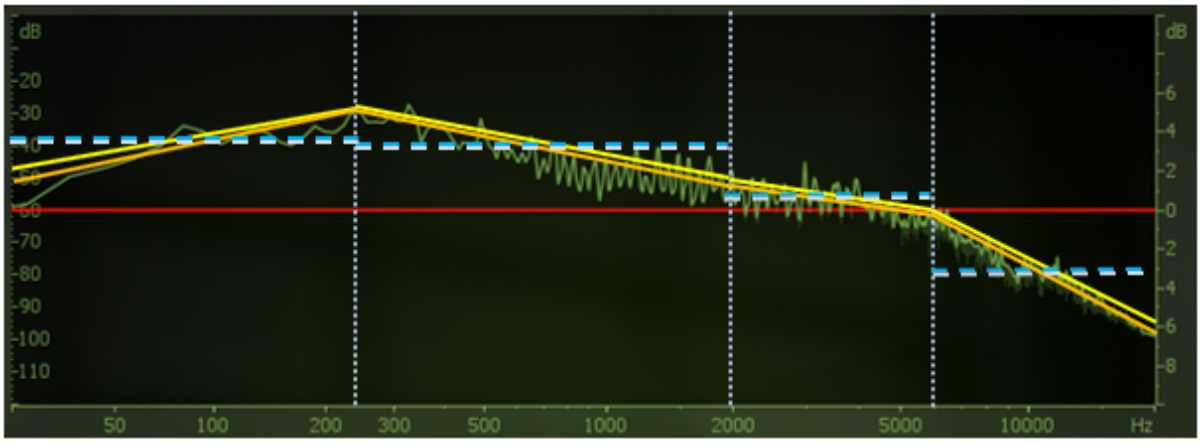


Graphs C6: Final Adapted Tracks (Third-Order Adaptions)

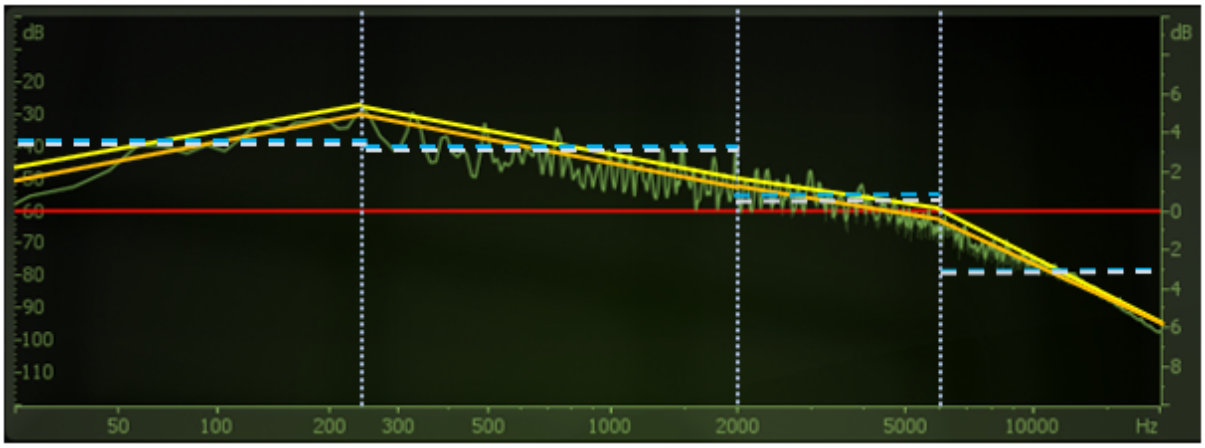
C6a: Adapted Masker



C6b: Adapted Musaion



C6c: Adapted Aula



Appendix D
Sound Sample Link

https://soundcloud.com/user-663745696/sets/mmus-sound-links

The link to the sound samples related to this dissertation is indicated above. These samples are for the reader to compare the different stages of the production and adaption.

01. Unmixed Recording_No Drumkit
02. Final Mix
03. Master_60-phon_No Exciter
04. Master_60-phon_with Exciter
05. Master_No Maximizer_No Post-EQ
06. Final Master
07. Cycled Studio_Normaliz[s]ed
08. Studio Benchmark_Adapted
09. Cycled Masker_ Normaliz[s]ed
10. Cycled Musaion_ Normaliz[s]ed
11. Cycled Aula_Normaliz[s]ed
12. SOA_Masker
13. SOA_Musaion
14. SOA_Aula
15. Final Adaption_Masker
16. Final Adaption_Musaion
17. Final Adaption_Aula

Please note that “06. Final Master” is embedded with ISRC. Any duplication, adaption or arrangement of the compositions contained in this musical require the written consent of the publisher or composer. Additionally, no part of the script or score may be photocopied or reproduced in any way without permission. Unauthorized use is an infringement of the South African Copyright Act No 98 (1978) and its amendments, and is punishable by law.

List of Figures

Figure 1.1: *Data collection* 5
Figure 1.2: *Practice-led research design* 6
Figure 1.3: *True-experimental research design* 7

Figure 2.1: *Sound propagation* 11
Figure 2.2: *Sound pulses* 12
Figure 2.3: *Human hearing* 13
Figure 2.4: *SLM weightings* 13
Figure 2.5: *The thrust theatre* 14
Figure 2.6: *The proscenium theatre* 15
Figure 2.7: *The inverse square law for sound intensity* 16
Figure 2. 8: *Absorption and diffusion* 19
Figure 2.9: *Reverberation* 19
Figure 2.10: *Frequency perception* 22
Figure 2.11: *Stereo arrays* 28
Figure 2.12: *Mixing actions* 34
Figure 2.13: *Fletcher-Munson and music frequency range* 37
Figure 2.14: *Inverted 60 Phon Fletcher-Munson contour (EQ Curve)* 37
Figure 2.15: *Stereo track of suspended cymbal in absolute mono and enhanced stereo* 38
Figure 2.16: *Uncompressed input signal peak (-1 dB) with a set threshold (-2 dB)* 41
Figure 2.17: *Dynamic processors using a ratio of 1:2 or 2:1* 41
Figure 2.18: *Analogue signal curve and digital signal steps* 44
Figure 2.19: *Audio spectrum analysis example* 48
Figure 2.20: *Filter Windows* 50

Figure 3.1: *Room modes* 54
Figure 3.2: *Gated pink noise sound waves* 56
Figure 3.3: *RT_{60} example with a decay time of less than 4 seconds* 56
Figure 3.4: *RT_{60} example with a decay time of exactly 4 seconds* 56
Figure 3.5: *RT_{60} example with a decay time of more than 4 seconds* 57
Figure 3.6: *Reduction of the acoustic instruments* 61

Figure 3.7: *Microphone selection* 66
Figure 3.8: *0 dB panning law* 73
Figure 3.9: *-3 dB panning law* 73
Figure 3.10: *Dynamic schematic of subject material* 80
Figure 3.11: *second-order filter* 87
Figure 3.12: *Implemented inverted 60-phon Fletcher-Munson curve* 87
Figure 3.13: *Harmonic excitement of the frequency bands* 88
Figure 3.14: *Multi-band compression on master track* 89
Figure 3.15: *Maximizer/limiter used on the master track* 90

Figure 4.1: *Signal chain of the cyclic process* 94
Figure 4.2: *Linear spectrum analysis of theatre environments* 95

Glossary

48V

Phantom power: external power needed to a power types of microphone.

A, B and C Weightings

Ranges of sound pressure levels to account for the human ear (see chapter 2.3.1).

AAC

(Advanced Audio Coding) Uncompressed high-quality audio format.

AD converter

A device that converts analogue signal into digital signal.

ADSR

(Attack, decay, sustain and release) Frequency time that forms the audio envelop of a sound signal.

Bounce

Total or sectional exported audio.

DAW

(Digital Audio Workstation) Platform for audio processing.

dB

(Decibel) Measurement of sound pressure level or loudness.

dBA

A-weighted sound pressure level.

DSP

Digital Signal Processing.

Entr'acte

A piece of music between the first and second act of a musical or opera.

EQ

Equalisation.

FIR Filter

Finite Impulse Response Filter.

Hz or kHz

(Hertz or Kilohertz) Measurement of frequency.

IIR Filter

Infinite Impulse Response Filter.

IRC

Intelligent Release Control (see dynamic signal processing in chapter 3.3.2).

ISRC

International Standard Recording Code.

L_p/SPL

Sound pressure level.

LUFS

(Loudness Units below Full Scale) Level of encoding in a digital format.

XY, NOS, OTRF

Microphone stereo array configurations (see chapter 2.4.1).

Mega-Musical

A through-composed stage production with very little or no dialogue.

MIDI

(Musical Instrument Digital Interface) A digital communication language.

Mono

(Monaural) One Source.

Overdubbing

Multi-track layering. (see chapter 3.3.1).

Overture

Generally, the opening piece in a collection of works such as a musical or opera.

Pa or µPa

(Pascal or Micro-Pascal) Measurement of pressure.

PA system

Public Address (PA) systems are used to control the amplified sound using, mixing consoles and other related equipment.

Phon

A subjective scale based on pressure at different frequencies (see chapter 2.4.3).

PL

Programme Loudness.

Profiling

A multi-staged audio process in which a mixing engineer imitates the timbre from an acoustic instrument to a digital instrument.

RE and DG

Random Efficiency and Distance Factor (see chapter 2.4.1).

RT_{60}/$RT_{duration}$

The time in which reverberation takes to decay to a solid state of energy.

Sample

Pre-recorded sound (see chapter 3.2.3).

SLM

(Sound Level Metre) A device or program used to analyse sound in an environment.

Stereo

(Stereophonic) Two sound sources.

Underscore

Instrumental music with dialogue.

VSL

(Vienna Symphonic Library) Digital instrument sample library.

WAV

(Waveform Audio File Format) Uncompressed high-quality audio format