Digital Media Technology Lab
Birmingham City University

Timbral Analysis and Recording Parameter Transformations of Snare Drums

Matthew Cheshire

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, January 2023

Abstract

A snare drum is capable of producing a wide range of timbres influenced by playing technique, its physical construction, and the recording methods used. When a recording engineer configures drums and studio equipment, they adjust a plethora of real-world recording parameters to achieve the desired timbre. These recording parameters impart their own timbral properties by varying amounts, and in most cases the only way to modify these properties is to re-record the audio with changes applied to the real-world variables.

This thesis examines methods for computational transformations of snare drum recordings to elicit perceptual changes that mimic modification of real-world recording variables. This is achieved through four main investigations, presented throughout this thesis, two which cover timbral analysis of snare drum recordings, and two which explore post-hoc recording parameter transformations.

Strike velocity and microphone selection are factors known to affect snare drum timbre, the first study analyses timbral differences associated with snare drum strike velocity. Results show that listeners are able to distinguish between high and low velocity strikes using timbral cues alone, with microphone selection having no influence on this perceptual identification. Audio analysis reveals distinct temporal and spectral features, with higher velocity strikes producing greater energy in the lower mid-range and significantly longer decay times. The second study aims to demystify the subjective preference of different microphones for snare drum recording. For the majority of microphones, preference does not change between isolated strikes and those with the presence of *bleed* from the hi-hat and kick drum. On average, preference is higher for condenser microphones compared to dynamic. Additionally, spectral centroid and an objective measure of brightness positively correlate with subjective scores.

The ability to perceptually modify drum recording parameters in a post-recording process would be of great benefit to engineers limited by time or equipment. The first post-hoc recording parameter transformation study focuses on microphone selection, mapping the spectral features from highly-preferred microphones onto a microphone with less favourable timbral characteristics. This investigation also details the development and evaluation of a robotic drum arm for consistent strike velocity. Subjective assessment reveals that participants show no preferences between recordings from highly-preferred microphones and those from a transformed least-preferred microphone. The last study employs a data-driven approach for post-recording modification of dampening and microphone position. The system consists of a autoencoder that analyses an audio input and predicts optimal parameters of one or more third-party audio effects, which process the audio to produce the desired transformations. Two novel audio effects are proposed and compared against existing audio plugins. Perceptual quality of transformations is assessed through a subjective listening test and an object evaluation is used to measure system performance, positive results demonstrate a capacity to emulate snare dampening.

Acknowledgements

I would like to thank my supervisors, Jason Hockman and Ryan Stables, for all the many years of guidance and support both in and out of the university. You have both helped to shape and develop my research with your own unique approaches. Thank you to my friends and colleges: Alan Dołhasz, Carl Southall, Dominic Ward, Jake Drysdale, Muadh Al-Kalbani, Mattia Colombo, Niccolò Granieri, Nicholas Jillings, Sam Smith, Sean Enderby, and Tychonas Michailidis. At some point in time, each of you have offered me advice, been someone to talk to, helped solve a problem, inspired me, and generally made the PhD journey more enjoyable. Thank you also to Ian Williams, Cham Athwal, Izzy Ali-MacLachlan, past and present members of the Digital Media Technology Lab, and all the great academics I have met from other universities during my time as a PhD student. A special thank you to Michelle Saunders for all the emotional support, proof reading, and providing me with motivation when I needed it most. Also thank you Mum and Dad for your support throughout.

Contents

Abstract **i**

Acknowledgements **iii**

Acronyms **xiii**

1 Introduction **1**

1.1 Motivation 1

1.2 Research Questions 2

1.3 Thesis Structure 2

1.4 Contributions 4

2 The Snare Drum **5**

2.1 Introduction 5

2.2 History and Variations 5

2.3 The Modern Drum Kit 7

2.4 Timbre 8

2.5 Construction 10

2.5.1 Shell 10

2.5.2 Bearing Edge 13

2.5.3 Drumheads 15

2.5.4 Tension Rods and Lugs 17

2.5.5 Hoops 18

2.5.6 Snare Wires 19

2.5.7 Snare Beds 21

2.6 Tuning 22

2.7 Dampening 25

2.8 Drumsticks 26

2.9 Conclusions 28

3 Recording the Snare Drum **29**

3.1 Recording Drums 29

3.1.1 Microphones 32

3.2 Microphones Preference for Snare Drum Recording 36

3.2.1 Previous Microphone Comparison Studies 36

3.2.2 Experimental Design 37

3.2.3 Microphones . . . 38
3.2.4 The Snare Drum . . . 38
3.2.5 Recordings . . . 39
3.2.6 Audio Pre-Processing . . . 40
3.2.7 Listening Test . . . 41
3.2.8 Participants . . . 42
3.2.9 Results . . . 42
3.2.10 Ranking the Data . . . 44
3.2.11 Brightness . . . 44
3.2.12 Correlation . . . 45
3.2.13 Signal-to-Bleed Ratio (SBR) . . . 47
3.2.14 Conclusions . . . 50

4 Snare Drum Strike Velocity Differences 51
4.1 Introduction . . . 51
4.2 Listening Test . . . 52
4.2.1 Methodology . . . 53
4.2.2 Recordings . . . 54
4.2.3 Results . . . 56
4.3 Feature Extraction and Analysis . . . 56
4.3.1 Methodology . . . 57
4.3.2 Statistical Tests . . . 58
4.4 Discussion . . . 60
4.5 Conclusions . . . 61

5 Microphone Transformation 63
5.1 Introduction . . . 63
5.2 Equalisation . . . 64
5.2.1 Drum Mixing . . . 67
5.3 Microphone Transformation . . . 70
5.4 Methods . . . 70
5.4.1 Microphone Selection . . . 70
5.4.2 Snare Drums . . . 70
5.4.3 Recording . . . 72
5.5 Robotic Drum Arm . . . 72
5.5.1 Construction . . . 72
5.5.2 Drum Arm Evaluation . . . 73
5.5.3 RDA Conclusion and Improvements . . . 74
5.6 Listening Test . . . 76
5.6.1 Methodology . . . 76
5.6.2 Results . . . 77
5.6.3 Pairwise Comparison . . . 77
5.7 Spectral Modification . . . 79
5.7.1 Frequency Response Analysis and Equalisation . . . 79
5.8 Pre- and Post-EQ AB Tests . . . 82

5.8.1 Methodology . . . 82
5.8.2 Results . . . 82
5.9 Conclusions . . . 83

6 Deep Audio FX for Snare Drum Recording Transformations 85
6.1 Introduction . . . 85
6.1.1 Background . . . 85
6.1.2 Motivation . . . 86
6.2 Methodology . . . 87
6.2.1 Network Architecture . . . 87
6.2.2 Audio Effects . . . 88
6.2.3 Network Training . . . 92
6.3 Snare Drum Data-Set (SDDS) . . . 92
6.3.1 Background . . . 93
6.3.2 Recording Configuration . . . 93
6.3.3 Dampening . . . 94
6.3.4 Performance . . . 98
6.3.5 SDDS Discussion and Reflections . . . 98
6.3.6 Sub-Sets of SDDS . . . 99
6.4 Evaluation Methodology . . . 99
6.4.1 Subjective Evaluation . . . 100
6.4.2 Timbral Reconstruction Metrics . . . 100
6.5 Results . . . 102
6.5.1 Subjective Results . . . 102
6.5.2 Objective Results . . . 104
6.6 Discussion . . . 106
6.7 Conclusions . . . 107

7 Conclusion 109
7.1 Timbre Analysis . . . 110
7.2 Microphone Comparisons . . . 111
7.3 Recording Parameter Transformations . . . 112
7.4 Further Work . . . 113

Appendices 129

A Publications 129

List of Figures

2.1 Example of a modern snare drum, showing batter side. 10
2.2 Example of different shape bearing edges. From left to right: 45° single, 45° double, 45° double with slight round-over, and round-over. (Courtesy Thomann). 14
2.3 Example of different lug styles. 18
2.4 Triple-flanged hoop design. 19
2.5 Different size snare wires, 20 strand (left) and 42 strand (right). 21
2.6 Example of different dampening methods. From left to right, moongel, dampening bracket, leather wallet, and tape and tissue paper. 26

3.1 Example of a standard modern drum kit configuration. 29
3.2 Microphone configuration showing close microphones, stereo and a mono overhead microphones. . 31
3.3 Close microphone placement on the batter and resonant head of the snare drum. 32
3.4 Different polar patterns, left to right: cardioid, hypercardioid, figure-of-eight, omnidirectional. (Courtesy Lewitt-Audio). 35
3.5 Calibration of the digital DrumDial used for tuning. 39
3.6 Position of M201 with triangle jig. 40
3.7 Score for drum beat used in Hits With Bleed recording experiment. 41
3.8 Shure SM57 recordings for Single Hits (Upper) and Hits With Bleed (Lower). 41
3.9 Mean score (x) and standard deviation (horizontal lines) for Single Hits listening test. 43
3.10 Mean score (x) and standard deviation (horizontal lines) for Hits With Bleed listening test. . . . 43
3.11 Spectral centroid and mean rank for Single Hits, with regression line. 45
3.12 Condenser brightness and condenser mean rank for Single Hits, with regression line. 46
3.13 Change in brightness and change of mean rank, with regression line. 47
3.14 Cardioid signal to bleed ratio and mean rank, with regression line. 48

4.1 Stimuli of Beta57 used for listening test, low velocity (Upper) and high velocity (Lower). 52
4.2 Interface used for the AB listening test. 53
4.3 Average frequency response of left and right channels of the AKG K240 headphones used in the listening test, shown with $\frac{1}{6}-$octave smoothing. 54
4.4 Snare drum with RØDE NT55 microphone (left) and Cirrus CK:162C SPL meter (right). 55
4.5 Recording of low velocity strike (100dBZ) and high velocity strike (125dBz). 55
4.6 Means and standard deviations of Bark band magnitudes for high and low velocity strikes. 58
4.7 Mean and standard deviation of normalised envelopes for high and low velocity strikes. 59
4.8 3D surface of spectrograms for high and low velocity snare strikes. 59

5.1 Example of a 30-band graphic equaliser. (Courtesy Klark Teknik). 65
5.2 Example of a 10-band graphic equaliser. (Courtesy Red Rock Sound). 65
5.3 Mäag Audio EQ4 with fixed frequency filters. (Courtesy uaudio.com). 66
5.4 Solid State Logic (SSL) AWS δelta 4-band equaliser. 66
5.5 Recording setup demonstrating robotic drum arm, triangular jig, and microphone. 73
5.6 Signal flow diagram of RDA configuration. 74
5.7 Striking position of the drum stick on the MIDI drum pad. 75
5.8 RDA and MIDI drum pad evaluation setup. 75
5.9 Signal flow of evaluation procedure. 76
5.10 Web Audio Evaluation Tool (WAET) interface used for multi-stimuli listening test. 77
5.11 Results from pairwise comparison test with microphones ordered by mean rank. Yellow squares indicate pairs of microphones that are significantly different from one other (p<0.05) and blue squares indicate pairs that are not significantly different (p>0.05). 78
5.12 Gain values for 30-band EQ applied to $\mathrm{D5}$ recording to minimise spectral difference with those of the Category-1 microphones. 81
5.13 Spectral difference between $\mathrm{NT55}$ and $\mathrm{D5}$ recordings of the Velvetone snare pre- and post-EQ. . 81
5.14 Results from AB test with 95% confidence interval. Top plot shows a comparison between the *unmodified* $\mathrm{D5}$ and Category-1 microphone recordings; Bottom plot shows *modified* $\mathrm{D5}$ and Category-1 microphone recordings. 83

6.1 System overview for snare dampening with DeepAFx with third-party audio effect. Solid lines depict flow of audio, the longer dashed line represents the predicted parameter values and shorter dashed lines depict gradient flow. 87
6.2 Example of a five-band dynamic EQ. (Courtesy Sonnox). 89
6.3 Example of a digital four-band multi-band compressor. (Courtesy Steinberg). 90
6.4 Architecture of DEQ10 and DEQ30 audio effects. 91
6.5 Close and overhead microphone positions. 96
6.6 Top, bottom and shell microphone positions. 97
6.7 One page of the testing interface used for the subjective evaluation. 101
6.8 Mel-scaled log frequency spectrograms for (a) U2D with DEQ10 and (b) D2U with TD. Input snare drums (left), target (centre), output transformations (right). 101
6.9 Violin plot of dampening results from listening test data. Means depicted by * symbol and medians denoted by black horizontal lines. Shape represents the distribution of scores for each variable. . . 103
6.10 Violin plots of positional results from listening test data. Means depicted by * symbol and medians denoted by black horizontal lines. Shape represents distribution of scores for each variable. 104
6.11 Mean smoothed Pearson correlation results computed with Mel-spectrograms for both dampening tasks (upper) and the positional tasks (lower). 106

List of Tables

3.1 List of microphones used in both recording experiments. 38
3.2 Percentage of rating scale used by each participant for Single Hits and Hits With Bleed test. . . 42
3.3 Paired t-test results, showing microphones with significant p-values. 44
3.4 All measurements of audio recordings. 49

4.1 Specifications of microphones used for recording, taken from manufacture websites. 54
4.2 Correct responses (%) across all participants with standard deviation (std) and p-values. 56
4.3 Mean (upper value) and standard deviation (lower value) features extracted from high and low velocity recordings for each microphone. All Mics presents analysis of all 88 recordings. 57
4.4 Top five ranked statistically-different critical bands for each microphone. 60
4.5 Critical bands found to be not statistically different. 60

5.1 Makes, models, types, and polar patterns of dynamic (D) and condenser (C) microphones used for the recordings. 71
5.2 Configurations and specifications of four snare drums used in experiments. 71
5.3 Microphone categorisation achieved through multi-stimuli listening test. 79
5.4 Frequency range and mid-band frequency of $\frac{1}{3}$-octave divisions used to segment DFT. 80

6.2 Specifications of all snare drums used. 94
6.1 All 53 microphones used, subdivided by preamp. Condenser (C), dynamic (D), ribbon (R). 95
6.3 Amount of strikes for each performance, showing total for both dampening method and snare. . . 98
6.4 Dampening task results using Mel-spectrograms: Mean multi-scale loss (MSL), spectral cosine distance (SCD), log-spectral distance (LSD), mean Pearson correlation (PC), and envelope cosine similarity (CS). Lower values indicate greater similarity,except for the PC and CS metrics where higher values do. Highest performing metrics shown in bold. 105
6.5 Positional task results, metrics are the same as those used in Table 6.4. Lower values indicate greater similarity, except for the PC and CS metrics where higher values do. Highest performing metrics shown in bold 105

Acronyms

ANOVA	Analysis Of Variance.
BFSD	Big Fat Snare Drum.
BPM	Beats Per Minute.
DAW	Digital Audio Workstation.
dBZ	Decibel Z-weighted.
EQ	Equaliser.
FFT	Fast Fourier transform.
HPF	High-Pass Filter.
KS	Kolmogorov-Smirnov.
LPF	Low-Pass Filter.
LUFS	Loudness Units Full Scale.
PEQ	Parametric Equaliser.
RDA	Robotic Drum Arm.
RT60	Reverberation Time 60.
SBR	Signal-to-Bleed Ratio.
SDDS	Snare Drum Data-Set.
SPL	Sound Pressure Level.
std	Standard Deviation.
TD	Transient Designer.
WAET	Web Audio Evaluation Tool.

Chapter 1

Introduction

1.1 Motivation

A song can be comprised of several synthesised and recorded acoustic elements, which may include human voice, piano, guitars, various string and wind instruments, and a drum set. It is the role of the recording engineer to facilitate the recording session to ensure that every element is captured with care and attention thus producing a set of separate recordings, which are referred to as stems. These stems must satisfy the requirement of the musicians, as well as potentially a producer and record label. They must be of sufficient quality that a mixing engineer will be able to take the separate stems and mix them down to a stereo or multi-channel file. Starting with good source material allows for the finished product to be of the highest quality possible; a recording session can therefore often be a lengthy and expensive process. Engineers must have technical knowledge of a range of analogue and digital audio equipment, and be able to communicate with the musician or band to discover what they are expecting the final song to sound like. Each small decision the engineer makes when recording a song can affect the characteristics of the recordings.

When it comes to recording the acoustic drum kit, further complexity is added by the mutli-timbral nature of the instrument. Several parts of the kit produce extremely distinct timbres that must all be considered. Different drummers may arrange their sets in unique configurations, including more or less toms and cymbals, and different playing styles and genres must be catered for. The main rhythmic element of the drum kit for most contemporary genres is the snare and kick drum. The snare drum is a recognisable instrument that has a distinct timbre, it can be found in genres such as jazz, hip-hop, rock, pop, and reggae, along with many others, all with timbral variations making it suitable for those genres. Recording engineers are often tasked with re-creating specific snare drum timbres from certain genres, bands, albums, or song, as well as blending together attributes and qualities from several styles to create a recording with a unique timbre. In order to achieve this, the recording engineer has at their disposal several tools and methods to manipulate the timbre of a given snare drum. This might include coaxing out the best performance from the musician, or changing the recording variables. These variables can include, deciding which drums the musician should record, where in the acoustic space the drum kit is placed, and the placement, amount, and models of microphones used.

In the case of the snare drum, the instrument itself contains several elements that can be modified or changed to achieve different timbres. The instrument can be tuned to produce a wide range of tonalities, as well as completely changing the drumheads for ones of different thicknesses or plies. The snare wires, responsible for the distinct timbre of the drum, can be loosened or tightened, and the material or amount of wires can also be specifically selected. Physical material can be added to the drum head to change both the spectral component of the sound as well as its temporal qualities. Often the drummers themselves would have meticulously configured their snare drum to obtain the timbre they are most satisfied with, and the recording engineer must utilise other variables if they wish to affect the timbre of the recordings. While the timbre of the recordings can be manipulated digitally once captured, the variables the recording engineer has access to are real-world physical changes. They must posses an understanding of the implication to the timbre when changing these real-world modifiable recording parameters. Audio production tools that allow for an engineer to virtually manipulate these real-world recording parameters once the recordings have already been created could save time and expense during a recording session. This dissertation focuses on analysing several factors that affect snare drum timbre and then proposes new post-hoc recording parameter transformations for snare drums, it follows a waterfall methodology where each chapter builds on the information in the preceding chapters.

1.2 Research Questions

The overarching research question of this thesis is to determine if it is possible to carry out post-hoc digital transformations of snare drum recordings in order to elicit a subjective change akin to modifying real-world recording parameters. In attempting this, several factors of the snare drum recording process which are known to influence timbre are investigated. The thesis is comprised of four case studies, where each study aims to answer the following sub-questions:

- What is the impact microphone selection has on subjective quality of snare drum recordings, and what are the difference in signal properties between them.
- Can listeners perceive timbral differences associated with snare drum strike velocity fluctuations, what is the role microphone selection has on the identification of velocity, and what are the spectral and temporal features associated with different velocity snare drum strikes.
- What is the impact that snare drum selection has on subjective ratings of studio microphones, and can snare drum recordings be perceptually transformed in order to emulate recordings from more preferred models of microphones.
- Can digital transformations emulate real-world snare drum recording parameter changes (i.e., the change to microphone position, and the act of physically dampening the batter head of the snare drum).

1.3 Thesis Structure

The aim of this work is to analyse various features of the snare drum's timbre and then to explore the potential of post-hoc perceptual recording parameter transformations related to the snare drum. Within the thesis there are three main themes running throughout, these are covered across all chapters, with some chapters containing more than one of these themes. These three themes are timbral analysis of snare drums; subjective preference of microphones for snare drum recording; and recording parameter transformations. The theme of timbral analysis of snare drums is present in several chapters, exploring how certain changes to the snare drum, the recording process, and playing technique all elicit a perceptual change to the character of the sound. The second theme of subjective preference of microphones for snare drum recording, is predominately explored across two chapters, which cover the subjective differences associated with using different microphones for the capture of the snare. Lastly, the theme of recording parameter transformation is examined over two chapters. One chapter explores the emulation of microphone selection, while the other investigates positional and dampening transformations. A synopsis of the subsequent chapters is as follows:

Chapter 2 covers the history of the snare drum, its many variations and the applications for which they are used. The process by which the snare drum came to be incorporated into the modern drum kit used in contemporary music is presented, followed by a synopsis of its physical construction and the implication certain elements have on timbre. Techniques and products used for dampening the drumheads and different type of drumsticks are then discussed.

In **Chapter 3** the recording process for the capture of the drum kit is presented, with the focus on the variables associated with selecting and positioning the microphones for the snare drum. The importance and difference between close microphone and ambient microphones is discussed, as well as different microphone topologies and their traditional uses within the recording studio. This information is used to inform a microphone comparison study, which examines the subjective difference between a broad range of studio microphones for use on snare drum. A novel metric is proposed to investigate the effect that the presence of bleed (the unwanted capture of other drum elements) has on the subjective ratings.

In **Chapter 4** the timbral effects associated with varied striking velocity intensities of the snare drum is investigated. Striking a snare drum at a range of velocities produces noticeable volume changes; however, there are a number of timbral attributes that result from lighter or more forceful strikes. A listening test is carried out in order to assess if experienced listeners could differentiate between high and low velocity snare drum strikes when loudness differences were normalised. Four common studio microphones were use to determine if there were microphone dependant results. As different microphones exhibit varied frequency responses and non-linear characteristics, that may affect the listeners ability to distinguish velocity variations. A number of common audio features were extracted from the high and low velocity recordings to discover which aspects of timbre were most disparate, as well as using human auditory models on the frequency spectrum.

In **Chapter 5** the use of equalisation as an audio production tool as it relates to the snare drum is explained. A second microphone comparison study is carried out using multiple snare drums. The results of the listening test allowed for the categorisation of least-preferred and highly-preferred microphones for snare drum recording. A digital equaliser was then used to automatically transform a least-preferred microphone recordings in a manner to emulate spectral features of four highly-preferred microphones. A listening test is then carried out to the evaluate the success of these emulations.

In **Chapter 6** a deep auto-encoder with embedded audio effects is utilised in order to emulate two real-world recording parameter transformations. The success of these transformations are then evaluated through subjective and objective measures. Two novel audio effects are proposed. A dataset is created with extreme timbral diversity of the snare drum. The use of dynamic equalisers and transient designers are explained as they relate to the enhancement of the snare drum.

The thesis is concluded in **Chapter 7** with a summary of findings across Chapters 3 to 6, and suggestions for further work in this area.

1.4 Contributions

The primary contribution of this thesis is the proposal of post-hoc transformations of real-world modifiable recording parameters of the snare drum. In achieving this, a number of other contributions are made, as follows:

- Snare drum construction overview and discussion of corresponding timbral variations associated with specific elements of the instrument (Chapter 2)
- Snare drum microphone assessment methodologies (Chapter 3 and 5)
- Understanding of relationship between snare drum striking velocity and timbre (Chapter 4)
- Methodology for spectral feature mapping of microphones (Chapter 5)
- Design, build, and evaluation of a robotic drum arm capable of consistent strike velocity (Chapter 5)
- New methods for recording parameter transformations using deep audio effects (Chapter 6)
- Creation of the Snare Drum Data Set (`dmtlab.bcu.ac.uk/matthewcheshire/audio/sdds`) published under a Creative Commons License (Chapter 6).

The following papers have been published as part of this work:

- Cheshire, M., Hockman, J. and Stables, R. (October 2018), Microphone Comparison for Snare Drum Recording, in *145th Conventions of the Audio Engineering Society*
- Cheshire, M., Stables, R. and Hockman, J. (October 2019), Microphone Comparison: Spectral Feature Mapping for Snare Drum Recording, in *147th Conventions of the Audio Engineering Society*
- Cheshire, M., Stables, R. and Hockman, J. (May 2020), Investigating timbral differences of varied velocity snare drum strikes, in *148th Audio Engineering Society Convention*
- Cheshire, M. (October 2020), Snare Drum Data Set (SDDS): More Snare Drums than you can Shake a Stick at, in *149th Convention*
- Cheshire, M., Drysdale, J., Enderby, S., Tomczak, M. and Hockman, J. (2022), Deep Audio Effects for Snare Drum Recording Transformations, in *Journal of the Audio Engineering Society*, volume 70, no. 9, pages 742–752

Chapter 2

The Snare Drum

2.1 Introduction

This chapter addresses the snare drum as an acoustic instrument and aims to provide a concise explanation of its history and construction, giving context to later chapters when discussing the snare drum in various recording and production scenarios. The chapter is structured as follows: Section 2.2 provides historical context of the snare drum and its variations. Section 2.3 focuses on the evolution of the modern drum kit as used in contemporary music. Section 2.4 defines what is meant by timbre, discusses various aspects of timbre related research, and provides definitions of commonly used semantic terms as they relate to the snare drum. Section 2.5 details the snare drum's physical construction, focusing on how specific parts are responsible for shaping and manipulating the drum's timbre. Section 2.6 discusses snare drum tuning, best practices, the impact of tuning, and alternative tuning methodologies. Section 2.7 explores the concept of drum dampening, highlighting various methods and products used, and lastly Section 2.8 reviews different types of drumsticks.

2.2 History and Variations

The snare drum is an instrument that is used in a large range of musical genres. In most modern music it is typically found as part of the drum set, being played simultaneously alongside other drums and percussion by a single drummer. The physical construction and function of the snare drum has progressively transformed and been redefined over a period of around 800 years. The origins of the snare drum are explained by Blades (1970), with the roots of the instruments being traced back to the medieval tabor. The tabor is a double-headed, rope-tensioned membranophone with a shallow wooden cylindrical body. It incorporates a single gut or cord snare that is placed on the batter side and is either nailed to the shell or threaded directly into the drumhead, made from animal skin. The tabor which rose to prominence in the 13^{th} century, becoming popular throughout medieval Europe, although it was potentially first introduced from various eastern countries during the the time of crusades. There are accounts of the tabor being included in the royal household band during King Edward III of England's reign in the the the 14^{th} century. Historically, a leather or rope strap was attached to the tabor, and was hung from the musician's arm allowing them to simultaneously strike the drum with a single beater in one hand, and play a cylindrical end-blown flute with three finger-holes simply called a 'pipe' with their other hand.

Brensilver (2015) further elaborates on the progression of the tabor into the modern snare drum, which began with the development of larger and deeper versions of the tabor during the the 15^{th} century. These larger tabors were played with two sticks and used to accompany a fife player for military music. This saw the change from a single musician playing the tabor and pipe simultaneously to fife and drum music requiring two players, as the two instruments were played by both hands of the musician. By the 16^{th} century, additional gut snares were stretched across the bottom heads of drums, becoming what was know as the field drum. This was introduced to North America by Europeans, first being used by the colonists as a signaling instrument to convey military orders and to call people to church or other gatherings. In military applications the drum allowed troops to communicate with one another over long distances. During this same time period this style of drum was also starting to be incorporated into classical music, notably being used by French composer Marin Marais in his 1706 opera 'Alcyone' to evoke the sound of a storm. Towards the end of the 19^{th} century in post-Civil War America the snare drum had began to become an instrument used for indoors entertainments, becoming popular for use in the vaudeville, Dixieland, and ragtime styles of the era. In the 1920s the snare drums underwent several innovations including re-designs of the hoops, tuning rods, and strainer that transformed the drum into an instrument that more closely resembles the contemporary snare drum used by today's musicians.

The drum set snare drum is used in a wide range of genres, including rock, jazz, blues, R&B, reggae, funk, and heavy metal (Meyer, 2014; Andertons, 2015). It is used in conjunction with a complete drum kit of varying configuration depending on the genre, song, and personal preference. Snare drums, typically made from wood or metal, with diameters of 14" and a depth of 5", 6.5", or 8" are most commonly used with the drum set. Although other materials can be used, there is no particular snare drum size or material that is specifically used for a given genre. Snare drums with diameters ranging from 6" to 8" are considered micro snares and can be used as an alternative to or used in conjunction with a 14" snare drum. Micro snares, which are most commonly used in Latin music, may feature a traditional two-head design while others will have only one head and use a fanned snare that contacts the underside of the batter head. It is also possible to get snare drums with diameters such as 10", 12", and 13". Snare drums with larger diameters than 14" are often referred to as ballad snares, and can be found in 15" and more rarely 16".

Concert or orchestral snare drums are similar to the drum kit snare, but will often be played as a singular instrument used in an orchestral context. These types of drums usually feature wood shells and a special strainer design that allows the user to select individual cables to be engaged against the resonant head. Although it is more common for synthetic gut snares to be used, metal snares are also available. Calfskin style batter heads are standard for concert snares, although coated heads will sometimes be used as an alternative. Piccolo snare drums are characterised by their shallower shell depth, usually around only 3" or 4.5", giving them a higher pitch and more pronounced high frequencies and attack. They are often used instead of a more traditional snare drum, but can also be used as an additional snare as part of a full drum kit. Soprano or *Popcorn* snare drums are similar to piccolo snare drums, generally producing a much higher pitched sound quality to them. Soprano snare drums typically have nonstandard shell dimensions, measuring between 5" and 7" deep and 10" or 12" in diameter.

Marching snares are typically deeper than orchestral and drum kit snares, usually around 14" diameter and 12" in depth. They are often fitted with reinforced, thicker batter heads made from kevlar, capable of withstanding extremely high tension tunings, and heavy playing styles. The hardware is commonly made from lightweight aluminium, allowing the drum to be more easily worn and carried for parades and drumlines. Most modern marching snares employ a free-floating design in which the hardware does not directly connect to the shell at any point, helping to protect the shell from damage due to the high drumhead tensions used on marching percussion. Field drums are primarily used for orchestral, concert band, and percussion-ensemble applications.

Designed to resemble the sound of military drums from the 19^{th} and early-20^{th} centuries, these models usually feature larger diameters and much greater depths than typical orchestral snare drums. Synthetic gut or cable snares are standard, and the drums are usually played at lower tunings. Pipe Band snare drums are similar to most marching snare drums except they feature a second set of snare wires. These additional wires are used in conjunction with the ones traditionally placed on the resonant head. They make contact with the underside of the batter head to reduce sustain, enhance attack, and produces a character with noticeable snare emphasise which help with the articulation of complex drum patterns. Although there are different types of snare drums with a range of uses, the type that has the largest diversity in regards to shell material, tuning, and other variables that affects timbral qualities is the drum set snare. It is used in a plethora of genres from jazz to heavy metal, and even within the same genre can have extremely diverse characteristics. It is the drum set style snare drum that the rest of the this chapter and the subsequent chapters will be focused on.

2.3 The Modern Drum Kit

Nicholls (2008) details the evolution of the modern drum kit, describing how separate instruments came together to be played by one drummer. At the beginning of the 20^{th} century drums and cymbals were traditionally played by separate musicians, typically standing up, or marching in the case of military bands. However, due to restricted and limited spaces in theatres and nightclubs it became increasingly common for a percussionists to play multiple instruments. *Double drumming* was a term used to describe the act of a percussionist playing both a bass drum and snare drum with each hand. The modern drum kit can first thought to have been developed when drummers began to sit down to play the drums and first started playing the bass drum with a pedal, initial experiments in this technique can be dated back to as early as 1890. In 1909 William F. Ludwig from Chicago, USA invented a foot pedal that incorporated a spring mechanism, his design encouraged percussionists to play seated, freeing their hands to play additional drums. In the 1920s jazz drummers began incorporating larger cymbals imported from China and Turkey into the rhythms of their drum beats, as previously only small cymbals were used for sporadic effects. Before the invention of the hi-hat, a cymbal was mounted near the bass drum, which was sometimes played by an additional arm connected to the bass drum pedal. Later, a pair of cymbals were mounted on their own short stand and called a *low-boy* or *low hat*. This pair of small cymbals were eventually extended with a vertical tube allowing the drummer to play them with their hands as well. The first hi-hat was created around 1926 by the leading drum accessory company of the time, Walberg & Auge.

The success of musicians such as Gene Krupa and Buddy Rich, and the rise in popularity of swing and jazz music originating from the United States in the 1930s saw the drum kit getting stripped away of various elements predominantly intended for sound effects, such as klaxons, triangles, rattles, bells, cowbells, woodblocks, temple blocks, and whistles (Dean, 2011). By the 1950s the drum kit was an integrated instrument, in its basic configuration it comprised: a kick drum, a snare drum, a tom mounted on the bass drum, a floor tom, a hi-hat, a ride cymbal, and a crash cymbal. This type of set up and its variations are what is commonly used in most contemporary music. Typical alterations include the addition of a second tom, usually somewhere in size between the first rack tom and the floor tom, as well as drummers using a broader range of cymbals. It is this common configuration that is implied when discussing and referring to the drum kit in following chapters.

2.4 Timbre

Timbre can be regarded as the perceived characteristic of a given instrument, voice, or sound, it relates to the frequency spectrum and the relative intensities of the harmonics (Huber and Runstein, 2010). The timbre of an instrument allows a listener to identify two related instruments playing the same note at identical volume, for example, a piano and a guitar both use strings, however they are easily distinguished by listeners familiar with either instrument. Subtle changes to the harmonic content and amplitude envelope allow listeners to perceive difference between two variations of the same instrument, for example two pianos. Frequent exposure and familiarity with particular instruments allow the listener or player to hear and recognise specific timbral variations associated with it, and thus a recording engineer may be able to recollect the timbre of a specific instrument from a certain song in order to recreate a similar quality during a recording session. In depth technical knowledge of how each recording variable affects timbre is essential for this to be achieved. This may include using a certain type of guitar or amplifier, or specifically selecting which microphones are used and where they are placed.

Analysing and quantifying how timbre is perceived and discussed by humans is an extensive and ongoing area of academic research. Saitis and Weinzierl (2019) provides insight into some of these different areas. A wide-ranging area of research includes how semantic descriptors are used to explain the often very subtle differences between sounds. Audio engineers, technicians, and music and sound scholars all rely on a shared vocabulary of verbal attributes when they are required to discuss and describe timbral qualities of sounds or music. Timbral qualities are often conceptualised and communicated through readily available attributes from different sensory modalities (e.g., *bright, warm, sweet*) but can also be expressed through the use of onomatopoeic terms (e.g., *ringing, buzzing, hissing*) or non-sensory attributes which relate to more abstract constructs (e.g., *rich, complex, harsh*). The terms used are not crucial for processing and understanding timbre as listeners have demonstrated the ability to compare, recognise, memorise, and imagine timbral qualities without being required to use specific words for them. The way in which people describe sensory experiences can be used to gain information about their perception of the stimuli. Individual terms can be thought of as representing micro-concepts, simple elements of a more generalisable semantic knowledge, that are not fully meaningful on their own, but instead yield meaning when assembled into broader semantic descriptions. Among vast timbre vocabulary, there exists many seemingly unassociated words that may share similar meaning and refer to the same perceptual experience. There are cases in which different words can be used to describe the same timbre—for example, it has been shown that sounds perceived as *rough* are also described as *harsh* where ratings on the latter were found to correlated with ratings of the former. Complete agreement for any semantic attribute does not exist, as with any subjective quality there will be always be some degree of agreement and disagreement between groups of individuals.

Aside from learning and understanding which words are used for discerning timbral differences, researchers aim to understand the concepts of semantic scales, such as *dark* to *bright* and *smooth* to *rough*, where a certain sound may exist on a scale between two verbal attributes, and may have more or less of a certain quality than another sound. This might mean that although one sound is not necessarily perceived as being bright in isolation the particular stimuli could still be *brighter* than another sound. Work has also been carried out to investigate how these semantic terms relate to specific features of the acoustic signal, such as its frequencies response and envelope characteristics. In timbre perception the impressions of *brightness* is typically found to be correlated with the spectral centroid—a scalar descriptor defined as the amplitude-weighted mean frequency of the audio spectrum. Frequency shifts in spectral envelope are systematically perceived as changes in brightness, additionally *sharpness* has also been found to strongly related to the frequency position with highest concentration of energy within the spectrum. Sharper, harder, and brighter sounds having more energy in higher frequency bands. Brighter percussive timbres have been shown to be associated with higher spectral centroid values during the attack portion of the envelope, while sharp and hard descriptors relates more closely with the attack time itself (i.e., sharper and harder percussive sounds feature faster attack times). Attack time refers specifically to the time needed by spectral components of the signal to stabilise into periodic oscillations. It is known to be a perceptually distinguishable impulsive separate from the sustain portion of the sounds.

Research has also been performed deciphering the impact that room acoustics have on timbre, as certain rooms will amplify and attenuate specific frequencies, as well as increase the attenuation of the spectral envelope toward higher frequencies due to air absorption. The timbre of an instrument or voice can vary substantially from one acoustic space to another, depending on the geometry and materials of the room, and any additional objects or furniture within the room. In most cases the direct sound from the instrument will merge with the characteristics of the performance space, and it is therefore difficult to predict the extent to which listeners can successfully segregate the audio source from the affects of the room when communicating timbral qualities. Certain verbal attributes have been used to describe the aspects of room acoustic qualities, such as *brilliance, brightness, boominess, roughness, coloration, warmth,* and *metallic*. Aspects of the room will either interact with, emphasise, or mask qualities of the instrument that is being played or recorded.

In general, an instrument has its own unique timbre (or range of timbral characteristics) which are inherent to the instrument itself. As such, the timbre of a snare drum is dictated by the size, shape, material, thickness, and design of the drum shell, the construction, design, and material of the hardware attached to the drum (including lugs, tuning rods, hoops, snare wires, or other attachments), the way the drum is mounted on the stand, and of course the type and design of the drumheads that are used (Toulson, 2021).

When discussing timbre specifically relating to the snare drum, Alldred (2019) provides definitions of terms that are commonly used. *Overtones* relate to the harmonics of the fundamental frequency, a theoretical pure sine wave would have no overtones, a snare drum can have many complex, interacting overtones which can be seen as both good and bad depending on the desired timbre. The specific overtones and their relative strengths can also be described as the *tone* of the drum; it is these frequencies that make up the individual timbral character of that particular drum. A snare drum that is said to have a *rich* tone, may have a more musically pleasing arrangements of harmonics compared to another drum. *Bright* and *dark* are commonly used to describe the amount of high frequency energy that the drum produces, but can also refer to perceived intensity of the strike, with a brighter sound having more defined attack, and a darker sound having a softer, less forceful attack. The attack is the initial portion of the transient that is produced when a drum is first struck, and the sustain describes the length of the drum hit that is produced. The *sensitivity* of a snare drum describes the amount the snare wires are activated when the drum is played at a range of velocity intensities. A highly sensitive drum would produce noticeable snare wire noise from the lightest strikes, whereas a drum with low sensitivity would require more striking force to excite the snare wires. *Resonance*, also referred to as *ring* or *ringing* relates to the head and drum shell working together which manifest as one or more noticeable higher frequency peaks in the spectrum which can be heard during the sustain portion of the strike. This is not necessarily a negative quality and depends on the required timbre for the song. Lastly, *wet* and *dry* are terms that are used to describe multiple timbral attributes; a wet timbre will typically feature longer sustain, more overtones, and more pronounced resonant frequencies, while a drier sound will be characterised by a more controlled strike, shorter sustain, less overtones, and less resonance. Overall a drier timbre will feature a more direct snare focused strike that does not ring out for as long and has a shorter amplitude envelope.

2.5 Construction

Vibrations of the drum heads, the shell, and even its stand all contribute to the sound of the snare drum, while the snares wires move in complex ways, initially being in contact with the resonant head but losing contact after a strike and then in turn returning to strike the head producing unique characteristic (Rossing et al., 1992). This section breaks down and explains each component of the snare drum, and where applicable discusses how changes to a particular part results in varied perceptual changes to the timbre. Figure 2.1 shows an example of a wooden snare drum with metal hoops.

Figure 2.1: Example of a modern snare drum, showing batter side.

2.5.1 Shell

The drum shell is responsible for supporting the drumheads and provides a physical contact point for hardware to attach, allowing the snares to be tensioned and the drumheads tuned. As previously mentioned, snare drums are produced in a range of shell diameters and depths which impact various timbral properites of the snare drum. Shell diameter affects the possible range of pitch a drum can be tuned to with larger diameters allowing for lowest pitches (Owsinski and Moody, 2009), whereas snares with smaller diameters can be tuned to a higher pitch and will also have a faster attack (Falk, 2019; Mitzner, 2021). The depth of the shell does not affect the pitch to the same extent as diameter, but is known to affect its sustain and lower frequency resonance. A deeper shell snare can produce more low frequencies whilst shallow snares such as piccolos will have less bass resonances (World Of Music, 2015; Andertons, 2015). Shell thickness also contributes to the resonant properties of the drum, with thinner shells being able to resonate much easier than thicker shells, thus producing shells that have a longer sustain, whereas thicker shells reduce the sustain (Owsinski and Moody, 2009; Drumhead Authority, 2020). Shell thickness can also affect fundamental frequency, given the same tuning, thinner shells will have a lower fundamental whereas thicker shells will tend to produce a higher fundamental (Azzarto, 2011).

Shell material is known to affect timbral properties of the overall snare drum character (Owsinski and Moody, 2009). Snare drums are predominantly made from two materials, either wood or metal. Between these two materials exists the greatest noticeable timbral difference with variability between different woods, and between different metals being far more subtle. Depending on the wood the sound can be bright and clear or more subdued and mellower sounding (a characteristic often associated with vintage drum kits) (Beattie, 2019). Snare drums can also be made from a range of other less common materials including, carbon fibre, fibreglass, or various types of plastics such as acrylic (Nichols, 2009; La Cerra, 2019). Some manufacturers also build hybrid drums, making a shell from a combination of materials in order to elicit unique timbres (Nichols, 2013). Hybrid shells may be made from two or more different materials, such as two different kinds of wood, wood and acrylic, or wood and metal. Each combination produces a unique and distinctive sound that can be better suited for a certain style of drumming . A popular hybrid construction is when acrylic is combined with wood, another common hybrid shells design is the combination of a softer wood such as maple with an inner ply of a hardwood, such as wenge (Newbold, 2022).

Various construction techniques are used to produce wooden snare drums. While some may have a solid shell, this is often costly and time consuming, more practical methods include shells being made from individual staves or segments of wood that are glued together, or sheets of wood are formed into a rigid shell through heating and compression, often bent using steam, and then clamped into a mould. The latter style shells may consist of a single thicker ply or be constructed from multiple thinner plies of wood that are glued. Wood snares may also require reinforcement hoops to provide structural stability to the drum, often added to the top and bottom, helping thinner shells to stay circular and not warp or bend due to environmental fluctuations. Reinforcement hoops can be made from either the same wood as the rest of the shell or from a different material, the thickness will impact the resonance of the shell and therefore its timbre (Azzarto, 2011). A snare drum made from a solid piece of wood tends to be more resonant, have longer sustain, and a more pronounced fundamental frequency than thin multi-ply construction, however multi-ply shells can often be more cost effective and offer unique timbres (Meyer, 2014). Using multiple plies also offers extra rigidity, making the shell less likely to crack or distort. Offsetting each ply so that shell is joined in several places creates a structurally stronger shell and eliminates the need for reinforcement hoops to be glued into the shell (Nicholls, 2008). Stave shells sound different to ply shells mainly due to ply shells using a larger amount of glue, which dampens resonances compared to stave construction (Drumhead Authority, 2020). Wooden drum shells that are warped or misshaped will have a detrimental effect on the timbral quality of the drum (Gibson, 2004).

There is some agreement between Azzarto (2011); Beattie (2019); Schroth (2020) and Newbold (2022) when it comes to describing the subtle timbral differences between various woods used for snare drum construction. Maple, one of the most commonly used woods, tends to be well rounded, being balanced across the whole frequency spectrum. Beech produces a similar timbral quality as maple, but with slightly more enhanced mid-range frequencies. Birch is thought to be comparable to both maple and beech with more pronounced lower frequencies, slightly more high frequencies than maple, and a reduced mid-range. Mahogany shells produce greater bass and low mid-range resonances, often thought to have a more subdued *vintage* character. Poplar is a softer wood with qualities comparable to mahogany, but noticeably more high frequencies, and is often used for the inner ply with either maple or mahogany for the outer plies of a multi-ply shell. Oak is thought to have a pronounced low and mid-range, be louder and have a noticeably distinct timbre from both maple and birch. Some other woods that are used to build snare drums include walnut, ash, basswood, hickory, eucalyptus, spruce, and wenge. More exotic woods also used by manufacturers include ebony, sycamore, myrtle, purpleheart, acacia, bubinga, cherry, and kapur (Donahue, 2018). Snare drums can also be constructed from multiple types of wood, either by using different woods for particular plies in a multi-ply design, or using different woods for individual staves on a stave style shell. This is not an exhaustive list of woods used for drum shell or the associated timbral qualities produced by these various woods. Many of the proposed timbral descriptors refer to specific models of snare drums using these woods, which will also have other contributing variations beside shell material. Timbral attributes will be heavily influenced by a combination of factors relating to the construction of the drum and wood material is not solely responsible for the timbre differences, although it may play a small role is emphasising certain frequency bands due to the way the wood resonates in response to the drumhead being struck.

Metal snares are typically louder than wooden snares and produce more high frequencies with pronounced transients making them sound brighter. The two main designs of metal shells are seamless and hammered. Seamless shells are typically machine-spun producing a smooth shell interior. The interior is a crucial feature in how sound waves interact and move around inside the drum once it has been struck. A seamless shell creates a smooth column for the sound waves to travel down from the batter head to the resonant head and does little to diffuse and interrupt the internal reflections, resulting in a brighter and louder drum strike. Hammered shells have a rough, irregular, and uneven internal surface which disrupts the air column and forces reflections to scatter inside of the shell. A rough shell interior will reduce reflections inside the drum, whereas smoother interiors will have greater reflections (Owsinski and Moody, 2009). These two distinct design choices are know to affect the timbral attributes of metal snares (Biancardi, 2015; MacEachran, 2019).

Just as different woods are used for snare drum shell construction, a range of different metals are also utilised. Azzarto (2011); MacEachran (2019); Beattie (2019) and Newbold (2022) all express a level of agreement between the different timbral characteristics associated with these metals. Steel is most commonly used for beginner to intermediate-level metal snares drums. Steel shells are thought to be very bright, having more high frequency energy than aluminium and brass, with an enhanced attack that helps them be heard in a mix. In some cases they may be plated with chrome or nickle. Aluminium shells are known for their shorter sustain compared to steel, this attribute often leads them to be described as *dry*. Generally brass shells tend to be louder than other metals used with more noticeable low end, and stronger low end attack. Like with steel shells, it is common for brass drums to be coated with either nickel or chrome which produces very slight variations of timbre. Bronze and copper have similar timbral properties, however bronze is thought to have more noticeable overtones producing resonances and ringing as well as a slight boost in the low end. The timbre of copper shells has been described as being more similar to a wooden snare than other metals.

There is some debate whether or not shell material choice actually has any perceptual difference to a listener. Westera (2018) believes that factors such as drumhead selection, tuning, bearing edges, shell-rigidity, snare-wires, hoops, and the acoustic environment all have a greater impact on the overall timbral characteristics of the snare drum than shell material alone. He conducted an informal study in which 23 drummers played and then answered questions relating to a snare drum made from Medium-density fibreboard (MDF). Question one asked; "What do you think it is made of?" only one response was "composite material" with all other participants answering more common materials such as bronze, aluminium, steel, birch, and maple. The highly inaccurate responses to the question suggests that there is no objective collectively-held stereotype of the timbre of either a metal or wooden shell snare drum, however due to the non-scientific manner in which this research was carried out it is different to draw definitive conclusions. Unfortunately no direct comparison was made between the MDF snare and other snare drums of any different material. An evaluation of this kind would have highlighted if participants were more or less likely to be able to correctly identify shell material in the presence of a reference, and if the preference scores would have been altered when compared to a snare drum made from a more traditional material. Waxman (2022c) also argues that the importance of different woods is overstated, stating that there are many other factors that contribute to defining the overall timbral quality of the drum more so than the wood species. These other factors are the quality of the drum shell construction, the bearing edge's angle and shape, the choice of drumheads and the way in which they are tuned. Other factors associated with construction will also likely overshadow any differences produced by which wood is used such as ply thickness, number of plies, and whether or not the shell is perfectly circular.

Toulson (2021) conducted a comparison between 3 snare drums constructed of different materials including oak, steel, and birch. There was not much noticeable difference between the snare drums in the acoustic environment and between the recordings from the overhead microphone placed around 1 meter from the drums, with Toulson stating that "the drums sound really quite similar with just subtle perceivable differences in sonic characteristics". However, a close microphone placed approximately 10cm away from each of the drumhead was also used, with this microphone the timbral differences were more apparent on the recordings. Although exact microphone position was not identical between all 3 drums, this finding suggests that depending on the microphone used and the position of the microphone in relation to the snare drum, the shell material differences may or may not be an important consideration, with only very close microphone positions being able to capture the subtle timbral changes elicited from the different shell and head combinations. A possible cause for the subtleness of the timbral differences is that any small variations that the shell material adds to the character of the sound is vastly overpowered by the volume of sound that is generated by the drumheads.

2.5.2 Bearing Edge

The bearing edge of a drum is the part of the shell that makes contact between the edge of the shell and the drumhead. It should be constructed in a manner that allows the drumhead to rest evenly across the entire circumference of the shell. Uneven, damaged, or pitted bearing edges will prevent the drum from performing optimally, not only negatively affecting the timbre, but also making the snare more difficult to accurately tune (Owsinski, 2005; D'Amico, 2015). As the bearing edge is the only location on the drum where the interaction between drumhead and shell occurs, the shape and condition of the bearing edge impacts how the drumhead vibrates and how vibrations are transferred from the drumhead into the shell. Bearing edges affect the range of possible tuning and the timbral characteristics of a drum strike, including how long the drum sustains for and the amount of overtones or harmonics are produced by the drum (Azzarto, 2011; World Of Music, 2015). A sharper bearing edge angle with less surface contact between the shell and the drumhead allows the head to vibrate more freely which results in longer sustain and more harmonics, producing a brighter timbre. A rounded bearing edge with a flatter angle has greater contact with the drumhead, this dampens some movement of the head producing shorter sustain and less higher frequency overtones which emphasise the fundamental frequency (Brown, 2017). The timbre associated with this effect is often described as being more *mellow* and sounding *warmer* or *softer* (Drumhead Authority, 2020; Owsinski and Moody, 2009; Howley, 2018). Additionally, the more surface area of the drumhead contacting the shell, the greater the amount of energy transferred into the shell of the drum to cause more shell resonance (D'Amico, 2015).

Brown (2017) and Howley (2018) describe several standardised bearing edge shapes, all which offer slight timbral variations, these include the 45° single, 45° single with counter-cut, 45° with round-over, 45° double, and various 30° alternatives to some of these shapes. There also exists hybrids between all the main types of bearing edges as well as some manufactures using different bearing edges on the batter and resonant side of the snare drums (World Of Music, 2015). One approach is to use a round-over edge for the batter side and a 45° single or double on the resonant side of the shell to utilise the advantages associated with both types of bearing edges (Modern Drummer, 2013). The 45° single is most commonly used on modern drums featuring a 45° angle cut from the outside edge towards the interior of the drum. It has a very low amount of surface

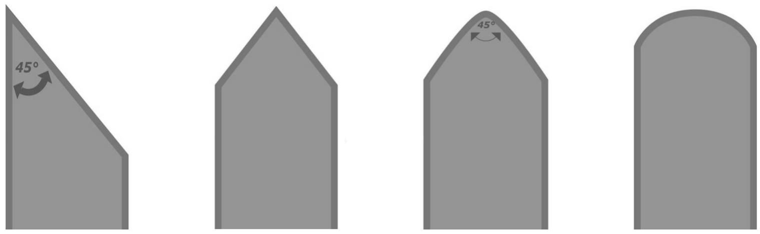

Figure 2.2: Example of different shape bearing edges. From left to right: 45° single, 45° double, 45° double with slight round-over, and round-over. (Courtesy Thomann).

area contact between the shell and the head to allow the drumhead to vibrate longer, and places the contact point towards the exterior of the shell. This type of bearing edge has become associated with a more modern sounding drum that has longer sustain, and brighter timbre produced by the increase in higher harmonics. This bearing edge can also make it harder to achieve a desired tuning as tuning becomes more sensitive to smaller changes in tension. The 45° counter-cut is very similar to the 45° single but has an additional cut from the outside edge intercepting the other 45° angle, therefore bringing the contact point slightly further towards the centre of the shell. A bearing edge the has its apex inline with the centre of the shell wall is the 45° double which produces a sharp peak that has a 45° angle on the inner and outer side of the shell. This moves the contact point even further from the outer edge of the drumhead where the head begins to curve. The 45° double is know to increase sustain and allow for a wider tuning range (Brown, 2017).

A round-over bearing edge can either be slightly smoothed over or be fully rounded off to create a much larger amount of surface contact between the shell and the drumhead and eliminating a clear point of contact between the two. This type of bearing edge dampens some of the vibrations of the drumhead, producing a shorter sustain and fewer overtones and resulting in proportionally more emphasis of the fundamental note (World Of Music, 2015; Brown, 2017). This type of bearing edge is typically associated with older or vintage style drums as this bearing edge design was traditionally used until the trend shifted toward much sharper edges around the 1980s (Modern Drummer, 2013). Similar to the round-over edge, the 30° bearing edge also creates a wider surface area than a 45° edge, with some arguing that this extra contact between the head and bearing edge increases vibrations to the shell and results in more shell resonance, thus emphasising the effect that shell material has on the timbre. As well as a 30° single, there are also some snare drums with a 30° single with round-over where the 30° inner cut is intercepted by a rounded profile coming from the outer shell wall (Howley, 2018). Figure 2.2 shows the shape of several of the bearing edges discussed.

Macaulay (2003) investigated the affects different bearing edges had on the sound of toms. Three toms of identical dimensions, 12" x 6", were used each with a different bearing edge. A 45° single, a 45° single with slight round-over, and a 45° double were used. Only the effects on the batter heads were tested, using Remo Ambassador clear heads. The drumheads were tuned using a DrumDial, with the tension being measured to be 75 by the DrumDial at each of the 6 lug positions. A 5A Vic Firth drumstick was used to strike the toms. It was found that the 45° double bearing edge allowed for the best head to shell energy transfer, although it was acknowledged that in order to state this with absolute certainty further testing using additional drums would need to be carried out. It was noted that the point of contact between the bearing edge and the drumhead was different between the 45° single and 45° double, with the 45° double being slightly closer to the centre of the drum. This meant that bearing edge applied pressure to a flatter part of the drumhead as opposed to the 45° single which was observed to be interacting with the curved part of the drumhead which may have made the drumhead more uneven and potentially wrinkle when under tension. The 45° double bearing edge was thought to be advantageous by allowing the middle of the shell to make contact with the drumhead. This would allow more energy from the drum strike to transfer directly into the centre of the shell rather than vibrations being transferred to the outer shell as is the case with the 45° single bearing edges. This effect is also described by D'Amico (2015), where the drumhead meets the bearing edge directly through the middle of the shell wall, this is known to focus vibrations generated by the batter head through the center shell wall to cause the shell to vibrate from its core. Although this experiment was carried out using toms instead of snares and only focused on the affect on the batter head, the observations are expected to be similar for a snare drum as the principles remain the same.

2.5.3 Drumheads

A large majority of the overall timbre of the snare drum comes from the drumhead, with drumhead selection having the biggest potential to dramatically change the timbre of the drum (Toulson, 2021). Drumheads are constructed from two elements including the skin or film that is struck, and the flesh hoop which is the rigid outer circle that provides structural support to the head (Alldred, 2019). Nicholls (2008); Ludwig and Cook (2001); Thibodeaux (2022), and Beck (2013) provide a brief history of drumheads and how the materials used radically changed in the 20th century. Before the 1950s the main material used for drumheads was animal skins. The animal used was predominately based on geographical location, with calfskin being most common in the United States. An issue with using animal skin was its sensitivity to changes based on humidity and temperature which affected the pitch of the drum, as well as the feel in terms of resistance when being played. When the weather was damp the calfskin stretched which required drummers to tighten them, however, leaving the heads tightened when they dried out resulted in them splitting. Calfskin heads were also prone to breaking when struck too hard and different skins had varied thicknesses making it difficult to get a consistent timbre.

Calfskin was eventually replace as the most popular material for drumheads by Mylar, a polyester film invented during the Second World War as a substitute for the more fragile cellulose based film used by reconnaissance aircraft photography (Nicholls, 2008). Experiments began in the early 1950s to manufacture drumheads using Mylar, initially the Mylar was stapled to the wooden flesh hoops that calfskin heads were originally attached to. In late 1956, Chick Evans completed a design of a Mylar drumhead that consisted of a drilled outer hoop that tacked a Mylar head to a smaller inner hoop. By early 1957 Remo Belli and Sam Muchnick improved on this design and developed the first successful plastic drumhead. A Mylar head, with a crowned edge with small holes punched out around the perimeter that was attached to a U-shaped aluminum hoop using a fast-setting liquid resin to bond it in place. This Mylar drumhead design was the first of its kind that did not involve tacking the film to a flesh hoop. On June 1st, 1957, Remo Inc. was established to market and sell these new aluminum channel Mylar drumheads (Beck, 2013). Kevlar is another synthetic material used for drumheads. Although not commonly used for drum kit heads it has found applications for use in snare drumheads for marching ensembles due to its ability to withstand extremely high tensions resulting in a much faster rebound. Remo became the market leaders with its Weather King drumhead and shortly after became the world's foremost manufacturer of synthetic drumheads (Thibodeaux, 2022).

There are many modern manufacturers of different drumheads featuring varied designs and constructions, the most common designs are either single- or double-ply and are either coated or uncoated. Single-ply heads are not as durable as double-ply, and tend to not last as long with heavier styles of playing, such as rock and metal music (D'Virgilio, 2018). The thicker the combination of plies, the lower the fundamental note will be when tensions are equal (Alldred, 2019). Additionally, heads produced with a higher density material will also result in lower frequencies. When coating is applied to a head this essentially makes the head thicker and heavier, and will also affect the pitch (Toulson, 2021). Thinner drumheads are commonly more suited to lighter playing styles of music, such as jazz, where a more open, resonant sound might be preferred (Azzarto, 2010b). Drumhead type, thickness, coatings, and its age and condition all have an effect on the timbre of the snare drum. Thinner, uncoated heads produce more sustain and are brighter, while thicker and double-ply heads, and those with coatings, emphasise the attack, dampen ringing, and reduce some high frequency harmonics (Major, 2014; Bartlett and Bartlett, 2016). Another way in which the coating on drumheads affects the timbre is by altering the way the drumstick interacts with the surface of the head. Applying a coating results in a harder surface than the uncoated drumhead, this reduces contact time between the stick and drumhead, making the initial impact more pronounced giving the perception of a faster sounding attack to the drum strike (Toulson, 2021).

Another category of drumheads are those that include varying forms of built-in dampening which will further decrease sustain and emphasise the timbre around the fundamental pitch (Owsinski and Moody, 2009; Meyer, 2014). In 1968, Remo first introduced the Control Sound Black Dot which featured an additional circle of Mylar affixed to the centre to further reinforce and strengthen the drumhead. Ludwig then introduced a similar design a few years later called the Silver Dot Rocker. The dot acts similar to a thick coating on the drumhead, adding increased attack to the sound and reducing sustain, as well as enabling drummers with a harder, more powerful style of playing to use the drumheads for a longer period of time before they need replacing as the dot creates a more robust head (Toulson, 2021). In 1977 Remo introduced the Pinstripe which consisted of two clear layers of Diplomat thickness Mylar bounded together around the outer collar. A similar design was produced by Evans called the Hydraulic, a 7-mil, double-ply head with a thin layer of oil between the plies. A different approach also intended to reduce overtones and shorten sustain is the Evans HD Dry drumhead which features 20 precision-drilled vents around the perimeter of the drumhead. These pinhole sized *Dry* vents allow small amounts of air to pass out of the drum when the head is struck, reducing the amount of time sound waves reflect back and forth between the batter and resonant head after the initial excitation of the batter head (Azzarto, 2010b). Other manufacturers have produced models that have varying degrees of built-in muffling, usually adding a layer of Mylar or other material to the top, underside, or between plies of the drumhead, and either in the center or around the outer edge depending on the amount of intended dampening.

Typically a thicker, more resilient head will be used on the batter side, while a thinner head is more commonly used on the resonant or snare side. For example a Remo Emperor is intended for use on the batter side and is constructed with 2-plies of 7-mil film, whilst the Remo Ambassador Hazy Snare Side, which is considered industry standard for the resonant head, is constructed with only 1-ply of 3-mil film. As the resonant head is not intended to be struck it does not need to be as durable (Alldred, 2019). Resonant snare drumheads usually range from 2 to 5-mil, thinner heads will produce a brighter timbre but will need more regular tuning as they are known to stretch more easily compared to thicker drumheads (Azzarto, 2010b). The resonant head is a key contributor in determining the amount of sustain. When the batter head is struck it vibrates and pushes air outwards, air inside the shell is forced to move down towards the resonant head, which in turn causes it to also vibrate. The vibrations from the resonant head once again move air towards the batter head until all kinetic energy has dissipated from the two heads. This back and forth keeps the heads vibrating which creates a longer sustain (Ritz, 2016). The thickness and tuning of the resonant head will impact the amount it can freely vibrate. Some resonant heads such as the Remo Ambassador Black Suede Snare Side and Renaissance Snare Side drumheads feature a light coating which give these resonant heads a textured surface. Just as with the batter heads these coatings act as subtle muffling, slightly dampening movement and reducing higher harmonics (D'Virgilio, 2018). Aside from the timbral changes produced by different drumheads, they also affect the range of possible tunings the head is capable of (Ritz, 2016). With so many options and combinations of drumheads available on the market, drummers may spend a long time deciding on their preferred drumheads based on personal taste, or will chose to select heads dependant on the song and genre with certain heads being used for one style of playing and different heads being specifically selected to complement another (Azzarto, 2010b).

2.5.4 Tension Rods and Lugs

Tension or tuning rods are the long threaded bolts that hold the drumheads under tension, allowing them to be tuned by varying the amount of force they apply. They are responsible for pulling the hoop down around the drumhead tighter to the shell. Tension rods are either slotted through holes in the hoop itself, or are used in conjunction with collar hooks, which are most commonly seen on kick drums. The more these rods are tightened the more tension will be placed on the drumhead, this allows the heads to be tuned up or down in pitch. The piece of hardware that is attached directly to the drum shell is call a lug. The tension rod either screw directly into the lug, such as on the tube lug design or into a swivel nut, which is a small piece of threaded tubing that sits inside the lug. Prior to 1923 most drum companies used tube lugs on their snare drums which commonly cause damage to the treads of the tension rods if the hoops were not perfectly aligned (Nicholls, 2008). Aside from the issues with the tube lug design, they place less metal in direct contact with the shell, thus improving sustain and providing a slightly different sound to other lug types which may dampen shell vibrations (Meyer, 2014). Around 1923, George Way invented the first non-tubular design that eliminated the problem of cross threading with the inclusion of the swivel nut, which he did whilst working for Leedy, an American drum manufacture (Dawson, 2014). This die-cast lug design was referred to as a *self-aligning lug*, which incorporated a threaded insert that was capable of moving around slightly inside of the lug that allowed for small adjustments to be made when lining up the hoops with the tension rods (Howley, 2020).

In 1929 and into the 1930s, Way's original design was heavily modified and the Leedy *X* lug or *Box* lug was introduced which featured a new style of casing that attached to the drum shell from the inside of the drum. These Lugs were made of cast aluminium, with the earlier versions having the swivel nut held in place using small screws but later changed to include the use of internal copper springs which allowed for even more flexibility of the insert. At the time, Leedy described the X lug as allowing the receiving tubes to move freely in any direction, which assures perfect alignment of the threads, eliminating binding and stripping, and permitting a more uniform head tension. The Leedy X lug was used on all top of the line snare drums from 1929 until 1938. The following year, the *beaver tail* lug was introduced. The beaver tail was a solid metal lug that featured a backing plate and were used from 1939 until the United States government placed restrictions on metal usages during the Second World War and drum companies were forced to develop lugs made from wood (Cooper, 2020). Since the 1950's many designs and variations of lugs have been produced which all aim to achieve the same goal of securely holding the tension rods in place.

On drums that use springs inside the lugs to hold the swivel nuts in place, the springs can cause extraneous noise when the drum it struck. La Cerra (2020) recommends removing the lugs from the shells and placing cotton wool or a soft foam like material inside of the lug to dampen any sympathetic vibrations of the internal lug springs. Lug choice is more of a practical and aesthetic consideration when drum manufacturers are producing lugs for use on their drums. However, additional mass added to the shell and the amount of surface area that makes contacts with shell could potentially impact the resonant properties and timbre of the drum. Figure 2.3 show two snare drum with different lug designs.

Figure 2.3: Example of different lug styles.

2.5.5 Hoops

Hoops or counterhoops are an integral part of the drum that allows the drumhead to be tuned and played, they are place on top of the drumhead and are tightened against the shell by the tension rods which causes the bearing edge to apply pressure to the underside of the drumheads. Nicholls (2008) provides insight into the history and development of drum hoops, before the 19^{th} century drums had predominantly been constructed with wooden hoops that were tensioned with rope. The first metal hoops were originally flat pieces of brass or steel that were bent into a circle, and were tensioned using threaded tension bolts that slotted into clips or claws, sometimes referred to as collar hooks, that clasped over the edge of the hoops. Hoops gradually evolved from a flat circle of metal to a flanged design, that allowed the hoop to sit lower on the drumhead. Another development was a second flange, simply know as double-flanged hoops. The double-flanged hoop allowed for tabs or ears to be included on the hoop for tension rods to be placed through, this design eliminated the need for collar hooks to be used. These flanged hoops were stronger and less likely to distort into an oval shape. The triple-flanged hoop was designed to prevent damage to drum sticks that would occur when playing a rim-shot on a double-flanged hoop. This design features three bends in the metal construction. An example of the triple-flanged design can be seen in Figure 2.4.

VandeStadt (2022) and Waxman (2022b) discuss the development of drum hoop design, stating that today the triple-flanged hoop had become the most popular of the flanged designs, with the other two main designs of hoops being die-cast hoops and wood hoops. Die-cast hoops were introduced around the 1940s as drum makers attempted to eliminate the need for reinforcement rings on the inside of wood shells by creating a stronger drum hoops, this led to the creation of the die-cast hoop design. A die-cast hoop is made by pouring molten metal, typically zinc, steel, aluminum, or brass, into a mold and leaving them to cool and harden, producing a harder and heavier drum hoop than triple-flanged hoops (Hollen, 2022). Wood hoops are typically found to be made of the same woods used for drum shells, notably, maple, oak, beech, birch, mahogany, and walnut. While some modern manufactures still produce wood hoops in the style of the original hoops from the 1900s that require collar hooks, the majority of wood hoops are designed to be tensioned with standard tuning pegs in the same manner as flanged and die-cast hoops, allowing hoops to be easily interchanged.

Hoops affect the amount of overtones generated by the drumhead (Azzarto, 2010a). Flanged hoops are lighter and less rigid than die-cast and wood hoops, allowing the hoop, shell, and drumhead to resonant more freely, resulting in more sustain and producing more overtones. The increased flexibility and less overall material contacting the drum accounts for these differences. Flanged hoops are manufactured in different thicknesses, such as 1.6mm and 2.3mm, thinner hoops will vibrate more than thicker ones. Die-cast hoops being less flexible and heavier than flanged hoops do not vibrate as freely and therefore reduce the amount of sustain produce by the drum and tend to produce fewer overtones, as the extra weight and rigidity dampens the vibrations

Figure 2.4: Triple-flanged hoop design.

of the shell and drumhead (Meyer, 2014). Wood hoops are known to have the biggest effect of shortening sustain and will also change the spectrum of the overtones produced. Hoops can be specifically selected to modify timbral characteristics, for example die-cast hoops could be used on an overly bright snare, to reduce unwanted higher harmonics, whereas triple flanged hoops would exaggerate this quality (Hollen, 2022). Some drummer may choose to use different types of hoops for the batter and resonant sides of their snare drum, for example using die-cast or wood hoops on the resonant head to control overtones and reduce excessive snare buzz, and a triple flanged hoop on batter side to enhance sustain (Azzarto, 2010a).

2.5.6 Snare Wires

The snares or snare wires of a snare drum refer to a set of tensioned spiral wires or cables most commonly placed in contact with the outer side of the resonant head, although on particular snare drums they may be place on the inside of the drum making contact with the batter head. The snares can be tightened against the drumhead or loosened to completely avoid contact with the drumhead, as well as finely adjusted to control the amount of tension (Beattie, 2019). The snares create a unique broadband buzz, or white noise burst heard in conjunction with the typical pitched note of a drumhead. This is caused by the snares vibrating across the surface of the resonant head which is excited when the batter head is struck. This buzzing property differentiates the sound of the snare drum from the other tuned percussion of the drum kit (i.e., the toms and kick drum) (Doerschuk, 2011; Alldred, 2019).

Vinson (2012) outlines the origins and evolution of the material used for snares. Historically snares were made from gut or *catgut*, a fibrous material from the wall lining of animal intestines, predominantly sourced from sheep, goats, cows, and horses. In the early 1900s drum companies began offering other materials as alternatives to natural gut snares. Braided linen, also referred to as *waxed string* by some drum companies, was used as a substitute for gut and was offered in a variety of gauges, it was less affected by changes in temperature and humidity which were problems for natural gut snares. Another alternative which was also more resilient to environmental fluctuations was wire wound silk, a thin strand of silk wrapped with wire. Various versions such as a core of linen or cord rather than silk were available and produce a brighter timbre than gut. The coiled wire or *snappy wire* are the snares that most resembled today's modern snares, they were several separate strands of coiled wire strung onto the drum in the same manner as gut or wire wound silk snares would have been. By the 1920s the use of gut snares had become less popular among drummers who favoured the brighter sound of wire snares for use in the popular music of the time.

Further development of the snare wire design came about in 1916 when Moulton Wheeler patented the James Snappi Wires, which saw the individual coiled metal wires soldered onto metal end plates, allowing them to be quickly and easily mounted onto a drum (Vinson, 2012). These became one of the most popular design due to their simplicity, as previously each wire would need to be individually affixed to the drum, they were also know to produce a much brighter timbral character, which was partly responsible for their large scale adoption (Brensilver, 2015). Although wire snares quickly became the most popular choice for jazz and other modern music of the time, symphonic percussionists as well as rudimental drummers continued to use gut snares, preferring the darker, less sustaining timbre they offered. Today, synthetic materials such as nylon cable or coated stainless steel cable have replaced natural gut, with many modern orchestral, concert, and marching band snare drums typically using imitation gut snares made from plastic (Meyer, 2014).

A snare throw-off or strainer is a mechanism on the shell of the snare drum that allows the snares to be decoupled from the resonant head, and the tension adjusted when engaged, a butt plate is placed on the opposite side of the shell to the strainer, anchoring the snares in place. Jones (2016) outlines the history of the strainer design and development; the first patented design of a snare release mechanism dates back 1889. Prior to this design it was common to have a wingnut or knob-type screw mechanism that could engage, disengage, and adjust the snares to the required tension. Fry's invention served two primary functions, the first being to release the snares after playing, allowing the snares to shrink back to their original length, as they would stretch if used in particularly humid conditions. The second function served as a protecting shield from the tension screw, preventing uniforms and clothing from being torn and damaged. The rise in popularity of vaudeville and jazz music increased the importance of the snare strainer due to the demand for drummers to quickly change from the typical snare sound to an *Indian tom-tom* or *snares off* sound instantaneously during a performance. Between 1889 and 1910 there were several adaptions of different designs which used various mechanisms to decouple and adjust the tension of the snares from the drumhead. In 1914 Robert Danley, from Ludwig introduce a *lever style* snare strainer, which featured a fine-tuning knob enabling performers to adjust the snares to the desired tension (Nicholls, 2008; Meyer, 2014). In 1920, Ludwig began to utilize this strainer as the standard attachment for its drums, this design became the model for continued variations and improvements. From the 1920s many developments were made to snare strainers, surpassing earlier inventions, although the Danley strainer design and its modifications are still used on a range of snare drums in the present day (Jones, 2016).

Modern snares are made up of several wires, these most commonly range from 12 to 24, with 20 strands being ubiquitous, however some manufacturers producer snares with 30 and 42 strand options (Wachtel, 2022), Figure 2.5 show the differences between a set of 20 and 42 strand snare wires. Fewer wires generally yields a more controlled, tighter, snappier timbre, which allow the drum shell sound to be more prominent compared to the snares, alternatively, more wires will make the effect of the snares more apparent, and produce a stronger noise component or *buzz* from the snare drum. The amount of strands on the snare wires will also change the amount of decay produced by the snare drum as the increase tension on the resonant head produced by the additional mass of the extra wires will act to dampen the vibrations and produce a shorter decay (Waxman, 2022a). There are also variations in designs such as split row snares, whereas most snares are comprised of a single row of wires and meeting on the same plane on the head, split row are split in the center to reduce head coupling and sympathetic vibrations. Split row snares are often a described as being drier and crisper and are also reported to be less susceptible to sympathetic resonance from other drums. The construction, material, and amount of strands used for the snare wires results in noticeable timbral differences (Hall, 2020).

The timbre of the snare drum can also be altered simply by changing the material of the snares, typically high-grade steel or brass will be used for snare wires (Doerschuk, 2011). Steel is considered to be neutral, providing a broadband response, with brass models offering a brighter more resonant timbre. As well as steel

Figure 2.5: Different size snare wires, 20 strand (left) and 42 strand (right).

and brass, there are a range of other materials which all offer slight variations in timbral qualities; a composite of carbon and steel can been used, where the higher amount of carbon added to the wires increases perceptual brightness. Bronze and nylon snares are also available, with bronze sometimes also being mixed with phosphor, these are known to produce a more subdued timbre with less higher harmonics (Alldred, 2019). The thickness of the wires corresponds directly to the sensitivity required by the player to excite the snares, more sensitive snares will require lower velocity strikes for the snares to be heard, whereas less sensitive wires will need more forceful strikes to initiate snare vibrations. If soft jazz is being played then a snare with thin wires that respond to lighter velocities may be required, while a heavier playing style may be better suited to thicker gauge wires that can withstand a more powerful impact (Doerschuk, 2011).

Although the snares are an integral component of the snare drum's timbral character they can present challenges when recording the drum kit. Snare wires will inevitably stretch and warp over time due to the constant tension being placed on them, diminishing the desired timbre and becoming looser, eventually the wires can become brittle and break (Schroedl, 2002). In order to obtain the optimum performance from the snares, the wires need to be evenly and centrally positioned, and tensioned so they are not loose and rattling but also not so tight that they prevent the head from vibrating properly (Toulson, 2021). If the wires are too tight the drum will sound choked and unnatural with no sensitivity, under tensioning can also have undesired effects on the sound such as an unwanted buzz caused by sympathetic resonance when the kick drum or toms are played (Alldred, 2019). Tuning both drumheads correctly can help eliminate snare buzz, this may require the resonant head to be tuned to a different pitch. It may also be required to dampen various parts of the resonant head if the unwanted buzzing is particularly problematic, however appropriate tuning, especially of the resonant head and correctly positioned snares should negate much of the issue (Owsinski and Moody, 2009).

2.5.7 Snare Beds

The snare beds on a snare drum shell are the shallow, gradual reliefs in the bottom bearing edge on the resonant side of the drum shell, located where the snare wires straps across the bearing edge to meet the throw-off and butt plate on opposing sides. When the snares are tightened, the snare beds allow them to lay flat again the resonant head, without them the snare would rattle, and be more susceptible to unwanted buzzing (Meyer, 2014). Older snare drums tend to have deeper snare beds. These deeper beds were intended for use with animal skin drumheads, which were tucked around wood hoops and would mould to follow the contour of the bearing edge on the shell without wrinkling and sagging near the beds. The deeper beds allowed the snares to sit flush against the heads without having to over tighten the snares (Nicholls, 2008). A modern plastic head is not pliable enough to stretch down into deeper snare beds, which would causes them to wrinkle around the beds. For this reason modern snare drums feature a more gradual bed that allows better contact between the head and the bearing edge. An uneven, misaligned, or poor quality snare bed can create issues when attempting to tune the batter head and will increase the likelihood of the snares buzzing and rattling from sympathetic vibrations, and put uneven strain on the snares causing them to stretch asymmetrically over time (D'Amico, 2001).

2.6 Tuning

Unlike other pitched instruments there is no agreed upon tuning for the snare drum in the same way that there is for instruments such as pianos and guitars. The tuning of the snare drum does not necessarily need to be in the same key as the other instruments or tuned to a particular note such as B♭. There are a wide range of acceptable tunings of the snare that will be suitable for different songs and different genres. Even within a given genre, such as metal, the snare's tuning can be broad from artist to artist, or across different albums or even different songs by the same artist. Mynett (2011) explains that there are no approaches or principles that can be deemed specific to snare tuning for a given genre, examples of low and high tunings can be found in almost all genres. The manner in which the two drumheads are tuned, and therefore interact with each other, has a highly significant impact on the drums timbral characteristics. Although there are several devices available to assist with tuning drums, they are most frequently only used to get a drum tuned into the correct general range, and then fine-tuning is completed by ear; making small adjustments of each of the tuning pegs using a drum-key until the drummer or engineer is satisfied with the tuning. These devices operate by providing a measurement of pressure reading at a specific distance from each drum lug for example the DrumDial, or by measuring the acoustic vibrations and representing that as a note on a display such as the Tune-Bot from Overtone Labs. Distinguishing the Overall pitch of the snare drum can be challenging as there are multiple and complex overtones associated with the snare drum, and the snare wires add a noise component that can mask the pitched elements of the drum. Drum tuning, in combination with re-heading and dampening, should be at the foundation of obtaining the timbre that the engineer is striving to capture prior to recording. Owsinski and Moody (2009) add that snare drums are often tuned in such a manner to best suite the genre of the song, but also notes that tuning will be heavily dependant on both the room that the drums are played in, as well as the exact location within the room itself, as certain nodes or anti-nodes of the room's acoustics may interact with the drums in a manner that requires tuning adjustments to be carried out. The most noticeable effects from tuning differences are produced by the batter head, and to a lesser extent the resonant head. Tuning has the potential to alter the timbre more than any other individual factor, as tuning will also effect the transient response and envelope characteristics of the drum strike. Tuning also effects the responsiveness of the stick rebound, higher pitched tunings will generate greater tension on the drumhead, causing the stick to have a faster more responsive rebound, which may consciously or subconsciously effect the player's performance (De Douvan, 2005; Wagner, 2006).

Schroedl (2002) notes that the number of lugs effects uniformity of tuning across the entirety of the drumhead and ultimately has an effect on the timbre of the snare drum. The fewer lugs on a drum, the coarser the tuning will inevitably be, with each lug having a greater impact on the change in pitch. With fewer lugs, the lengthened distance between them will results in larger sections of the hoop applying less tension between the edge of the drumhead and the bearing edge of the shell, which results in a darker timbre with less high frequencies, and more variability in timbre across the drumhead. It is common for more high-end snare drums to use as many as ten lugs, the added lugs incur additional cost, as well as mass on the drum, cheaper or vintage drums may have as little as 4 lugs, with most having 6 or 8. Schroedl expresses that in order to properly manipulate the attack and sustain characteristics through tuning alone, the tuning process must first be be carried out without any form of dampening applied. Once the drums have been tuned appropriately then they can be dampened accordingly. In addition, if possible the snare should be tuned away from the other drums and cymbals, as they will resonate when the snare is hit, clouding ones pitch reference and making it more difficult to tune accurately. The head that is not being tuned should be muted, so that only the head that is being adjusted is heard when played, this can be done by setting the drum on a towel preventing the bottom head from vibrating. This can help to mitigate some problems associated with tuning the snare drum, such as snare wire noise making it difficult to hear the subtle tuning differences between each lug positions. Toulson (2021) agrees that tuning analysis should be carried out with the wires disengaged, as it allows the resonant head to vibrate more freely and gives a truer representation of the pitch of the drumhead. However, it is advised that the final assessment of subjective quality be conducted with the snares engaged, as this is how the snare will inevitably be played. Schroedl (2002) suggests that the snare drum should typically be tuned slightly higher than the toms and in general have a shorter sustain. The higher the snare is tuned the more prominent and noticeable the fundamental pitch will be. Uneven tuning can cause unwanted ringing or resonances when the drum is struck, it is preferable to address this issue with careful tuning modifications than to simply apply excessive dampening to the drum head in an attempt to fix these problem frequencies. Although dampening may remove the ringing, it will also reduce the high frequencies, this may or may not be desirable. Toulson (2021) notes that dampening is regularly applied to snare drums if the desired tuning can not be achieved, however, suggests that if the tuning is performed correctly, and the drumheads have been carefully selected, dampening should not actually be required. If excessive dampening is needed in order to achieve the desired timbre, then this is typically compensating for poor tuning, and instead a replacement drumheads with some form of built in dampening system should be used. Too much external dampening will impact the evenness of the drumhead's vibration, particularly if only applied to one location of the drumhead.

There is some agreement for a general approach to obtaining a good overall tuning, that serves as a baseline for the engineer or drummer to make tuning modifications to as desired. This technique seeks to ensure that there is the same tuning at every lug position around the perimeter of the drumhead, this enables the drumhead to vibrates evenly across its entire surface (Gibson, 2004; Major, 2014; Bartlett and Bartlett, 2016). Toulson (2021) explains that this process is referred to as *clearing* the drumhead, Lug tuning, or equalising the drumhead. Uncleared drumheads can produce undesirable effects, if one point around the drum has a slightly different frequency to the rest of the head, this will cause the harmonics to negatively interact with one another. This manifests as an audible pulsing or warble, this type of modulation is called beating, which is particularly apparent after the initial attack as the drum strike beings to decay. This beating occurs when two or more frequencies are close but not identical and are present at the same time. The better and more even the tuning at every lug position, the more likely it will be for the drum to have a single strong pitch that is easily identifiable without several conflicting frequencies. Even tuning implies that the drumhead vibrates with uniform overtones at each lug position, and that there are no beating frequencies. Tuning devices can be utilised to aid when clearing the head by providing a diagnostic method of evening the pitch or tension at each lug, making the process more exact, and removing subjective differences.

Gatzen (1994) offers some additional recommendations to ensure optimum tuning can be achieved, which starts by ensuring that the head is seated evenly against the bearing edge. It is recommended to tighten each tension rod until they are *finger-tight*, meaning they can no longer be tighten any more by hand, and would require a drum key to apply additional tension. This is to ensure that each tension rod has roughly equal tension before using a drum key to tighten them further. Once this has been achieved the drummer should begin tuning in half-turn drum-key increments in a crossing pattern between the tension rods, this ensures that one side is not pulled down further, causing the opposite side to raise up and resulting in an uneven drumhead. The drummer can tap using their finger or lightly use a drumstick to play the drumhead near the edge at the lug position to compare the pitch of the overtone, then an assessment needs to be made to apply more or less tension to ensure that the pitch appears to be the same all the way around the drum. As the drumhead approaches uniformity the wobbling overtones will begin to disappear as they all become closer in frequency. Most drummers will either tune the batter and resonant heads to the same pitch, or tune the resonant head of the snare drum higher or lower than the batter head depending on the intended goal. This may include tuning the snare batter head a minor third or a perfect fourth above the resonant.

Toulson (2021) further explains tuning principles and presents a method for drumhead tuning. The drumhead is able to vibrates at many different frequencies all at the same time. The fundamental frequency produced is dependant on the relationship in tuning between the batter and resonant heads, manipulating this tuning ratio allows for control over the range modal frequencies (Richardson et al., 2012). Certain frequencies and modes may resonant more strongly than others depending where on the drumhead it is struck. The fundamental frequency, referred to as F0, is excited most when the drum is hit at the centre of the drumhead and the first overtone frequency, referred to as F1, is excited most when the drum is hit at the edge. If the drum is hit somewhere between the centre and the edge, then both frequencies can be excited evenly at the same time. There are many different ways a vibrating drumhead can physically deform and vibrate, the drumhead can vibrate on a circular axis (around the drumhead) or in a diagonal axis (across the diameter of the drumhead). All snare drums will have several tunings that they will sound best at, typically a snare will perform optimally at 2 or 3 different pitches for lower and higher tunings. For a 14" snare, a typical low tuning F0 pitch may be around 160Hz, while a higher tuned snare may be closer 200Hz. Toulson introduces the concept of the Resonant Tuning Factor (RTF), which refers to the relationship between the the batter head's F0 and F1 frequencies. It is important to note that the resonant head is actually responsible for controlling and adjusting the RTF. In order to increase the RTF: increase the batter head tension and lower resonant head tension, to decrease RTF: lower the batter head tension and increase resonant head tension. An RTF of around 1.5 relates to a musical 5^{th}, as the frequencies of a the fifth are 1.5 times the root note, with the octave being 2 times the root. This means that with an RTF of 1.5 when the snare drum's F0 is 180Hz when struck in the centre, the F1 frequency would be 270Hz when struck near the edge. This particular frequency ratio is related to the psychoacoustic phenomenon know as the phantom sub-harmonic or a missing fundamental. This occurs when there are two frequencies presented to a listener where one is 1.5 times the frequency of the other, because these two tones are musically related, the listener will anticipate a lower frequency of 0.5 times the first and perceive an additional tone that is not present, this gives the sense of low-frequency power to the strike. The RTF could be calculated using a tuning device or spectrum analyser however Toulson has specially design an application called the iDrumTune app, which measures the pitch at the centre of the drumhead and the edge and will automatically calculate the RTF. It is advised to ensure that the pitch at each lug position is within 1 to 2 Hz of each other. The optimal RTF value is recommended to be around 1.5 although for certain drums they may perform better with values of 1.6 or higher. This allows for a repeatable method to achieve the same tuning when changing drumheads, as well as being able to accurately match tuning of two different drums for subjective assessment of other variables such as shell material and type of drumhead. Different RTF values will produce a unique balance between the fundamental and the overtones of the snare drum.

2.7 Dampening

Certain resonant frequency produced by the snare drum can have noticeably longer decay times than others, these persistent frequencies are typically referred to as *ringing* (Charlton, 2017). Ringing can often be an undesirable characteristic, thus there are several different methods to dampen the vibrating drumhead in such a way to minimise or remove this unwanted ringing (Huber and Runstein, 2010). Drums that have a shell depth that is larger than the diameter of the drum will suffer from noticeably more ringing issues (Pedersen and Grimshaw-Aagaard, 2018). If a shorter sustain or a timbre with less overtones is required, Toulson (2021) suggests that better results can be achieved by first selected a more appropriate drumheads rather than relying on heavy amounts of dampening. However changing drumheads may not always be an option, due to financial or time constraints, and therefore the drummer or engineer may need to modify the sound of the drumheads through tuning and dampening to achieved the desired timbre. Although dampening the drumhead will reduce the decay of the unwanted frequency any mass that inhibits movement of the head will shorten the decay times of some of the desirable frequencies as well, the drummer and engineer must carefully apply the correct amount of dampening to remove the necessary amount of overtones without detrimental results to the overall timbre of the snare drum. The extent to which a snare drum is dampened may also be heavily influenced by the genre of music, with genres such as jazz using a snare drum with little or no dampening, and genres such as rock or hip-hop using snare sounds that are more heavily dampened and with much shorter decay times (Owsinski and Moody, 2009). Excessive dampening could result in the snare drum being completely devoid of any sustain, overly reduce high frequencies, and sounding unnatural. In some cases careful adjustments to the tuning of either the batter or resonant head can help to mitigate some ringing, an evenly tuned snare may not require any dampening. If tuning is unable to address the problem frequencies, then providing the availability, an engineer may consider changing from a single-ply head to a double-ply head, or using a coated head as appose to a uncoated head, however if the ring is still pervasive some form of dampening may be essential (Dowsett, 2015). In order to achieve professional sounding recordings the drums must not only be correctly tuned but also correctly dampened to the appropriate amount to reduce unwanted ringing (Gibson, 2004).

Damping can dramatically alter the sound of the snare and is most often used on the batter head. If the engineer is not satisfied with the timbre of the snare drum, before microphone selection and placement are considered, the engineer may modify a combination of the drumheads, the tuning, and the dampening used (Savage, 2011). There are several commercially available products specifically designed to reduce unwanted ringing, one of which is RTOM MoonGel, a self-adhesive gel rectangle (3.5cm x 2.5cm) that affixes directly to the drumhead, it has the benefit of being able to be place any where on the drumhead to specifically target an area responsible of ringing. Placing the MoonGel closer to edge of the drum drumhead will reduce its dampening effect, whereas placing it closer to the centre will increase its effects, in can also be easily cut down to a more appropriate size if needed, and any amount can be applied to achieve the desired result (Dowsett, 2015). Although versatile, any dampening method that adds mass to only one location of the drumhead, can cause the head to vibrate in an uneven and unequal manner, and possibly have deleterious results to the timbre (Toulson, 2021). Another style of dampening products include dampening rings, *control* rings, or *O-rings* which are thin plastic rings that fit around the periphery of the batter head and simply rest on top, these can be between 1/2" to 3" wide, with the width determining the extent to which they dampen the snare (Owsinski, 2005; Major, 2014). This style of dampening affects the vibrations towards the edge of the drumhead, which allows the fundamental to remain somewhat unaffected, the heavier and wider a dampening ring is the more they will also dampen and shorten the fundamental frequency (Toulson, 2021). Other products include; Snareweight, a small brass weight that attaches to the rim of the drum via a magnet with interchangeable leather inserts; Big Fat Snare Drum, a circular disk made from a blend of rubber and plastic, designed to cover the entirety of the snare drumhead and apply a large degree of dampening; Remo Weckl, a free floating adjustable dampening system that attaches to the rim of the drum and features an adjustable plunger mechanism that can apply varying amounts of dampening; and various other designs from a range of manufacturers. Some drums are installed with internal dampeners that allows the user to adjust the amount of dampening via an external control on the shell of the drum (Cesarz, 2022).

As well as commercial devices a variety of alternative methods can also be implemented. Old drumheads can be repurposed to create dampening devices, either by simply placing one upside down on the batter head, or by cutting a hole cut in the middle to reduce the effect of dampening, leaving a 1" or 2" ring around the outside similar to control rings (Crich, 2010). A heavy leather wallet can be simply placed on the drumhead to also provide dampening, the size, weight, and placement will dictate the characteristics of the dampening (Savage, 2011; Allen, 2020). Gaffer tape can also be effective at removing overtones, due to its adhesive nature it can also be utilised on the resonant head as well, folding sections of the gaffer tape increases the mass on the head without taking up additional surface area, this can be used to customise the amount of dampening (Borwick, 1980). Any kind of cloth material can be placed on the edge of the drumhead, similar to where MoonGel would be placed, taping this in place will reduce movement for consistent dampening (Dowsett, 2015). For even more dampening the snare drumhead can be treated by taping a folded paper towel, gauze pads, tissue paper, or folded handkerchief to the top of a drumhead, a few inches off its edge, taping down 3 sides of the pad, leaving one edge to vibrate and dampen the head in motion (Gibson, 2004; Owsinski and Moody, 2009; Bartlett and Bartlett, 2016). A few examples of different dampening methods and products can be seen in Figure 2.6. Any of these techniques or combination of techniques can be utilised by drummers and engineers to achieve the designed timbral qualities they are striving for prior to recording the drums. The methods described above are not an exhaustive list, as there are many other methods and products used to dampen and manipulate the timbre of the snare drum. The amount and type of dampening required will be dependant on the drum heads used, tuning, song, tempo, genre, where in the room the drums are positioned, how the drums are played, and ultimately personal preference.

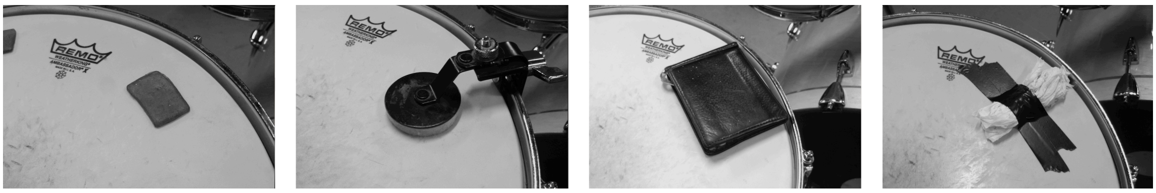

Figure 2.6: Example of different dampening methods. From left to right, moongel, dampening bracket, leather wallet, and tape and tissue paper.

2.8 Drumsticks

In contemporary music the snare drum is most typically played with a wooden drumstick, although a variety of other implements exist and are utilised by drummers, there also exists variations of drumstick design. Libman (2019) states that drumsticks affect the timbre of the drums when struck, but the majority of drummer will select a drumsticks based on the way they feels and react whilst playing. Factors such as stick length, diameter, weight, balance, and texture are all considerations the drummer can make to best suit their hand size and playing style. The different timbral qualities associated with different sticks specification is often overlooked or an after thought.

Dierstein et al. (2018) described the different implements that have been and are still currently used to play drums. Traditionally and most notably the drumstick or drum stick is a rounded wooden mallet which tapers in a conical shape towards the tip just before the head of the stick, the whole stick is made from one piece of wood. The most common wood types used for drumsticks are maple, hickory, white beech, and rosewood. The head of the drumstick is a spherical or droplet shape and sometimes is covered in hard plastic. The stick's conical form enables an optimal balance of weight in the hand, which is not the case for mallets that will be weighted closer to the beater. The length of the conical transition between the stick's head and the part where the taper beings, called the shoulder, varies across different models, the shorter the transition the more head-heavy the stick will be. Other types of drum striking implements include rutes, or rods, and various types of brushes. Rutes can be made from around 5 to 20 pliable sticks that are tied together at the non striking end. Various materials can be used for this, such as small, rough wooden branches or twigs, bamboo, straw, yucca leaves or plastic sticks. Today, multiple forms of rutes, often simply called *rods* or *bound rutes*, can be made of birch dowels, other thin canes, or synthetic materials. Modern day rods tend to be bound in several places, to compacted them more tightly together that traditional rutes, because of this tight binding, they are more versatile to use and, in terms of playing technique, are more similar to drumsticks. Brushes, jazz brushes, or percussion brushes, are striking implements specifically developed for playing the drum set, and their origin and main area of use today is in the jazz music genre. The majority of brushes belong to two main types: those with bundled steel wire and those with nylon wires, each of which has different timbral qualities. There are various models of brushes that have wires of varying thickness and firmness. Thin, soft wires produce a relatively quiet and delicate sound. The thicker and harder the wires are, the louder and coarser the timbre. Steel wire brushes tend to have a much louder and brighter sound compared to nylon brushes. Most nylon models consist of wires that can be pulled inwards and outwards. Similarly to steel brushes, the sound can be varied by changing the length of the strands. Using brushes has predominantly been developed for use with rough, textured drumhead, which achieves a sweeping sound. The relationship between how course the drumhead is and the material and thickness of the brushes affects the range of possible timbres.

Azzarto (2010c) highlights some of the differences between contemporary drumsticks. Many drumstick manufacturers denote different specification of their drumsticks with a letter/number system. Originally, the letters were intended to denote particular styles of drumming, for example the letter A stood for orchestra and the letter B stood for marching and concert bands. The number related to the diameter of the drumstick, with the lower numbers relating to thicker diameters, and higher numbers used for thinner drumsticks. Some manufactures still produce classic modules of some of their earlier styles of sticks such as 2B, 5A, 5B, and 7A. However, most companies have adopted their own individual standards for naming and numbering their drumsticks that do not directly relate to the size or shape of the drumstick and are not consistent across different manufacturers. Hickory and maple are the two most common woods that modern drumsticks are made from, however other woods such as oak is also widely used, as well as other resins and plastics, these are know to all have different timbral and tactile properties, that may be favoured by drummers playing a certain style of music. Some companies produce aluminium drumsticks with a protective polyurethane coating and replaceable nylon heads, they are designed to offer extreme durability and better rebound properties than wooden sticks. In addition to different materials, drumsticks can also feature different tip shapes. There are five basic shapes of drumstick tips, and each one produces a subtly different timbre and rebound characteristic, these tactile differences are most apparent when playing cymbals. The most common head shapes are oval, teardrop or arrow, round or ball, acorn, and barrel, the shape and size of the tip will change the amount of surface area that makes contact with the drumheads and cymbals, thus slightly affecting timbre.

As well as tip shape, whether the tips are uncoated or have a nylon coating will also produce slightly different timbral results due to their different densities and rebound characteristics. The harder nylon tips generally produce more high frequencies and a brighter sound than wooden tips (Gibson, 2004; Libman, 2019). All the variables associated with drumstick design, while subtle, can have an impact on the timbre, which may inform or influence the way the drummer interacts with the drum set to varying degrees, for example, brighter sounding sticks may encourage a drummer to play softer. For the remaining chapters any recordings of snare drums are made with the snare being struck by a drumstick; no brushes or rods are used. This is for consistency as these other implements are considered special use cases, used for certain applications, genres, and effects.

2.9 Conclusions

As has been shown in this chapter various properties of the snare drum's physical construction can be changed and manipulated to effect timbre, and there are innumerable combination of these variables that the engineer and drummer can modify to achieve the desired result. Some of these changes will have a more noticeable and dramatic affect than others, and not all variables will be accessible for changing or adjustment during a recording session. In some cases the easiest variable to change for the greatest impact could be to simply use a different snare drum, another scenario might see the snare drum only requiring slight tuning adjustments in order to achieve the ideal timbre. The next chapter will discuss recording the entire drum kit and explore the role that different microphones have on the timbre and subjective preference of snare drum recording.

Chapter 3

Recording the Snare Drum

3.1 Recording Drums

The standard modern drum kit is commonly comprised of a kick drum, snare drum, one or more rack toms (also known as tom-toms), a floor tom, a hi-hat, and a variety of cymbals, most typically a crash and ride (Huber and Runstein, 2010). An example of a drum kit can be seen in Figure 3.1. Drum arrangement acts as the rhythmic foundation for the other instruments to be recorded over the top of and thus influence how the other instruments should be recorded and subsequently mixed. Genre can often dictate the timbral characteristics of the drums and thus the recording techniques used, for example genres such as heavy rock may need to be hyper-realistic, while the drums for jazz music may simply be an accurate acoustic representation (Major, 2014). The drum kit can be one of the most difficult instruments to record, as it combines many different sounds into one instrument (Pedersen and Grimshaw-Aagaard, 2018). However, before microphones are set up, a number of other practical and creative decisions need to be made in order to optimise recording quality.

One major consideration is the room in which the drums are to be recorded. When drums are played, the acoustic energy produced excites the room, which resonates in response, and thus in turn re-excites the drums; this interaction modifies the resonances of the kit (Newell, 2012). The natural ambience of a recording space helps to consolidate the individual elements of the drum kit together, helping to contextualise it as one complete instrument. Factors such as the size of the room, its geometry, construction, surface coverings, and amount of furniture will all influence how the drums are perceived in the acoustic space. The acoustics of a

Figure 3.1: Example of a standard modern drum kit configuration.

recording environment have the ability to transform the timbre of any instrument dramatically and will dictate the overall character of the recordings (Savage, 2011; Senior, 2018). Hard surfaces like tiles and glass will be highly reflective, whereas softer materials, such as carpet, curtains, and acoustic foam, will be absorbent. Different material in the room will reflect and absorb certain frequencies by varying amounts (Parsons and Van Horn, 1996). High levels of reverb in a room can be detrimental to the character of the drum recordings, and rooms with excessive high frequency reflections can make localisation of specific parts on the kit difficult (Major, 2014). In order to enhance the characteristics of a recording space, bass traps, absorbent panels, and diffusers can be added to reduce problems such as flutter echo and frequency cancellation in a studio (Owsinski and Moody, 2009). Once the room has been selected the drums need to be placed in the optimal position within the recording space. If the studio has a large window between the live room and the control room, it is advised to position the drums away from the glass as it will produce a lot of unwanted reflections (Huber and Runstein, 2010). Positions extremely close to a corner should also avoided as this can cause the bass frequency to be over-emphasised although this may used by an engineer specifically seeking this effect (Owsinski and Moody, 2009).

The drums most appropriate for the song, genre, and desired timbral characteristics need to be selected; including how many drums and cymbals are to be used, along with the brand, dimensions, and material of each. Drums that are not played for a particular song may be removed from the drum kit completely to reduce sympathetic resonances from occurring, whereby a drum may produce sound when not being played by resonating in response to other drums near being struck.

Once the configuration of the drum kit has been decided upon, both batter and resonant drum heads will need to be chosen for the snare, kick, and toms. When selecting drum heads for a recordings session the timbre they produced should be the main factor, whereas, heads chosen for live performance may be selected based on their durability, cost, or volume (Beck, 2004). Worn drum heads may be stretched or dented, diminishing their ability to be correctly tuned and preventing a consistent pitch and timbre across the entire head (Gibson, 2004). Old heads may also lose their ability to produce high frequencies leading them to sound dull and lifeless; it is often recommended to fit new drums heads prior to a recording session to optimise sound quality (Bartlett and Bartlett, 2016). Depending on the desired character specifics such as single-ply, multi-ply, coated, or uncoated heads may be chosen, as well as different combination on different drums. Tuning and dampening the drum heads has a significant impact on the pitch and timbre of the drum respectively, these can both be used in creative ways to further modify the drums for preferred results (Parsons and Van Horn, 1996; Beck, 2004).

If not correctly attended to certain areas of the drum kit can produce unwanted extraneous noises. Loose tuning rods, cymbal stands, or other poorly fitted hardware have the potential to cause buzzes and rattles, mechanical parts such as the kick drum pedal or hi-hat pedal can cause other non-musical noises, great care and attention is required to ensure no part of the kit is resonating or producing unintended sound of any kind (Parsons and Van Horn, 1996). In some circumstances the process to correctly set up and position the drums in the room, apply and tune new drum heads, and eliminate ringing and rattling noises may take several hours (Pedersen and Grimshaw-Aagaard, 2018), only once the drums are configured to a satisfactory standard will an engineer begin the process of positioning microphones around the drum kit in order to record them (Beck, 2004; Owsinski, 2005).

Drums are capable of producing very high sound pressure levels and are therefore routinely recorded in isolation from the other musicians. This is in order to avoid the drums being captured by the microphones intended for the other instruments in the studio (Eargle, 2004). When recording a full drum set, engineers may choose to

Figure 3.2: Microphone configuration showing close microphones, stereo and a mono overhead microphones.

use any combination of microphones placed closely to individual drums referred to as close or *spot* microphones as well as microphones placed at various distances away from the drums in order to capture the sound of the entire drum kit. Microphones placed above the drums, know as *overhead* microphones, are often a matched stereo pair—panned left and right—positioned to primarily capture the cymbals and toms (Gibson, 2004; Owsinski and Moody, 2009). This technique can emulate how the stereo spread of the full drum kit may be heard by a listener in the recording space. Panning the two microphones to different speakers by varying amounts determines the width of the perceived stereo image. Alternatively some engineers may opt to record a monophonic signal by using a single overhead positioned either centrally above the whole kit or positioned to emphasize the snare drum, others may use this mono overhead in conjunction with a stereo pair (Major, 2014). Microphones placed further away specifically set up in a way to capture the natural ambience of the recording space are know as *room* microphones, any type of microphone or amount of microphones could be chosen for this role. Room microphones are typically placed around 6' away from the front of the kit and around 6—7' from the floor. As an alternative two microphones can be placed one either side of the kit, around 10' apart from each other, or in larger rooms, microphones can be placed near to the furthest wall in order to capture the greatest possible amount of ambient sound (Owsinski and Moody, 2009). Microphones placed further away from the kit will emphasise more of the ambience, where closer positions will capture more of the direct sound of the drums. Factors such as the band's or engineer's personal taste, the genre of the song, or the tempo may dictate how the microphones are arranged around the kit, as well as how many microphones are selected to emphasise specific elements. Depending on the size and configuration of the drum kit, engineers may use a microphone setup ranging from only one or two microphones up to 15 or more, made up of a combination of close and ambient microphones (Pedersen and Grimshaw-Aagaard, 2018). Figure 3.2 shows close microphones, a stereo pair of overheads, and a central mono overhead microphone set up for recording a full drum kit.

Capturing individual elements of the drum kit with close microphones gives the mixing engineer additional control over how they manipulate the subsequent recordings with audio effects such as equalisation and

Figure 3.3: Close microphone placement on the batter and resonant head of the snare drum.

compression whilst not affecting the tonality of the other drum recordings (Huber and Runstein, 2010). The overall balance of the entire drum kit can also be adjusted; if the engineer wishes to have the snare drum louder, the associated close microphone recording can be raised in volume, thus retroactively modifying the recorded performance of the drums. Close microphones are most notably used for the snare and kick drum (Owsinski and Moody, 2009), however, close microphones can be place on any drum, including the cymbals and toms. Engineers may place multiple microphones on a single elements to capture specific timbral qualities produced by different parts of the instruments; such as the batter and resonant drum heads of the snare (Owsinski, 2005; Tingen, 2005; Senior, 2008), as can be seen in Figure 3.3. Multiple microphones can also allow the engineer to combine the complementary timbral characteristics of two different microphones, pointing at nearly the same, or slightly different parts of the drum head which allows for timbral manipulation without the need for an EQ (Major, 2014). When an additional bottom microphones is used on the snare drum it is often necessary to reverse the polarity of one of the microphones due to destructive phase cancellation, this is typically carried out on the bottom microphone (Huber and Runstein, 2010).

3.1.1 Microphones

Microphones are one of the most important elements in the recording chain, with each imparting its own unique characteristics. Engineers will choose specific microphones for certain parts of the drum kit, in order to highlight the strengths and mask the weaknesses of the drums being recorded (Major, 2014). As microphone selection plays a key role in achieving the desired timbre some manufacturers produce microphones that are specifically designed for the recording of certain instruments, such as the kick and snare drum. Snare specific microphones may feature tailored frequency responses or casings and attachments, allowing the microphone to be connected directly to the drum without hindering the drummer's playing. Different microphones are know to produce varied timbral qualities based on several factors, including but not limited to: topology (dynamic, condenser, ribbon), size, shape, weight, and material of the diaphragm and housing of the microphone, material and arrangement of protective grills and meshes, polar pattern, frequency response, and non-linear characteristics associated with the active circuit and passive components such as transformers (Ballou, 2008).

These design considerations all impact the resulting recordings in varying amounts and can be selected to better suite the particular timbre the engineer is aiming to achieve. Some microphones may noticeably modify the original acoustic signal, referred to as *colouring*, whilst other manufactures aim to capture the most accurate representation of the original acoustic signal (Pedersen and Grimshaw-Aagaard, 2018). Different microphones can then be used as creative tools to sculpt and manipulate recordings to better fit the rest of the song and reduce time intensive post processing.

Microphones are transducers, they convert the mechanical motion of their diaphragm into an alternating voltage, the diaphragm is sensitive to changes in air pressure which are caused by a sound waves (Borwick, 1980). The three main topologies of microphones used for recording drums are dynamic, condenser, and ribbon, although others such as piezoelectric or pressure zone microphones may also be utilised (Bartlett and Bartlett, 2016). Dynamic microphones, also referred to as the electrodynamic, electromagnetic, or moving coil microphones, use a coil of wire attached to a thin, lightweight, flexible, usually circular diaphragm, suspended within a magnetic field. Sound pressure waves cause the diaphragm and coil to moves backwards and forwards producing a voltage proportional to the strength of the magnetic field which represents the audio signal (Eargle, 2004; Crich, 2010; Corbett, 2020). Ribbon microphones use the same principle as dynamic microphones; however, instead of a coil of wire attached to a diaphragm they implement a folded ribbon of aluminum, which is fixed at the top and bottom and surrounded by a magnet. As the the ribbon is excited by sound waves, it moves backwards and forwards within the magnetic field which generates an electrical representing the audio signal (Corbett, 2020).

Condenser or capacitor microphone work differently to dynamic microphones, they make use of two very close adjacent metal plates, one stationary and the other acting as a diaphragm which vibrates in response to sound pressure waves, the two plates are charged with a constant voltage known as phantom power, which forms a capacitor (Bartlett and Bartlett, 2016). The sound pressure fluctuations causes the distances between the fixed plate and the diaphragm to change, varying the capacitance between them and resulting in an electrical output that varies with the acoustical signal. Condenser microphones have substantially lower mass diagrams than dynamic microphones, allowing them to respond faster to transient information (Ballou, 2008). Valve or vacuum tube microphones are types of condenser microphones that use valves within their internal preamplification stage instead of transistors, in order to charge the valves an additional external power supply is required (Crich, 2010).

Although any microphone is capable of recording a snare drum, engineers will have their preferred choices for specific situations depending on the desired timbre. The Shure SM57 is often selected for the task of capturing the snare drum and has become ubiquitous among recording engineers for this application, this is due to its frequency response, accentuating the attack of the transient, its off-axis response, polar pattern, and its proximity effect (Gibson, 2004). The Shure SM57 has a low frequency roll off that can help to prevent some of the kick drum bleed from being captured on the snare channel, without having to use a low pass filter (Major, 2014). Some alternatives to the SM57 include; Sennheiser MD-441, Audix i5, AKG C414, audio technical ATM650, Neumann KM84, AKG C451, and Shure Beta 57a (Senior, 2008; Owsinski and Moody, 2009; Pedersen and Grimshaw-Aagaard, 2018).

The distance between the sound source and the microphone is know to affect its frequency response. When the microphone is placed very closely the sound source and an effect known as the *proximity effect* occurs, this results in a noticeable increase of low frequency energy being captured on the recording (Crich, 2010). Directional microphone produces an output signal in relation to the difference in sound pressure levels between the front and the rear of its diaphragm, because of this design all directional microphones exhibit proximity effect of varying amounts. Lower frequencies can still be regarded as consisting of a spherical wave near the point of generation, this is because for low frequencies the amplitude drops more rapidly with distance compared to high frequencies (Burroughs and Woram, 1974). This effect results in low frequency air pressure differences between the front and rear of the diaphragm being greatly exaggerated (Eargle, 2004). Omnidirectional microphones are not affected by the proximity effect (Borwick, 1980). The engineer can use this to add low frequencies to the instrument by varying the distance between the microphone and the sound source, increasing the bass as distance decreases (Ballou, 2008). Polar pattern, or *pickup pattern* refers to sensitivity of the microphone from different angles. When acoustic signals approach the microphone perpendicular to its diaphragm, this is known as being *on-axis*. The signals reaching the microphone from other angles will be *off-axis* (Crich, 2010). The sensitivity of the microphone from all possible angles will dictate its polar pattern.

Pressure-operated microphones are ones that have their diaphragm open to the air only on one side. The movement of the diaphragm will response to air pressure fluctuations above and below the normal atmospheric level (Borwick, 1980). These types of microphones have an omnidirectional polar pattern (Tashev, 2009) and theoretically capture sound from a sphere around the diaphragm. Most microphones will exhibit omnidirectional behavior at lower frequencies due to the longer wave lengths; however, the physical casing of the microphone can act as an acoustic baffle to higher frequency sound waves causing aberrations in the directional response, this occurs at frequencies where their wavelengths become a significant portion the microphones size (Eargle, 2004). This means that even omnidirectional microphones will become somewhat more directional at higher frequencies (Corbett, 2020).

Pressure gradient microphones have a diaphragm open to the air on both sides to varying amounts. Microphones completely open on both sides will produce a *figure-of-eight* polar pattern where maximum sensitivity is 0° and 180°, and sensitivity reduces close to 0 for 90° and 270° (Borwick, 1980). The amount to which the rear of the microphones diaphragm is open dictates the properties of its polar pattern. A unidirectional microphone is most sensitive to sound coming from one direction, the 3 main types of unidirectional microphone patterns are cardioid, supercardioid, and hypercaridiod. A cardioid microphone has a null (i.e., the angle at which the microphone is least sensitive) at 180° behind the microphone (Tashev, 2009; Corbett, 2020). Microphones with a cardioid pick up pattern are sensitive to sound pressure fluctuations across a broad angle in front of its diaphragm, roughly around 6dB less sensitive from 90° and 270° from the front, and 15—25dB less sensitive from the rear (Bartlett and Bartlett, 2016). Both hypercardioid and supercardioid give better off-axis rejection of sound at 90° and 270° than a cardioid pick up pattern (Borwick, 1980; Huber and Runstein, 2010). Supercardioid is around 8.7dB less sensitive from the sides and hypercardioid is around 12dB less sensitive from the sides. Both hypercardioid and supercardioid patterns reject sound from the side of the microphone more so than cardioid; however, they do also exhibit some sensitivity from the rear of the diapharagm. Hypercardioid microphones have more side rejection but also higher sensitivity at the rear than supercardioid (Bartlett and Bartlett, 2016). Hypercardioid microphones can be thought of as being a mixture between figure-of-eight and cardioid, although similar to hypercardioid, supercardioid microphones have less of a rear lobe at 180° (Tashev, 2009; Talbot-Smith, 2017). A 3D representation of cardioid, hypercardioid, figure-of-eight, and omnidirectional pickup patterns is shown in Figure 3.4.

The polar pattern of a microphone needs to be taken into consideration when positioning close microphones around the drum kit as sound from other elements will be captured by a microphone off-axis. The presence of extraneous sounds from other signals than those of the intended instrument is often referred to as *spill* or *bleed* (Pedersen and Grimshaw-Aagaard, 2018). For example, when recording the snare drum, the hi-hat can be in close proximity to the microphone, the hi-hat signal will also be present through the snare microphone. An

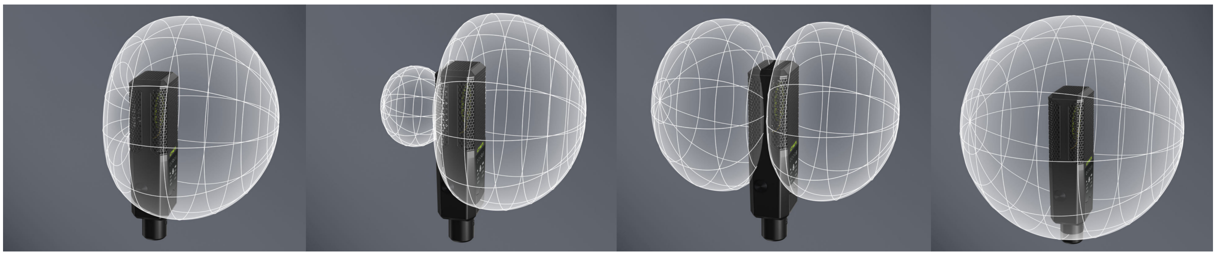

Figure 3.4: Different polar patterns, left to right: cardioid, hypercardioid, figure-of-eight, omnidirectional. (Courtesy Lewitt-Audio).

omnidirectional microphone would capture a relatively stronger hi-hat signal than a unidirectional microphone positioned towards the snare. Engineers will position close microphones in order to minimise bleed and ambience, attempting to direct the least sensitive angle of the microphone towards elements of the drum kit that are not intended to be recorded, such as the hi-hat into the snare microphone, or ride cymbal into the floor tom microphone (Gibson, 2004; Major, 2014).

There are three main drawbacks of excessive bleed. First, the timbral characteristics of the instrument captured from the side of the microphone may not be to the preference of the engineer due to the microphones off-axis frequency response (Beck, 2004). Second, the addition of this unwanted signal could negatively interact with the recording of the intended instrument. In the case of the hi-hat and snare drum, there may also be a close microphone placed on the hi-hat. The hi-hat bleed on the snare recording may cause destructive phase interference with the isolated recording of the hi-hat, due to time arrival differences between the two microphones at varied distances (Crich, 2010). Third, when attempting to mix the individual elements an engineer may wish to enhance the recording of the isolated instrument, such as amplifying the high frequencies of the snare. This same processing would also be applied to the bleed, amplifying the high frequencies of the unwanted hi-hat, causing an undesired result and further exacerbating the other problems associated with bleed (Gibson, 2004).

Non-linear characteristics define how the microphone distorts the incoming audio. At high SPL the microphone will add additional harmonics to the spectrum, but this is not uniformed across the full frequency range of the microphone, and is dependant on the particular microphone. These additional harmonics are caused by exceeding the upper limit that the microphone can handle without distorting the audio signal. In most studio microphones, the distortion present at very high levels is not a result of non-linearities from diaphragm motion but from electrical overload of the internal amplification stage directly following the diaphragm (Eargle, 2004). The particular way in which a microphone distorts the audio recordings could be desired or undesired by the engineer depending on the overall timbral quality desired for the song. Some genres may be better suited for drum recordings with more obvious distortion, whilst other may benefit from a *natural* sound, characterised by the similarity to how the drums are heard in the recording space. The intensity at which the drummer strikes the drum head may also be a consideration the engineer will have to take into account when selecting a microphone. A high sensitivity microphone will output a higher voltage signal than a low sensitivity microphone when both are capturing a sound of equal SPL (Bartlett and Bartlett, 2016). A song with a very softly played snare drum may be better suited to a more sensitive microphone that isn't capable of handling extreme SPL, whereas a song with much higher velocity snare strikes throughout, may encourage the engineer to select a less sensitive microphone (Gibson, 2004). However, the inverse may be true if the engineer specifically wants to use a microphone to shape the timbre of the snare recording by using its non-linear properties to generate harmonics. The degree to which a microphone will distort from excessive SPL is based on the its physical construction, design choices, and trade-offs made by the designers and manufacturers.

Another way an engineer can alter the timbral properties of the recordings is through microphone placement. By knowing a particular microphone's on- and off-axis frequency response, polar pattern, and the amount to which it is affected by proximity effect, the engineer can modify the timbre simply by adjusting the location, angle, and distance of the microphone in relation to the drum being recorded (Major, 2014). The exact position and angle of the snare microphone are important factors that the engineer must consider, as both have an impact of the timbre and content of the resultant recordings. If for example, a drummer is playing very lightly, the engineer may direct the microphone towards the contact point of the snare drum. However, if the drummer is playing with greater strike velocity and exciting the entire drum, the microphone may have to be moved further away to avoid overloading it (Senior, 2008). Although the position will be driven by the engineers personal preference there is a practical consideration involved too, as the microphone can not interfere or obstruct the drummers playing, and a compromise to timbral quality may need to be made (Gibson, 2004). Common positions of the snare drum microphone range from 1—4" away from the batter head, angled 30—45° toward the center of the drum (Owsinski and Moody, 2009; Huber and Runstein, 2010; Crich, 2010). Placing the microphone closer to the outer rim will result in capturing more resonance and overtones, which may be better suited for genres such as jazz and acoustic music. Positioning the microphone more towards the centre of the head will emphasise the initial transient and fundamental frequency of the drum resulting in a timbral quality better suited for rock music. For any microphone, a *sweet spot* can be located somewhere between the two positions where the drum will sound most balanced and natural (Gibson, 2004; Major, 2014; Pedersen and Grimshaw-Aagaard, 2018). A correctly placed microphone can help to minimise the need for overly corrective equalisation and additional processing which will speed up the mixing process.

3.2 Microphones Preference for Snare Drum Recording

As has been previously discussed, microphones selection plays a key role in obtaining the desired timbral qualities for a recording. This section presents a microphone comparison study which investigates the subjective and objective differences between microphones for snare drum recording. Two other microphone comparison studies are discussed in Section 3.2.1, the methodology is presented in Sections 3.2.2 to 3.2.8. The results and a discussion are provided in Sections 3.2.9 to 3.2.13, and conclusions are presented in Section 3.2.14.

3.2.1 Previous Microphone Comparison Studies

De Man and Reiss (2013) carried out a perceptual evaluation of microphones for a female singer using both a multi-stimuli and pairwise approach. The aim was to determine if participants showed a consistent preference of the 6 microphones that were selected. The microphones consisted of 2 condensers, 1 ribbon and 3 dynamic microphones commonly used for vocal recording. Due to the variations in performance of the vocalist, it was determined that all microphones be recorded simultaneously to avoid subjective preference being affected by performance variation. However, this meant that not every microphone could be placed in an optimal location (i.e., closer and directly in front of the vocalist), it was noted that some microphones may have benefited from position changes. If each microphone had been positioned optimally then different preferences may have been discovered. Subjects were asked to rate the perceived quality of a four-second sample from a rock song and jazz song both recorded with the 6 different microphones. The ribbon microphone was shown to have a significantly lower preference across both tests, participants reported this was due to its lack of high frequency content, analysis of the recordings supported this claim. When analysing the results of the two songs separately, some slight preference differences were found, indicating that certain microphones may be more preferred for different songs and genres even when performed by the same vocalist. The multi-stimuli approach yielded more significant preferences than the pairwise test and was also found to take substantially less time to complete. Although microphone preference was investigated using human voice and not a snare drum, the process of subjective evaluation of timbral differences associated with microphones could easily be applied to other instruments using similar methods.

Quiroga et al. (2015) investigated different recording techniques for tom drums, which featured different drum dimensions, batter heads, microphones, and microphone position. Although subjective evaluation was not undertaken, the authors conducted frequency, amplitude, and time analysis of the recordings. In total, the authors tested 81 different combinations of recording parameters including, 3 different sized toms (12x9", 13x10", 16x16"); 3 drum types of batter heads (*Clear*, *Coated*, *Hydraulic*); 3 microphones (2 dynamics, one small diaphragm condenser); and 3 different microphones positions (*Center*, *Middle* and *Edge*), all raised 3" from the head. An electromechanical trigger system that allowed the tom to be struck in the center of the drum head and with the same mechanical force was used to avoid variation associated with strike position and velocity. The samples were evaluated based on their *attack* frequency (i.e., the frequency band with the highest dBFS values from the FFT of the whole audio signal), *tone* frequency (i.e., the FFT band that decays less than other bands), and *decay* time (i.e., the time it took the signal to drop from its highest peak to -60 dBFS). The tom samples were also segmented into 6 time bands, for frequency analysis to be carried out on each of the segments. Some of the main findings of this study showed that for 100% of the samples the frequency band identified as the *attack* was also the band that had the highest decay level, i.e., the *tone* frequency band. Independent of microphone or microphone position, the *clear* drum heads had the longest *decay*, followed by *coated* and then *hydraulic* heads. For 98.17% of the samples, independent of the microphone, drum head, or tom size, the *edge* microphone position had the longest longest decay of the *tone* frequency band, as well the *edge* also having the longest overall *decay* time across each tom. This study helps to highlight how variables such as microphone position and drum head type can be carefully selected by a recording engineer to manipulate certain timbral attributes such as decay time. Although measurable changes can be found between the different parameters used, it is not understood how these differences relate to a perceptual change for a listener.

Both De Man and Reiss (2013) and Quiroga et al. (2015) can be used to inform a methodology for a microphone comparison study for snare drum recording, which would seek to identify if listeners are able to detect differences of recordings of snare drums using a range of different microphones and if those differences elicit any subjective preferences that is agreed upon by all listeners.

3.2.2 Experimental Design

In order to determine whether preference plays a role in the selection of snare drum microphones, two experiments were carried out: the snare drum played on its own (Single Hits) and the snare drum played as part of a beat involving a hi-hat and kick drum (Hits With Bleed). The latter scenario being more relevant to real world applications of snare drum recording. There may however be situations where the snare is recorded on its own, for example, when capturing an isolated snare strike to either replace or enhance an unideal snare drum recording as part of a full drum kit, or to use as a sample in sequenced computer based music. It is also common for engineers in live or studio settings to request the drummer play each element of the drum kit separately to assess the quality and timbre. Comparison of these two tests would show if microphone preference for snare drum changes with the presence of bleed from the other elements of the drum kit.

3.2.3 Microphones

Table 3.1 presents a list of the microphones used in the recording experiments. The total number of microphones used was 25, comprised of 15 dynamic (D) microphones and 10 condenser (C) microphones. Out of these microphones 14 had a cardioid polar pattern, eight had supercardioid and three had hypercardioid. Specification were taken from the manufacturers websites. Microphones were selected based on availability and appropriateness, only small diaphragm condensers were used as large diaphragm condensers are often difficult to position between the hi-hat and the rack tom without obstructing the drummer. Several of the microphones had built in filters, such as the Sennheiser MD421 which has a five position high pass filter however the recordings were carried out with all filters switched off and no additional processing.

Brand	Model	Type	Polar Pattern
AKG	C451B	C	Cardioid
Audix	ADX51	C	Cardioid
Audix	D2	D	Hypercardioid
Audix	D4	D	Hypercardioid
Audix	i5	D	Cardioid
Beyerdynamic	M201	D	Hypercardioid
DPA	4099	C	Supercardioid
Electro Voice	PL80	D	Supercardioid
Electro Voice	RE20	D	Cardioid
Neumann	KM184	C	Cardioid
RØDE	M2	C	Supercardioid
RØDE	M3	C	Cardioid
RØDE	NT5	C	Cardioid
RØDE	NT55	C	Cardioid
Sennheiser	e609	D	Supercardioid
Sennheiser	e614	C	Supercardioid
Sennheiser	MD421	D	Cardioid
Sennheiser	MD441	D	Supercardioid
Shure	Beta57a	D	Supercardioid
Shure	SM57	D	Cardioid
Shure	SM7B	D	Cardioid
T.Bone	CC100	C	Cardioid
T.Bone	CD55	D	Cardioid
T.Bone	MB75	D	Cardioid
Telefunken	M80	D	Supercardioid

Table 3.1: List of microphones used in both recording experiments.

3.2.4 The Snare Drum

The snare drum selected and other variables that affect the timbral qualities were carefully considered to produce a sound representative of typical snare drum characteristics. Common drum heads were used as well as tuning the heads appropriately for a wide range of genres. This provided a generalisable and realistic scenario of drum recording. A Mapex Black Panther Velvetone 14" x 5.5" snare drum was used. It had an

Figure 3.5: Calibration of the digital DrumDial used for tuning.

8.1mm shell consisting of a 3 mm exterior burl maple outer layer, enclosing a 3.4 mm walnut wood middle layer and a 1.7 mm maple interior layer. The drum had 10 tension rods for the batter head and 10 for the resonant head, as well using PureSound 20 strand coiled steel medium-gauge wires. For the batter head an Evans B14HBG Hydraulic was used, this is a coated 2-ply head, between the plies is a thin layer of oil and a 1-ply Remo Ambassador Black Suede Snare side was used for the resonant head.

A digital DrumDial was used to tune both the batter and resonant heads. This device ensured that the tension of the heads were uniform around the drum and allowed for accurate, repeatable tuning. The DrumDial device is calibrated by placing it onto a small piece of glass, pressing and holding the *cal* button on the device for 3 seconds produces a reading of 100, a drum head with a lower tuning will produce values closer to 0 and a drum head with a higher tuning will result in values closer to 100, the calibrated DrumDial can be seen in Figure 3.5. The tension was set to 90 for the batter head at every tuning rod position and set to 80 for the resonant head. These tension values were suggested by a tuning chart provided with the DrumDial based on the dimensions of the snare drum and the types of heads used. Once the drum was tuned a Big Fat Snare Drum (BFSD) dampening disk was placed on the batter head, this was to reduce excessive overtones of the drum. This device was chosen over other products such as MoonGel dampening pads as the placement of the BFSD takes up the entirety of the drum head, ensuring placement repeatability unlike smaller devices that could potentially be placed anywhere on the head. The drum sticks used were Vic Firth 5B Nova Hickory wood tip sticks.

3.2.5 Recordings

All recordings were undertaken in an acoustically treated studio control room, with an ambient noise level of ~20 dBA. This was opted for over a larger studio live room for its shorter reverberation time and flatter frequency

Figure 3.6: Position of M201 with triangle jig.

response, which minimized the impact of the room on the character of the recordings. The microphone was positioned at 60° with the diaphragm of each microphone placed 10 cm above the rim of the drum, pointing directly at the the centre of the drum head. This position was chosen for consistency as it was easy to replicate with every microphone and was found to produce a reliable recording without overloading any of the microphones when the snare drum was played with a medium to high velocity by the drummer. The utmost care was taken to ensure each microphones position was matched as accurately as possible. A triangular jig was used to aid the alignment of the microphones (Figure 3.6), This measured 10 cm x 17.78 cm x 20.4 cm, where applicable the distance of the diaphragms location in relation to the external grill was compensated for.

The drummer (with over 9 years of professional drumming experience) was instructed to maintain consistency of velocity and striking position throughout. A recording was made by each microphone consisting of four individual hits of the snare drum (Single Hits). In addition to this, without moving or re-positioning the microphone and without adjusting the gain of the microphone amplifier, four beats of the kick drum and four strikes of the hi-hat hits were also recorded, as well as four hits of kick drum and hi-hat played simultaneously (Bleed). This recording would be used to assess the amount of bleed captured by the microphone from the additional elements of the drum kit. Lastly a four-bar phrase was recorded (Hits With Bleed), played to a 110 BPM metronome, again maintaining the same position for individual snare hits. No additional microphones were used for the kick drum or hi-hat, and these were captured solely through the snare close microphone. The musical score for the drum beat is shown in Figure 3.7. The recordings were captured using a Metric Halo ULN-2 2D analogue to digital converter into Pro Tools 12 running at 32 bit-float and 44.1 kHz. The level of the microphone preamplifier was set so that no clipping occurred for any recording. The tuning of the drums was checked with the DrumDial after every recording. The recordings from the Shure SM57 can be seen in Figure 3.8, showing Single Hits and Hits With Bleed.

3.2.6 Audio Pre-Processing

Before the samples were used in the listening test they required some amount of pre-processing. The four-bar phrase was manually edited to ensure quantization of all drum hits to the beat. This aided in removing any of the player's timing variation from the recordings. The whole phrase was then normalised to -23 LUFS

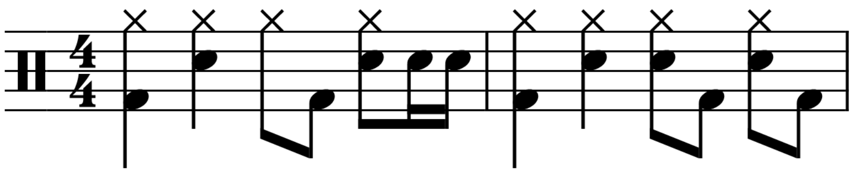

Figure 3.7: Score for drum beat used in Hits With Bleed recording experiment.

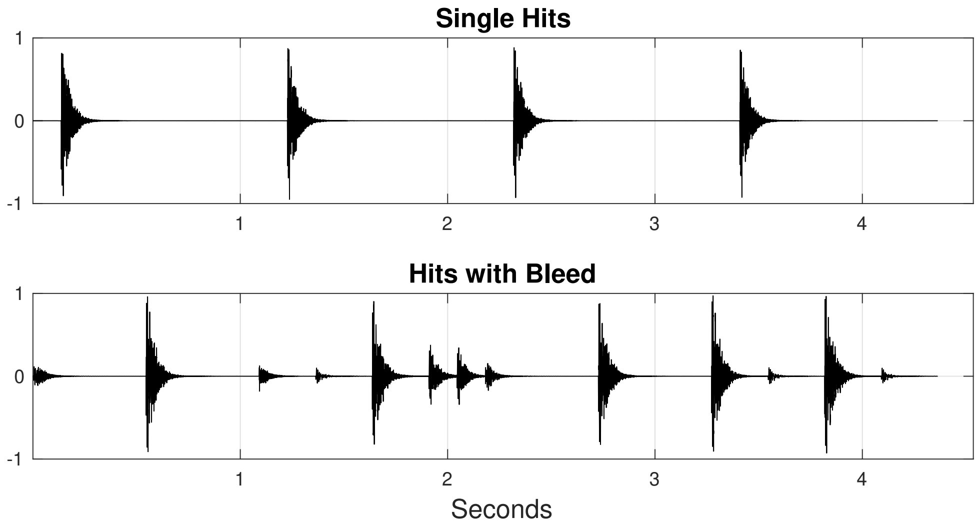

Figure 3.8: Shure SM57 recordings for Single Hits (Upper) and Hits With Bleed (Lower).

which removed any loudness variation between samples (EBU-R-128, 2014). The individual hits were also first quantized and each hit was then separately normalised to -23 LUFS, to ensure that the perceived loudness between hits and different microphones was as consistent as possible for the listening test.

3.2.7 Listening Test

The listening test was carried out in an acoustically treated studio control room, the speakers used were PMC IB1S, with a Bryston 2B-SST2 amplifier and an RME Fireface 802 audio interface. The test was conducted using the Web Audio Evaluation Tool (WAET) (Jillings et al., 2015) with the APE interface (Man and Reiss, 2014). A multi-stimuli approach was used over an AB comparison to minimise the duration of the tests as in a previous study (De Man and Reiss, 2013) results from multi-stimuli and AB test were found to produce comparable results. Two listening tests were carried out, one evaluating the Single Hits recordings and one evaluating the Hits With Bleed, these two tests were presented to participants in a random order. It required participants to position 25 markers corresponding to each audio sample along a one-dimensional axis, leftmost representing least-preferred and rightmost for most-preferred. Participants were instructed to "rate the audio samples based on the quality of just the snare drum. Using the full range of the scale". Selecting a marker would play the sample on a loop, and the loop position was maintained when switching between samples for uninterrupted playback. Participants could not complete the test until every sample had been played at least once and the marker had been moved from its original position. Starting position and marker number were randomised for every participant and for each of the two tests.

3.2.8 Participants

Twelve participants took part in the listening test, all of which had previous experience using both condenser and dynamic microphones for studio or live sound applications. The range of the subjects age was 22—48 (mean: 27). The participants were asked how many years experience they had in sound recording/audio production related fields (range: 3—27 years, mean: 10 years). The participants took on average 9 min 14 s to complete the Single Hits test and 8 min 54 s to for the Hits With Bleed test. Not every participant made full use of the rating scale, the range was calculated by subtracting the lowest score from the highest, only 6 participants used more than 90% of the rating scale with one participant using as little as 51.4% of the scale for the Single Hits listening test, the range of the scores for each participant for both tests can be seen in Table 3.2. It can be seen that in general a wider range of scores were used during the Hits With Bleed, this could potentially indicate that some participants had a greater preference difference for the lowest and highest rated stimuli when there was the presence of bleed in the recordings.

Single Hits												
% of scale used	52.5	94.8	96.0	99.5	94.0	51.4	57.9	95.1	65.9	96.7	85.6	84.9
Hits With Bleed												
% of scale used	76.2	98.3	97.2	99.3	97.9	69.9	87.7	96.3	74.5	94.6	84.2	79.3

Table 3.2: Percentage of rating scale used by each participant for Single Hits and Hits With Bleed test.

3.2.9 Results

A one-way analysis of variance (ANOVA) was carried out on the results of the listening test to determine if the differences between the mean scores of any of microphone were significantly different from one another. Low p-values ($p > 0.05$) for both Single Hits ($p = 5.581\text{e-}8$) and Hits With Bleed ($p = 0.0045$) showed that among all the microphones some of them did have significantly different means from others. This indicated that not only could participants perceive a difference between the recordings when microphone selection was the only variable, but also that there was some agreement of preference between the listeners. Had there been no agreement an ANOVA would have revealed there to be no significant differences between any of the microphones scores. In Figure 3.9 and Figure 3.10 results are presented in order of their mean score, where the horizontal line shows the standard deviation for each microphone and the cross shows the mean across participants. Despite its reputation as the industry standard snare drum microphone, the Shure SM57 was not scored notably high for either listening tests, with its ranked score being 14^{th} for Single Hits and 12^{th} for Hits With Bleed. Although not statistically significant it was rank higher than several more expensive microphones in both tests, including the MD421, MD441, and SM7B, all considerably more costly, this potentially indicates that its wide ranging use is a trade-off between performance and cost. A large disparity between the highest scored microphone and lowest scored microphones can be seen for both Single Hits and the Hits With Bleed test, however it is clear that the highest and lowest scored microphones do not remain consistent between the two tests. A paired t-test was used to compare the results from the Single Hits and the Hits With Bleed test, in order to determine if scores for any microphone were significantly reduced or improved between the two conditions. As can be seen in Table 3.3, three microphones had significant differences; the DPA 4099 and the Audix ADX51, which had the two highest mean scores for the Single Hits test both received significantly lower score for the Hits With Bleed test, with the 4099 receiving the 4th lowest mean score for Hits With Bleed and the ADX51 being scored 10th highest. However the Audix D4 significantly improved with the addition of bleed, having received the second lowest mean score for Single Hits and the 8th highest mean socre for Hits With Bleed. This indicates that the majority of the microphones received statistically similar scores with and without the presence of bleed.

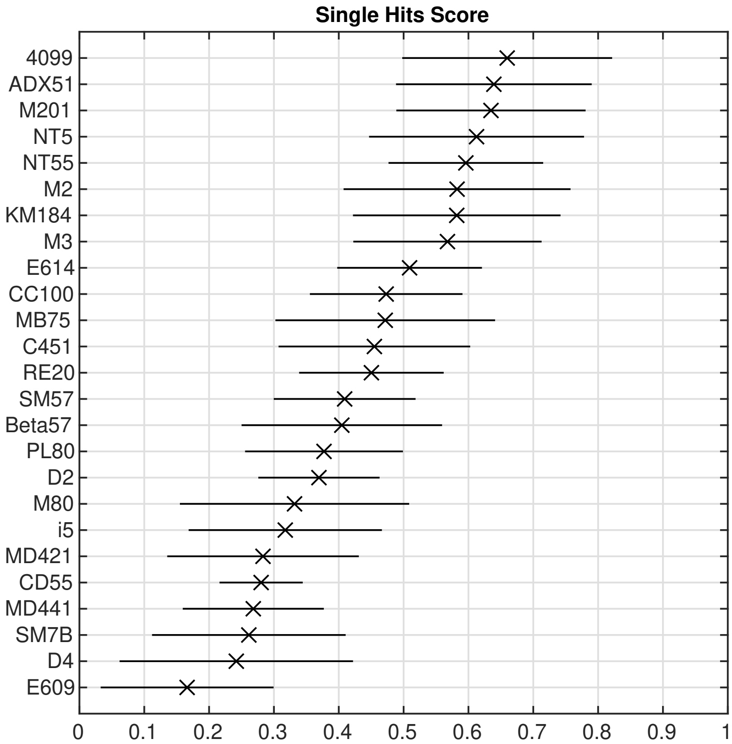

Figure 3.9: Mean score (x) and standard deviation (horizontal lines) for Single Hits listening test.

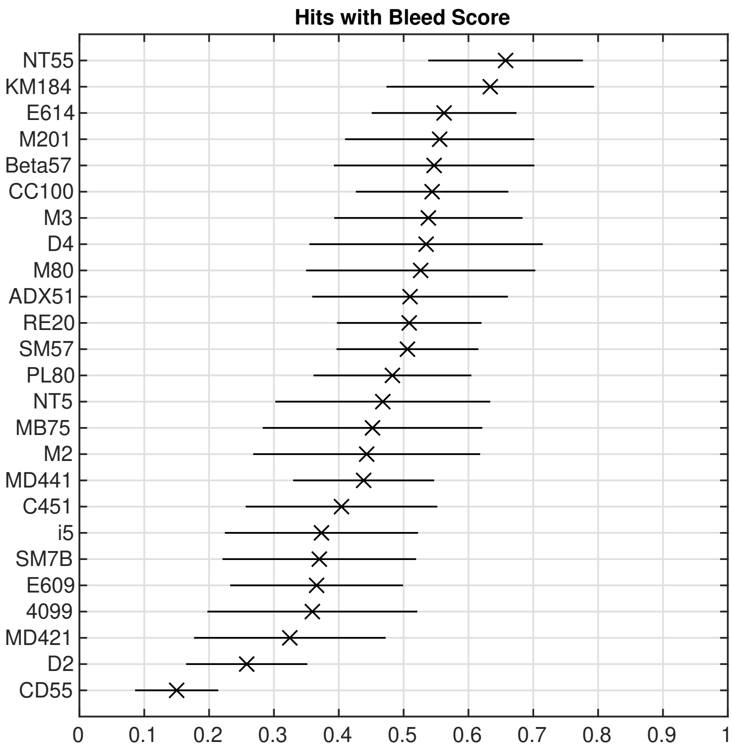

Figure 3.10: Mean score (x) and standard deviation (horizontal lines) for Hits With Bleed listening test.

Microphone	Single Hits mean score	Hits With Bleed mean score	Paired t-test p value
ADX51	0.63	0.51	0.04
Audix D4	0.24	0.53	0.02
DPA 4099	0.65	0.35	0.01

Table 3.3: Paired t-test results, showing microphones with significant p-values.

3.2.10 Ranking the Data

Although participants were asked to use the full range of the scale (i.e., placing their least-preferred at 0 and their most-preferred at 1), only one participant used close to the maximum rating scale for both tests (99.5% and 99.5%). Normalisation of data is a common procedure used to compare results between participants, in this case however, normalisation to the entire scale might misrepresent the intention of the participant (e.g., moving similarly scored microphones farther away from each other). Alternatively, we chose to assess the rank order as it is a robust against normalisation, and thereby a more comparable measure between participants than the raw data. Once the data was ranked, a clear preference for condenser microphones over dynamics was observed. The top eight out of ten ranked microphones in the Single Hits test were condensers, with the average rank being five places above the average rank of the dynamic microphones. For the Hits With Bleed test, condensers only made up five of the top ten ranked microphones and the average rank was two places above the average rank of dynamics.

3.2.11 Brightness

Once a participant had completed the test they were asked "What qualities of the samples were you comparing?" the answers included: *resonant frequencies on the snare, depth, clarity, brightness, fullness, punch, crispness, sharp transient, bright, how hard the top-end sounded, the attack and frequency content, punch, warmth, highs, the tone of the snare*, and *the snap of the impact*. From the responses, frequency content—in particular the high-frequency energy—seemed to be an attribute to which participants were basing subjective responses on. Two spectral features, spectral centroid and brightness are used to measure the high frequency characteristics mentioned by participants in the post-test survey. The spectral centroid refers to the center of gravity of the frequency spectrum. This can be a good indication of how bright a sound is perceived, as a higher spectral centroid contains more energy within higher frequencies than in lower frequencies (Grey and Gordon, 1978). The Juslin (2000) definition of brightness was also used, which measures the amount of spectral energy above 1.5 kHz, the result is expressed as a number between 0 and 1.

The spectral centroid for every microphone was higher (mean: 5kHz) for Hits With Bleed than for Single Hits (mean: 2.6kHz). The brightness was also found to to be higher for every microphone for Hits With Bleed in comparison to the Single Hits, with a mean increase of 0.13 across all microphones. This increase is likely caused by the addition of the hi-hat cymbal placed to the side of the snare. The mean increase in brightness for the 14 cardioid microphones was 0.14, while the mean for the 11 hypercardioid and supercardioid microphones together was a 0.12 increase. This demonstrates the effect of off-axis rejection of the directional pickup patterns (i.e., hypercardioid, supercardioid) on amplitude of hi-hat bleed from the side of the microphones.

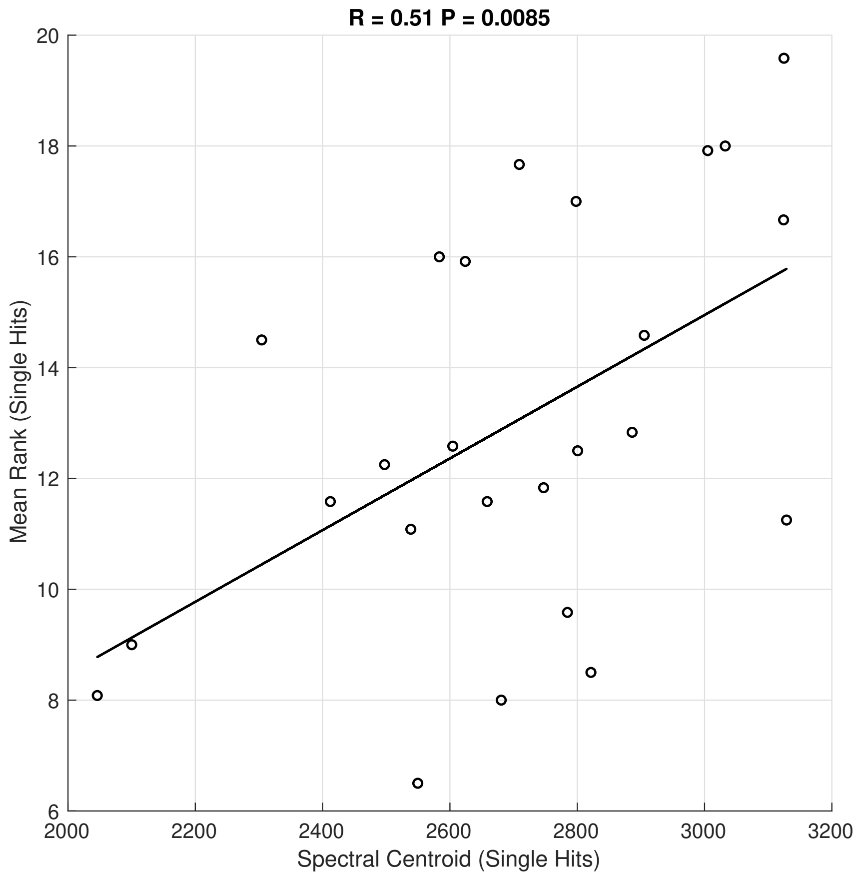

Figure 3.11: Spectral centroid and mean rank for Single Hits, with regression line.

3.2.12 Correlation

Spearman correlation, with a significant *p*-value ($p > 0.05$) was used to observe both the relationship between the spectral centroid of a sample and its mean rank, and between brightness and mean rank. A positive correlation was found between the centroid and the mean rank (R = 0.51, p = 0.008) for the Single Hits (Figure 3.11). There was no correlation for the Hits With Bleed microphones mean rank and its spectral centroid (R = 0.28, p = 0.17).

When taking all the samples together no correlation was found between brightness and mean rank for either Single Hits, or Hits With Bleed. However when the condenser microphones were analysed separately, positive correlation was found for condenser microphone brightness for Single Hits and the mean rank (R = 0.74, p = 0.015) (Figure 3.12). This indicates that condensers used for Single Hits the *brighter* microphones received higher rank scores.

When taking only the hypercardioid and supercardioid microphones into account, a negative correlation (R = -0.65) was observed between the change in mean rank and the change in brightness across the two tests (Figure 3.13). The change in brightness can be described as the influence of the hi-hat on the microphones brightness. A relatively small increase in brightness from the Single Hits to Hits With Bleed means the high frequency bleed from the hi-hat was not as prevalent than had there been a much larger relative increase in brightness. If an omnidirectional microphone had been used it would be expected that this would show the greatest increase in change of the brightness. The change in mean rank could be described as how much better or worse the rank was with the addition of the hi-hat and kick drum. Positive values show higher mean rank for the Hits With Bleed, whereas values below zero show where microphones mean rank reduced, zero

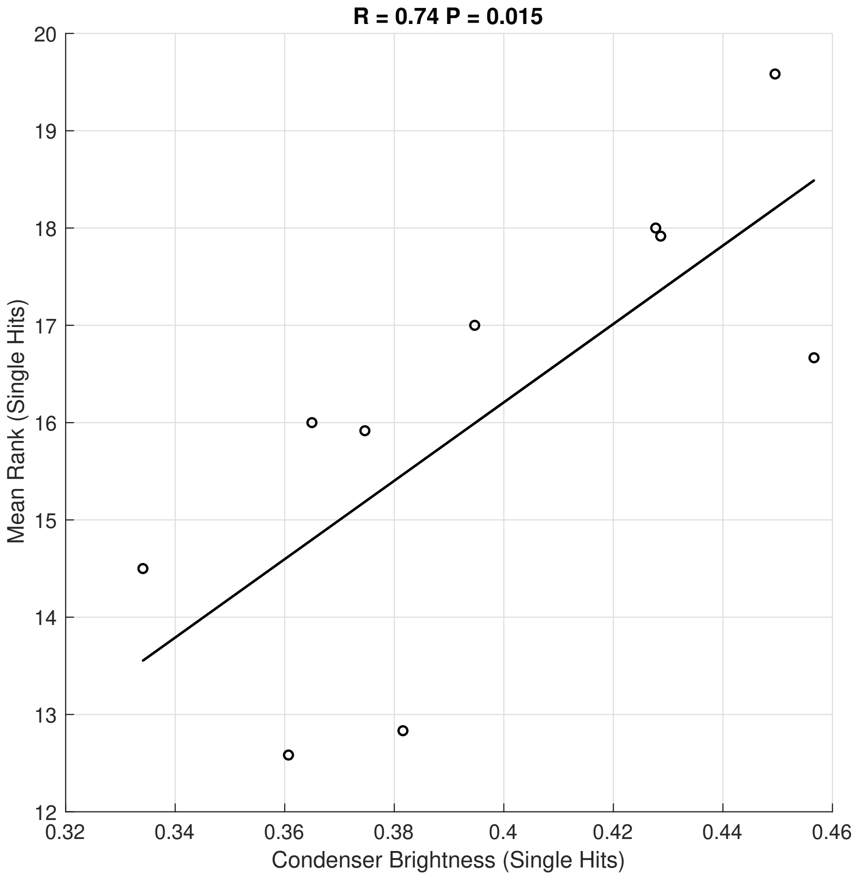

Figure 3.12: Condenser brightness and condenser mean rank for Single Hits, with regression line.

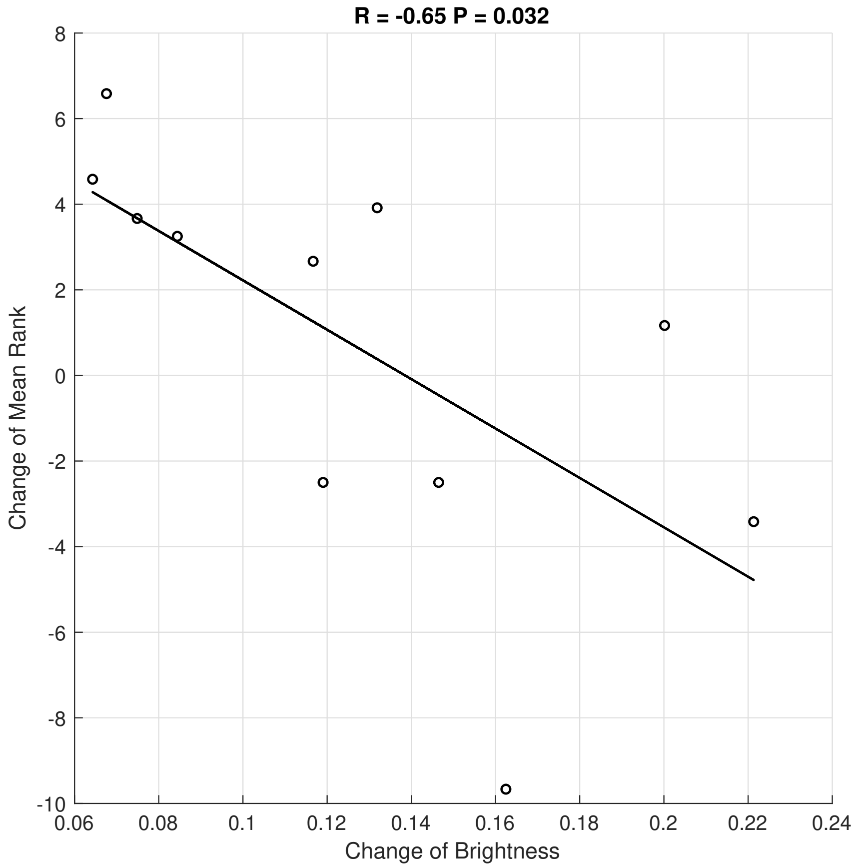

Figure 3.13: Change in brightness and change of mean rank, with regression line.

would show where the microphones rank was the same for both tests. The negative correlation (R = -0.65, p = 0.03) in Figure 3.13 shows that as the change in brightness increases, the change in mean rank decreases. This would indicate that hypercardioid and supercardioid microphones that have a higher off-axis rejection of high frequencies are more likely to be ranked higher when used for $\mathrm{Hits\ With\ Bleed}$ over $\mathrm{Single\ Hits}$.

3.2.13 Signal-to-Bleed Ratio (SBR)

The most common use of a snare microphone is when the rest of the drum kit is also being played. For this reason, it is important to examine the microphones behaviour when used in a real world application. The amount by which the microphone captures these other drums as well as the snare is likely to affect listener preference to some degree. As previously mentioned, as well as isolated snare hits, the kick drum and hi-hat were also recorded through the snare microphone without the position being changed. These recordings were used to quantify the amount of bleed picked up by every microphone. A ratio was taken between the RMS of the snare hits and the RMS of the bleed. To calculate the signal-to-bleed ratio (SBR), we calculate the sum RMS of each individual solo snare drum strike (Sn) of a given microphone, and for the corresponding Bleed recording (Bn) of the same microphone.

$$SBR = \frac{\sum_{n=0}^{N-1} RMS(Sn)}{\sum_{n=0}^{N-1} RMS(Bn)} \tag{3.1}$$

A high ratio indicated that the signal captured in front of the microphone is stronger than the signal off axis, while a low ratio shows the bleed is close to or as strong as the snare signal. The mean SBR for all hypercardioid

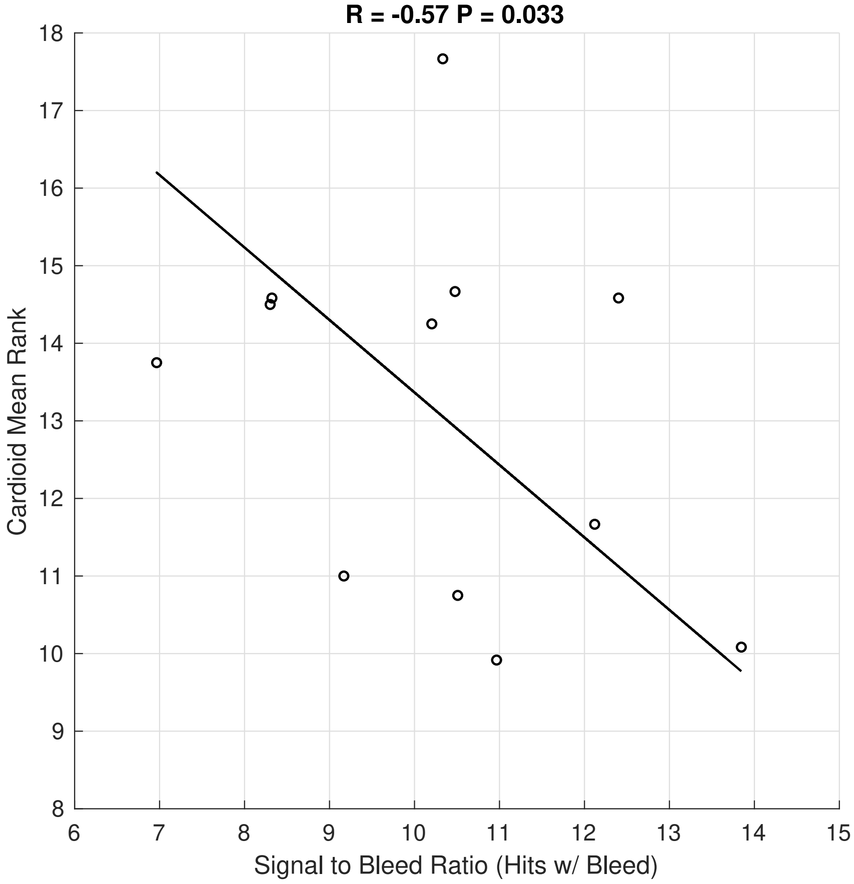

Figure 3.14: Cardioid signal to bleed ratio and mean rank, with regression line.

and supercardioid microphones was 12.04, whereas the mean for all cardioid microphones was 10.43, this is to be expected as the broader polar pattern of a cardioid microphone is more susceptible to record a higher relative signal off-axis than either a hypercardioid and supercardioid microphone. Spearman correlation was used to examine the relationship between the SBR and the mean rank of the microphones. Interestingly, a negative correlation was observed between the mean rank and the SBR of the cardioid microphones (Figure 3.14). This shows that as the SBR decreases (i.e., the bleed becomes proportionally louder compared to the snare drum), the mean rank increases, or the better ranked cardioid microphones are those with worse off-axis rejection.

A potential explanation for this could be the participants preferred a more complete sounding drum beat that included relatively more prominent hi-hat and kick drum, resembling produced music which would also include close microphones on both of these elements. Therefore the stronger the presence of bleed, the more preferred the cardioid microphones were. This effect could be investigated further through the additional of close microphones on the hi-hat and kick drum to explore if the presence of bleed on the snare microphone still played a factor in preference, however the interaction of the different microphones used would also need to be taken into account.

microphone	SBR	Single hits Brightness	Single hits Centroid (Hz)	Hits with Bleed Brightness	Hits with Bleed Centroid (Hz)
AKG C451B	9.168	0.381	2886	0.603	6760
Audix ADX51	8.301	0.428	3005	0.630	6616
Audix D2	10.963	0.426	2658	0.543	4444
Audix D4	11.448	0.460	2680	0.527	4016
Audix i5	10.508	0.448	2747	0.543	4296
Beyerdynamic M201	9.807	0.410	2708	0.556	4881
DPA 4099	8.184	0.449	3124	0.612	6222
Electro Voice PL80	14.045	0.381	2412	0.498	3965
Electro Voice RE20	10.203	0.368	2497	0.474	4254
Neumann KM184	10.332	0.365	2583	0.539	5620
RØDE M2	0.446	0.374	2624	0.596	6594
RØDE M3	8.322	0.456	3124	0.611	6498
RØDE NT5	6.965	0.427	3032	0.610	6379
RØDE NT55	8.738	0.394	2798	0.575	6113
Sennheiser e609	14.250	0.421	2549	0.496	3648
Sennheiser e614	11.721	0.334	2304	0.534	5311
Sennheiser MD421	10.965	0.440	2784	0.547	4670
Sennheiser MD441	15.161	0.305	2046	0.437	3975
Shure Beta57a	16.239	0.405	2538	0.469	3576
Shure SM57	12.401	0.451	2800	0.511	3893
Shure SM7B	13.846	0.318	2100	0.464	4145
T.Bone CC100	10.477	0.360	2604	0.606	7095
T.Bone CD55	13.663	0.481	2821	0.538	3704
T.Bone MB75	12.120	0.457	2905	0.528	4064
Telefunken M80	11.227	0.495	3128	0.569	4468

Table 3.4: All measurements of audio recordings.

3.2.14 Conclusions

In order to determine if microphone selection plays a role in the preference of snare drum recording, two experiments with 25 different microphones were performed. These experiments were designed to mimic real world recording scenarios. The results of the listening test demonstrate a clear disparity in score between the highest and lowest rated microphones. However, due to the broad standard deviation of some of the scores, providing conclusions regarding the preference of microphones with close mean scores is not possible. A paired t-test revealed a significant change in score for three microphones between the two recording experiments, with the Audix D4 microphone receiving a better score for Hits With Bleed, and the Audix ADX51 and the DPA4099 getting significantly lower scores. The other 22 microphones did not show a significant change in their scores between the two tests, with most microphones maintaining their ranked position or only moving up or down a few ranks, this would suggest that although the preference of some microphones may be heavily dependant on whether they are to be used for isolate snare recording or a complete drum kit, most of the microphones tested will perform equally in both scenarios.

Of the subsets assessed (i.e., polar pattern, type), the condenser microphones demonstrated the strongest correlation with the mean rank, with only one dynamic microphone (M201) making the top 10 ranked microphones for Single Hits and only 3 (M80, D4, M201) being in the top 10 ranked microphones for Hits With Bleed. For Single Hits, spectral centroid for all microphones correlated positively with rank. A consideration outside the scope of this investigation was microphone positioning, which may have also dictated microphone performance, as every microphone might theoretically have an optimised position that would make it out perform other microphones simply by moving it closer or further away or changing its angle. Additionally, during the recording process velocity fluctuations were intentionally minimised as much as possible so that subjective evaluations could be based on microphone variables alone.

This chapter has detailed the process of recording an acoustic drum kit and the implication of various recording techniques such as microphone choice and placement. Different microphone topologies were discussed as well as features of microphones such as proximity effect and polar patterns as they relate to the capture of the snare drum. From there an investigation into microphone selection for snare drum recording was presented, showing both subjective and objective differences. It was clearly seen that different microphones produced subjectively better quality recordings as was shown by the results of a listening test. In the next chapter the timbral differences associated with snare drum strike velocity is explored as well as determining if microphone choice has an effect on the perception of velocity.

Chapter 4

Snare Drum Strike Velocity Differences

4.1 Introduction

In the previous chapter the recording process was outlined and the effect that different microphones had on the subjective quality of the recordings on snare drums was explored. This chapter investigates the role of timbral difference in distinguishing between high and low velocity snare drum strikes, in the extreme case where stimuli have been loudness normalised, to completely eliminate any perceptual loudness cues. Timbral differences are then assessed through signal analysis to characterise the high and low velocity strikes. These effects are investigated across several common studio microphones in order to observe the interaction between different microphones and velocities.

Variation in the striking velocity of a percussion instrument results in modification of the sound output, with the main effects being related to the volume envelope and timbre. Velocity can be described as the amount of physical force the musician applies in order to the excite the instrument. An obvious example of this phenomenon is the difference between a high velocity full strike, which requires higher levels of physical exertion, and a substantially less energetic stick bounce, requiring little effort on the part of the drummer. A strike applied to the drumhead causes it to vibrate, which in turn causes various parts of the drum to vibrate, causing additional sound waves to be produced. When the force of the strike is increased, it is expected that not only volume increases but also more energy is transferred into the shell, resonant head, and snare wires, producing a perceptually distinct drum strike. In addition, the duration for the instrument to cease vibration and return to homeostasis is also expected to increase. While loudness models have been used to define the relationship between sound pressure level and perception of complex sounds (Pestana et al., 2013), it has yet to be fully determined if timbral difference alone is a sufficient factor in the perceptual identification of snare strike velocity variation, with the absence of loudness cues.

Timbral differences between snare drum recordings are derived from the response of the instrument to different striking velocities, dependant on its physical construction (Tindale, 2004; Wagner et al., 2005; Richardson, 2010), in conjunction with the non-linear properties of the microphone used for recording and any additional electrical circuits in the signal path. A snare drum is capable of producing sound pressure level (SPL) above 140dB (DPA Microphones, 2015), which exceeds the maximum SPL-handling capability of many studio microphones that may only have a total dynamic range of 125—130dB, this overload of the microphone introduces harmonic distortion (Eargle, 2004). The amount of distortion produced will be dependant on the specification of the microphone and the amount by which its maximum SPL tolerance is exceeded.

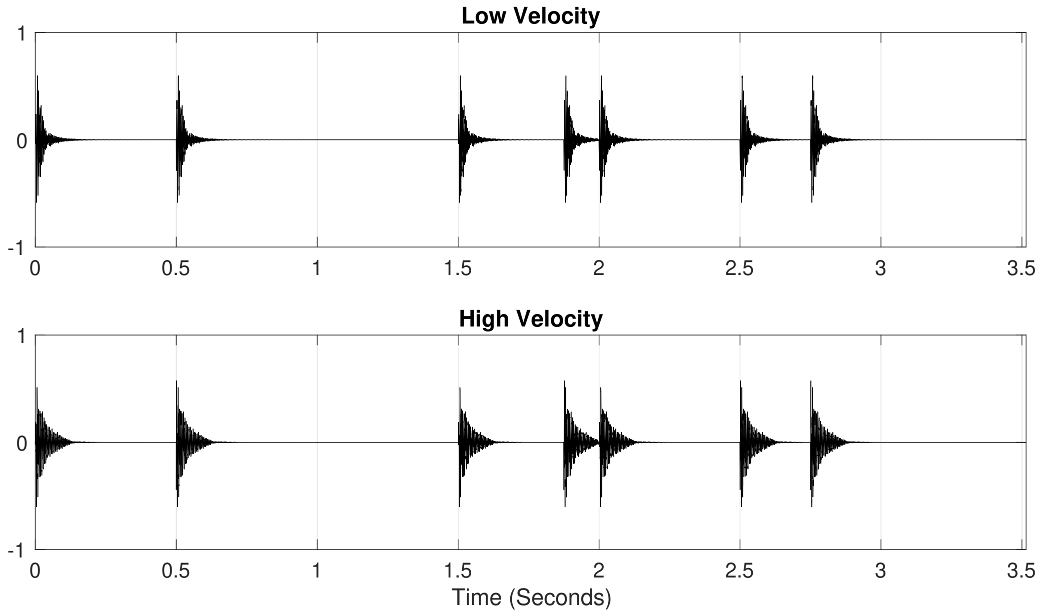

Figure 4.1: Stimuli of Beta57 used for listening test, low velocity (Upper) and high velocity (Lower).

While loudness and timbre are intrinsically linked, it is important to know how the timbral character of a snare drum changes with varying velocity and the effect this has on listener perception. This information helps to detangle the mixing preferences of engineers that utilise dynamic range compression to minimise the relative volume differences between high and low velocity strikes. Often this can be a technical consideration, ensuring all hits sound as though they are played with the same intensity, or used creatively to modify the volume envelope and enhance the perceived excitement of the performance (Owsinski, 2017). In the context of audio production, there are potential applications where linking timbral differences to audio features may prove beneficial, such a process could afford new tools for subtle timbre modification of recorded drums, correct a highly dynamic performance, or provide a humanisation effect to sample-based production without a adjusting volume. In the information retrieval domain, this would allow for sorting and searching of sample libraries by perceived velocity. Other uses might include a novel mixing task in which high and low velocity strikes are processed independently.

This chapter first presents a listening test methodology in Section 4.2 and the results in Section 4.2.3. Section 4.3 presents signal analysis methods for characterising the timbral differences between strikes of varied velocity. Section 4.4 provides a discussion on the implications of the results from the previous two sections and Section 4.5 presents conclusions, highlighting some of the key findings of the investigation.

4.2 Listening Test

A listening test was conducted to evaluate whether participants with sound engineering training could identity loudness normalised snare strikes of varied velocities. In removing volume cues, participants would be required to evaluate differences between strikes based on inherent attributes in the recordings other than loudness. To determine if microphone selection played a significant role in listener perception of velocity, multiple commonly used microphones for snare drum recording were used in the evaluation.

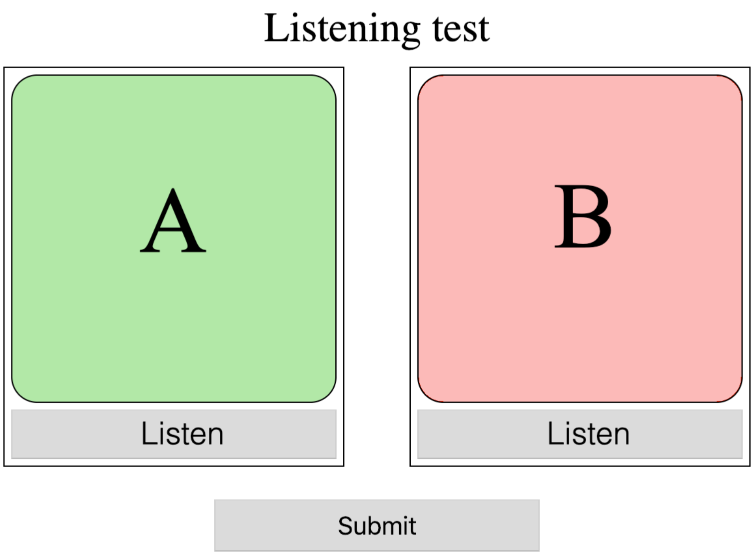

Figure 4.2: Interface used for the AB listening test.

4.2.1 Methodology

An AB listening test was used to evaluate if participants could distinguish between high and low velocity strikes from each microphone. All recordings were loudness adjusted to -23 LUFS (EBU-R-128, 2014), which removed any loudness variation between samples to ensure that perceived loudness between strikes and different microphones was as consistent as possible. The accuracy of this loudness adjustment stage was evaluated by 3 listeners with more than 5 years audio production experience prior to the listening test being carried out to ensure that there were no noticeable volume differences between any of the samples. By removing the cue of loudness, participants would have to evaluate differences based on any temporal and spectral variations between velocities. To create more engaging stimuli for the listening test, high and low velocity strikes from each microphone were sequenced into a two-bar drum phrase, the stimuli used can be seen in Figure 4.1. For a given microphone, participants were presented with the high and low velocity phrases 10 times each in a random order. For 5 pairs, participants were asked to select the phrase that had lower velocity and for the other 5, participants were asked to select the phrase with higher velocity, resulting in a total of 40 comparisons. Participants could not proceed to the next evaluation until they had played both phrases and made a selection.

The test interface used can be seen in Figure 4.2, the order of the high and low velocity stimuli were randomised on every test page. When first loading a test page both boxes are initially green. Selecting one of the boxes changes its colour to pink, selecting *Listen* underneath either of the boxes plays the corresponding sample on a loop, for uninterrupted playback the loop position is maintained when selecting between samples. 15 participants aged 21—50 years (mean: 26.8 years) took part in the listening test, and their experience in audio related fields was 3—30 years (mean: 8.5 years). AKG K240 studio headphones were used for playback, and participants were encouraged to adjust the volume to a comfortable level. Frequency response of the headphones was measured with an Earthworks M30 omnidirection measurement microphone while the headphones were placed on a Sennheiser MKE2002 dummy head. Figure 4.3 shows the average frequency response of the left and right channels, it should be noted that both channels were nearly identical.

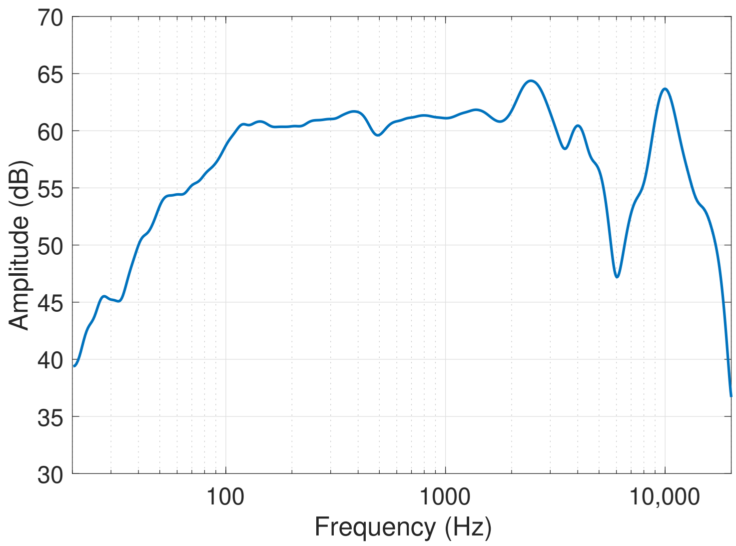

Figure 4.3: Average frequency response of left and right channels of the AKG K240 headphones used in the listening test, shown with $\frac{1}{6}$—octave smoothing.

Brand	Model	Type	Polar Pattern	Frequency Range	Sensitivity	Impedance	Maximum SPL
Neumann	KM184	Condenser	Cardioid	20Hz to 20kHz	15mV	50 Ω	138 dB
RØDE	NT55	Condenser	Cardioid	20Hz to 20kHz	12.6mV	100 Ω	136 dB
Shure	Beta57a	Dynamic	Supercardioid	50Hz to 16kHz	2.8mV	290 Ω	not specified
Shure	SM57	Dynamic	Cardioid	40Hz to 15kHz	1.6mV	310 Ω	not specified

Table 4.1: Specifications of microphones used for recording, taken from manufacture websites.

4.2.2 Recordings

In order to evaluate different velocity snare, a dataset of recordings that reflect professional standards (e.g., microphone position, microphone selection) was required. In addition, objective measurements of velocity were essential for the categorisation of high and low strikes—achieved through the use of an SPL meter. Four common studio microphones (i.e., 2 dynamic microphones and 2 condenser microphones shown in Table 4.1) were selected, common studio microphones were used to investigate if microphone selection altered perception of strike velocity. All stimuli were captured as 16-bit 44.1kHz sample rate recording using an RME Fireface 802 audio interface in an acoustically treated recording studio.

As microphone position impacts timbral characteristics (Bartlett, 1981; Senior, 2008; Quiroga et al., 2015), a consistent and generalisable placement was utilised. A close microphone technique minimised the effects of room acoustics. Typical close microphone placement range from 1—4" from the drumhead (Henshall, 2014; Gonzalez, 2022; Worrell, 2015; Fuston, 2017b). However, as both proximity effect and the potential to overload the microphones internal circuity were potential issues (Owsinski, 2005; Ballou, 2008; Major, 2014), the distance of the microphone was on the further side of the typical close mic recommendations (Huber and Runstein, 2010; Kokkinis et al., 2012). As can be seen in Figure 4.4, the microphone was position 4" above the drumhead and 2" over the rim, pointing directly at the centre of the drum, This technique provided professional grade recordings of quasi-isolated snare strikes (Owsinski and Moody, 2009). Simultaneous recording and continuous SPL measurements of the snare drum were captured, this produce varied velocity audio recordings with a corresponding measurable SPL reading. Lower velocities corresponded to lower SPL measurements,

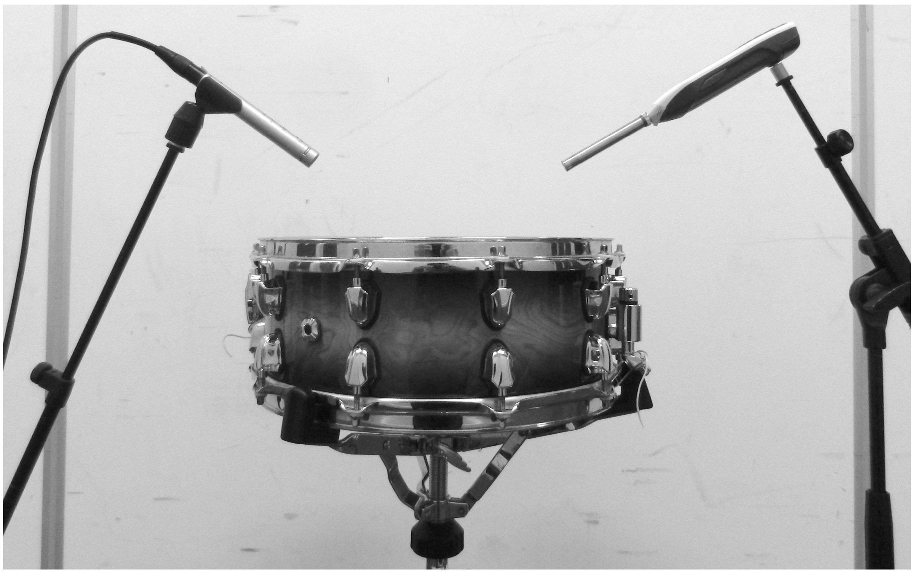

Figure 4.4: Snare drum with RØDE NT55 microphone (left) and Cirrus CK:162C SPL meter (right).

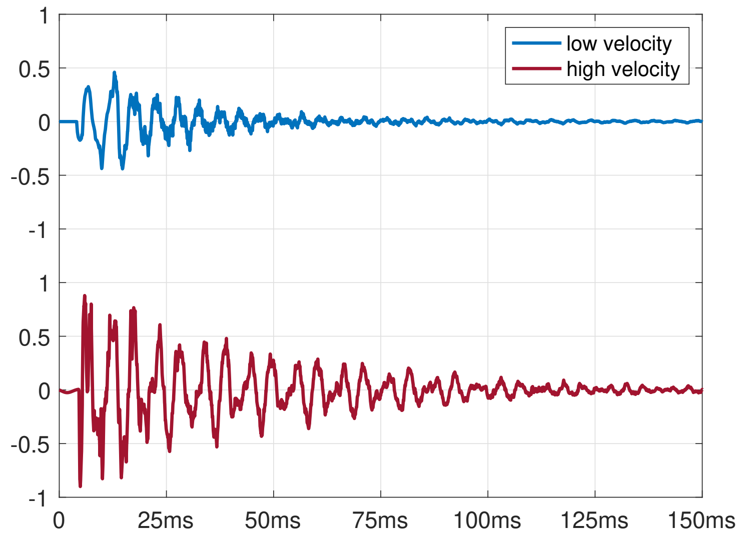

Figure 4.5: Recording of low velocity strike (100dBZ) and high velocity strike (125dBz).

and higher velocity with high SPL. A Cirrus CK:162C Optimus Red sound level meter[1] was placed at the same angle as the microphones used for recording, this can also be seen in Figure 4.4.

SPL was measured in LZFMAX, which measures the Z-weighted, fast-response maximum sound level. Z-weighted refers to no weighting across frequency response between 10Hz and 20kHz $\pm$1.5dB. The fast response has a 125ms rise and decay time. In order to minimise the effects of room acoustics, the recordings were carried out in an acoustically treated isolation booth (dimensions: 2.5m x 3m x 4.5m), with a RT60 of 112ms (mean of $\frac{1}{3}$-octave measurements). The snare drum was struck in the centre of the head, such that it produced SPL of 100dBZ and 125dBZ. These values were selected as they were consistently playable by the drummer, when asked to play high and low velocity strikes. Multiple recordings were captured until a strike was within $\pm$0.5dBZ of the target SPL, as depicted in Figure 4.5.

[1]www.cirrusresearch.co.uk

Model	Correct (%)	std	p
Beta57	88.67	1.73	8.6e-24
KM184	89.26	1.36	1.9e-24
NT55	93.29	1.16	1.6e-30
SM57	91.95	0.99	2.4e-28

Table 4.2: Correct responses (%) across all participants with standard deviation (std) and p-values.

To produce a generalisable and realistic scenario of drum recording, the same snare drum specification and tuning method were used in Chapter 3. For snare recording it is typical to dampen the batter head (Seymour, 2010; D'Virgilio, 2014). For this study a single piece of MoonGel was placed directly on the batter head 1" from the top edge of the rim. This option was chosen to reduce ringing by only a minimal amount. As temporal properties such as the decay time were to be investigated it was not favourable to heavily dampen the drum and thus modify the volume envelope by a noticeable amount.

4.2.3 Results

A binominal test was used to determine if the percentage of correct responses was significant with the null hypothesis being that participants could not distinguish between varied velocity strikes when perceptual loudness differences were removed. The null hypothesis would be accepted if the percentage of correct answers was below the significance level of 58.67% based on 150 trials (i.e., 10 trials for each participant per microphone; 15 total participants). Table 4.2 shows the percentages of correct responses for each microphone (correct), with the standard deviation (std) and the p-value from the binomial test (p). Small p-values ($<$0.05) for each microphone suggest that the percentage of correct responses was significant, thus the null hypothesis is rejected. This shows that the participants were able to distinguish between, and successfully identify varied velocity snare drum recordings. In order to determine if microphone selection played a significant role in listener perception of velocity a one-way ANOVA test was used to evaluate if the results from each microphone were statistically different from each other. The p-value from the ANOVA was $p = 0.79967$, this indicates that there is no significant difference between the results produced by the different microphones, p-values $<$0.05 would indicate that the difference between microphones was significant. The influence that microphone selection had on the perception and identification of velocity was not statistically significant, with any small differences between the results likely being due to chance or the natural variation inherent in the test design. All participants were able to successfully complete the listening test regardless of the microphone used, and no microphone made the task easier or harder for the participant.

When striking the snare drum at different velocities there is a loudness disparity, which was demonstrated by the SPL measurements taken at the time of recording. When perceptual loudness was normalised between recordings of varied velocity strikes, experienced participants with sound engineering training were able to successfully identify which velocities corresponded to the different recordings. This highlights that there are cues other than loudness variation that participants use to distinguish between striking velocities.

4.3 Feature Extraction and Analysis

In order to characterise the perceptual differences experienced between the varied velocity stimuli in Section 4.2, feature extraction methods are applied to the recordings. Statistical analysis is then performed on the extracted features to identify if features from the high and low velocity strikes are statistically different.

	Beta57a		KM184		NT55		SM57		**All Mics**	
Features	Low	High	Low	High	Low	High	Low	High	Low	High
Attack (ms)	30.99	33.26	30.28	35.80	30.41	32.75	30.72	33.08	30.60	33.73
	0.29	0.28	0.42	2.27	0.36	0.32	0.23	0.21	0.43	1.67
Decay (ms)	99.49	129.13	99.61	126.06	100.55	128.53	98.47	121.67	99.54	126.35
	2.85	4.08	2.59	3.12	2.31	3.98	2.41	2.99	2.62	4.59
$f0$ (Hz)	226.38	206.79	226.47	206.49	226.42	206.92	226.49	206.72	226.45	206.72
	0.72	0.59	0.69	0.41	0.70	0.58	0.70	0.56	0.69	0.56
Centroid (Hz)	2694.70	2410.70	2691.00	2471.10	2818.00	2498.00	269.4.80	2410.70	2641.00	2360.70
	65.35	53.30	53.01	92.12	59.74	61.28	65.35	53.27	180.59	187.86
Spread (Hz)	3215.60	2881.50	3891.50	3654.30	3982.40	3652.70	3339.80	3093.80	3607.30	3320.60
	33.63	41.43	34.85	71.99	35.83	44.57	32.21	46.18	337.67	347.09
Rolloff (Hz)	5548.60	4674.00	6504.70	5705.00	6933.00	5959.90	6358.50	5636.10	6336.20	5491.20
	118.85	155.28	130.88	277.92	142.96	193.80	110.02	151.93	519.26	528.84
Entropy	0.82	0.81	0.83	0.83	0.84	0.83	0.84	0.83	0.81	0.83
	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01
Flatness	0.14	0.11	0.23	0.20	0.23	0.20	0.15	0.14	0.18	0.16
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.05	0.04
Irregularity	1.09	0.58	1.04	1.13	1.08	1.02	1.17	1.09	1.09	0.95
	0.14	0.25	0.17	0.07	0.13	0.07	0.11	0.05	0.15	0.26
Kurtosis	6.89	8.87	6.91	8.16	6.15	7.53	5.31	7.29	6.33	7.96
	0.18	0.40	0.19	0.48	0.16	0.32	0.15	0.30	0.66	0.72
Roughness	50.21	396.93	31.86	765.98	45.33	398.92	34.19	339.94	42.39	475.44
	30.04	174.31	23.71	99.59	27.13	161.30	28.58	112.22	27.65	219.26
Skewness	1.94	2.28	2.00	2.24	1.86	2.13	1.58	1.94	1.85	2.15
	0.04	0.06	0.04	0.09	0.04	0.06	0.04	0.06	0.17	0.15
Brightness	0.39	0.35	0.38	0.36	0.32	0.36	0.44	0.41	0.40	0.37
	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.03	0.03

Table 4.3: Mean (upper value) and standard deviation (lower value) features extracted from high and low velocity recordings for each microphone. All Mics presents analysis of all 88 recordings.

4.3.1 Methodology

To conduct timbral analysis of varied velocity strikes, a second set of recordings were created, which comprised 22 high velocity strikes (125dBZ $\pm$2dBZ) and 22 low velocity strikes (100dBZ $\pm$2dBZ), for each of the four microphones, resulting in 88 recordings of each of the velocity intensities. Recordings were captured in the same manner as in Section 4.2.2. Prior to any feature extraction, all samples were peak normalised, truncated to 1 second and synchronised using cross correlation.

Spectral and temporal analysis was undertaken to examine the different properties of the recordings that made identification of velocity possible by participants when the cue of loudness was removed. A variety of features from the MIRtoolbox (Lartillot and Toiviainen, 2007) were selected to reflect features relevant to the spectral and temporal domain of a snare drum (Table 4.3), including *attack time* and *decay time* to define temporal envelope characteristics; *fundamental frequency* ($f0$); *entropy*, *flatness*, and *kurtosis* to describe the peakiness of a spectrum; and *spectral rolloff* and *brightness* to estimate high frequency energy. As all the recordings were monophonic, any spatial features were excluded.

The frequency spectrum of the recordings was divided into 24 Bark scale critical bands as in Figure 4.6. These perceptual subdivisions of the spectrum are based on the natural division of the audible range by the human ear, and are known to correlate closely to cochlear mechanics (Zwicker, 1961). Comparing perceptually relevant frequency bands allows observations of significantly different bands, thus aiding in explaining which characteristics contribute to perception of timbre-related velocity variation.

In order to visually compare temporal differences between velocities the envelope of each recording was extracted using the Bark-band decomposition of the signal, and the resultant envelopes were then normalised.

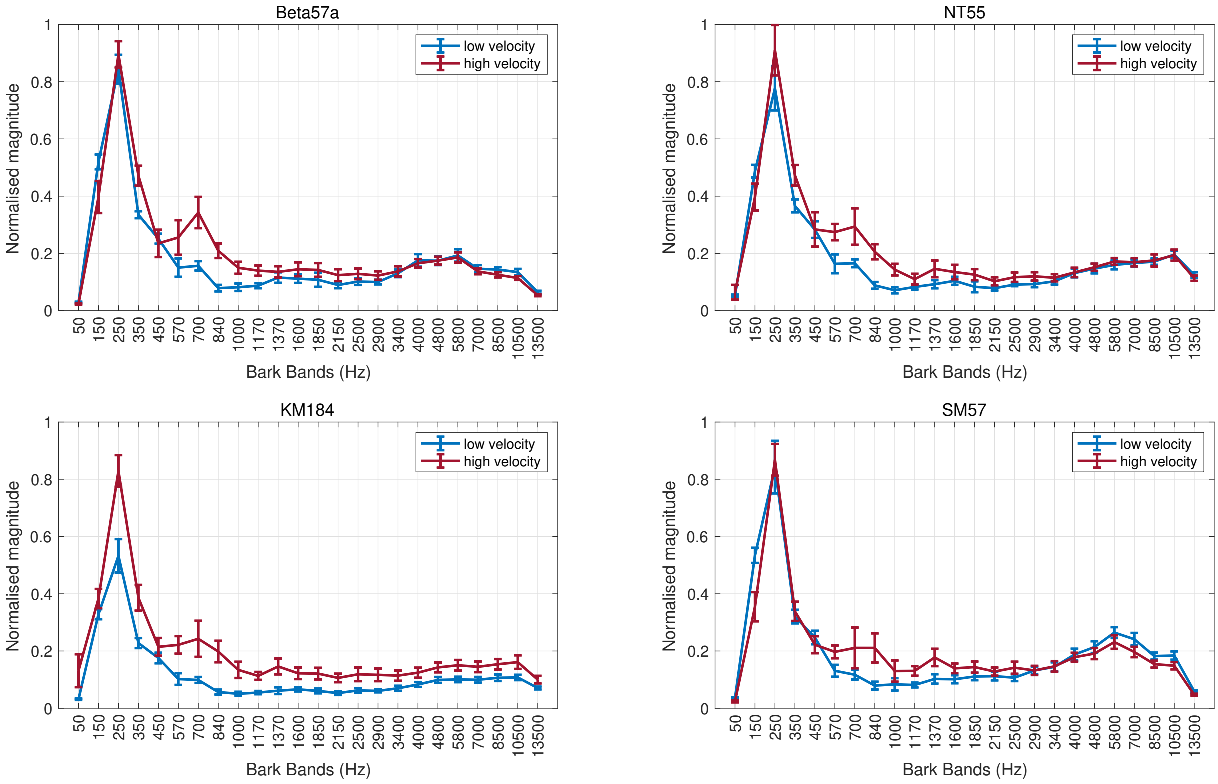

Figure 4.6: Means and standard deviations of Bark band magnitudes for high and low velocity strikes.

The normalised envelopes for all 88 low velocity and 88 high velocity strikes can be seen in Figure 4.7. The mean of all 88 recordings for each velocity intensity is shown by the solid lines and the vertical bars depict the standard deviation between all hits. When envelopes from each microphone were analysed separately there were no observable differences between any of them. It can be seen that high velocity strikes have more sustain than lower velocity strikes, whereas the low velocity strikes have a slightly faster attack time (as in Table 4.3).

Separate means were computed for high velocity recordings and low velocity recordings and the spectrograms calculated (window size: 300, with a 50% overlapping Hamming window between segments), this allowed for visualisation of the spectral and temporal differences between the two velocity intensities. In Figure 4.8 the low velocity spectrogram is shown in the upper plot and high velocity in the bottom plot, it can be seen that high velocity strikes generate greater harmonic content, as well as the energy taking a longer time to decay for the fundamental and its corresponding harmonics.

4.3.2 Statistical Tests

In order to test whether the features extracted from the varied velocity recordings were significantly different, a two-sample Kolmogorov-Smirnov (KS) test was used, as the data was from a non-normal continuous distribution. The Anderson-Darling test was used to check for normality. The null hypothesis for the KS test is that the data in two vectors originate from the same continuous distribution. The two-sample KS test was used to evaluate each feature pair in Table 4.3 from the 22 low and 22 high velocity recordings. All microphones were evaluated separately as well as pooling all 88 low and 88 high velocity recordings. The test revealed that all features for low velocity recordings were significantly different from those of the high velocity recordings (p <0.05). The only features which showed no significant difference were for the KM184 microphone; these were entropy (p = 0.56) and irregularity (p = 0.08).

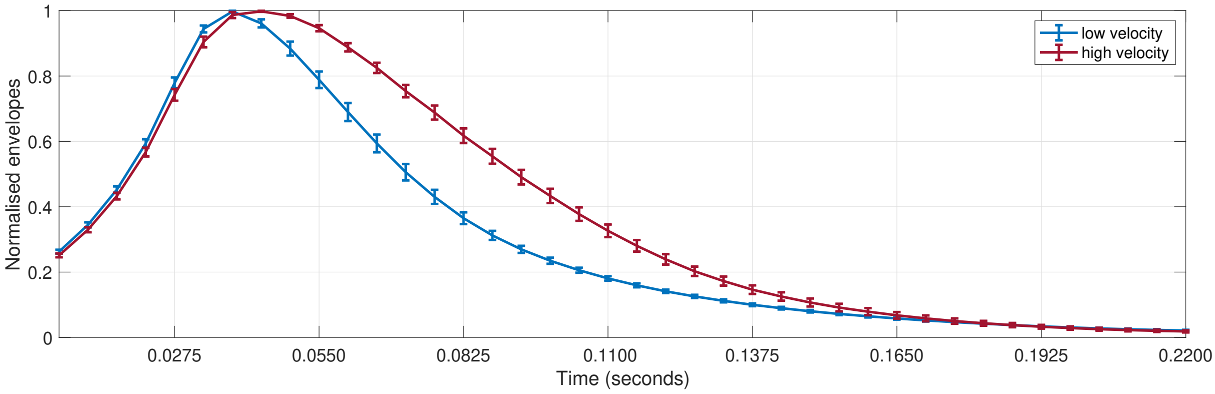

Figure 4.7: Mean and standard deviation of normalised envelopes for high and low velocity strikes.

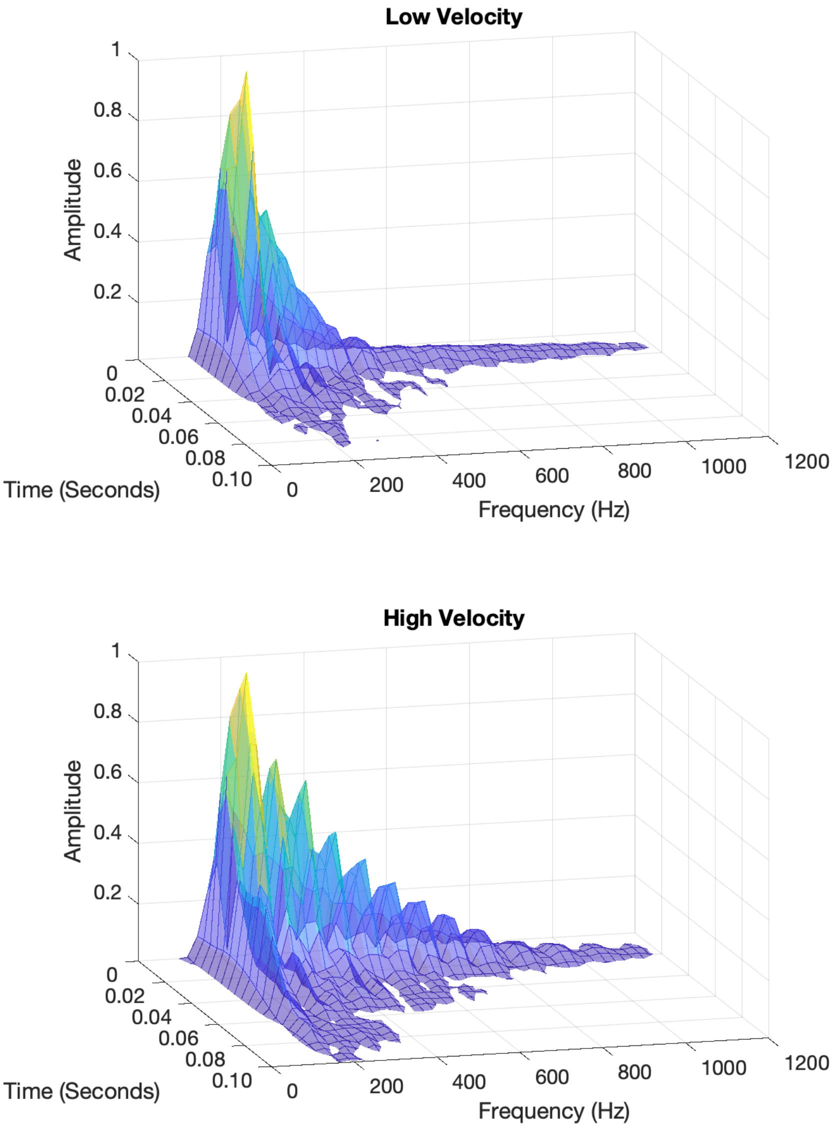

Figure 4.8: 3D surface of spectrograms for high and low velocity snare strikes.

Rank	Beta57	NT55	KM184	SM57
1	700Hz	250Hz	250Hz	150Hz
2	350Hz	700Hz	350Hz	840Hz
3	840Hz	840Hz	700Hz	700Hz
4	150Hz	570Hz	840Hz	1370Hz
5	570Hz	350Hz	570Hz	570Hz

Table 4.4: Top five ranked statistically-different critical bands for each microphone.

Beta57	NT55	SM57
450Hz	450Hz	250Hz
3.4kHz	4.0kHz	2.9kHz
4.0kHz	4.8kHz	3.4kHz
4.8kHz	7.0kHz	4.0kHz
5.8kHz	8.5kHz	
	10.5kHz	

Table 4.5: Critical bands found to be not statistically different.

The Bark scale critical bands from the high and low velocity recordings for each microphone were also evaluated to identify if there were significant differences between any of the bands. The two-sample t-test was used to compare the distributions of each critical band, and the Anderson-Darling test was used to check normality of the distributions. The null hypothesis of the two-sample t-test is that two normal distributions have equal means and equal but unknown variances and the alternative hypothesis is that the distributions come from populations with unequal means. Table 4.4 shows the rank of the top 5 significantly different critical bands ordered by their level of significance from the t-test results. Table 4.5 shows which Bark bands had no significant difference between the high and low velocity strikes. All Bark bands for the KM184 recordings were significantly different between velocities.

4.4 Discussion

Analysis of the high and low velocity recordings show various timbral differences. These differences made it possible for participants to distinguish between velocities when the cue of loudness was removed. All but two features extracted from the high and low velocity recordings were significantly different to each other. A notable feature which demonstrated significant difference was decay time, which was slower for high velocity strikes with an average time of 27ms. This confirms intuition, as the higher velocity strikes excite the drum skin with more energy and thus more time is required for energy to dissipate. Interestingly, attack time was found to be approximately 3ms quicker for the lower velocity strikes. Although statistically significant, the degree to which this was perceivable by participants was out of the scope of this study.

The fundamental frequency was significantly lower by roughly 20Hz when the snare was struck harder. Centroid, rolloff, and brightness were all lower for high velocity strikes, indicating that the low velocity strikes have proportionally more high frequency energy, and the high velocity strikes have proportionally more low frequency energy. High velocity strikes have additional energy below 1000Hz compared to the low velocity recordings, across all microphones, this can be seen in Figure 4.6. Further analysis of the Bark critical bands revealed high velocity strikes have significantly increased energy between the 570 and 840Hz critical bands (as seen in Figure 4.6). This suggests that although the low velocity strikes have proportionally more high frequency energy across the spectrum, which may be perceived as being brighter. The additional low frequency energy created by higher velocities is responsible for the reduction in brightness and centroid measurements and is not due to a lack of high frequency energy.

It was found that all critical bands were significantly different for the KM184. Although many of the same critical bands across all microphones were significantly different between velocity intensities, the strength of these differences varied between all microphones, which indicates that the unique frequency responses of the microphones had a non-linear effect. Out of the top 5 ranked statistically different critical bands, all 4 microphones had differences between the bands centred at 570Hz, 700Hz, 840Hz, and three of the microphones all having a 350Hz significantly different critical band. Both the NT55 and the KM184 had the highest ranked difference for the 250Hz critical band.

In nearly all cases where critical bands are significantly different between the high and low velocity strikes, the energy in that band is significantly higher for the higher velocity strike with a few exceptions. For the Beta57a, NT55, and SM57, the critical band centered at 150Hz had more energy for the low velocity strikes as seen in Figure 4.6, this was only not seen for the KM184, likely caused by features associated with this microphones non-linear characteristics.

4.5 Conclusions

This chapter has explored subjective and objective timbral differences associated with snare drum strike velocity. A listening test was carried out in order to assess if participants could distinguish between high and low velocity snare strikes when loudness disparity had been removed. It was discovered that all participants could identify the velocities with the absence of loudness cues with a high degree of accuracy. This indicated that participants were using temporal and spectral differences to select the correct velocity recordings. To determine if microphone choice played any significant role in affecting listener perception of velocity, 4 common studio microphones were selected for the test. The lowest (88.67%) amount of correct responses were for the Beta57a and the highest (93.29%) were for the NT55. Statistical evaluation revealed no significant differences between any of the score from the 4 microphones, indicating that the listeners ability to accurately distinguish between velocity intensities was not affected by any timbral difference associated with each microphone.

Nearly all features extracted from the recordings were significantly different between high and low velocity strikes, showing that attack time was shorter for the low velocity strikes, whilst decay time was longer for the high velocity strikes. Additionally, fundamental frequency was also shown to vary with change in velocity, with high velocity strikes producing on average a 20Hz lower fundamental. Statistical analysis of the Bark scale critical bands using a two-sample t-test showed that the largest disparity for velocity intensities was exhibited between the bands centred at 570Hz to 840Hz.

A consideration to take into account is that, while efforts were made to select a snare drum with generalisable timbre and tuning, the exact measurements only directly apply to the specific snare drum used in the study. Although much can be extrapolated from these findings, a more comprehensive understanding of timbral differences between high and low velocity strikes could be obtained through the use of additional recordings from a range of snare drums. This could include snare drums of different shell material and dimensions, as well as a range of drumhead type and tunings. Other properties such as the tension of the snare wires and number of snare strands may result in velocity dependant spectral variation. Drum stick material (e.g., nylon, wood) may even play a role in timbral differences. Player technique and location of strike are also likely to produce measurable variations. In the next chapter, the use of audio affects is explored, then snare drum microphone preference is evaluated in order to categorise least-preferred and and highly-preferred microphones. From this categorisation an attempt is made to transform the spectral features of one least-preferred microphone in order to mimic the features of highly-preferred microphones.

Chapter 5

Microphone Transformation

5.1 Introduction

Microphone selection is a method used by recording engineers to tailor the timbre of the recordings for their specific requirements. Additionally, audio effects allow mixing engineers to manipulate the timbre of the individual elements that comprise a full song in order to combine them together and improve the overall subjective quality (White, 2006a; Messitte, 2022). This chapter first introduces two audio effects, the graphic and parametric equaliser (EQ). The application of EQ is then discussed for the task of shaping and transforming the spectrum of the snare drum when mixing a song comprised of multi-track recordings. Literature is explored that provides specific recommendations for dealing with technical issues and enhancing subjective quality of the timbre of the snare drum. An investigation is then presented which explores the feasibility of spectrally transforming one microphone's characteristics to mimic those of another when used for snare drum recording. Transformations are carried out through the use of a digital graphic EQ. The investigation makes use of a robotic drum arm for consistent playing, the development and evaluation of the drum arm's performance are detailed.

As discussed in Chapter 3, differences between microphones occur based on their physical construction, which affects properties such as frequency response and polar pattern. For this reason, particular microphones are specifically chosen for their ability to produce favourable results when used to capture certain instruments or for certain styles of music (Bartlett, 1987; White, 2006b; De Man and Reiss, 2013). With a plethora of microphones for the recording engineer to choose from, microphone selection is often based on personal experience acquired from many years of experimentation, recommendations from other engineers, or personal preference (Eargle, 2004; Owsinski, 2005; Houghton, 2010). Chapter 3 highlighted the importance of microphone selection for snare drum recording. Positive correlation was observed between subjective listening test preference scores of snare drum recordings from multiple microphones and the spectral energy above 1.5 kHz. This indicates that the frequency response of the microphones is in part responsible for preference. Another study by Pearce et al. (2016) found that the perceptual attributes of *brightness, harshness* and *clarity* contributed the most to describing inter-microphone differences—descriptors closely related to frequency content. McKinnie (2006) suggests that when microphones of similar build type and polar-pattern are equalised to have near identical on-axis frequency response they would still exhibit some variation in timbral qualities, yet this claim was not investigated. The study instead aimed to identify the most salient perceived differences between the nine condenser microphones under evaluation. This study found that listeners could not distinguish between many of the stimuli recorded with the different microphones. Hebrock et al. (1997) developed a method for measuring time domain responses of 25 microphones to understand why microphones with similar performance features were perceived differently by listeners. The results proved inconclusive due to the large amount of variables, however, it was noted that the deviation between the frequency responses of the microphones under evaluation was vital for listener characterisation of the sound.

In previous studies, frequency response differences of microphones was shown to be one of the most important factor influencing subjective preference and the perception of timbre. This chapter first introduces the equaliser and its traditional uses for mixing and modifying drum recordings in Section 5.2, this is then followed by a microphone transformation investigation from Section 5.3 onwards, then Section 5.9 presents conclusions, highlighting some of the key findings of the investigation.

5.2 Equalisation

Audio effects refers to a range of processors, both analogue and digital, that transform a live audio input or audio recording in order to elicit some desirable change. These changes could impact the frequency, amplitude, and phase of the signal, and some effects may introduce additional artifacts, such as reverbs and delays which prolong the duration of the original signal. One of the most ubiquitous and important audio effects available to the engineer when mixing is the EQ (Aisher, 2012). In its simplest form EQ is designed to attenuate or amplify a specific range of the frequency spectrum of the audio signal. EQs are found built into small and large-scale mixing consoles designed for both live sound and studio applications, and are often featured in many home Hi-Fi systems. In addition, there exists standalone analogue and digital EQs specially designed for audio mixing and production tasks. It is not uncommon for an EQ to be used on every individual instrument of a multi-track mixing session (Hahn, 2018).

A common type of EQ found on nearly all mixing consoles is the parametric EQ which allows the user to specify the centre frequency, the amount of amplification also called *boosting*, the amount attenuation often referred to as *cutting*, and the bandwidth of the frequency range that is being affected, know as the Q, Quality-factor, or peak shape (Messitte, 2021). To be considered a true parametric equaliser all 3 parameters should be continuously variable. Before the invention of the parametric EQ, earlier designs from around the 1930s—1950s featured a set of selectable frequencies with boost or attenuation. The parametric EQ was first invented in the early 1970s (Massenburg, 1972), before this design became widespread the graphic equaliser was typically relied upon for various frequency sculpting applications (Mellor, 2018).

Mellor (2018) provides further insight into the graphic EQ, which is comprised of fixed frequency overlapping peaking filters with a slider for each filter controlling the amount of amplification or attenuation. When the slider of the filter is set in the middle position the gain of the filter is set to 0 dB. Graphic EQs may feature either a filter per octave, a filter per $\frac{1}{3}$-octave, or filters with tailored centre frequencies chosen for specific applications. The larger the number of bands, the smaller the bandwidth of the individual filters and therefore the greater the control over more subtle aspects of the frequency spectrum. A 30-band and a 10-band analogue graphic EQ can be seen in Figures 5.1 and 5.2 respectively. In certain graphic EQ designs, the filters will have a constant Q, where others will feature proportional or variable Q meaning that as more amplification or attenuation is applied, the narrower the bandwidth of the filter becomes so that at the most extreme settings the Q is at its tightest. On a constant-Q graphic EQ the Q always stays the same regardless of the amount of gain applied.

The graphic EQ gets its name from the fact that the many sliders found on the device act as a graphical representation of the frequency curve that is being applied, allowing the user to easily obtain visual feedback

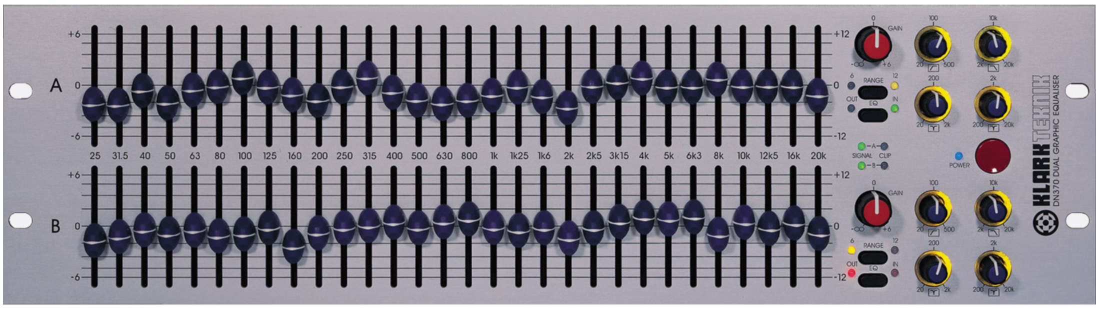

Figure 5.1: Example of a 30-band graphic equaliser. (Courtesy Klark Teknik).

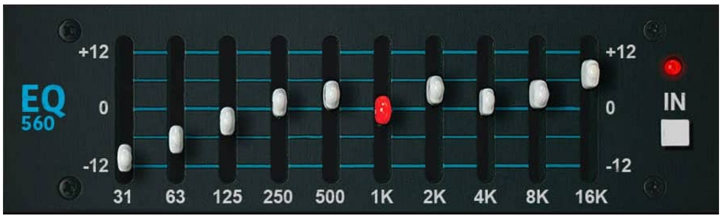

Figure 5.2: Example of a 10-band graphic equaliser. (Courtesy Red Rock Sound).

of the transformation being applied. This aspect is particularly useful in live sound applications where the ability to quickly identify and attenuate a problematic frequency cause by feedback is essential. Although graphic EQs are still extensively used in live sound they are less frequently used for studio application with the parametric EQ being the primary tool for frequency adjustment when recording and mixing. Massenburg (1972) describes some possible benefits associated with parametric EQs; they can be faster to use than a graphic EQ in that one can hear as the peak of the EQ is being swept through the frequency spectrum and hear the necessary point where correction is needed, one can then quickly judge the frequency and amount of correction that is required. Although a common three-band parametric EQ cannot construct as complex a characteristic as a graphic EQ, its variable shape and frequency let it produce a peak at any frequency and contour its effect to match an anomaly that may need to be removed.

Certain EQs only feature parameters for frequency and gain and will have a fixed Q values, others models may only allow for adjustment of gain and have fixed frequency and Q values, such as in the case of the Mäag Audio EQ4 shown in Figure 5.3, which features 5 filters with fixed frequencies bands and one filter with an adjustable frequency range. The EQ section of the SSL AWS924 mixing console is a 4-band design, shown in Figure 5.4, it features 4 filters with overlapping frequency ranges that cover the majority of audible spectrum, the bands of each filter are; high frequencies (HF) 1.5kHz—22 kHz, high mid frequencies (HMF) 600 Hz—7 kHz, low mid frequencies (LMF) 200 Hz—2 kHz, and the low frequencies (LF) 40 Hz—600 Hz. Only the HMF and LMF are fully parametric with a variable gain, frequency, and Q parameters. The HF and LF bands do not incorporate a Q parameter but do allow the user to select between a bell and shelf filter type. The bell filter will boost or attenuate frequencies above and below the centre frequency by equal amounts in a curve that resembles a bell shape. A shelf filter will boost or attenuates all the content above the specified frequency for a high-shelf, and below the specified frequency for a low-shelf. In addition to the 4 EQ bands, the SSL console also features a high pass filter (HPF), this filter simply attempts to removes all the content below the specified cut-off frequencies, there is a gradual roll-off of attenuation based on the topology and design of the filter (Solid State Logic, 2010). An ideal filter that is capable of removing all audio content above or below a certain frequency is referred to as a brick-wall filter (Biswas, 1998). A low pass filter (LPF) will roll off the

Figure 5.3: Mäag Audio EQ4 with fixed frequency filters. (Courtesy uaudio.com).

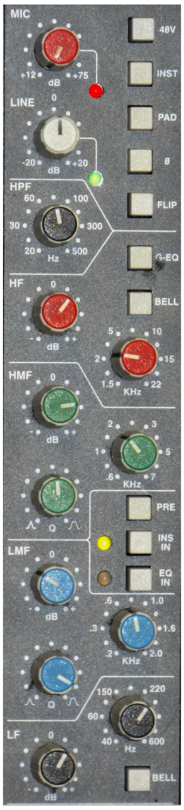

Figure 5.4: Solid State Logic (SSL) AWS δelta 4-band equaliser.

content above a specified cut-off frequency allowing the content below to be unaffected. Although the exact frequency ranges may vary between manufactures, this style of EQ is ubiquitous among studio and live sound mixing consoles (Mellor, 1995; Inglis, 2022).

Owsinski (2017) states that the EQ is the primary tool for mixing engineers to make any instrument sound *clearer, bigger, brighter*, and more *defined*. The desired results are often obtained by removing obtrusive frequencies and emphasising the predominant ones. EQ is a tool that can be utilised for both correcting variations and technical flaws in a recording or playback system as well as being used creatively in order to improve subjective timbral qualities (Senior and White, 2001). An issue when equalising musical signals is deciding which frequencies correspond to particular elements of a sound's timbre; Senior and White (2001) outline a common technique used to identify frequencies associated with timbral characteristics when utilising a parametric EQ—by applying a high amount gain with a peaking filter and the Q parameter set in its middle range, one can sweep the frequency control through the spectrum whilst critically listening as certain properties of the sound are accentuated by the EQ. This will help reveal what subjective changes will occur at certain frequency bands of that particular sound. When applying an EQ the pitch associated with a given frequency should not be of primary concern, but rather the timbral change that arises from boosting or attenuating the content in that specific region.

5.2.1 Drum Mixing

Toulson (2021) provides an overview of several techniques used specifically for drum mixing and optimising the timbral quality of multi-track drum recordings. It is noted that one must first start with good quality recordings which will alleviate the need to rely on excessive EQ which may produce noticeable and unwanted artefacts generated from extreme boosts, cuts, and Q values. It is also suggested to apply EQ while listing to a full mix or a sub-mix of instruments all at once to hear the changes being made in the context of the song, rather than listening and adjusting each element in isolation. Mixing has the potential to make a good drum recording sound even better, hyper-realistic, impactful, and more exciting to the listener. EQ can be used to manipulate the volume of specific frequencies of the signal's spectrum, emphasising certain characteristics, allowing instruments to stand out, avoid frequency competing and clashing between different instruments, and manage issues of bleed. Before any effects are added an initial volume balance of all instrument should be carried out, as well as panning each track to an appropriate stereo location. There are many valid cases where effects may need to be applied to an individual instrument, for example dynamic processing tools such as compression and gates are needed for shaping and controlling the volume envelopes of specific signals.

When mixing drums Toulson (2021) presents four main uses of EQ; cutting low frequencies, treating the fundamental and overtones of each drum, adding attack and presence, and controlling high frequencies. Each of these aspects are discussed in more detail. Firstly, low frequency energy can build up from multiple microphones all capturing different parts of the drum kit, but also capturing room ambience and spill from the kick drum. This has the affect of reducing clarity of each individual element and often produces a mix that is referred to as sounding *muddy*, caused by destructive phase problems in the low frequency range and low frequency reverberations overpowering the direct signal. By cutting low frequencies using a high-pass filter one can help alleviate this problem. In general, the cut off frequency should be set lower than the fundamental frequency to target only the problematic frequencies and not negatively affect drum timbre. For a snare drum channel this could be set around 100Hz to help remove some of the kick drum bleed that has been captured by the snare microphone. This technique can be applied to all microphone channels including overhead and room microphones, resulting in enhanced definition of the drum kit. Secondly, EQ can be utilised to exaggerate and emphasise certain drums in the mix by boosting the fundamental frequency, rather than simply increasing the volume of the whole channel. This produces a more focused drum sound that is less likely to be masked by other instruments. Overtones of the snare can also be reduced if they ring for too long, are distracting, or are overpowering the fundamental. EQ allows for complete control to reshape and manipulate the balance between the fundamental and overtones, thereby transforming the drum's characteristics. Thirdly, the perceptual impact of the attack and presence can be enhanced through the use of EQ, this is because high frequency content relates closely with these attributes. The overall character of the drum can be changed by increasing or decreasing certain higher frequency bands of the recording. To make the drum sound brighter and clearer the 2—8 kHz frequency range can be boosted; the exact frequency and amount will depend on the properties of original drum recording. Finally, Toulson (2021) states that EQ can be implemented to control the high frequencies that build up as a result of loud cymbal and hi-hat bleed being captured by other microphones which can be problematic. It is common for the snare microphone to capture a lot of unwanted bleed from the hi-hat, making it difficult to enhance the snare's high-frequency content without over emphasising the hi-hat. It can often be appropriate to reduce high frequencies in drum tracks that do not need as much high-frequency content. An example of a common mixing strategy is to apply a high-frequency shelf at 12 kHz, applying 3 dB of attenuation to reduce the impact of the hi-hat on the snare channel, thus allowing the snare to be compressed and boosted without causing an increased presence of the hi-hat in the mix. This is a trade-off between allowing the snare drum to be enhanced with compression, but sacrificing some of the natural characteristics of the timbre by applying the shelf EQ.

Major (2014) reiterates the importance of addressing and correcting the low frequencies when mixing drums, suggesting the removal of all unnecessary low frequencies from instruments that do not have very much musical information in that range. This task can be carried out through the use of a high-pass filter in order to enhance multi-track drum recordings, in particular the snare drum. When recording the underside of the snare drum, a large amount of the kick drum signal will be captured. Assuming one is using a dedicated kick drum microphone, this bleed is often undesirable as the microphone used and its position are not optimal for capturing the kick drum, this bleed can negatively interact with the signal captured from the kick drum microphone. A simple solution is to remove the majority of the low frequency energy from the underside snare microphone thus removing the kick drum bleed. Weekhout (2019) suggests using a high-pass filter with a cut-off frequency between 80—150 Hz with a roll-off of around 12—24 dB/oct for this purpose. Gibson (2004) also suggested utilising a high pass filter, specifically around 100—200 Hz, on all microphone channels that do not require a lot of low frequency energy, such as the snare, hi-hat, rack toms, and overheads, in order to remove bleed caused by the kick drum.

Senior and White (2001) acknowledges that not all low frequency content is solely a result of bleed from the kick drum, as there will also be unwanted noise, low frequency resonance, and rumble potentially from other instruments, as well as extraneous sources from the surroundings, such as air conditioning or traffic. This noise may be imperceivable until exaggerated with an excessive boost from the EQ. Due to this low frequency information being unwanted it is advised to use a high-pass filter to remove any unused low-frequency energy, which may have a negative impact on the overall mix. This can become particularly problematic when many separate recordings with the same issue are layered together which causes a build up of undesirable energy in this low frequency range, reducing perceived clarity. Owsinski (2017) agrees that a crucial part of the mixing stage involves *cleaning up* all the individual tracks, removing rumbles, thumps, creaks, and any other extraneous noises that detract from the recording; these tasks can be achieved through the use of an EQ. Additionally, EQ can be used to help when two elements have conflicting and competing energy at the same frequencies. By removing energy from the lesser important element around the problematic frequency bands one can prioritise the more important element by creating space in the spectrum for it to be heard more easily.

Major (2014) suggests that when applying EQ to the snare drum, boosting frequencies between between 2.5—6 kHz will emphasise articulation, allowing strikes to be more clearly heard above other instruments. However, he notes that a possible drawback to this approach is the potential to further exacerbate any unwanted hi-hat bleed. An alternative he suggests is to apply EQ the bottom microphone channel instead, as microphones placed underneath the snare drum are less susceptible to bleed from the hi-hat. Pedersen and Grimshaw-Aagaard (2018) suggest the use of a noise gate as a possible solution to dealing with hi-hat bleed, whereby the threshold is set to exclude the quieter hi-hat strikes, allowing only the snare strikes to be heard. One downside to this approach is that softer played snare drum strikes may also be removed by the noise gate, this has lead to some engineers opting to manually remove sections of audio in-between the snare strikes within the digital audio workstation (Senior, 2018).

In order to improving the timbral quality of the snare drum, Pedersen and Grimshaw-Aagaard (2018) recommends boosting frequencies between 100—300 Hz, and slightly attenuating frequencies around 500—700 Hz.

Senior and White (2001) also provide generalisable guidelines for three frequency ranges where EQ should be focused, these are: 120—400 Hz to emphasise lower mid-range frequencies of the snare drum, described as increasing the perceived or *weight* or *body*; 2—4 kHz where the resonance of the snare drum produces ringing, which can either be reduced or further increased as appropriate; 4—8 kHz a range that helps to create *brightness* or *crispness* to the timbre and is responsible for articulating the drum's attack. Owsinski and Moody (2009) suggests slightly different frequency ranges to address when mixing a snare drum; 80—110 Hz for low frequencies, 3.5—5 kHz for mid-range frequencies, and 10—12 kHz for high frequencies. Weekhout (2019) defines the low range to be slightly higher in frequency, between 125—250 Hz, which is described as providing *warmth* to the sound. He defines the mid-range to be centred around 1.5 kHz which can be boosted to increase perceived *aggression*, and the high frequency range is considered to be all information above 4 kHz which is responsible for a *sizzle* like quality of the timbre.

Gibson (2004) suggests that the use of EQ during recording should only be considered once the drums have been appropriately tuned, and microphone selection and position have been finalised. In additional, only very subtle EQ should be applied to drums during the recording stage as all timbral modifications are imprinted on the recordings and may be difficult to correct if not required or appropriate later. More dramatic and noticeable alterations can therefore be carried out during the mixing stage once all the instruments have been recorded and can be judged in the context of the whole song. Gibson (2004) also recommends reducing energy between 200—600 Hz, which he states has a tendency of being overabundant in multi-track drum recordings, due to the proximity effect of multiple cardioid microphones positioned closely to the drums.

Case (2012) discusses some of the technical challenges when attempting to EQ a snare drum, stating that snares are difficult to EQ in isolation as they react strongly to almost any spectral change due to their broadband spectrum. Therefore, the context of the other instruments is essential when choosing specific spectral regions to either emphasise or de-emphasise. Typical snare drum recordings will often have predominant mid-range energy which can benefit from some gentle attenuation in order to create space in the frequency spectrum for other instruments that have important information in this range. Case (2012) also recommends that if the snare recording has a noticeable unpleasant sharp metallic *ringing* sound, attenuation of 6—12 dB can be focused around 1—2 kHz, which will help alleviate this issue and emphasise other desirable spectral features without the ring overpowering the sound. A high-Q notch filter can be used over other problematic frequencies where the filter is set wide enough to remove the unwanted content, but narrow enough to avoid diminishing the snare's timbre. It is suggested that issues of ringing should first be addressed by means of tuning and then dampening before relying on EQ. It is also noted that EQ is typically used in conjunction with compression, gating, reverb, and other effects to create an impactful and exciting sounding snare drum.

Although there are slight differences between the suggested frequency ranges intended to target specific aspects of snare timbre, there is some consensus when applying EQ to the snare drum in three general frequency bands: low-range (80—400 Hz), mid-range (2—5 kHz), and high-range (4—12 kHz). EQ plays a pivotal role in modifying the recorded audio to further tailor its frequency spectrum in ways that are difficult or impossible to achieve through the real-world adjustments of the recording parameters such as tuning, dampening, microphone selection and placement. EQ also allows for issues to be addressed and corrected that arise during the recording stage, eliminating the need to re-record the drums which could cause delays to a recording project or incur additional expense.

5.3 Microphone Transformation

An area of investigation that has not been fully explored is the ability to utilise EQ to modify the spectrum of a signal in order to emulate modifiable recording parameters, such as microphone selection. The rest of this chapter explores the potential to carry out such a transformation. Firstly, a comprehensive multi-stimulus listening test is conducted with 12 microphones across four distinct snare drums to identify a ranked categorisation of microphones. In order to evaluate the effects of spectral modification, the least-preferred microphone is transformed to take on the frequency characteristics of the most highly-preferred microphones. A second listening experiment is then conducted to determine the extent to which the preference of the least-preferred microphone has been improved.

5.4 Methods

The listening tests in this study utilise professional quality recordings of consistent snare drum performances. In total, 12 microphones are selected across a range of variables including cost and manufacturer (Section 5.4.1). To ensure that listener preference is not an effect of snare drum selection, multiple snare drums are selected with varying configurations and specifications (Section 5.4.2). Great care is taken to ensure that the recording equipment and procedure are of a professional level (Section 5.4.3) and that snare drum excitation is as consistent as possible (Section 5.5).

5.4.1 Microphone Selection

In this investigation six small diaphragm condenser microphones and six dynamic microphones were selected. Table 5.1 shows the full list of microphones used. Microphones were chosen from a range of available manufacturers, as well as recommendations from recording engineers and online articles, (Elliott, 2014; Fuston, 2017a). Only microphones that were commercially available at the time and deemed appropriate for snare recording were selected. Microphones that could not be positioned without obstructing a drummer or specialist microphones such as kick drum microphones were excluded. In Chapter 3, solo snare drum recordings, and snare drum recordings including a kick drum and hi-hat found that listener preference for the majority of microphones did not significantly change between the two recording scenarios. However, as the preference for three microphones changed significantly, these three microphones were excluded from this study as only recordings of solo snare drums were investigated. These microphones were: Audix ADX51, Audix D4, and DPA 4099.

5.4.2 Snare Drums

Table 5.2 presents the four snare drums selected for this investigation, including two steel shell drums, one maple shell drum, and one maple and walnut shell drum. All four drums were 14" in diameter, used a Remo Weatherking Ambassador Hazy Snare Side[1] for the resonant head, and were fitted with 20 strand PureSound snare wires. The resonant and batter heads of the snare drums were tuned with the aid of a digital DrumDial,[2] following the same protocol as in Chapter 3, in order to ensure uniform head tension at every lug position. A 1" Evans E-ring[3] was placed on the batter head to slightly dampen overtone and create a more realistic recording scenario.

[1] www.remo.com
[2] www.drumdial.com
[3] www.evansdrumheads.com

Brand	Model	Type	Polar Pattern
AKG	D5	D	Supercardioid
Beyerdynamic	M201	D	Hypercardioid
DPA	4011A	C	Cardioid
Neumann	KM184	C	Cardioid
RØDE	M5	C	Cardioid
RØDE	NT55	C	Cardioid
SE	V7X	D	Supercardioid
Sennheiser	e614	C	Supercardioid
Sennheiser	MKH40	C	Cardioid
Shure	Beta57a	D	Supercardioid
Shure	SM57	D	Cardioid
Telefunken	M80	D	Supercardioid

Table 5.1: Makes, models, types, and polar patterns of dynamic (D) and condenser (C) microphones used for the recordings.

Snare	Batter Head	Shell	Depth
Black Panther Machete	Evans Hydraulic	Steel	6.5"
Black Panther Velvetone	Evans Hydraulic	Maple/Walnut	5.5"
Premier Artist Maple	Evans HD Dry	Maple	5.5"
Tama Rockstar	Tama Power Craft II	Steel	6.5"

Table 5.2: Configurations and specifications of four snare drums used in experiments.

5.4.3 Recording

The recordings were carried out in a sound-treated isolation recording booth measuring H2.5 x W3.0 x L4.5 metres, with an ambient noise level of ~40dBA. Each snare drum and microphone was recorded separately using a Metric Halo ULN-2 into Apple Logic Pro X at 32-bit resolution and 44.1kHz sample rate. The gain of the preamplifier was set to avoid clipping during any recording. The option to record all microphones concurrently as performed by De Man and Reiss (2013) and Pearce et al. (2016) was considered. For human performances where inconsistencies might occur between recordings this may be an optimal approach, however, a recording methodology was selected that allowed for accurately repeatable performances and thus allowed each microphone to be recorded separately. One issue associated with simultaneous recording of all microphones is that microphone positioning is strongly linked to variance in timbral characteristics (Bartlett, 1981; Senior, 2008; Quiroga et al., 2015). Pearce et al. (2015) suggests that for an ideal microphone comparison test, one should locate microphones under observation at the exact point in space, maintaining an identical pressure-gradient or soundfield. Recordings were therefore made serially, at a near identical position, as opposed to equally spacing the 12 microphones around the rim of the snare drum. Great care was taken to ensure the position was matched as accurately as possible. This was achieved by aligning the microphones to a triangular jig (removed from the drum prior to any recording), measuring H10 x W17 x L20 cm, see Figure 5.5.

In order to avoid listener fatigue in the subsequent evaluations, a 4-bar rhythmic pattern consisting of 16 snare hits was played at 120 beats per minute (BPM). This aimed to produce a more engaging stimuli than isochronous events for listeners and to provide a more realistic application of snare drum recording.

5.5 Robotic Drum Arm

This section discusses the design and evaluation of the robotic drum arm (RDA) used for the collection of the snare drum recordings. As shown in Chapter 4, the velocity at which a drum head is struck strongly impacts the tonality of the resultant sound, therefore the RDA was built to provide consistent excitation in the serial recording with each microphone (see Figure 5.5). An additional advantage of using the RDA over a human drummer is the consistency of strike location upon the drum head, another variable which could cause discrepancies in timbre.

The RDA is controlled through a MIDI interface with events sequenced in Logic Pro X. An Arduino Uno is used to convert MIDI messages into voltages, thereby switching a relay connected to an actuator that triggered the RDA to strike the drum. An elastic band is used to initialise the actuator position after each hit. The striking distance of the drumstick was calibrated to ensure that the stick would not dampen any resonance after it had excited the drum head and would not prevent the drumstick tip from reaching the drum head. A striking distance of 5 cm above the centre of the drum head was chosen to meet this criteria, producing hits of approximately 90dB SPL. The excitation consistency of the RDA was assessed by measuring the MIDI velocities achieved from 500 hits on a MIDI drum pad (*mean*: 118, *std*: 2).

5.5.1 Construction

The objective of the design process was to produce a machine capable of striking a drum head at a consistent point in space, i.e., not change striking position on the drum head between hits. It was also paramount that striking velocity remained near identical for each hit and to be performed with accurate timing. For ease of use and a practical interface, the desired machine was to be controlled via MIDI messages, allowing a DAW

Figure 5.5: Recording setup demonstrating robotic drum arm, triangular jig, and microphone.

to trigger the drum strikes and simultaneously record the resulting audio. The drum hits that were produced by the RDA had to be analogous to that of a human player in terms of timbral characteristics.

The main striking element of the RDA is a 40cm Vic Firth 5B hickory drum stick, this was chosen to produce a realistic sounding drum hit. A hole is drilled through the drum stick 10cm from the handle end, a thin metal rod is placed through this hole which acts a pivot point for the drum stick to revolve around, the metal rod is held in place by a microphone clip. The distance for the pivot was chosen to approximately emulate where a human drummer would place their hand. A microphone clip was chosen to allow easy mounting onto a standard 3/8" microphone stand, or 5/8" with an adapter. The mechanism that powers the movement of the RDA is a 12 volt DC actuator. When a 12 volt signal is applied to the actuator, a shaft is extended by 2cm. This shaft is connected via a bolt to the drum stick, 1cm from the handle end. This action causes the tip of the drum stick to move 7cm from its initial starting position. A -12 volt signal is able to cause the actuator to retract the shaft, however for simplicity the retraction of the shaft is achieved through the use of an elastic band. Thickness and length of the elastic band was chosen such that the force of the drum stick was dampened as little as possible whilst still providing sufficient retraction time. A mains adapter supplies the actuator with 12 volts DC by way of a 12V 30A HEF555 relay. The relay is triggered by a 5 volt control signal from pin output 13 of an Arduino UNO, switching the relay and allowing the 12 volts to pass to the actuator. The UNO receives an incoming MIDI *Note On* message from an M-Audio USB MIDI interface. MIDI notes are sequenced in Logic Pro X, however any DAW with MIDI output capabilities could be used. Figure 5.6 shows the signal flow of the system.

5.5.2 Drum Arm Evaluation

The excitation consistency of the RDA was assessed by analysing the MIDI velocities achieved from recording 500 hits played on an AKAI MPD24 velocity sensitive MIDI drum pad (*mean*: 118, *std*: 2). The RDA was set up so the tip of the drum stick was 5 cm above the the center of a 3 x 3 cm rubber drum pad (Figure 5.7). A

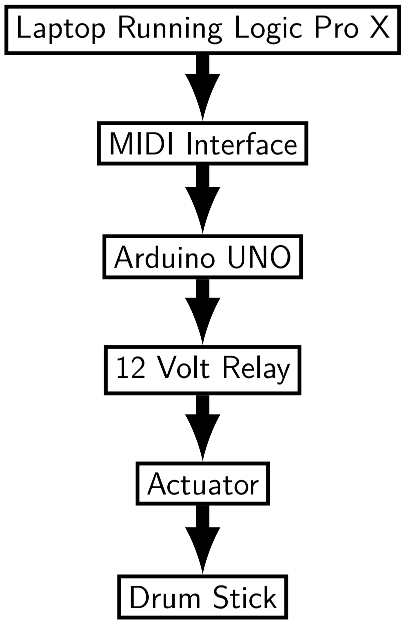

Figure 5.6: Signal flow diagram of RDA configuration.

further distance prevented the drum stick tip from reaching the MIDI drum pad, a closer distance produces a striking velocity so hard that every recorded MIDI velocity was 127, preventing any measurement of velocity variation. Figure 5.8 shows the set up used for evaluation. The small striking area allowed observations of any movement that the drum arm might make over the 500 notes. Pre- and post-test measurements revealed that the drum stick tip was still positioned in the centre of the pad once all 500 notes had been played, visual monitoring throughout revealed the RDA did not deviate position during the testing procedure.

Asynchrony was evaluated by measuring the variance in MIDI delay times (in milliseconds, corrected for buffer time) between the original sequenced MIDI and the resulting MIDI recorded from the RDA excitation. The onset times of the MIDI events from the original MIDI track that triggered the RDA were subtracted from the onset times of the newly created MIDI notes, which were generated by the RDA striking the drum pad. This resulted in delay times for each of the 500 MIDI notes, as the delay times change with the buffer size of the DAW, absolute values are not relevant, however of the 500 recorded MIDI notes 471 had identical delay times, with the other 29 notes having a $<$1.7ms shorter delay time than the other notes. Figure 5.9 shows the configurations for the evaluation setup.

5.5.3 RDA Conclusion and Improvements

Testing of the RDA revealed it was suitable for the intended application, and met the design considerations. The RDA's inconsistencies were not tested against that of human players, however it is expected that as a drummer's experience level increases they may be able to better maintain consistent velocity, timing, and striking position. By carrying out such comparisons of drummers across a range of experience levels, it may be determined how much experience is required for a human player to be on par with or even outperform the consistency of the RDA. The RDA used a conventional drum stick, powered by a 12 volt actuator and was triggered via MIDI. The RDA was able to remain in a fixed position and repeatedly struck the same location 500 times without deviating from starting position. Although the speed at which this excited a drum head could not be adjusted, analysis showed that the velocity had a small error margin over 500 hits. Analysis of the timing information captured from the recorded MIDI messages showed that of the 500 notes only 29 had

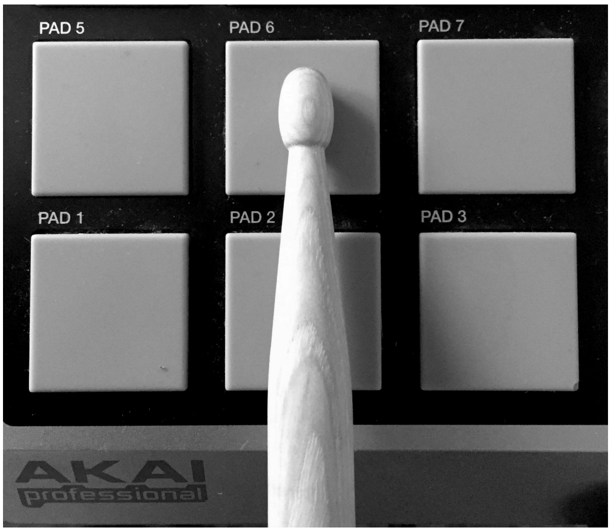

Figure 5.7: Striking position of the drum stick on the MIDI drum pad.

Figure 5.8: RDA and MIDI drum pad evaluation setup.

slight fluctuation in consistency which measured <1.7ms of variation. Despite its limitations the RDA was able to perform to a level of accuracy needed for the desired application.

One potential limitation of the system is the elastic bands used to retract the actuator shaft wearing out over time and producing different striking intensities. A lightweight metal spring may be better suited for this purpose. The RDA's capabilities could further be improved by the addition of varied striking velocities. The actuator used was not capable of variable speeds, however a mechanism with control of this parameter could be mapped to MIDI velocities. This would better facilitate real world drumming applications, and be able to further mimic a human player. As drum timbre changes based on striking position, a secondary motor could be installed to rotate the RDA to allow it to strike various locations across the drum head and produce a range of different timbres, or a secondary drum stick with independent control could be used to produce flams, drags, and rolls. Instead of directly modifying the drum stick for use on the RDA, a system for interchangeable sticks would allow different stick types to be quickly and easily changed for comparison, such as brushes and rods. However, these additional features were not required for the intended application.

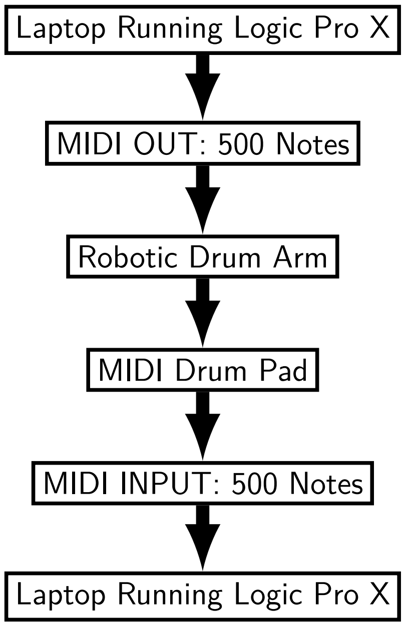

Figure 5.9: Signal flow of evaluation procedure.

5.6 Listening Test

A multi-stimuli listening test was used to evaluate listener preference of recordings captured with a range of different microphones. While previous studies have evaluated listener perception based on semantic descriptors (e.g., *warm*, *bright*) (C. Disley et al., 2006), the presented listening test was undertaken to produce a categorisation of highly-preferred and least-preferred microphones for snare drum recording prior to spectral modification.

5.6.1 Methodology

The listening test was performed in an acoustically-treated mastering studio using a pair of PMC IB1S speakers with a Bryston 2B-SST2 amplifier and an RME Fireface 802 digital-to-analogue converter. The tests were conducted using the Web Audio Evaluation Tool (WAET) (Jillings et al., 2015) with the APE interface (Man and Reiss, 2014). The samples were loudness normalised to -23 LUFS using the EBU-R-128 (2014) specification to remove any perceived loudness disparity between samples. De Man and Reiss (2013) found results from a multi-stimuli and an AB test produced comparable findings, a multi-stimuli approach was implemented to minimise test duration. In total, 42 participants took part in the test (*range*: 19 to 49 years, *mean*: 23.7 years, *std*: 6.2 years) with an average of 6.9 years of music production/recording/mixing experience (*range*: 2 to 25 years, *std*: 5.1 years).

The recordings from the 12 microphones were presented one snare drum at a time, resulting in four separate listening tests presented in a random order. Participants were instructed to rank recordings from least-preferred to most-preferred, from left to right. The 12 samples were randomised and represented by large green boxes labelled alphabetically as seen in Figure 5.10. These were ranked by the listener based on personal preference and moved using a select and drag method. When switching between samples, loop position was maintained for uninterrupted playback. Participants were not able to complete the test until every sample had been played at least once. The average test duration for all four listening tests was 13 min 36 s, with participants spending 3 min 24 s on average on each test page.

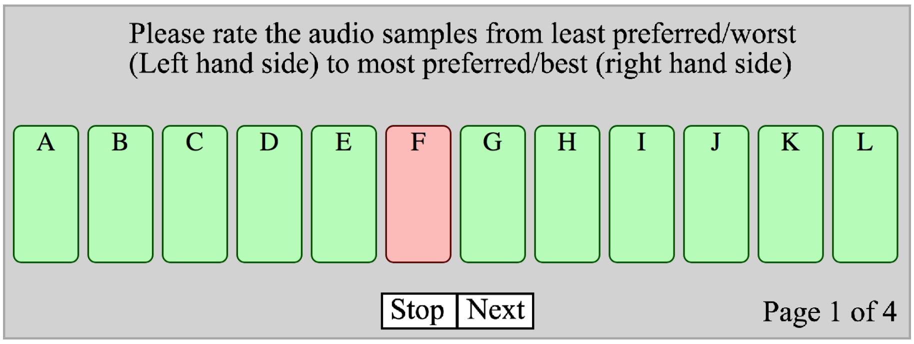

Figure 5.10: Web Audio Evaluation Tool (WAET) interface used for multi-stimuli listening test.

5.6.2 Results

As the participants ranked the stimuli from 1 to 12, the resulting data was ordinal. The Kruskal-Wallis test is used to test the null hypothesis—that is, listeners did not collectively show a preference between stimuli recorded from different microphones. The results from the Kruskal-Wallis test (p<0.05) reject this null hypothesis, indicating that listeners do indeed show preference between the different microphone stimuli. This is essential to ascertain prior to any further investigation on the preference of different stimuli.

A certain microphone might be better suited to a particular timbral characteristic inherent in the snare drum by emphasising desired spectral features. By recording multiple snare drums with the same 12 microphones and recording procedure, the impact of snare drum selection on microphone preference can be established. The Kruskal-Wallis test was also used to investigate if the rank of microphones was significantly different for the four drums. For 11 of the 12 microphones there is no statistical significance between rankings across snare drums (p>0.05). The only microphone exhibiting snare drum dependent results is the Sennheiser e614, (p<0.05). Although the test cannot be used to differentiate which specific snare rankings are significantly different from each other, post-hoc analysis indicates that the mean rank for the Machete snare was significantly lower than for the other three snare drums (p<0.05). As a result, the Sennheiser e614 is excluded from the remainder of this investigation, leaving a total of 11 microphones (i.e., five condenser microphones and six dynamic microphones).

5.6.3 Pairwise Comparison

To determine if the subjective rank of each microphone exhibits a significant difference to that of the other microphones, a non-parametric pairwise multiple comparisons test was carried out using Dunn's test (Dinno, 2015). Figure 5.11 presents the results of the pairwise comparison test, with microphones ordered by the mean rank across participants. Here, the yellow squares depict pairs of microphones which are significantly different, based on the result of Dunn's test (p<0.05). The matrices show that the top four ranked microphones consistently exhibit significant differences to the lowest four ranked microphones. The top four microphones cannot be considered to have significantly different ranks from each other, so no single microphone out of these may be considered optimal. Additionally, the lowest four microphones cannot be considered to be ranked differently from each other, so no single microphones should be interpreted as the least-preferred, with one exception. For the Tama snare drum, the SM57 can be considered more highly-preferred than the V7X. The remaining three middle-ranked microphones did not have consistent results across all four snares.

Using the results of the multiple pairwise comparison test, the microphones can be classified into three categories as seen in Table 5.3. The Category-1 classification denotes highly-preferred microphones that ranked

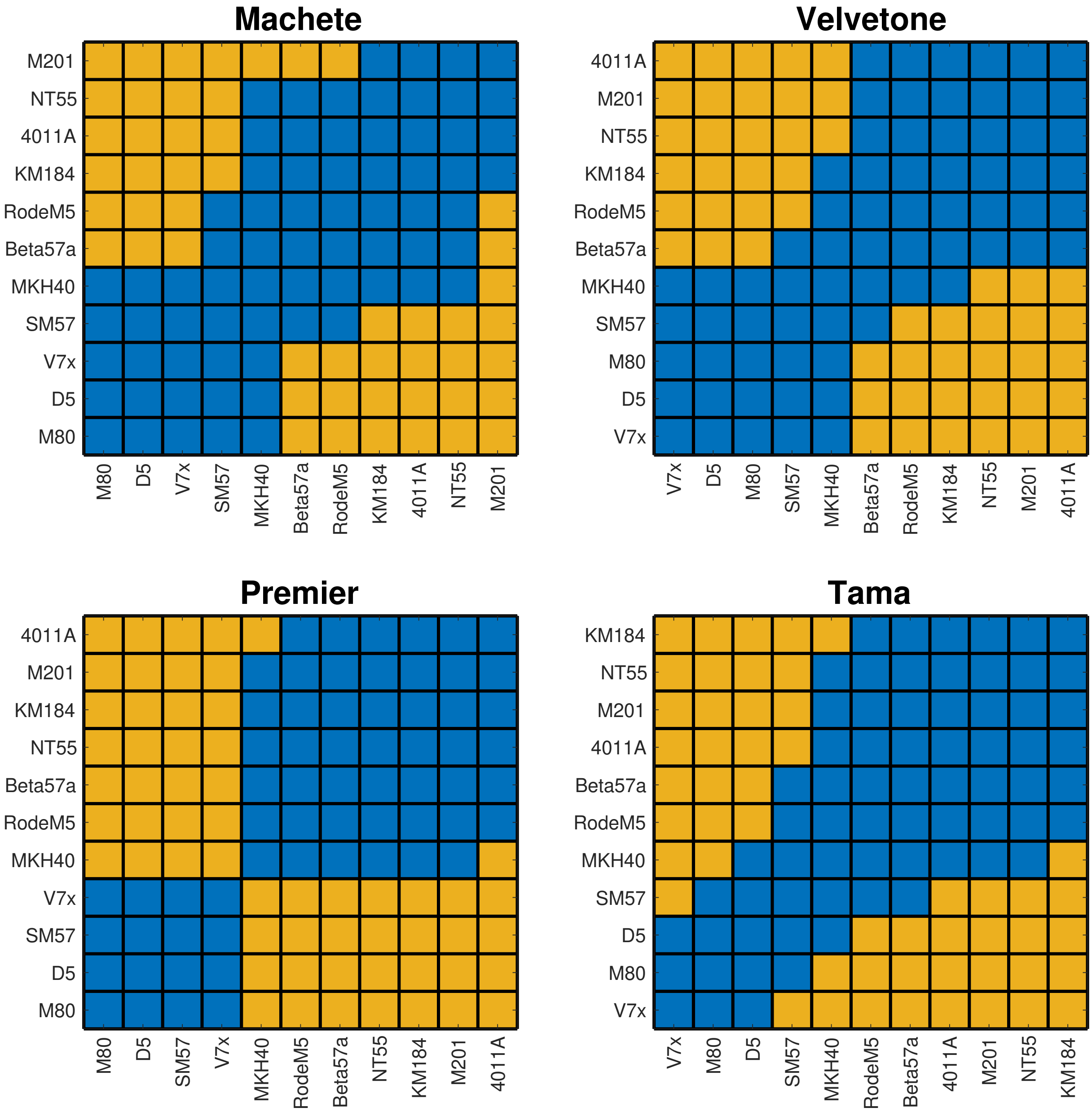

Figure 5.11: Results from pairwise comparison test with microphones ordered by mean rank. Yellow squares indicate pairs of microphones that are significantly different from one other ($p<0.05$) and blue squares indicate pairs that are not significantly different ($p>0.05$).

in the top four positions for all snares. These were significantly different from the microphones in Category-2, which are the least-preferred microphones ranked in the bottom four positions for all snares. Category-3 includes the remaining microphones which did not fall into either category, these microphones are of no relevance to the rest of the study and are not discussed further.

Category-1	**Category-2**	**Category-3**
4011A	D5	Beta 57a
KM184	M80	MKH40
NT55	SM57	M5
M201	V7X	

Table 5.3: Microphone categorisation achieved through multi-stimuli listening test.

5.7 Spectral Modification

Towards improvement of listener ratings for the least-preferred microphones, an exploratory audio effect is introduced to modify the frequency response of a Category-2 recording such that it mimics those of the Category-1 microphones. The adopted approach is to perform spectral analysis followed by a timbral transformation through a graphic EQ stage. A graphic EQ is preferred over a parametric EQ due to the inclusion of a standardised frequency division (ANSI, 2009). The Category-2 microphone chosen for transformation is the D5, which is the least expensive Category-2 microphone, allowing for the greatest monetary disparity between the microphone categories.

5.7.1 Frequency Response Analysis and Equalisation

The discrete Fourier transform (DFT) is used to extract the frequency response from the snare drum recordings under evaluation. To minimise differences between the drum excitations, the average of three individual excitations is used once these are aligned in the time-domain by cross correlation. While this extra step is not necessary with recordings made with the RDA, it provides consistency in performances with more variability. The DFT bins are then mapped to gain values associated with a 30-band graphic EQ. The centre frequencies for the 30 bands were selected based on the ANSI standard for fractional-octave-band digital filters (ANSI, 2009)—the same bands used by the graphic EQ. The $\frac{1}{3}$-octave bands used with mid-band frequencies ranging from 25—20,000 Hz are shown in Table 5.4. The energy within the DFT frequency bins associated with the bands of the $\frac{1}{3}$-octave EQ were first summed and then converted to decibels (dB).

The difference between all 30 frequency bands of a Category-1 microphone recording and that of the D5 is calculated. These 30 values are used to set the gains of a 30-band digital graphic EQ implemented by Oliver and Jot (2015), using a 24$^{\text{th}}$ order cascaded design. The large order of filters is required to minimise the difference between microphones. Figure 5.12 displays the gain used for each band of the EQ in the modification of the D5 recording. This process is repeated for all Category-1 microphones and applied to the D5 recording. Figure 5.13 shows the difference between the D5 and the NT55 recordings of the Velvetone snare, as well as showing the D5 recording post-EQ having been equalised to match the NT55 recording.

Octave	Lowerband Frequency (Hz)	Midband Frequency (Hz)	Upperband Frequency (Hz)
1	22.4	25	28.2
	28.2	31.5	35.5
	35.5	40	44.7
2	44.7	50	56.2
	56.2	63	70.8
	70.8	80	89.1
3	89.1	100	112
	112	125	141
	141	160	178
4	178	200	224
	224	250	282
	282	320	355
5	355	400	447
	447	500	562
	562	640	708
6	708	800	891
	891	1000	1122
	1122	1280	1413
7	1413	1600	1778
	1778	2000	2239
	2239	2560	2818
8	2818	3200	3548
	3548	4000	4467
	4467	5000	5623
9	5623	6400	11220
	11220	8000	8913
	8913	10000	11220
10	11220	12800	14130
	14130	16000	17780
	17780	20000	22050

Table 5.4: Frequency range and mid-band frequency of $\frac{1}{3}$-octave divisions used to segment DFT.

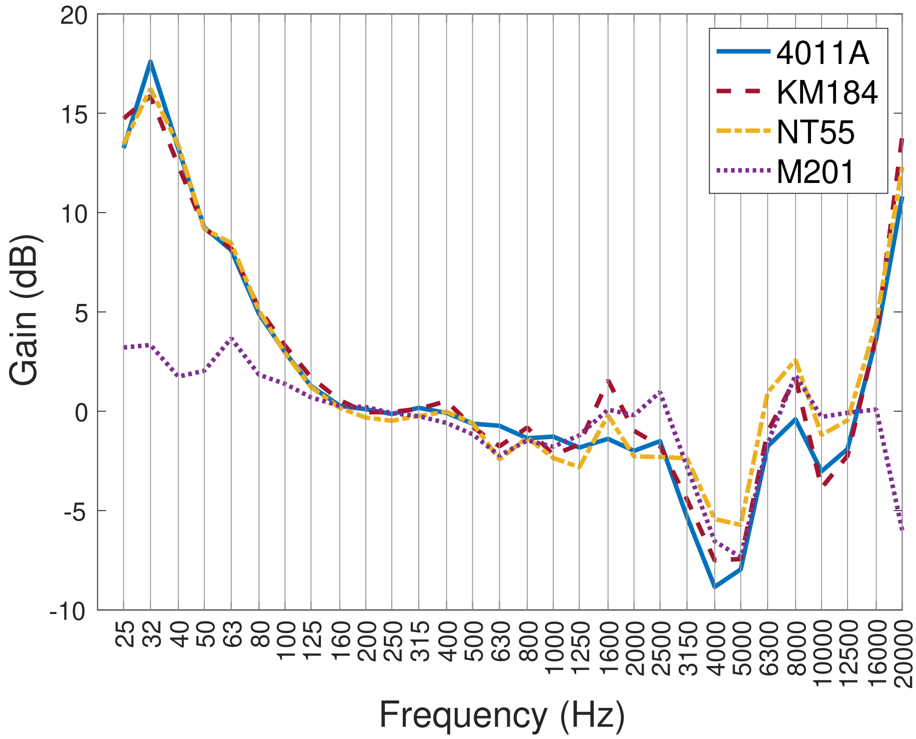

Figure 5.12: Gain values for 30-band EQ applied to D5 recording to minimise spectral difference with those of the Category-1 microphones.

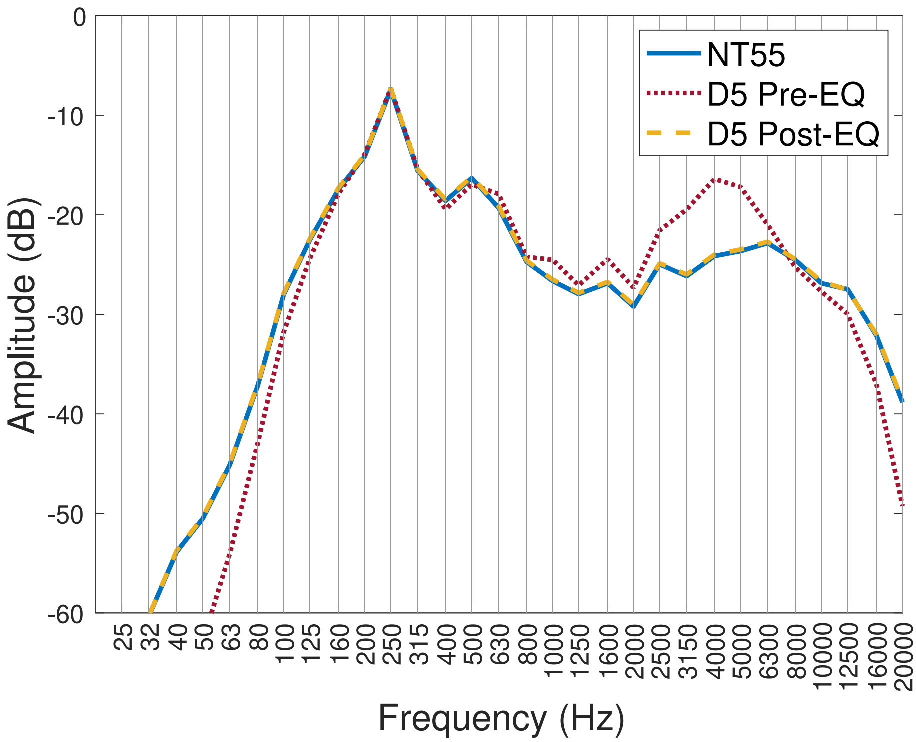

Figure 5.13: Spectral difference between NT55 and D5 recordings of the Velvetone snare pre- and post-EQ.

5.8 Pre- and Post-EQ AB Tests

To assess the success of the spectral modification in reducing the bias towards Category-1 microphones, a comparison is made between pre- and post-EQ using an AB listening tests with expert listeners all of whom had over seven years of music production experience.

5.8.1 Methodology

A pre-EQ AB test compares the *unmodified* D5 recording and those of the four Category-1 microphones, while a post-EQ AB test compares the *modified* D5 recording to those of the four Category-1 microphones. These tests use the same listening environment and equipment as in Section 5.6.1 and are conducted with the WAET software in an AB comparison configuration. In both tests, the sample order is randomised and repeated in 10 trials for each microphone pair, resulting in a total of 40 AB comparisons. As it was found that microphones did not exhibit snare-dependent results (Section 5.6.2), only recordings from the Velvetone snare drum are used. The pre-EQ AB test was completed by 10 participants (range: 24 to 55 years, mean: 35 years, std: 10.4 years) with at least 8 years experience (*range*: 8 to 29 years, *mean*: 16.7 years, *std*: 9 years). The post-EQ AB test was completed by 10 participants (*range*: 20 to 49 years, *mean*: 29.4 years, *std*: 9.5 years) with at least 7 years experience (*range*: 7 to 27 years, *mean*: 13 years, *std*: 6.7 years).

5.8.2 Results

Observations of consistency in selection for both the pre- and post-EQ AB listening tests were possible due to participants repeating each pairwise comparison for 10 trials. The binomial test is first used to check for consistency of each participant's scoring, with the null hypothesis being that two categories are equally likely to occur, such as in a fair coin flip where the outcome will be close to 50%. Due to the small trial size, only participants who preferred the same sample for $>$80% of the repeated trials can be considered to produce statistically consistent results. For any ratings lower than 80%, i.e., 6/10 and 7/10, the preference would statistically be no better than random chance. For the pre-EQ AB test, all participants were consistent in their preference for Category-1 microphones; however, for the post-EQ AB test, participants were rarely consistent. No participant was consistent for the 4011A, and only two participants had a consistent preference for the KM184 over the *modified* D5. In the M201 comparison, one participant had a consistent preference for the *modified* D5, and one participant had a consistent preference for the NT55.

The binomial test is once again used to interpret observations of preference across all participants. The upper plot in Figure 5.14 depicts the results of the pre-EQ AB test with 95% confidence intervals, indicating a very apparent statistically significant preference for all of the Category-1 microphones over the D5. The lower plot of Figure 5.14 shows that there is no significant preference for any microphone in the post-EQ AB test with all microphones scoring no better than random chance. This indicates that participants showed no preference between Category-1 microphones and the *modified* D5, suggesting that matching frequency responses of recordings through equalisation is an effective approach to improve the preference of recordings made with a less-preferred microphone. The preference for the recordings of the *unmodified* D5 over the Category-1 microphones is: M201 — 2%, NT55 — 4%, KM184 — 3%, 4011A — 2%, all significantly worse ratings. The preference for the *modified* D5 over the Category-1 microphones is: M201 — 54%, NT55 — 57%, KM184 — 43%, 4011A — 48%, which were neither statistically better or worse.

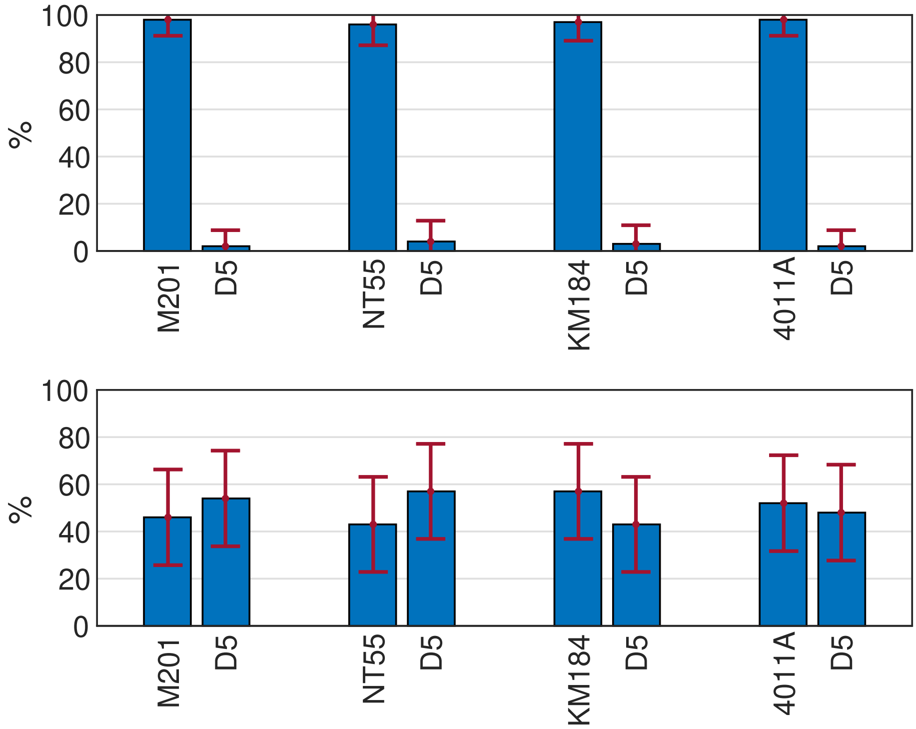

Figure 5.14: Results from AB test with 95% confidence interval. Top plot shows a comparison between the *unmodified* D5 and Category-1 microphone recordings; Bottom plot shows *modified* D5 and Category-1 microphone recordings.

5.9 Conclusions

This chapter first details the audio equaliser for its spectral shaping ability, this is followed by specific EQ techniques, presented by several authors, for snare drum enhancement in the context of mixing a multi-track recording project. They suggest specific frequency ranges that when modified aim to improve subjective quality and eliminate technical flaws. As different microphones are known to change the timbre of recordings due to their individual frequency responses, it was propose that EQ could be utilised in a non-traditional manner to spectrally transform recordings from a microphone with a poor subjective preference rating to mimic the timbral properties of recordings from more highly-rated microphones. An investigation was carried out to determine the feasibility of such transformations. Initially, a multi-stimuli listening test was carried out to categorise highly-preferred and least-preferred microphones for recording snare drums. Multiple timbrally distinct snare drums were used, and the results of this revealed that for 11 of the 12 microphones, listener preference did not significantly change between the four snare drums. From the results, classification of least-preferred and highly-preferred microphones was determined. Recordings from one of the least-preferred microphones was equalised to have the same frequency characteristics as that of the top four highly-preferred microphones. Two AB listening tests were then carried out to compare preference of pre- and post-transformation. Listener consistency was observed across repeated comparisons in both tests, which revealed that very few participants could choose the same sample consistently in the post-EQ AB test. This demonstrated a trend for no preference between the recordings of the highly-preferred microphones and those of the modified least-preferred microphone.

These findings suggest that frequency response is the biggest variable between microphones and is predominately responsible for the preference of the resultant recordings. Although there exists other properties of microphones that may be responsible for listener preference, these seemed to be negligible in the context of snare drum recording. When frequency characteristics of the least-preferred microphone was matched to the highly-preferred microphones, listeners had no preference for either recording. These results are promising as it illustrates that the best microphone may not be required when recording snare drum, as the audio can potentially be modified to emulate a more desirable microphone during the mixing stage. All the microphones used had polar pattern responses that were either cardioid, hypercardioid, or supercardioid, which are common choices for snare drum recording as it is optimal to minimise bleed from the other elements of the drum kit. Therefore, omni-directional and figure-8 pick up types are seldom used for this application. The transformations applied to the AKG D5 may have not performed so favourably if there was the added complexity of attempting to emulate microphones with vastly different polar patterns. In addition, microphones exhibit non-linear characteristics which may also influence preference. As the snare strikes used were of a consistent velocity, the variation of harmonics generated by excessive volumes was not a factor that the participants of the listening tests were able to assess. Had a range of striking velocities been used perhaps microphones with more favourable distortions characteristics would have been highly preferred. The microphones were also all positioned in an identical position for each of the recordings, whilst this minimised variations between microphones associated with its position related to the drum head, it is possible that some microphones may have performed better being closer or further back, positioned in a theoretical optimal position. While this may have implications in the classification order for some of the microphones, it does not invalidate the transformation process, as it was shown that regardless of which preferred microphones the AKG D5 was emulating there was no difference in outcome.

Towards an improvement of the spectral transformation, finer adjustment of spectral characteristics could be achieved through a higher-resolution EQ, and formalised testing would validate the suitability of the modification process across a range of microphones and stimuli. Additionally, transformations could be applied in the context of a full drum kit recording, to determine how the presence of bleed may affect the spectral feature mapping. A clear limitation of this transformation approach is the requirement of recordings from the target microphones. This issue could be somewhat negated if a data-set of ideal recordings existed that one could use to transform their recordings if particular microphones were unavailable. This chapter specifically looked into the use of EQ to retroactively modify the real-world recording parameter that is microphone selection. In the following chapter, two other modifiable real-world recording parameters will be emulated using a range digital audio effects. The two parameters under investigations are dampening amount and microphone placement, a deep learning approach will be investigated for this application.

Chapter 6

Deep Audio FX for Snare Drum Recording Transformations

6.1 Introduction

In Chapter 5 the spectral characteristics of snare drum recordings were modified through the use of a graphic EQ to transform a lesser-preferred microphone to be comparable to a highly-preferred microphone as judged by human listeners. Although successful, a draw back of this approach was the requirement of the target audio. In this chapter, two types of recording parameter transformations are attempted, dampening amount and microphone placement. As discussed in Chapter 2 and 3, these are two variables that are commonly adjusted during the recording stage, for which varied parameterisation will result in specific timbral modification of the recordings. This chapter explores training a deep neural network with a large dataset of source-target pairs, in order to investigate if a transformation can be carried out to elicit a perceptual change that mimics a real world alteration. These transformations are then evaluated by listeners as well as by several objective metrics.

6.1.1 Background

During a recording session, the positioning and recording of a standard acoustic drum kit—comprising of kick, snare, toms, and an assortment of hi-hats and other cymbals, is a technical and time-consuming endeavour. Recording drums may account for as much as 25% of the whole recording project (Toulson, 2009). During a typical session, an engineer must modify a large number of recording parameters to achieve a desired result. Key considerations include the selection of drums, drumheads, tuning, dampening, room and in-room placement, and the selection, arrangement and positioning of microphones. These decisions impact the overall timbral quality of a recording, with certain modifications producing greater effects than others (Bartlett and Bartlett, 2009; Owsinski and Moody, 2009).

Snare batter head dampening is a common practice in drum recording (Seymour, 2010; D'Virgilio, 2014), which involves adding mass to the drumhead to remove unwanted overtones and shorten decay time to produce a perceptually tighter, more controlled sound. Further, it reduces high frequencies, removing unwanted overtones that may prove difficult during the mixing stage (Major, 2014; Parsons and Van Horn, 1996). Once dampening has been applied, those timbral properties are then committed to the recording, and one loses the ability to apply additional dampening if later required, or to remove any if too much was used. Time permitting, an engineer may test different parameter options to identify an appropriate configuration for a song before committing to the final recording; however, with many variables this can easily become a lengthy process. As such, the ability to perceptually modify these recording parameters in a post-recording process would be of great benefit to engineers limited by time or equipment, especially during sessions where compromises may need to be made.

Several methods for the automatic mixing of drums have been proposed (Yoshii et al., 2005; Moffat and Sandler, 2019; Martínez-Ramírez et al., 2021a). Although these focus on emulating processes of the digital mixing stage, the proposed system in this chapter attempts to emulate techniques that are carried out prior to the recording stage. Two notable techniques an engineer can use to modify snare drum timbre include treating the drum heads directly through dampening, or by varying the position of the microphones around the drum in order to emphasise or subdue certain timbral characteristics. Audio effects (Fx) are an integral part of the music production workflow that can be used to modify sound characteristics such as dynamics, frequency, and timbre. Utilising audio effects for a predefined audio transformation can be a laborious task that often requires mastery over a large number of parameters. As a result, there has been an increasing focus on audio effect modeling and intelligent audio effects within the field of music information retrieval (MIR).

6.1.2 Motivation

Martínez-Ramírez et al. (2021b) developed a system that facilitates training of audio plugin parameters or a series of plugins for any desired transformation given the appropriate training data. In this chapter, the ability to modify the timbre of an undampened snare recording in order to elicit a perceptual change that corresponds to that of a dampened snare, referred to as Undampened-to-Dampened (U2D), will be explored through the use of multiple audio effects by utilising the tools presented by Martínez-Ramírez et al. (2021b). The inverse transformation is also examined, whereby a dampened snare recording is modified to perceptually emulate qualities of an undampened snare recording, referred to as Dampened-to-Undampened (D2U). In addition to these dampening transformations, two positional recording parameter changes are explored: Bottom-to-Top (B2T) microphone position as well as Top-to-Bottom (T2B). The purpose of these transformations is to be able to change or modify recording parameters only available for consideration in the real world prior to any recordings taking place. Alteration of the timbre to mimic physical changes to the recording configuration would allow recording engineers to retroactively manipulate these recording variables for creative or technical reasons, and to overcome time or resource limitations during a recording session, essentially virtually applying or removing dampening, and re-positioning the microphones of the recordings. These two real-world modifiable recording parameters were specifically chosen as they are well known to produce distinct timbral differences and are frequently altered by recording engineers to achieve desired characteristics. Dampening is used for addressing technical issues associated with unwanted resonant frequencies, but is also used as a creative tool to shape the volume envelope and vary the amount of high-frequency energy produced by the snare drum. Bottom-to-Top and Top-to-Bottom modification were chosen as these two positions are the most disparate and therefore most dissimilar in timbral quality. If this transformation could be accurately mimicked, it would be reasonable to assume that more subtle changes to positions could also be emulated.

The rest of this chapter is structured as follows: section 6.2 discusses the methodology used, including the network architecture and network training, as well the audio effects chosen and their traditional applications. Section 6.3 covers the creation of the Snare Drum Data-Set which was used for training the system, it details the recording techniques used, including information about the microphones, their positioning, the specifications of the snare drums, and the types of dampening, as well as how specific sub-sets were created specifically for training purposes. Following on from this, Section 6.4 is segmented into two parts including the subjective evaluation which describes the listening test used, and the timbral reconstruction metrics used

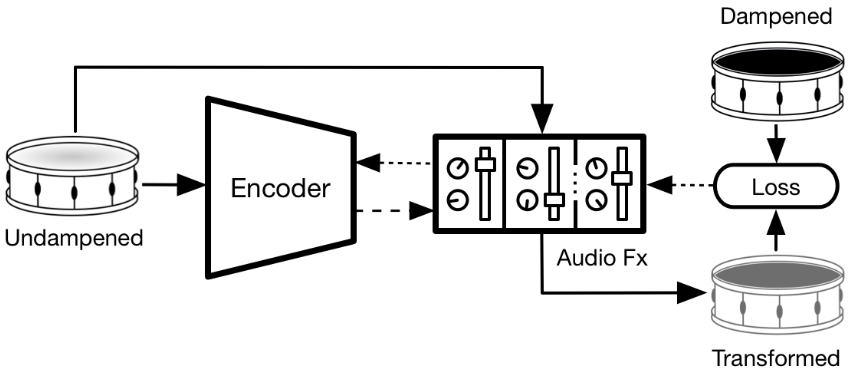

Figure 6.1: System overview for snare dampening with DeepAFx with third-party audio effect. Solid lines depict flow of audio, the longer dashed line represents the predicted parameter values and shorter dashed lines depict gradient flow.

for objective evaluation. The results are presented in Section 6.5, broken down into subjective and objective results. A discussion of the investigation takes place in Section 6.6 and finally the chapter is concluded in Section 6.7.

6.2 Methodology

In order to automatically carry out different perceptual transformations, DeepAFx (Martínez-Ramírez et al., 2021b) is utilised for its powerful parameter learning and audio processing capabilities. An overview of the system configuration for transforming an undampended snare drum into a dampened snare is provided in Figure 6.1. DeepAFx consists of a deep encoder that first analyses the input audio and then predicts the optimum parameters of one or more effect, these effects then process the audio, producing the desired transformed output audio. It does this by comparing its transformed, predicted output to that of the desired target output, and updating the parameters based on its performance. Once it has decreased the similarity between its transformation and the target audio to a point where it can longer improve, it can produce transformations of audio that is has not been previously trained on, i.e., audio samples where no desired target is required. The system makes use of the LV2 audio plugin open standard,[1] and incorporates third-party audio effects as a black box layer within a deep neural network.

6.2.1 Network Architecture

Following Martínez-Ramírez et al. (2021b), an inception-based encoder network (Lee et al., 2020) is implemented to predict the audio effect parameter values required for a desired snare drum transformation. The input to the network is a log-scaled Mel-spectrogram represented as a four-dimensional tensor $t \in \mathbb{R}^{b \times w \times h \times c}$, with batch size b, number of frames w, number of frequency bins h, and channels c. The batch in this case is a collection of different snare drum recordings. The model consists of 64 convolutional filters with a 5×5 sized kernel followed by 2×2 strided max-pooling. A stride of 2 was used and is the distance that the kernel moves over the input matrix. Pooling is used to reduce the computational cost by reducing the size of tensor, therefore, reducing the number of parameters and computations required in the network. This also helps to make the network more generic because it combines several pixel values into a single one. Strided max pooling refers to the filters moving across the input and selecting the pixel within 2×2 matrix that has the maximum value and sending this to an output array. Rectified Linear Unit (ReLU) activations, which

[1] `https://lv2plug.in/`

removes negative values setting them to 0, are used for all layers apart from the network's last layer, which is a fully-connected output layer consisting of r output nodes and a sigmoid activation function where r is the number of parameters associated with a particular audio effect. The Sigmoid functions scales the outputs to between 0 and 1 in order to directly control the corresponding parameters of the audio effect. The network outputs estimate audio effect parameter values for each snare drum transformation under observation. Instead of having a decoder that reconstructs the audio, the decoded output is the audio processed through the audio effects with the new estimated parameters values applied.

6.2.2 Audio Effects

For this study, two novel LV2 audio effects are specifically developed to take advantage of DeepAFx's multiple parameter learning abilities, both effects have high parameter count that would make it tedious and time-consuming for a human engineer to fine tune each control. Typically, audio production tools are designed with the audio engineer in mind. Graphic user interfaces (GUI) are implemented, and variables such as parameter amount, layout, size, and colour are considered in order to enhance the experience of the user. By allowing DeepAFx to learn the parameters of an audio effect a GUI is not required, nor are any considerations to the impracticality to a human user.

The two novel audio effects investigated for their timbre transforming abilities are a 10-band dynamic EQ (DEQ10) and a 30-band dynamic EQ (DEQ30). Dynamic EQ is similar to parametric EQ however, the different frequency bands are altered dynamically as the energy in those bands surpass a specified threshold. An example of a digital dynamic EQ can be seen in Figure 6.2. Similar to a dynamic range compressor, dynamic EQ also features threshold, attack, and release parameters in order to modify the ballistic characteristic of the amount of gain applied (Fox, 2022). Each band of the traditional EQ has a fixed gain value, but with dynamic EQ, the gain changes according to the intensity of the input signal. When the incoming signal goes above the threshold, the dynamics portion of the EQ is engaged. A dynamic EQ can behave like a compressor, attenuating the signal level of the frequency band selected, it can also act similar to an expander, amplifying that specific band. Dynamic EQ feature many of the same parameters as a standard parametric EQ, it has various filter types, such as bell and shelf shapes, and allows the user to adjust the frequency, slope and Q values of the filters (Hahn, 2020). Dynamic EQ is typically considered to offer detailed control over specific frequencies unlike multi-band compression which will cover a much broader range of the spectrum. Due to its ability to address problematic frequencies, it is used for mixing when a high degree of precision is required and is not achievable with a traditional static EQ. They are commonly used to control or enhance the fundamental frequency of a snare, subdue an overly emphasised bass note, suppress low-mid range vocal resonance, eliminate excessive sibilance on bright vocal tracks, and to reduce harsh sounding high frequencies from hi-hats (Stewart, 2021).

While similar to dynamic EQ, multi-band compression has several distinct differences that influences its application and outcome. Brown (2020) details how both effects process the audio signal, as well as their uses for mixing. The use of dynamic EQ is recommended when transparent corrective alteration is required, while multi-band compression generally imparts a more pronounced timbral change on the character of the sound. Dynamic EQ does not divide a signal's spectrum into individual frequency bands, and therefore does not impact the signal until its amplitude crosses that band's threshold, this lowers the chance of introducing unwanted artifacts. It is mostly used in place of a standard, static EQ for balancing a solo instrument or adjusting a full mix. Although EQ can also be used to attenuate problem frequencies, the dynamic EQ allows the frequencies to be controlled only when appropriate and required, rather than simply eliminating those frequencies all

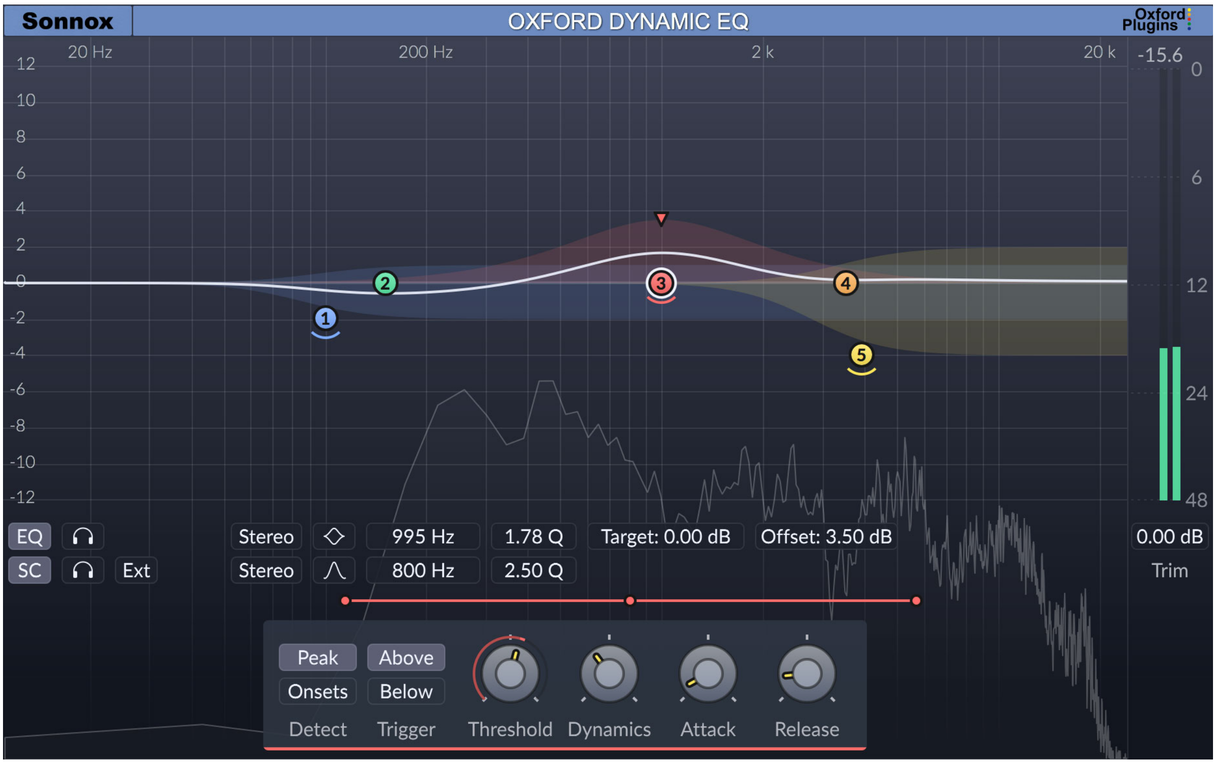

Figure 6.2: Example of a five-band dynamic EQ. (Courtesy Sonnox).

together. Dynamic EQ affects signal proportionally according to how far it crosses the threshold. When the signal's level falls below the threshold, the EQ band returns to its default position based on the release time. Proportional attenuation of this kind would be very difficult to replicate by using automation with a static EQ. This can be useful when dealing with recordings of instruments in unideal acoustic environments, as certain notes or parts of the performance may have distracting and uneven resonances.

Digital multi-band compressors are designed to function as their analog counterparts; the full signal's frequency content is divided into separate bands using several crossover filters, typically either three or four (Brown, 2020). A digital four-band multi-band compressor can be seen in Figure 6.3. When the audio signal's frequency spectrum is divided with band-pass filters, such as in a multi-band compressor, a phase shift occurs at the crossovers of those filters. Conversely, the points where the filters overlap within a dynamic EQ will remain free of phase distortion when there is no control signal triggering the parametric EQ band (Hoffman, 2022). Brown (2020) elaborates that once the spectrum has been divided, each frequency band is then passed to its own dedicated compressor. Signal within a band may cross the band's threshold, causing its compressor to engage and signal within the band to be attenuated. The outputs from the individual compressors will then be summed together at the output. Dynamic EQs tend to have more available bands than multi-band compressors, often four to six, sometimes as high as eight. As a result they can affect a much narrower range, allowing the user to control specific frequencies more easily compared to a multi-band compressor. If one desires to control larger areas of the frequency spectrum, a multi-band compressor would be ideal for this application, however if a small range, or a very specific frequency needs to be attenuated, a dynamic EQ would be the preferred choice.

Unlike traditional dynamic EQs that typically consist of four to eight parametric frequency bands which allow the user to specify centre frequency, Q-factor, and between shelf or bell filter types, DEQ10 and DEQ30 are implemented as fixed-band graphic equalisers, with fixed centre frequencies based on the specification for

Figure 6.3: Example of a digital four-band multi-band compressor. (Courtesy Steinberg).

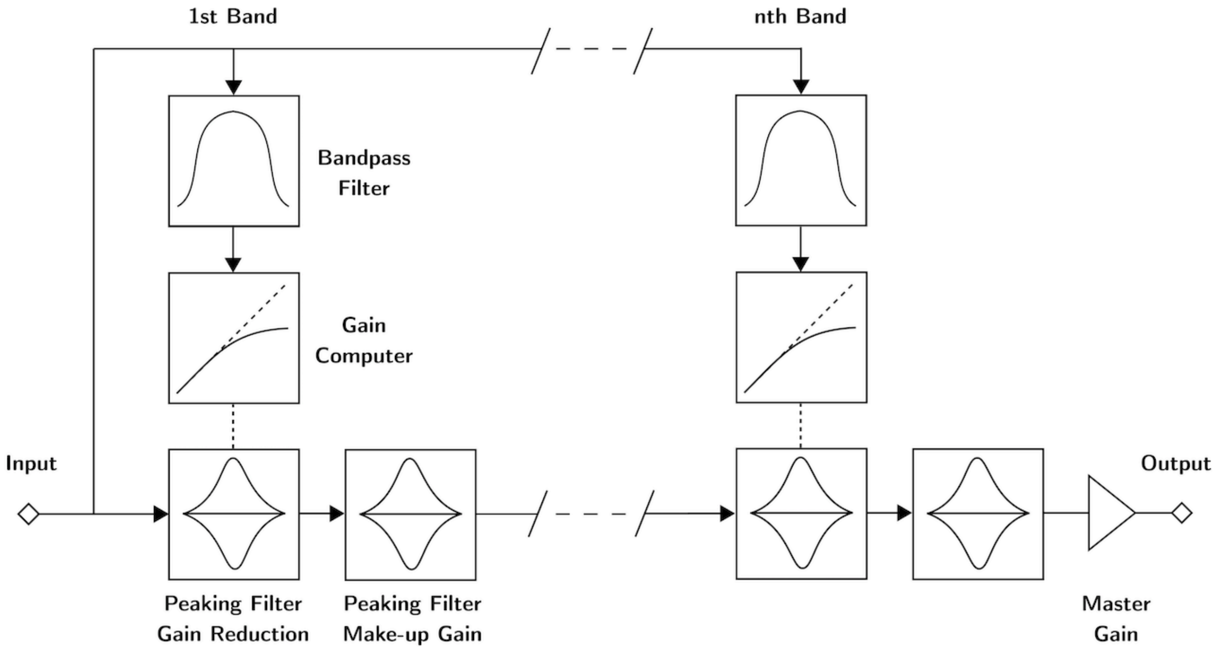

Figure 6.4: Architecture of DEQ10 and DEQ30 audio effects.

octave-bands and fractional-octave-bands described in ANSI (2009). This allows for complete dynamic control over the full frequency spectrum. Dynamic EQ was specifically chosen in order to provide both spectral and temporal manipulation within one audio effect (Wise, 2009), often used in mastering application for correcting time varying frequency imbalances (Izhaki, 2017). The ability to control specified frequency bands over time lends itself to transformations where some frequencies may be similar where others are disparate, such as in the case of dampening a snare; where high frequencies are both attenuated and their associated envelopes shortened, while the lower frequencies remain mostly unaffected. This would be difficult to achieve through the use of a standard full spectrum compressor, thus dynamic EQ having the potential to perform better than a standard EQ and compressor combined for particular production tasks. Both DEQ10 and DEQ30 have the same architecture, the signal path consisting of cascaded biquad peaking filters. Each frequency band comprises of two such filters, the gain of the first being controlled dynamically and that of the second through the *make-up gain* parameter for that band. Dynamic control of each band is achieved through a standard feed-forward compressor architecture. Within the side chain for each band the signal first passes through a biquad band-pass filter, with centre frequency and bandwidth matching that of the corresponding peaking filter in the signal path. Level detection and ballistics are carried out within the gain computer of the compressor's side chain. The output of this filter undergoes peak amplitude detection and then feeds a gain computer with the following parameters: *threshold, attack, release, ratio,* and *knee*. Each effect has a *master gain* parameter at the end of the signal path. A graphical representation of this architecture is given in Figure 6.4. The principal difference between DEQ10 and DEQ30 is that the first uses an octave band layout, while the second uses $\frac{1}{3}$-octave increments. With six parameters per band and a master gain, this gives 61 trainable parameters for DEQ10 and 181 for DEQ30.

In addition to the two novel effects, two open-source plugins were used.[2] Firstly, an eight-band parametric equaliser (PEQ) was chosen for its frequency sculpting ability and for the ubiquitous nature of parametric EQs in audio engineering. Secondly, because applying dampening to a snare drum alters its envelope characteristic, a transient designer (TD) was chosen as a possible candidate for a tool that might perform well at emulating

[2] `http://calf-studio-gear.org/`

this feature. A transient designer, also referred to as transient *shaper* (McAllister, 2022) provides level-independent processing of the signals envelope by using envelope followers to control output dynamics, this allows transients to be accelerated or slowed down and sustain to be prolonged or shortened (Gier and White, 1999). This tool is commonly used to emphasise the attack on snare and kick drums, helping it be heard in a dense track, without needing to use excessive EQ (Major, 2014). Not only can one accentuate and exaggerate the initial transient of drums, one can simultaneously extend the sustain of each strike, helping them to sound more *explosive* and *powerful* (Gibson, 2004). In general, transient designers enable sounds to have a sharper or softer attack, or a longer or shorter sustain and decay. This has the potential to make drums sound *bigger, bolder, stronger*, more *impactful*, more *exciting*, *hyper-realistic*, and fit the context of the song more appropriately (Toulson, 2021). Additionally, by shortening the sustain of drum strikes it is possible to minimise microphone bleed or reduce the amount of room tone captured on the recordings. This then allows the mixing engineer to use an artificial reverb to specify the desired amount of ambience they require (Owsinski and Moody, 2009). DeepAFx also has the ability to train multiple plugins in series. Chaining multiple effects together is a common practice among mixing engineers (Owsinski, 2017), and for this reason this aspect was also investigated. The parametric equaliser and transient designer (PEQ+TD) were used in conjunction with one another to determine if they were able to perform better together, providing both spectral and temporal manipulations. The order of PEQ and TD were tested in both configurations, placing TD before PEQ as well as after. This was found to have very little audible difference on the processed audio, therefore only the PEQ+TD configuration was chosen for investigation.

6.2.3 Network Training

The deep encoder takes as input data x and outputs parameters λ. Audio is pre-processed through resampling and conversion to a spectrogram representation. Following Martínez-Ramírez et al. (2021b), the short-time Fourier transform (STFT) of each snare drum recording is calculated using a Hanning window with a size of 1024 samples and a hop size of 256 samples to facilitate the desired temporal resolution of the network input. The magnitudes of STFT are transformed to log-scaled Mel-spectrograms with 128 Mel-frequency bands.

The model is trained using the Adam optimiser (Kingma and Ba, 2015). This method computes individual adaptive learning rates for different parameters; the learning rate is set to $1e$–4, where each iteration takes a mini-batch of 100 examples. Network weights are initialised using He's constant, as opposed to random weights drawn from Gaussian distributions with fixed standard deviations. This method allows for extremely deep models to converge, as bad initialisation can hinder learning (He et al., 2015). Once model performance ceases to improve over 25 epochs, early stopping is applied to complete training, and the epoch that achieves the best accuracy on the validation set is used for testing. Training was carried out on a Nvidia TESLA M40.

6.3 Snare Drum Data-Set (SDDS)

In this section the creation of the Snare Drum Data-Set (SDDS) is presented. SDDS is a comprehensive acoustic snare drum dataset, comprised of 212,795 multi-velocity recordings of 10 tonally distinct snare drums over 48 performances, using 53 professional microphones. Real-world scenarios were mimicked, such as using various commercial dampening techniques, whilst maintaining consistency over several variables, i.e., the recording space, location of microphones, and using the same drummer for all performances. The dataset's intended use is for machine learning and MIR tasks as well as providing insight into timbral changes that occur when snare type, dampening method, microphone model, and placement are altered.

6.3.1 Background

For various machine learning tasks, such as automatic drum transcription (ADT) and instrument or playing style classification, large datasets containing multiple varied examples of the problem can help avoid overfitting. These multiple examples will be task specific, reflecting variations that will likely be encountered when attempting to solve real-world problems. A dataset designed for snare drum specific tasks should include a broad range of examples as a snare's timbral characteristics are heavily genre dependent. The timbral properties of a snare are affected by several factors, including shell material and dimension, head type and tuning, amount and method of dampening, amount of snare wires, hoop material, and playing method (i.e., sticks, rods, or brushes). Distinct timbral alterations can also be achieved through modification of the recording process, including adjustments of the acoustic environment, microphone make, model, and placement (Bartlett, 1981; Quiroga et al., 2015; Senior, 2018).

Tindale et al. (2004) used a feed-forward backpropagated neural network for realtime classification of different snare drum playing techniques based on their spectral properties. In total 20 examples of 5 different playing techniques were recorded using a single Neumann U-87. Although this addressed a specific classification problem, similar research requiring varied timbres of acoustic drum samples would benefit from SDDS which includes multiple microphones and several timbraly distinct snare drums.

Several datasets exist containing acoustic drum samples, however very few offer extensive timbral variations. MDB Drums (Southall et al., 2017), a subset of MedleyDB (Bittner et al., 2014) was constructed for ADT, it consists of drum annotations and audio files for 23 tracks from various genres. IDMT-SMT-Drums (Dittmar and Gärtner, 2014) also used for ADT and source separation, is comprised of 104 polyphonic drum loops containing kick, snare, and hi-hat. Alongside acoustic drums it includes synthesised drums and loops created from sampled drums, eliminating velocity fluctuations and the recording process entirely. The ENST drum dataset (Gillet and Richard, 2006) is an audio-visual database for signal processing and ADT. It contains annotated drum recordings from 3 drummers with 3 different drum kits. Although recorded on 8 audio channels, only one close microphone was used to capture the snare. Due to the intended use of these datasets, using a range of snare recording techniques was not prioritised. Limitations for timbral analysis include sparse or absent metadata of microphones, positions, and snare drums, as well as a lack of multiple recordings of snares captured by several microphones simultaneously, preventing comparisons of identical strikes.

Commercially available drum sample software such as FXpansion's [3] BFD3 and Toontrack's [4] Superior Drummer 3, offer varied velocity recordings of acoustic drum kits. Designed to emulate a live studio drummer, they are not ideal for research application due to a traditional microphone set-up that utilises only a few close microphones, as well as their optimised and enhanced timbral properties. A sample library recorded with the extensive selection of microphones used by SDDS has yet to be found. Superior Drummer 3 utilises a very minimal, 3 close microphone configuration for the capture of the snare drum. FXpasion offers several snare dedicated expansions to the software, using a variety of different make and material of snare drum. However, recordings use traditional microphone set-up, utilising only a few close microphones.

6.3.2 Recording Configuration

The recordings were captured in an acoustically treated live room at 44.1kHz, 16 bit. Out of all 53 microphones, 24 were recorded through a SSL AWS 924 mixing console, 15 through a MIDAS M32R console, and a further 14 channels were connected to the MIDAS via a Behringer S16 digital stage box, see Table 6.1. All equaliser

[3] www.fxpansion.com
[4] www.toontrack.com

and dynamic processing effects on the mixing consoles were disengaged. Gain was set so that the loudest strike produced would not cause clipping. Originally 3 additional channels were recorded, however due to poor signal quality from faulty cables, the recordings from these 3 channels were discarded.

The recordings were undertaken using 53 microphones (32 condenser, 18 dynamic, and 3 ribbon), see Table 6.1, comprised of microphones used in previous chapters and various sources (Elliott, 2014; Fuston, 2017a). Built in high pass filters or frequency emphasis selectors, such as those found in the Shure SM7B and Electro Voice RE20, were switched off on all microphones that featured them. Microphones with variable polar pattern were all set to cardioid. During recording 35 close, 14 overhead (OH), and 4 room positions were utilised. The close and overhead microphone placements used for recording the dataset are shown in Figure 6.5. The close positions are further subdivide into 17 Top, 9 Shell, and 9 Bottom. The 3 close positions can be seen in Figure 6.6, every microphone's position is listed in Table 6.1. Multiple positions were chosen to emulate several real world recording techniques. The dataset can be easily augmented by combining multiple microphone positions together, which is a common mixing technique used to emphasise desired timbral attributes that can not be achieved through a single microphone (Albano, 2016).

In total 10 snare drums of varied dimensions, shell material, and head type were chosen to represent a broad range of timbral properties. The batter heads were tuned to different fundamental frequencies to further emphasise timbral variation, whilst maintaining traditional characteristics (i.e., not extremely low or high). A digital drum dial was used to ensure even tension across all lugs for both batter and resonant heads. Table 6.2 shows specifications of the snares used. Eight of the drums had a diameter of 14", whilst 2 drums were chosen due to there less common diameters of 13" and 15". While 15" snare drums are commercially available, the drum used in this study was custom built made from mixed woods. The snare wires for all snare drums were fitted and tightened such to minimise any rattling and buzzing issues.

Brand	Material	Dimensions	Lug Amt.	Batter Head	Resonant Head
Gretsch	Birch	14 x 5.5"	8	Remo Emperor Smooth White	Remo Weatherking Ambassador
Mapex	Maple	13 x 6.0"	8	Mapex Remo UX Coated	Mapex Remo UX Resonant
Mapex	Steel	14 x 6.5"	10	Evans Hydraulic	Remo Weatherking Ambassador
Mapex	Walnut	14 x 5.5"	10	Evans Hydraulic	Remo Weatherking Ambassador
Premier	Maple	14 x 5.5"	10	Evans Level 360 HD Dry	Remo Weatherking Ambassador
Tama	Steel	14 x 5.5"	10	Evans Hydraulic	Remo Ambassador Black Suede
Tama	Steel	14 x 6.5"	8	Remo Ambassador X	Remo Weatherking Ambassador
Yamaha	Birch	14 x 5.5"	8	Remo Emperor Smooth White	Tama 200 Hazy Snare Side
Yamaha	Maple	14 x 6.5	8	Remo Emperor Smooth White	Evans Level 360 Snare Side 300
n/a	Mixed	15 x 8.0"	8	Remo Emperor Smooth White	Remo Weatherking Ambassador

Table 6.2: Specifications of all snare drums used.

6.3.3 Dampening

As discussed in Chapter 2, it is a common practice to dampen the batter head in order to modify the temporal and spectral components of a snare drum strike. For SDDS, 4 different dampening products were employed to reduced unwanted overtones and shorten sustain by varying amounts. Firstly, *Moon Gel*, a self-adhesive gel rectangle (3.5cm x 2.5cm), dampened the least of the 4 methods. Secondly, an *Evans E-ring 1"* dampened more so than *Moon Gel*. Thirdly, an *Evans E-ring 2"* which covered more area and reduced overtones to a greater degree than the 1" version, and lastly a *Big Fat Snare Drum* (BFSD) which covered the entirety of the batter head, producing the greatest dampening effect. The 14" versions of the E-Rings and BFSD were used for all 14" drums as well as for the 15" snare. The 13" snare was only dampened using *Moon Gel* and a

SSL AWS 924							
Brand	**Model**	**Type**	**Position**	**Brand**	**Model**	**Type**	**Position**
AKG	C414	C	OH	Electro Voice	RE20	D	Top
AKG	C414	C	Shell	Neumann	KM184	C	OH
AKG	C414	C	Top	Neumann	KM184	C	Top
AKG	C451B	C	OH	RØDE	NT55	C	Top
AKG	C451B	C	Top	RØDE	NTG2	C	OH
Audix	i5	D	Top	Royer	R-121	R	OH
Beyerdynamic	M201	D	Top	Royer	R-121	R	Shell
Coles	4038	R	OH	Sennheiser	MD421	D	Top
DPA	4090	C	OH	Sennheiser	e614	D	Top
DPA	4090	C	Top	Shure	Beta57A	D	Top
DPA	4099	C	Top	Shure	SM57	D	Top
Earthworks	M30	C	OH	Shure	SM7B	D	Top
Midas M32R				**Behringer S16**			
Brand	**Model**	**Type**	**Position**	**Brand**	**Model**	**Type**	**Position**
AKG	C414	C	Bottom	AKG	C451B	C	Bottom
AKG	C414	C	Room	AKG	C451B	C	Shell
Audix	D2	D	Top	AKG	D5	D	Bottom
Audix	ADX51	C	OH	Audix	ADX51	C	Shell
Brauner	Phantera	C	Shell	Audix	i5	D	Bottom
Lauten Audio	FC-357	C	Room	Audix	i5	D	Shell
Neumann	TLM 103	C	OH	DPA	4090	C	Bottom
RØDE	M3	C	OH	DPA	4099	C	Bottom
RØDE	M3	C	Top	DPA	4099	C	Shell
RØDE	NT1-A	C	OH	Electro Voice	RE20	D	Bottom
RØDE	NT5	C	OH	Sennheiser	MD421	D	Bottom
SE	2200	C	OH	Sennheiser	MD421	D	Shell
Sennheiser	e901	C	Room	Shure	SM57	D	Bottom
Shure	Beta91A	C	Room	Shure	SM57	D	Shell
Telefunken	M80	D	Top				

Table 6.1: All 53 microphones used, subdivided by preamp. Condenser (C), dynamic (D), ribbon (R).

Figure 6.5: Close and overhead microphone positions.

Figure 6.6: Top, bottom and shell microphone positions.

Snare	Un-dampened	Moon Gel	E-Ring 1"	E-Ring 2"	BFSD	Total
Gretsch	73	104	85	93	97	452
Mapex Maple	69	91	72	n/a	n/a	232
Mapex Steel	71	95	102	75	72	415
Mapex Walnut	105	79	120	91	84	479
Premier	76	97	98	98	100	469
Tama 5.5	85	110	93	92	85	465
Tama 6.5	62	87	93	79	85	406
Yamaha Birch	58	73	75	82	87	375
Yamaha Maple	50	74	87	96	93	400
Mixed Wood	49	68	74	52	79	322
Total	698	878	899	758	782	4,015

Table 6.3: Amount of strikes for each performance, showing total for both dampening method and snare.

13" version of the *Evans E-ring 1"*, whilst all other snares were dampened with the 4 products. Additionally, all snares were recorded without any dampening applied, resulting in 5 recording scenarios for 9 of the snares and 3 for the 13" drum.

6.3.4 Performance

One performance took place for each of the unique recording configurations (i.e., one snare drum with one dampening technique), resulting in 48 performances. The drummer was instructed to strike using the full range of velocity intensities they were capable of, to repeat the same velocity several times, and to allow each strike to ring out before playing the next. They were not limited to where on the drum head they could strike and could play with both hands. The length of the performances ranged from 171secs to 246secs (mean: 208secs), with a range of 69 to 120 (mean: 86) varied velocity strikes being played per performance. The amount of strikes for each performance can be seen in Table 6.3. The whole dataset features 4,015 unique strikes captured with 53 microphones, equating to 212,795 strikes from all audio files. A limitation of the methodology prevents direct comparison of velocities from different snares or dampening methods due to the sound pressure level (SPL) being dependent on those factors which vary between scenarios. However, the relational difference of velocity can be examined between strikes from the same performance. The range of velocities during each performance will be identical in all 53 microphones due to simultaneous recording, therefore allowing analysis of timbral variation that occurs as a result of velocity fluctuations.

6.3.5 SDDS Discussion and Reflections

SDDS allows researchers to group the recordings by snare drum, dampening method, microphone manufacturer, microphone model, type, or placement. The aim was to create diverse recordings that emulate real world recording scenarios and techniques, whilst also allowing for a level of scientific control. The absence of acoustic bleed facilitates investigation of the snare drum's timbral properties in isolation. As microphones can be an expensive commodity, the plethora of microphones were selected to uncover the expanse of characteristics associated with industry standard and specialised microphones for recording snare drums. Although it is not possible to directly compare microphone to microphone due to the difference in position in relation to the snare drum, which will also affect timbre, the audio files may serve as a reference for future research.

There are several potential research topics SDDS may be useful for. The dataset could aid in classification tasks for automatic mixing scenarios. By first classifying the microphone position, EQ or compression settings could then be applied automatically to suit the distinct timbral differences associated with top, bottom, and overhead location. Classifying snare microphones into condenser or dynamic would provide useful information to a mixing engineer when little or no metadata is provided, informing their mix decisions. Classifying based on velocity would facilitate sorting and searching sample libraries based on perceived velocity, or allow novel mixing tasks where high and low velocity strikes are processed independently. Timbral variation linked to velocity could be extracted and mapped onto programmed drum samples, creating a humanisation effect without a compromise in volume.

In addition to the intelligent music production applications, SDDS also provides recording engineers a resource to hear how various microphones perform across different snares, positions, and velocities. This will serve as a reference for microphone characteristics when access to specific models may be limited. The methodology presented here will allow others to emulate the recording configuration in order to compare timbral properties of other microphones, positions, or snares to those in this dataset.

6.3.6 Sub-Sets of SDDS

In order to train DeepAFx to learn the most suitable parameters for any given audio processing task it requires input-target paired audio as supervision. The training data for each of the four transformation tasks is comprised of specific sub-sets from SDDS. For the four sub-sets, 3000 input-target pairs were randomly selected to create the test set.

One of the dampening methods used in SDDS was a BFSD, a specialised device designed to dampen a snare or tom, placed directly on top of the batter head without needing any form of adhesive. This allows for exact repeatability as it covers the entirety of the drumhead and could only be placed in one position unlike other products. Although SDDS included other dampening methods such as MoonGel, BSFD was chosen to be used for the dampening transformations as it produces a distinct timbral change. The BFSD is also used for the D2U transformation. For each U2D and D2U input-target pair, the snare drum, microphone, and microphone position were all identical, with the only variable being the dampening, either undampened or dampened with a BFSD. Individual strikes from each pair were matched based on closest peak amplitude levels, and time aligned using cross correlation. The dataset included a 13" snare drum that was not recorded with the BFSD, this drum was excluded from the sub-set.

For the positional transformation of T2B and B2T, only eight of the same microphones were used in both top and bottom positions, these were: AKG C414, AKG C451, Audix i5, DPA 4090, DPA 4099, Electro Voice RE20, Sennheiser MD421, and the Shure SM57. These pairs were used on all 10 snare drums and for all dampening methods; the paired strikes were identical performances, as the top and bottom microphones were recorded simultaneously. For each subset, 80% was used for training, 10% for validation, and the remaining 10% for test data for later evaluation. Once processed by the trained models, the evaluation data was used for the comparative metrics, as well as providing stimuli for the subjective listening tests.

6.4 Evaluation Methodology

The system presented in Section 6.2 is assessed through two evaluations to determine: 1) perceptual quality of the transformations through a subjective listening test, and 2) the similarity of the transformed audio compared to the target audio through an objective evaluation using various comparative metrics. For each type of transformation under investigation, the unprocessed snare drums from the test dataset of input-target pairs are transformed using the proposed audio effect configurations, where parameter values for each audio effect are inferred from the trained encoder network. In this section the experimental methodologies for subjective and objective evaluations are presented.

6.4.1 Subjective Evaluation

A subjective listening test was carried out using a multiple stimulus approach in order to determine if participants would perceive the transformed audio as comparable to the real world recording parameter adjustments it was emulating. The test was implemented using the Web Audio Evaluation Tool (Jillings et al., 2015) and was carried out by 25 participants between the ages of 20–42 (*mean*: 27), and their experience in audio related fields ranged from 1 to 25 years (*mean*: 9). Participants were instructed to use the highest quality playback system available to them. Two participants reported using loudspeakers, whereas the other 23 reported using headphones. They were required to provide the specification of equipment used and all systems reported were deemed to be suitably professional.

The four transformations were evaluated on separated pages of the listening test. On each page, participants were presented with seven sliders, each corresponding to a different audio sample. For each participant the page and slider order were randomised, as well as slider starting position. The seven audio stimuli were comprised of the unprocessed input used as a baseline for similarity, with the target acting as a *hidden reference*, and the five samples of the input processed by the five different plugin chains. Participants were instructed to arrange stimuli based on their similarity to the *reference*, and to use the full range of the scale, placing the most similar at the top and the least similar at the bottom. The *hidden reference* was used to ensure participants could accurately identify the identical sample to the *reference*. No low anchor was used in order to allow participants to rate the perceptually least similar stimuli lowest on the rating scale. Stimuli were loudness normalised to -23LUFs. The testing interface can be seen in Figure 6.7, when playing a sample the corresponding slider would change colour to red. The input-target pairs for each transformation were randomly selected from the test data subset (Section 6.3). Participants could not move onto the next page until all stimuli, including the reference, were played at least once and all sliders were moved from initialised random positions. Figure 6.8 shows an example of (a) U2D snare transformation using DEQ10 and (b) D2U snare transformation using TD.

6.4.2 Timbral Reconstruction Metrics

In order to evaluate the ability of the model to produce desired transformations of snare recordings, we compare how accurately the transformed examples $\hat{S}$ match the target examples S. Recording pairs in the test set introduced in Section 6.3 are evaluated using reconstruction metrics. Each transformation type is grouped into two tasks: 1) dampening (i.e., U2D and D2U) and 2) positional (i.e., T2B and B2T) and is evaluated with a range of spectral representations and metrics focused on timbral reconstruction capabilities of the model. To extract the selected comparative metrics, a magnitude spectrogram S_{stft} is computed using the STFT for each file using an n-length Hann window ($n = 2048$), hop size of $\frac{n}{4}$. Additionally, S_{stft} is mapped onto the Mel-scale or converted to Mel-frequency cepstral coefficients (MFCCs), resulting in S_{Mel} and S_{mfcc}, respectively.

The timbral reconstruction metrics include multi-scale spectrogram loss (MSL) and spectral cosine distance (SCD), using the the implementation by Martínez-Ramírez et al. (2021b), where the former uses STFT magnitudes and latter uses 13 MFCCs. The log-spectral distance (LSD) (Bitton et al., 2019) and Pearson correlation (PC) coefficients are used as an objective measure of audio quality, both previously employed in evaluations of deep generative models for music signals (Kim et al., 2019; Tomczak et al., 2019). Additionally, the cosine similarity (CS) metric based on spectral difference functions (SDFs) used in research on automatic event detection (Bello

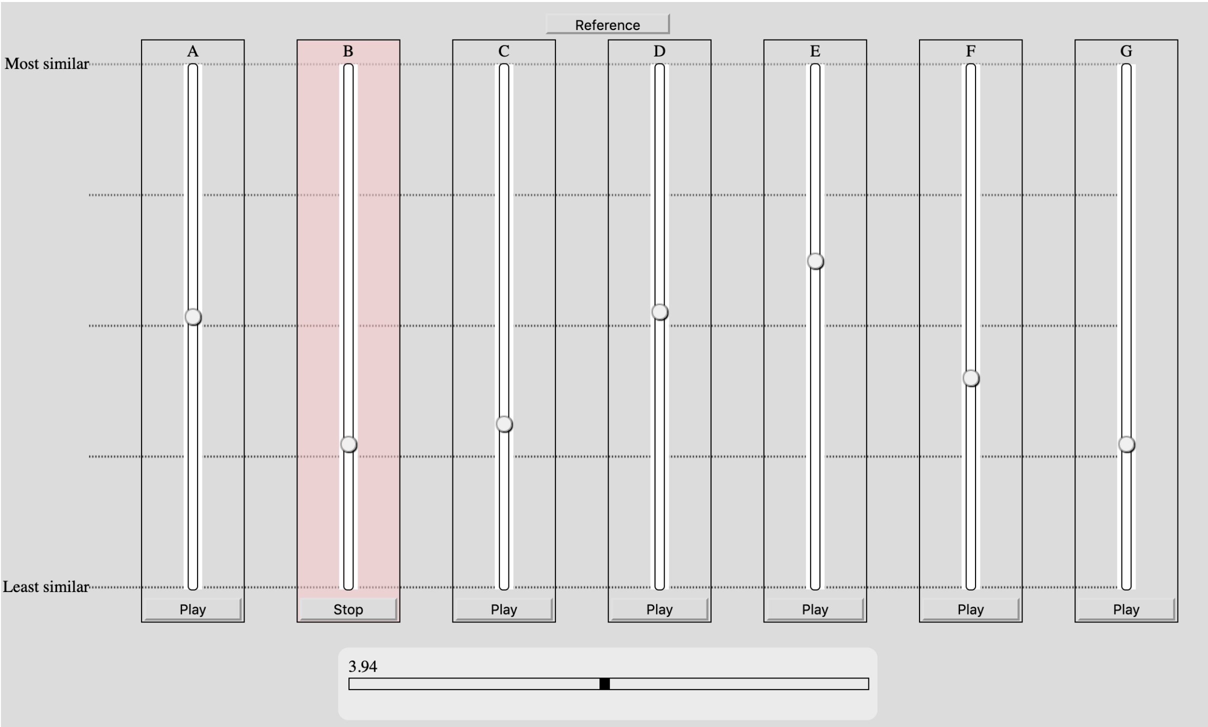

Figure 6.7: One page of the testing interface used for the subjective evaluation.

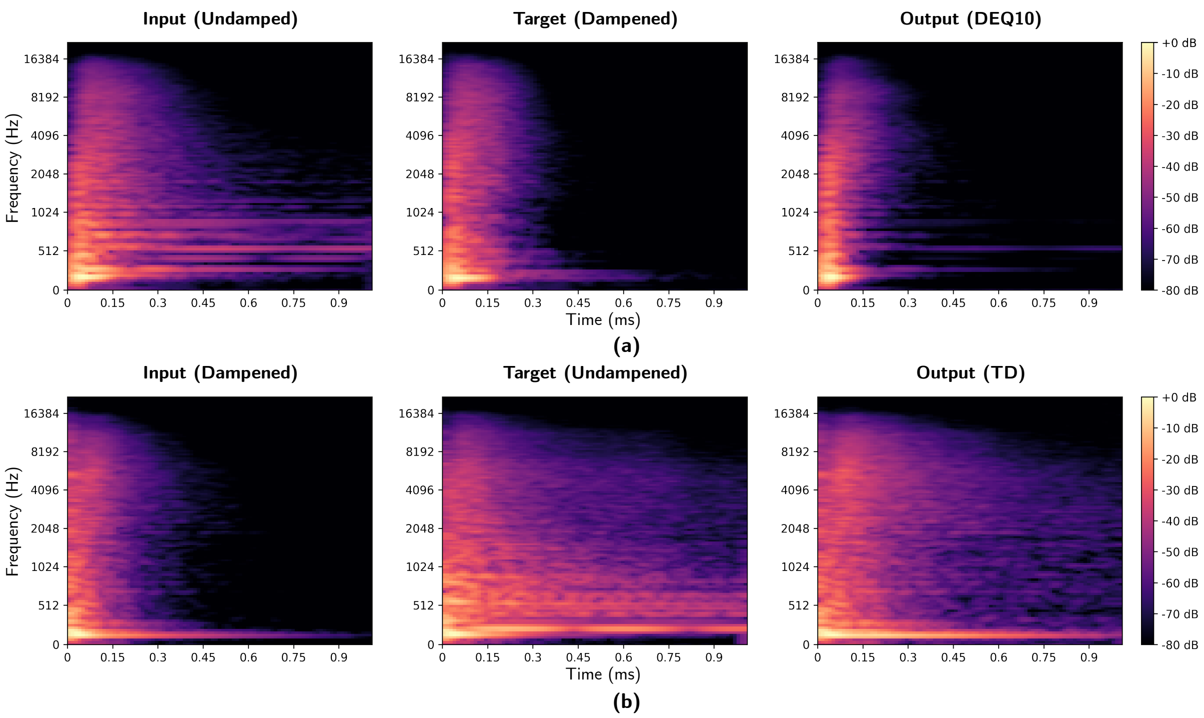

Figure 6.8: Mel-scaled log frequency spectrograms for (a) U2D with DEQ10 and (b) D2U with TD. Input snare drums (left), target (centre), output transformations (right).

et al., 2005) and automatic music remixing (Davies et al., 2014) is utilised for measuring temporal similarity between drum recordings.

The LSD is calculated using Mel-spectrograms as follows:

$$LSD_{Mel}(S, \hat{S}) = \sqrt{\sum[10 \log_{10}(|S|/|\hat{S}|)]^2}. \tag{6.1}$$

Following Bello et al. (2005), spectral difference envelopes E are computed as:

$$E_S(t) = \sum_{k=0}^{K-1} \{H(|S_k(t+1)| - |S_k(t)|)\}, \tag{6.2}$$

where S represents a Mel-spectrogram with K bins. The $H(x) = (x + |x|)/2$ is a half-wave rectifier, which returns zero for negative arguments. The calculations of the E_S envelopes is the same for $E_{\hat{S}}$. Envelope reconstruction of the transformations is evaluated with CS calculated between envelopes extracted from target and transformed recordings as follows:

$$CS_{sdf}(S, \hat{S}) = \frac{E_S \cdot E_{\hat{S}}}{\|E_S\| \|E_{\hat{S}}\|}, \tag{6.3}$$

where $\cdot$ represents a dot product between E. CS_{sdf} will be close to unity for very similar drum envelopes and nearer to zero for dissimilar ones. Spectral difference functions are then calculated as the sum of the first-order difference between each spectrogram (Dixon et al., 2004). Resulting envelopes are normalised between $[0, 1]$.

All reported timbral reconstruction experiments are presented as means calculated over the test set, while PC coefficients are averaged over the frequency axis.

6.5 Results

6.5.1 Subjective Results

Dampening

Figure 6.9 presents normalised violin plots showing the dampening transformation results for the subjective listening test. A one-way ANOVA was used to determine if distributions of the responses have a common mean—that is, if the plugin chains under evaluation had a different effect on the subjective scores of similarity. U2D ($p =$ 3.12e-14) and D2U ($p =$ 4.81e-14) both had p values of <0.05. The small p values allows us to reject the hypothesis that all group means are equal and indicates that the different ratings are not statistically the same as each other.

A post-hoc multiple pairwise comparison was used to establish which of the ratings were significant based on the results from the ANOVA test. As per the recommendations in ITU-R (2015), Tukey's Honestly Significant Difference (HSD) test was used for this comparison. The U2D subjective listening test showed promising results. It can be seen in Figure 6.9 that DEQ10 (*mean*: 0.66) and DEQ30 (*mean*: 0.58) are rated more similar to the hidden reference (*mean*: 1) than the input (*mean*: 0.3). It should be noted that all participants correctly identified the hidden reference and placed it at the top of the rating scale for all four test pages. The ratings for DEQ10 and DEQ30 were both statistically higher than the input ($p =$ 2.07e-08 and $p =$ 9.84e-06, respectively) using HSD. This suggests that both of these effects moved the processed input perceptually

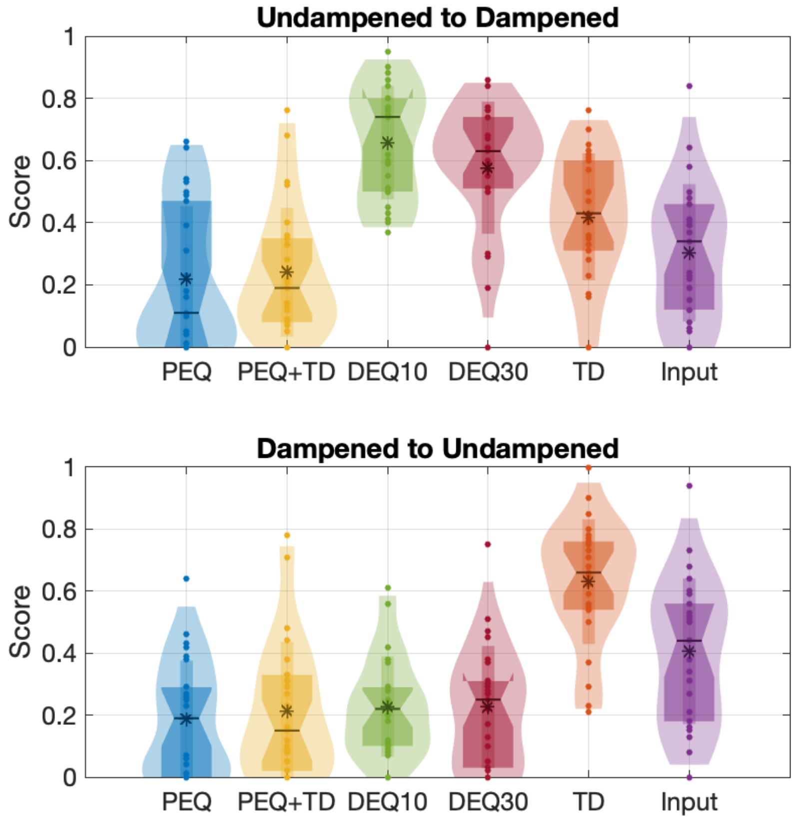

Figure 6.9: Violin plot of dampening results from listening test data. Means depicted by * symbol and medians denoted by black horizontal lines. Shape represents the distribution of scores for each variable.

closer in similarity to the reference, which in this instance was a snare drum recording dampened with a BFSD. Although not able to completely emulate the real dampening effect, these results indicate that the transformation is indeed creating a more dampened sound compared to the undampened recording. Although TD (*mean*: 0.42) was rated higher than the input overall, the ratings were not significantly higher ($p = 0.054$). Likewise, although PEQ and PEQ+TD do have lower overall ratings than the input, they are not statistically different. For D2U the only effect that had a significantly higher rating ($p = 0.0012$) than the input (*mean*: 0.4) was TD (*mean*: 0.63) based on HSD, which can be seen in Figure 6.9.

Positional

The listening test results for the positional transformation are presented in Figure 6.10, all participants correctly identified the hidden reference (*mean*: 1). An ANOVA was used again to determine if any of the ratings were significantly different, it was found for B2T transformations that there were no statistical differences between any of the scores ($p = 0.42$). This can be seen by the relatively close means and overlapping ranges of the different ratings. Although DEQ10 has a higher rating (*mean*: 0.49) than the input (*mean*: 0.36), these ratings were not statistically different from each other when the HSD test was conducted. For T2B, some significant differences were shown based on the results from an ANOVA ($p = 1.54\text{e-}11$). The HSD test revealed that the performance of both PEQ and PEQ+TD (*mean*: 0.34 and 0.16, respectively) were statistically lower than the input (mean 0.58), with PEQ+TD being rated least similar to the target. TD had slightly higher ratings (mean 0.64) than the input, but again these ratings were not statistically different from each other. This showed that for T2B positional changes, no method was successful at moving the input perceptually closer to the target, with both PEQ and PEQ+TD statistically worsening similarity. For B2T, no significant effects were seen, either positively or negatively by any of the transformations.

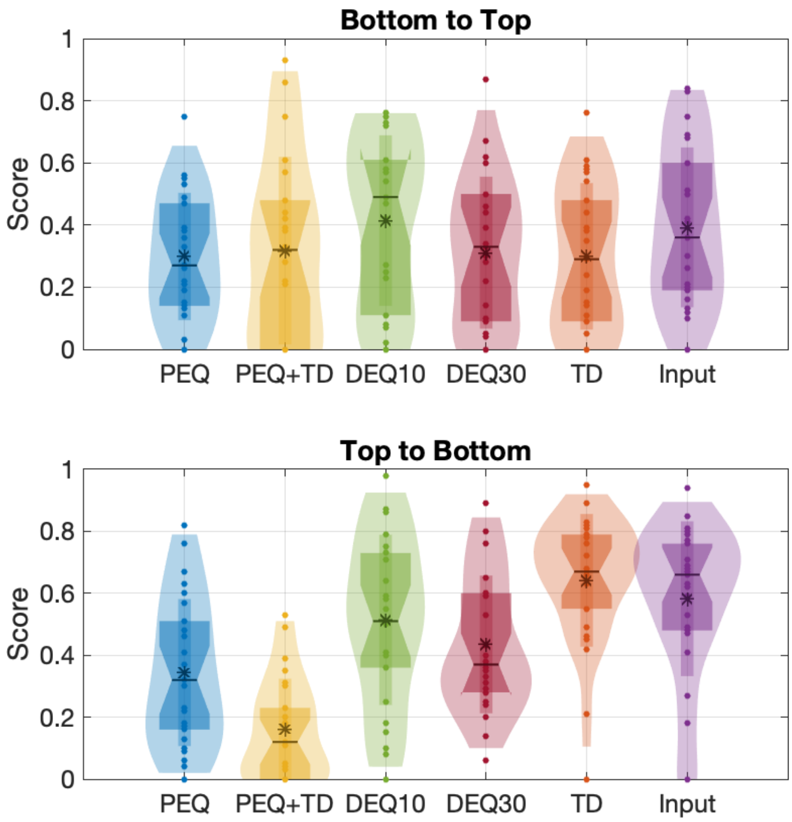

Figure 6.10: Violin plots of positional results from listening test data. Means depicted by * symbol and medians denoted by black horizontal lines. Shape represents distribution of scores for each variable.

6.5.2 Objective Results

Several of the objective metrics for U2D shown in Table 6.4 display similar trends to the subjective evaluations. For U2D all metrics showed DEQ10 to be most similar to the target. For D2U, TD rated most similar in the subjective evaluation and measured most similar when using SCD, LSD, and CS; however, unlike the subjective ratings when using MSL and PC, DEQ30 was the best performing.

The objective metrics for the positional tasks can be seen in Table 6.5, DEQ10 had the highest similarity for T2B and B2T when measured with MSL and PC respectively. PEQ also showed favourable results for T2B when using LSD and B2T when using both MSL and CS. TD was another effect that performed well across different metrics, as it displayed the highest similarity with both SCD and PC for the T2B transformation. PEQ+TD showed strong similarity for B2T, scoring most similar when using SCD and LSD.

	MSL_{stft}		SCD_{mfcc}		LSD_{Mel}		PC_{Mel}		CS_{sdf}	
Name	U2D	D2U	U2D	D2U	U2D	D2U	U2D	D2U	U2D	D2U
PEQ	8.31	65.57	0.75	0.90	2.53	3.09	0.68	0.52	0.86	0.69
TD	6.92	12.90	0.73	**0.85**	2.78	**2.72**	0.64	0.60	0.70	**0.91**
PEQ+TD	8.91	39.96	0.64	0.87	2.45	3.49	0.62	0.45	0.61	0.52
DEQ10	**4.77**	11.83	**0.55**	0.80	**2.13**	4.32	**0.70**	**0.68**	**0.89**	0.90
DEQ30	5.46	**8.01**	0.63	0.87	2.25	4.71	0.69	**0.68**	0.86	0.90

Table 6.4: Dampening task results using Mel-spectrograms: Mean multi-scale loss (MSL), spectral cosine distance (SCD), log-spectral distance (LSD), mean Pearson correlation (PC), and envelope cosine similarity (CS). Lower values indicate greater similarity,except for the PC and CS metrics where higher values do. Highest performing metrics shown in bold.

	MSL_{stft}		SCD_{mfcc}		LSD_{Mel}		PC_{Mel}		CS_{sdf}	
Name	B2T	T2B	B2T	T2B	B2T	T2B	B2T	T2B	B2T	T2B
PEQ	**7.86**	10.63	0.39	0.43	2.09	**2.11**	0.64	0.53	**0.91**	0.87
TD	10.16	7.35	0.40	**0.39**	2.34	2.07	0.61	**0.64**	0.89	**0.92**
PEQ+TD	17.86	23.09	**0.35**	0.42	**1.81**	2.45	0.52	0.38	0.48	0.57
DEQ10	8.17	**5.83**	0.54	0.50	2.39	2.54	**0.66**	0.54	0.83	0.87
DEQ30	8.27	6.33	0.68	0.62	2.61	3.01	0.65	0.54	0.81	0.88

Table 6.5: Positional task results, metrics are the same as those used in Table 6.4. Lower values indicate greater similarity, except for the PC and CS metrics where higher values do. Highest performing metrics shown in bold

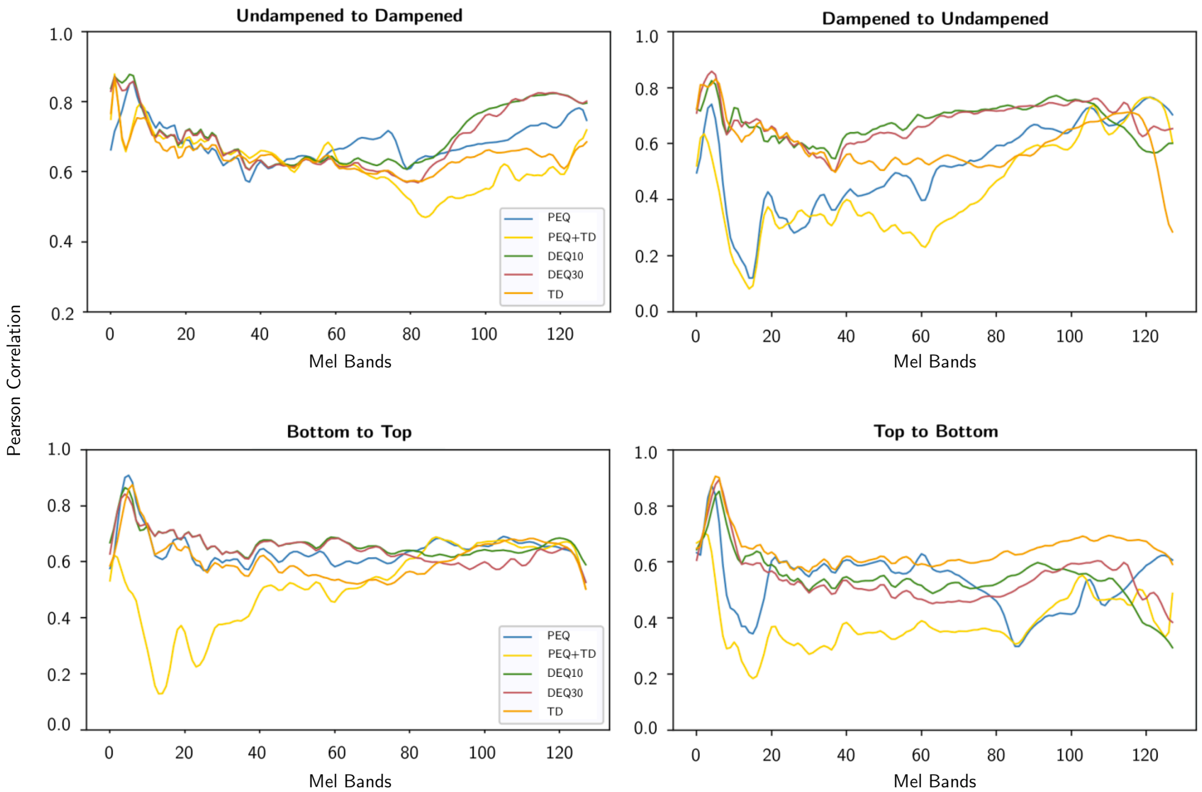

Figure 6.11: Mean smoothed Pearson correlation results computed with Mel-spectrograms for both dampening tasks (upper) and the positional tasks (lower).

6.6 Discussion

The results from the listening test indicate that D2U may be a harder transformation to emulate than U2D, with both DEQ10 and DEQ30 being rated statistically more similar to the target for U2D, but whose ratings were not significantly different to the input when used for D2U. Dampening a snare drum removes high frequency energy, whereas removing dampening increases higher frequency content. When dealing with a heavily dampened snare recording the high frequency content has already been removed and it shows that DeepAFx was not able to learn optimal parameters for the effects to enhance the missing information.

TD was most successful for the D2U transformation, likely due to TD's *release boost* parameter shaping the envelope of the drum recording to better emulate an undampened strike. Figure 6.11 displays the mean smoothed PC results for the dampening tasks. High degrees of similarity to the target can be observed by both DEQ10 and DEQ30 only for the higher frequency ranges for U2D. Little difference is seen between any of the plugins for the lower frequency bands. As high frequencies are most effected by dampening, the high measure of similarity in these important bands is likely responsible for the significantly higher ratings in the subjective evaluation.

For the D2U transformation, DEQ10 and DEQ30 had the highest measures of similarity in the mid- and upper mid-range frequency bands; however, this similarity is not reflected in the subjective responses. Although TD was subjectively the most similar to the target, the PC in Figure 6.11 shows that it does not outperform DEQ10 or DEQ30, suggesting that envelope similarity is more important for D2U than spectral similarity. The subjective evaluation for B2T did not show any effect chain to statistically produce different ratings, with only DEQ10 producing slightly higher ratings. In the case of T2B, PEQ and PEQ+TD produced ratings that were statistically lower than the input. A possible cause for this may be due to the input being rated similar to the target. With little timbral disparity between input and target, it may be more difficult for DeepAFx to use the provided plugins to make the necessary improvement, converging on parameter settings that are more extreme than required. A potential solution may be to constrain the ranges of certain parameters when more subtle transformations are to be carried out. PEQ and PEQ+TD also showed very low similarity for the mean smoothed PC results for T2B seen in Figure 6.11, with the most notable dissimilarity being in the lower frequency ranges and upper mid-range. PEQ+TD also showed very poor similarity in the lower frequencies for B2T; however, this dissimilarities was not reflected in the subjective evaluations.

The stimuli selected for each of the transformations in the listening tests may not best exemplify the ideal transformation, as the input-target pairs were chosen randomly from the available evaluation data. Thus, there may exist more representative samples that were not able to be assessed during the subjective evaluations. Other variables such as timbral differences associated with velocity disparity could also play an important part in the subjective perception of similarity. These interactions may affect the performance of the model if for example very soft hits do not elicit the same discernible traits as harder played strikes when dampening is applied. This could potentially be mitigated through only training with strikes of similar velocities. Certain microphones may be more adept at capturing timbral subtleties, making it easier for a listener to distinguish changes, or particular snare drums may emphasis the effects of parameter changes more so than others. The relationship between subjective ratings and the objective metrics cannot be strongly linked, as the objective measures made use of all samples from the evaluation data. In most cases DEQ10 outperformed DEQ30, which indicates that octave-band control (i.e., DEQ10) had sufficient timbral shaping abilities and $\frac{1}{3}$-octave band (i.e., DEQ30) provided no additional benefits. There may exist a point where additional frequency bands produces better results, and exceeding this amount beings to decrease performance.

6.7 Conclusions

This chapter has explored the potential of retroactively modifying two snare drum recording parameters—batter head dampening and microphone position, through the use of various audio effects within a layer of a deep neural network. This was aided through the creation of a large dataset consiting of acoustic snare drum recordings, which provided training data for the network. Two novel audio effects—an octave band and $\frac{1}{3}$-octave band dynamic EQ, with fixed center frequency bands and trainable parameters were created specifically for use within the deep learning system. Results from a subjective evaluation demonstrated that with particular audio effects, the system was able to move perceptually closer to the real world targets for dampening tasks, it was unsuccessful in any microphone positional transformations (i.e., perceptually making a bottom snare microphone sound like a top position microphone and vice versa). Objective metrics also revealed a tendency towards improvements in similarity for certain transformations. Most notably, DEQ10 performed best at Undampened-to-Dampened in all measures.

Dampening and microphone position were just two recording parameter transformation that were evaluated based on several recording variables associated with SDDS, although more aspects of the recording process and their related timbral transformations could be explored—for example, transformations between different drum shell materials (e.g., maple to steel), microphone type (e.g., dynamic to condenser), or velocity intensity (e.g., hard to soft). An investigation into other audio effects could be carried out using similar methods, such as distortions or reverbs for their timbral shaping capabilities. Training DeepAFX with a reverb effect could potentially allow for the transformation from a close microphone position to an overhead position. Further investigations might explore how timbral characteristics associated with velocity interacts with different recording parameter changes. This could be used to train a system that allows for the perceptual modulation of strike velocity without a change in loudness, such an effect could be useful for drum humanisation of sampled drums sequences.

A possible direction for future research in this area could be to assess the benefits of additional computational power, as training was only carried out with a relatively small sub-set of the SDDS available library; larger subsets and alternative architectures could improve the quality of the desired recording parameter transformations. Additionally, subsequent studies could investigate methods for navigating the networks latent space. Navigation controls could be provided as a GUI to creatively interpolate between transformations or to refine the estimated parameters.

Chapter 7

Conclusion

This thesis has investigated timbral analysis of snare drum recordings and presented methods and results for several proposed recording parameter transformations all with the aim of answering the primary research question: can post-hoc digital transformations of snare drum recordings elicit a subjective change akin to modifying real-world recording parameters. Through this process, timbral variation associated with various aspects of the snare drum were studied, this included its physical construction, the velocity intensities at which it is played, and recording parameters (i.e., microphone selection, microphone position, and dampening amount). The earlier chapters addressed several sub-questions, such as determining the impact that microphone selection had on subjective quality of snare drum recordings, identifying if listeners could perceive timbral differences associated with snare drum strike velocity fluctuations, and if snare drum selection affected subjective ratings of different microphones.

From these findings it was then explored if specific transformations could be applied to elicit a perceptual change akin to altering common real-world recording parameters that might typically be carried out prior to recording. After an investigation into the subjective differences between a range of commonly used studio microphones for snare drum recording, differences between the sensitivity to higher frequencies seemed to correlate to preference. For this reason it was proposed that modifications to a recording's frequency spectrum through the use of a common audio production tool, the graphic EQ, it might be possible to spectrally transform a least-preferred microphone to emulate the qualities of a more highly-preferred microphones. This study showed promising results, however, access to the target audio was paramount for the success of the transformation, meaning it had limited application. From this, a large scale acoustic snare drum dataset was created, that focused on extreme timbral variation and was made publicly available to encourage further snare drum based research. The dataset included a range of timbrally diverse snare drums from different manufacturers, made of different materials, and fitted with different drumheads. These drums were recorded at multiple velocity intensities and with multiple microphones. Additionally, several commercial dampening products were used to offer real-world studio recording scenarios. This dataset would also serve to train a deep neural network in order to carry out perceptual parameter transformations of the snare drum recording process. The network attempted to emulate the timbral modification associated with microphone position and batter head dampening. Greater success was found with the dampening emulation than with positional changes.

The intended purpose of this work is to highlight how chaining together a series of smaller recording decisions ultimately accumulate to produce the desired timbral attributes and to lay the ground-work for the development and refinement of techniques for post-hoc recording parameters transformations. The success of the recording session is determined by how satisfactory the resultant recordings are at the end. Many decisions need to be made on the spot, under time constraints, or without the context of other instruments to base ones decisions on. Often engineers will rely on past experience to trust the choices they make, however for newer, less experienced engineers the recording process can be daunting. Post-hoc recording parameter modifications would allow for an audio engineer to virtually move or change a microphone, apply more or less dampening to a snare, undo poor or inappropriate choices that might be made in haste or due to requirements of the song evolving as more instruments are added. Additionally, this work presents several methodologies for the perceptual comparison of different microphones, velocities, and audio effect transformations. These methods could be adapted to suit a range of audio production tasks, not solely related to drums or percussion. The work undertaken in this thesis can be divided into three main themes: timbral analysis of snare drums (Section 7.1); subjective preference of microphones for snare drum recording (Section 7.2); and recording parameter transformations (Section 7.2). These three topics are covered from different perspectives across all the chapters.

7.1 Timbre Analysis

Timbral analysis of the snare drum is explored across chapters 2 to 4, with each chapter focusing on different elements. Chapter 2 focused on the snare drum alone, while Chapter 3 investigated timbral differences associated with microphones, then in Chapter 4 the impact that velocity intensity had on timbre was presented.

Chapter 2 was intended to present a detailed description of the variables of the snare drum that can be altered before recording takes place, and explored literature associated with the physical construction of the snare drum and discussed how each component plays a role in defining drum timbre. Alterations to certain components of the snare drum result in noticeable timbral changes, by describing these properties one could select, modify, tune, and dampen the snare drum in such a manner to achieve the desired timbral character. However, not all variables are readily available for modification or adjustment during the recording session and there are intractable combinations of variables that the engineer and drummer can modify. Understanding the importance of each element may allow for better and more informed choices to be made. This chapter aimed to provide sufficient depth to the elements most responsible for alteration to timbre, and presented a range of technologies designed to enhance and tailor timbre for specific needs. Chapter 2 also discussed the history of the snare drum, the incremental development of certain components, and its incorporation into the modern drum kit. With the features of the physical snare drum addressed, Chapter 3 described industry standard methods of the recording process that are used to manipulate drum timbre and in particular the timbre of the snare drum. Recognised approaches and guidelines for timbral modification were presented from various recording engineers which focused primarily on different microphones techniques such as microphone selection and placement, two recording variables responsible for shaping the sound to the desired specification.

When a snare drum is struck at varying velocities, louder or quieter strikes are produced. Striking the drum with a more forceful intensity excites the batter head more and in turn causes greater vibrations to occur within the shell, hoops, and resonant head, resulting in increased sound pressure levels. In Chapter 4 it was investigated if participants could distinguish between high and low velocity snare strikes when loudness disparity had been removed from recordings made with four common studio microphones. It was discovered that all participants in the study could successfully identify the velocities with the absence of loudness cues to a very high degree of accuracy. This indicated that participants were relying on temporal and spectral variations to differentiate between the different velocity intensities. Once this had been established the next stage was to ascertain which aspects of the spectrum and envelope may be responsible for the identification of velocity. A range of commonly used audio features were extracted from a small dataset of carefully recorded snare drums at two velocity levels. Nearly all features extracted from the recordings were significantly different between high and low velocity strikes, notably the attack time was shorter for the low velocity strikes, whilst decay time was longer for the high velocity strikes. Fundamental frequency was also shown to vary with change in velocity, with high velocity strikes producing on average a 20Hz lower fundamental, likely due to the oscillations of the batter head travelling a greater distance due to the further displacement caused by a more forceful strike. Bark scale critical bands between 570Hz to 840Hz showed the biggest disparity for velocity intensities.

The information in these chapters help to demonstrate how many factors are associated with controlling and changing the timbre of the snare drum. Through manipulation of the drum itself and the recording practices used, an engineer may utilise their knowledge of this process to produce timbres appropriate for certain styles of playing, or for different genres or songs.

7.2 Microphone Comparisons

Physical properties of microphones impart certain characteristics on to the recordings they create, these characteristics can be favourable or detrimental to the outcome that a recording engineer is trying to achieve. In Chapter 3 and 5, subjective evaluation of different microphones for the application of snare drum recording was carried out. Chapter 3 included a microphone comparison study in order to determine if microphone selection plays a role in the preference of snare drum recording. Twenty-five microphones were selected and the recordings were captured in a manner to mimic real-world recording scenarios. Both isolated snare strikes were captured as well as strikes with the addition of bleed generated from the kick drum and hi-hat. The results of the listening test revealed a clear disparity in the scores between the highest and lowest rated microphones. The finding suggested that although the preference of some microphones may be heavily dependent on whether they are used for isolated snare recording or a complete drum kit, most of the microphones tested would perform equally in both scenarios. Of the subsets assessed (i.e., polar pattern, type), the condenser microphones demonstrated the strongest correlation with the rank. The subjective scores of the isolated strikes had positive correlation with spectral centroid, indicating listeners had a preference for microphones with relatively more high frequency content.

In Chapter 5 a second microphone comparison study was carried out, for this, four distinct snare drums and 12 microphones were selected to investigate if there was snare dependent preference for any of the microphones. To further reduce any timbral variation associated with strike velocity or strike location on the drumhead, the human drummer was eliminated from the data capture stage, opting for a specifically designed robot drum arm capable of consistent strike placement and velocity. Statistical evaluation of the results from the listening test allowed for the classification of microphones into least-preferred and highly-preferred microphones, additionally, only one microphone was shown to exhibit snare drum dependent results. The results from this microphone comparison study were used to inform a microphone transformation task where it was attempted to map the spectral features of a highly-preferred microphone on to a least-preferred microphone.

These two microphone comparison studies showed that certain microphones were subjectively better for recording snare drums compared to other choices. This indicates the importance of selecting an appropriate microphone during the recording process. Time permitting, an engineer may test several potential microphone choices before committing to the one that elicits the most desirable results. This data can be inferred to other instruments, suggesting that there may exist particular microphones that are more favourable for kick drum, toms, hi-hats or string or wind instruments. Through the exposure to recording from a vast array of different microphones, an engineer may build up knowledge of which microphones they prefer for certain applications to achieve specific timbre which may be more suited to a particular genre. When time is limited, one can then rely on previous experience to successfully select the best microphone for the intended outcome.

7.3 Recording Parameter Transformations

In Chapters 5 and 6 various real-world recording parameters were attempted to be perceptually modified post-recording. Building upon the findings from the microphone comparisons, further investigations were carried out to determine if spectral content alone was the most important contributing variable to preference. It was hypothesised that changes to a microphone's spectrum would allow it to take on the characteristics of a subjectively more preferred microphone. In Chapter 5 a listening test was carried out to classify microphones based on subjective preference scores. Analysis of the results allowed for the selection of a least-preferred microphone. In addition, 4 highly-preferred microphones were also chosen. The spectral content of the recordings produced from the least-preferred microphone and the 4 highly-preferred microphones was analysed, and through the use of a digital 32-band graphic equaliser, the necessary frequency bands were either attenuated or amplified so that the least-preferred microphone took on the spectral features of the 4 highly-preferred microphones. A subsequent listening test was carried out to assess listener preference between the 4 highly-preferred microphones and the least-preferred microphone with the spectral transformation applied to mimic the corresponding highly-preferred microphone. The results revealed that very few participants could distinguish between the real highly-preferred microphone and the post-EQ least-preferred microphone. These findings suggest that the most prominent subjective variable between microphones is likely the frequency response, which is predominately responsible for the preference of the resultant recordings. These results are promising as they illustrate that one can potentially modify an unideal microphone to emulate a more desirable one during the mixing stage if so required. Being able to emulate a better microphone is certainly a useful audio production tool, as it would save one time and expense when recording, as less thought to microphone selection would be required and more care and attention could be given to microphone placement in order to minimise bleed from surrounding drums. An engineer could then virtually select from a range of appropriate microphones. While commercially available products that allow for this do exist, they require a specialist microphone and pre-amp system. The proposed system could theoretically work with any microphone, however an obvious drawback of this is the need for the target audio. If one were to have a dataset of all microphone frequency responses at specified distances, it would be possible for an engineer to remove or flatten their microphone's response by applying an EQ curve with the inverse frequency response of the microphone they used for a given recording. They would then be able to choose any other microphone and apply the necessary EQ curve to emulate that microphone's frequency characteristics. This is most likely to be effective with microphones that have similar polar patterns and topologies, as these variables may be more difficult to emulate using only EQ.

The successful transformation from one microphone to another through simple graphic EQ, led to two further recording parameter transformations being investigated, including microphone position and dampening amount. In Chapter 6, the Snare Drum Data-Set was presented for general purpose acoustic snare drum research, with the aim of providing usable training sets for a range of machine learning tasks. From this large dataset, various sub-sets were collated in order to provide source-target training pairs for a deep encoder, capable of learning the necessary parameter values of a given audios effect, to mimic the proposed real-world transformations. Four proposed transformations were attempted using this system, two positional changes, from a top snare microphone position to a bottom snare position, and the inverse transform, from bottom to top, as well as two dampening transforms, from an undampened snare to a dampened one, and from dampened to undampened. A parametric EQ and a transient designer effect were chosen for the system, as well as using the two effects in series with the EQs passing its output to the input of the transient designer. Along with these two common audio productions tools, two novel audio effects were also investigated for the potential to carry out these transformations; a 10- and 30-band dynamic graphic EQ. The network attempted to learn from the provided training data the optimised parameters for these various effects. Experienced sound engineers evaluated the post-hoc dampening and undampening transformations and it was shown that the perceptual similarity was improved by certain effects. The transient designer performed best for the dampened to undampened transformation, outperforming the other effects, suggesting that envelope characteristics are more important for emulating the removal of dampening, as the transient designer had no spectral shaping abilities. For the undampened to dampened transformation task, both novel audio effects performed the best, being rated statistically more similar to the real-world target than other effects. In most cases the octave-band dynamic EQ outperformed the third-octave band version, indicating that the additional bands provided no greater benefits. No success was found attempting to carry out microphone positional changes. This is not to say that this task cannot be achieved, as limitations with the network or the subsets may have been responsible for the results. Due to larger training data resulting in lengthy, impractical times for convergence, training data size was chosen as a trade-off between training time and accuracy of the models. Had more time or more powerful computing been available, more training pairs could have been implemented from the original Snare Drum Data-Set which may have yielded more impressive transformations.

7.4 Further Work

There is large scope for further research to extend and develop upon the work presented in this thesis. In Chapters 3 and 5 microphone comparisons were carried out on either solo snare drum, or snare drum with the presence of bleed from the hi-hat and kick drum. Although subjective preference only changed for a few microphones with the addition of bleed, this did not investigate how the addition of other close, overhead, and room microphones may affect the perceived quality of the snare drum. When a drum kit is recorded with multiple microphones, the snare drum is captured by every microphone in varying amounts. This snare bleed will interact favourably or negatively with the direct signal captured from the snare's close microphones. This interaction is a complex problem to explore, as the combinations and placement of microphones around a drum kit are innumerable. By simply investigating which microphones interact best for snare drum, hi-hat, kick drum, and overhead, when selecting from six different microphones for each element, there would be a possible 1296 microphone permutations, which is far too many for a single listener to critically assess. This number is generated without consideration of microphone position variations. In addition, it is common practice to capture the snare drum with both a microphone on the batter and resonant drumheads, as well as the kick drum with more than one microphone, further adding to the complexity of possible interactions. While it is expected that an engineer may attempt to reduce the amount of snare bleed into the other microphones, or use tools such as noise gates to eliminate it from close microphones placed on the rack and floor toms for example, other microphones are specifically selected to capture the entirety of the drum kit—notably the overhead and room microphones. Attempting to reduce the amount of snare capture by these microphones would be counter intuitive, and thus examining the subjective quality of interaction between different overhead microphones and close microphones may produce results that have real-world implications. An engineer may wish to select a particular overhead or room microphone to complement the timbral qualities of the close snare microphones. Conversely, if they have satisfactory results from their overhead microphone they may opt to select a snare microphone that complements the overhead. Further investigations could be carried out to address which underlying spectral components interact to produce subjectively more favourable combinations of microphones. It could be hypothesised that the two microphones fill in missing information not present in the other microphone recordings, thereby creating a fuller spectrum.

Additional research focused on the subjective evaluation of microphone comparison for multi-microphone drum recording should consider the impact of constructive and destructive phase issues that arise when using multiple microphones around the drum kit. It is not expected that one will find the ideal microphone combination for drum kit recording, but rather a set of general purpose configurations that will suit a variety of recording scenarios, perhaps as a good starting point for a junior recording engineer. Once it is understood what constitutes good microphone selection, one can make informed decisions when it comes to using the available equipment. Certain topologies may be better suited to particular applications, such as the use of dynamic microphones for the batter head paired with condensers on the resonant head. Further to this research, the role that audio effects have on microphone choice could also be studied. While microphone choice has been shown to produce perceptual differences in audio quality, it is not expected that the audio engineer will simply leave the recordings unaffected. The use of compression, EQ, and other audio effects, such as reverb is commonly utilised to enhance the recordings. With this is mind, one may seek to determine if equally satisfactory results can be achieved via additional processing of an unideal microphone choice. It was shown in Chapter 5 that it is possible for one microphone to take on the timbral characteristics of a more highly-preferred model. This suggests that the importance of microphone choice is overstated; however, this has yet to be validated. If a recording engineer is inevitably going to change and manipulate the timbre of the recordings to achieve a desired outcome, the question could be raised if they would be able to do so with inferior source material. A subjective comparison of pre- and post-processed recordings of a range of microphones would illustrate if processing by an experience engineer is enough to negate microphone choice. A study could be conducted to determine if mixing engineers require greater amounts of processing and invest more time to produce satisfactory results when presented with least-preferred and highly-preferred recordings, thereby ascertaining the importance of microphone selection when mixing is carried out.

In Chapter 6, an acoustic snare drum dataset was presented which offers researchers multi-velocity recordings of a range of timbraly distinct snare drums, captured with 53 simultaneous microphone, and recorded with different dampening products. This work was specifically undertaken to encourage snare drum based research where large datasets may be required, such as in the case of machine learning tasks. This dataset was used to attempt to train a model to learn the appropriate parameters to transformer between a top and bottom snare microphone position, as well as between dampened and undampened snare recordings. An additional way this dataset could be utilised is for the task of velocity modelling. As shown in Chapter 4, striking the snare drum at lower and higher velocity intensities produces diverse timbres, regardless of any volume differences, meaning that listeners can perceive the timbre of a light and hard strike. There are several applications where manipulating the recording's timbre to perceptually modify the velocity might be useful. When programming a drum sequence using sampled recordings, often mulit-velocity recordings of the drum strikes are not always available, such as when the sample is taken from a fully-produced song. In such cases the engineer may wish to humanise the sequenced drum parts, which can be done with slight timing adjustments, but also through the use of modulation of the amplitude to emulate the dynamic playing of a human drummer. In addition to timing and level alteration, a tool could be developed that can performs spectral modifications in a manner that elicit the perceptual characteristics of real-world velocity variations. This may also be used where level fluctuations are not necessary, but one wishes to change the intensity of playing by changing the spectral components of velocity alone, such as in the chorus of a song. An effect of this kind could be used creatively or to correct technical issues, where one strike is noticeably lighter or harder than others, its perceived velocity could be adjusted.

Chapters 5 and 6 saw the investigations into post-hoc recording parameter transformations, firstly with microphone choice, and then followed by microphone position and dampening amount using more sophisticated methods. Although little success was seen for positional changes, the other transformations both showed promising results. By further developing the strategies for these kinds of transformations, exploring alternative audio effects, and utilising larger available training data, it may be possible one day to create audio production tools capable of perceptually transforming any real-world recording parameter, allowing for any virtual adjustment to be made to the recordings. Engineers may have the ability to virtually swap out the snare drum for different models, or try a range of shell sizes, virtually move or even add microphones that were never there during the recordings. It may even be possible to virtually try out different locations for the drum kit within the recording space, or perceptually change or re-tune the drum heads for different songs on an album, all while the recording studio is closed and the lights are off.

Bibliography

Aisher, B. (Decemeber 2012), How To EQ A Mix, URL https://www.attackmagazine.com/technique/tutorials/how-to-eq-mix/, [Online; accessed 4/3/2022]. [Cited on Page 64.]

Albano, J. (September 2016), Mixing Snare Drums By Blending Multiple Microphones, URL https://ask.audio/articles/mixing-snare-drums-by-blending-multiple-microphones, [Online; accessed 27/1/2022]. [Cited on Page 94.]

Alldred, B. (December 2019), Drum Heads with Evans (D'Addario UK), URL https://www.leedsconservatoire.ac.uk/about-us/class-of-2020/career-support-advice/musicians-survival-guide-articles/drum-heads-with-evans-d-addario-uk/, [Online; accessed 3/2/2022]. [Cited on Pages 9, 15, 16, 19, and 21.]

Allen, M. (August 2020), 10 Quick Ways to Dampen Your Snare Drum, URL https://drummingreview.com/how-to-dampen-a-snare-drum/, [Online; accessed 24/6/2022]. [Cited on Page 26.]

Andertons (October 2015), Ultimate Guide to Snare Drums, URL https://www.andertons.co.uk/snare-drums-guide, [Online; accessed 11/3/2022]. [Cited on Pages 6 and 10.]

ANSI (June 2009), Specification for Octave-band and Fractional-Octave-band Analog and Digital Filters, Technical Report S1.11-2004, American National Standards Institute. [Cited on Pages 79 and 91.]

Azzarto, F. (July 2010a), What You Need to Know About...Counterhoops, URL https://www.moderndrummer.com/2015/03/need-know-counterhoops/, [Online; accessed 18/12/2022]. [Cited on Pages 18 and 19.]

Azzarto, F. (November 2010b), What You Need to Know About...Drumheads, URL https://www.moderndrummer.com/2011/10/what-you-need-to-know-about-drumheads/, [Online; accessed 1/11/2022]. [Cited on Pages 16 and 17.]

Azzarto, F. (May 2010c), What You Need to Know About...Drumsticks, URL https://www.moderndrummer.com/2014/12/need-know-drumsticks/, [Online; accessed 15/6/2022]. [Cited on Page 27.]

Azzarto, F. (October 2011), What You Need to Know About...Drum Shells, URL https://www.moderndrummer.com/2015/02/need-know-drum-shells/, [Online; accessed 23/3/2022]. [Cited on Pages 10, 11, 12, and 13.]

Ballou, G. (2008), *Handbook for Sound Engineers*, Audio Engineering Society Presents Series, Focal, ISBN 9780240809694. [Cited on Pages 32, 33, 34, and 54.]

Bartlett, B. (1987), Choosing the Right Microphone by Understanding Design Tradeoffs, in *35th Journal of the Audio Engineering Society*. [Cited on Page 63.]

Bartlett, B. and Bartlett, J. (September 2009), *Practical Recording Techniques: The Step-by-step Approach to Professional Audio Recording*, Focal Press, ISBN 9780240811444. [Cited on Page 85.]

Bartlett, B. and Bartlett, J. (2016), *Practical Recording Techniques: The step-by-step approach to professional audio recording*, Routledge. [Cited on Pages 16, 23, 26, 30, 33, 34, and 35.]

Bartlett, B. A. (1981), Tonal Effects of Close Microphone Placement, in *29th Journal of the Audio Engineering Society*. [Cited on Pages 54, 72, and 93.]

Beattie, S. (February 2019), Wood vs. Metal Snare Drums, URL `https://blog.andertons.co.uk/labs/wood-vs-metal-snare-drums`, [Online; accessed 13/1/2022]. [Cited on Pages 11, 12, and 19.]

Beck, D. (2004), *The Musician's Guide to Recording Drums*, Musicians Institute Press, ISBN 9781476859279. [Cited on Pages 30 and 35.]

Beck, J. (2013), *Encyclopedia of Percussion*, Taylor & Francis, ISBN 9781317747673. [Cited on Page 15.]

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M. and Sandler, M. B. (August 2005), A Tutorial on Onset Detection in Music Signals, in *Speech and Audio Processing, IEEE Transactions on*, volume 13, no. 5, pages 1035–1047. [Cited on Pages 100 and 102.]

Biancardi, M. (November 2015), Tonal Properties of 5 Metals Used in Snare Drums, URL `https://reverb.com/uk/news/tonal-properties-of-5-metals-used-in-snare-drums`, [Online; accessed 22/3/2022]. [Cited on Page 12.]

Biswas, A. (1998), Novel brick-wall filters based on the auditory system, in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, volume 6, pages 3637–3640 vol.6. [Cited on Page 65.]

Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C. and Bello, J. (October 2014), MedleyDB: A Multitrack Dataset for Annotation-intensive MIR Research. [Cited on Page 93.]

Bitton, A., Esling, P., Caillon, A. and Fouilleul, M. (September 2019), Assisted Sound Sample Generation with Musical Conditioning in Adversarial Auto-encoders, in *Proceedings of the International Conference on Digital Audio Effects*. [Cited on Page 100.]

Blades, J. (1970), *Percussion Instruments and Their History*, F. A. Praeger, ISBN 9780571088584. [Cited on Page 5.]

Borwick, J. (1980), *Sound Recording Practice: A Handbook*, Oxford University Press, ISBN 9780193119208. [Cited on Pages 26, 33, and 34.]

Brensilver, D. A. (September 2015), History of the Snare Drum: Eight Centuries of Innovation & Ingenuity, URL `https://drummagazine.com/history-of-the-snare-drum/`, [Online; accessed 20/9/2022]. [Cited on Pages 5 and 20.]

Brown, C. (December 2017), How Bearing Edges Affect the Sound of Your Drums, URL `https://www.moderndrummer.com/2014/12/need-know-bearing-edges/`, [Online; accessed 7/8/2022]. [Cited on Pages 13 and 14.]

Brown, G. (April 2020), Multiband Compressors vs. Dynamic EQs: Differences and Uses, URL `https://www.izotope.com/en/learn/multiband-compressors-vs-dynamic-eqs.html`, [Online; accessed 8/3/2022]. [Cited on Pages 88 and 89.]

Burroughs, L. and Woram, J. (1974), *Microphones: Design and Application*, Sagamore Publishing Company, ISBN 9780914130000. [Cited on Page 34.]

C. Disley, A., Howard, D. and Hunt, A. (January 2006), Timbral description of musical instruments, in . [Cited on Page 76.]

Case, A. (2012), *Mix Smart: professional techniques for the home studio*, Taylor & Francis, ISBN 9781136114533. [Cited on Page 69.]

Cesarz, N. (July 2022), 9 Best Drum Dampeners to Stop Snare Drum Ring, URL `https://drummingreview.com/best-drum-dampeners/`, [Online; accessed 17/8/2022]. [Cited on Page 26.]

Charlton, S. (April 2017), Why Your Snare Drum Rings (+How To Reduce It & Embrace It), URL `https://producerhive.com/buyer-guides/percussion/why-your-snare-drum-rings-how-to-reduce-it-embrace-it/`, [Online; accessed 10/2/2022]. [Cited on Page 25.]

Cheshire, M. (October 2020), Snare Drum Data Set (SDDS): More Snare Drums than you can Shake a Stick at, in *149th Convention*. [Not Cited.]

Cheshire, M., Drysdale, J., Enderby, S., Tomczak, M. and Hockman, J. (2022), Deep Audio Effects for Snare Drum Recording Transformations, in *Journal of the Audio Engineering Society*, volume 70, no. 9, pages 742–752. [Not Cited.]

Cheshire, M., Hockman, J. and Stables, R. (October 2018), Microphone Comparison for Snare Drum Recording, in *145th Conventions of the Audio Engineering Society*. [Not Cited.]

Cheshire, M., Stables, R. and Hockman, J. (October 2019), Microphone Comparison: Spectral Feature Mapping for Snare Drum Recording, in *147th Conventions of the Audio Engineering Society*. [Not Cited.]

Cheshire, M., Stables, R. and Hockman, J. (May 2020), Investigating timbral differences of varied velocity snare drum strikes, in *148th Audio Engineering Society Convention*. [Not Cited.]

Cooper, M. (March 2020), Guide to Vintage Drum Identification - Leedy Lugs, URL `https://www.notsomoderndrummer.com/not-so-modern-drummer/2020/3/11/guide-to-vintage-drum-identification-leedy-lugs`, [Online; accessed 5/1/2022]. [Cited on Page 17.]

Corbett, I. (2020), *Mic It!: Microphones, Microphone Techniques, and Their Impact on the Final Mix*, Taylor & Francis, ISBN 9781000289176. [Cited on Pages 33 and 34.]

Crich, T. (2010), *Recording Tips for Engineers: For Cleaner, Brighter Tracks*, Mastering music, Focal Press/Elsevier, ISBN 9780240521763. [Cited on Pages 26, 33, 34, 35, and 36.]

Davies, M. E. P., Hamel, P., Yoshii, K. and Goto, M. (August 2014), AutoMashUpper: Automatic Creation of Multi-song Music Mashups, in *In IEEE/ACM Transactions on Audio, Speech, and Language Processing*, volume 22, no. 12, pages 1726–1737. [Cited on Page 102.]

Dawson, M. (June 2014), Product Close-Up: Geo. H. Way Tradition Snare Drums (From the June 2014 Issue), URL `https://www.moderndrummer.com/2014/04/audio-product-close-geo-h-way-tradition-snare-drums-june-2014-issue/`, [Online; accessed 17/10/2022]. [Cited on Page 17.]

De Douvan, M. (Decemeber 2005), How to tension your drum skins (or "drum heads"), URL http://www.marcdedouvan.com/en/lesson.php?id=2, [Online; accessed 10/9/2022]. [Cited on Page 22.]

De Man, B. and Reiss, J. (May 2013), A Pairwise and Multiple Stimuli Approach to Perceptual Evaluation of Microphone Types, in *134th Convention of the Audio Engineering Society, Rome*. [Cited on Pages 36, 37, 41, 63, 72, and 76.]

Dean, M. (2011), *The Drum: A History*, Scarecrow Press, ISBN 9780810881716. [Cited on Page 7.]

Dierstein, C., Roth, M. and Ruland, J. (2018), *The Techniques of Percussion Playing / Die Spieltechnik des Schlagzeugs: Mallets, Implements and Applications / Schlägel, Anreger und Anwendungen*, Bärenreiter-Verlag, ISBN 9783761870648. [Cited on Page 26.]

Dinno, A. (2015), Nonparametric pairwise multiple comparisons in independent groups using Dunn's test, in *Stata Journal*. [Cited on Page 77.]

Dittmar, C. and Gärtner, D. (2014), Real-Time Transcription and Separation of Drum Recordings Based on NMF Decomposition, in *DAFx*. [Cited on Page 93.]

Dixon, S., Gouyon, F. and Widmer, G. (January 2004), Towards Characterisation of Music via Rhythmic Patterns, in *Proceedings of the International Conference on Music Information Retrieval*. [Cited on Page 102.]

Doerschuk, A. (October 2011), A Few Words About Snare Wires, URL https://drummagazine.com/a-few-words-about-snare-wires/, [Online; accessed 25/12/2022]. [Cited on Pages 19, 20, and 21.]

Donahue, A. (September 2018), Timber Timbre: Exploring Sonic Traits of Exotic Wood Snares, URL https://drummagazine.com/timber-timbre-exploring-sonic-traits-of-exotic-wood-snares/, [Online; accessed 25/12/2022]. [Cited on Page 11.]

Dowsett, P. (2015), *Audio Production Tips: Getting the Sound Right at the Source*, Focal Press, ISBN 9781317614197. [Cited on Pages 25 and 26.]

DPA Microphones (January 2015), How To Mic A Snare Drum, URL https://www.dpamicrophones.com/mic-university/how-to-mic-a-snare-drum, [Online; accessed 7/6/2022]. [Cited on Page 51.]

Drumhead Authority (June 2020), The best wood for drums? Forget drum wood, URL https://drumheadauthority.com/articles/forget-the-best-drum-wood/, [Online; accessed 24/3/2022]. [Cited on Pages 10, 11, and 13.]

D'Virgilio, N. (September 2014), How to Control Drum Sustain with Dampening, URL https://www.sweetwater.com/insync/how-to-control-drum-sustain-with-dampening/, [Online; accessed 25/6/2022]. [Cited on Pages 56 and 85.]

D'Virgilio, N. (September 2018), How to Choose the Best Drumheads for You, URL https://www.sweetwater.com/insync/choose-best-drumheads/, [Online; accessed 15/7/2022]. [Cited on Pages 15 and 16.]

D'Amico, G. (January 2001), Snare Beds – What They Are And What They Do, URL https://drummagazine.com/snare-beds-what-they-are-and-what-they-do/, [Online; accessed 22/7/2022]. [Cited on Page 22.]

D'Amico, G. (April 2015), Bearing Edges: Sustain Vs. Attack, URL https://drummagazine.com/bearing-edges-sustain-vs-attack/, [Online; accessed 25/12/2022]. [Cited on Pages 13 and 15.]

Eargle, J. (2004), *The Microphone Book*, Audio Engineering Society Presents Series, Focal Press, ISBN 9780240519616. [Cited on Pages 30, 33, 34, 35, 51, and 63.]

EBU-R-128 (2014), *Loudness Normalisation and Permitted maximum Level of Audio Signals*. [Cited on Pages 41, 53, and 76.]

Elliott, B. (February 2014), Mic Comparison: A Snare Drum Microphone Shootout, URL https://www.drummercafe.com/music-industry/featured-article/mic-comparison-snare-drum-microphone-shootout.html, [Online; accessed 4/1/2022]. [Cited on Pages 70 and 94.]

Falk, J. (July 2019), How To Tune Your Drums, URL https://www.drumeo.com/beat/how-to-tune-your-drums/, [Online; accessed 9/2/2022]. [Cited on Page 10.]

Fox, A. (2022), Top 10 Best Dynamic EQ Plugins For Your DAW In 2022, URL https://mynewmicrophone.com/top-best-dynamic-eq-plugins-for-your-daw/, [Online; accessed 24/12/2022]. [Cited on Page 88.]

Fuston, L. (July 2017a), Best Mics for Recording Snare Drums, URL https://www.sweetwater.com/insync/best-mics-recording-snare-drums/, [Online; accessed 12/6/2022]. [Cited on Pages 70 and 94.]

Fuston, L. (July 2017b), How to Mic a Snare Drum, URL https://www.sweetwater.com/insync/how-to-mic-snare-drum/, [Online; accessed 6/2/2022]. [Cited on Page 54.]

Gatzen, B. (1994), *Drum Tuning: Sound and Design*, Dci Music Video, Alfred Publishing Company, ISBN 9780769220499. [Cited on Page 23.]

Gibson, B. (2004), *Sound Advice on Recording & Mixing Drums*, Artistpro Series, ProAudio Press, ISBN 9781931140379. [Cited on Pages 11, 23, 25, 26, 28, 30, 31, 33, 35, 36, 68, 69, and 92.]

Gier, H. and White, P. (July 1999), Transient Designer - Dual-Channel, Model 9946, Technical Report 1.0, SPL electronics GmbH. [Cited on Page 92.]

Gillet, O. and Richard, G. (2006), ENST-Drums: An Extensive Audio-visual Database For Drum Signals Processing, in *ISMIR*. [Cited on Page 93.]

Gonzalez, D. (May 2022), Nine mic placements to record a snare drum in your home studio, URL https://blog.discmakers.com/2022/05/mic-placements-to-record-a-snare-drum/, [Online; accessed 2/10/2022]. [Cited on Page 54.]

Grey, J. and Gordon, J. (January 1978), Perceptual effects of spectral modifications on musical timbres, in , volume 63, pages 1493–1500. [Cited on Page 44.]

Hahn, M. (October 2018), Equalization 101: Everything Musicians Need to Know about EQ, URL https://blog.landr.com/eq-basics-everything-musicians-need-know-eq/, [Online; accessed 5/3/2022]. [Cited on Page 64.]

Hahn, M. (July 2020), How to Use Dynamic EQ for A Better Mix, URL https://blog.landr.com/dynamic-eq/, [Online; accessed 16/8/2022]. [Cited on Page 88.]

Hall, R. (July 2020), Best Snare Wires, URL https://thestepkidsband.com/best-snare-wires/, [Online; accessed 6/5/2022]. [Cited on Page 20.]

He, K., Zhang, X., Ren, S. and Sun, J. (December 2015), Delving Deep into Rectifiers: Surpassing Human-level Performance on Imagenet Classification, in *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034. [Cited on Page 92.]

Hebrock, J., Statham, K. and Kraft, C. (September 1997), Psychoacoustic Effects of Ringing in Microphones, in *103rd Conventions of the Audio Engineering Society Convention*. [Cited on Page 63.]

Henshall, M. (March 2014), How to Mic a Snare Drum with a Shure SM57, URL `https://www.shure.com/en-US/performance-production/louder/how-to-mic-a-snare-drum-with-a-shure-sm57`, [Online; accessed 27/12/2022]. [Cited on Page 54.]

Hoffman, C. (May 2022), The Difference Between Dynamic EQs and Multiband Compressors, URL `https://www.blackghostaudio.com/blog/the-difference-between-dynamic-eqs-and-multiband-compressors`, [Online; accessed 4/6/2022]. [Cited on Page 89.]

Hollen, J. L. (Decemeber 2022), What You Need to Know About Counterhoops, URL `https://oldtimemusic.com/what-you-need-to-know-about-counterhoops/`, [Online; accessed 23/12/2022]. [Cited on Pages 18 and 19.]

Houghton, M. (July 2010), Vocal Mics, Matching Mics & Voices, URL `https://www.soundonsound.com/techniques/vocal-mics`, [Online; accessed 11/10/2022]. [Cited on Page 63.]

Howley, A. (July 2018), What is a bearing edge?, URL `www.lonestarpercussion.com`, [Online; accessed 3/8/2022]. [Cited on Pages 13 and 14.]

Howley, A. (February 2020), What is the Difference Between a Tension Rod and a Lug?, URL `www.lonestarpercussion.com`, [Online; accessed 19/9/2022]. [Cited on Page 17.]

Huber, D. and Runstein, R. (2010), *Modern Recording Techniques*, Focal Press/Elsevier, ISBN 9780240810690. [Cited on Pages 8, 25, 29, 30, 32, 34, 36, and 54.]

Inglis, S. (February 2022), EQ: What Do All Those Knobs Do?, URL `https://www.soundonsound.com/sound-advice/eq-what-do-all-those-knobs-do`, [Online; accessed 4/7/2022]. [Cited on Page 66.]

ITU-R (October 2015), Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems, Technical Report Recommendation ITU-R BS.1534-3, ITU-R. [Cited on Page 102.]

Izhaki, R. (October 2017), *Mixing Audio: Concepts, Practices, and Tools*, Taylor & Francis, ISBN 9781317508502. [Cited on Page 91.]

Jillings, N., De Man, B., Moffat, D. and Reiss, J. D. (July 2015), Web Audio Evaluation Tool: A Browser-based Listening Test Environment, in *12th Sound and Music Computing Conference*. [Cited on Pages 41, 76, and 100.]

Jones, T. (January 2016), History and Development of the Early Snare Strainer, 1889 - 1920, URL `https://www.notsomoderndrummer.com/not-so-modern-drummer/2016/1/20/history-and-development-of-the-early-snare-strainer-1889-1920`, [Online; accessed 27/9/2022]. [Cited on Page 20.]

Juslin, N. (January 2000), Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception, in , volume 26, pages 1797–813. [Cited on Page 44.]

Kim, J. W., Bittner, R., Kumar, A. and Bello, J. P. (May 2019), Neural Music Synthesis for Flexible Timbre Control, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 176–180, IEEE. [Cited on Page 100.]

Kingma, D. P. and Ba, J. (May 2015), Adam: A Method for Stochastic Optimization, in *Proceedings of the International Conference on Learning Representations*. [Cited on Page 92.]

Kokkinis, E. K., Georganti, E. and Mourjopoulos, J. (April 2012), Statistical Properties of the Close-Microphone Responses, in *Audio Engineering Society Convention 132*. [Cited on Page 54.]

La Cerra, S. (October 2019), The Modern Drum Set, Part 1: The Snare Drum. The backbeat of contemporary music., URL `https://hub.yamaha.com/drums/stage/the-modern-drum-set-part-1-the-snare-drum/`, [Online; accessed 10/6/2022]. [Cited on Page 11.]

La Cerra, S. (Decemeber 2020), How to Prepare Your Drum Kit for Recording, URL `https://www.sweetwater.com/insync/how-to-prepare-your-drum-kit-for-recording/`, [Online; accessed 20/6/2022]. [Cited on Page 17.]

Lartillot, O. and Toiviainen, P. (January 2007), MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio., in *Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France*. [Cited on Page 57.]

Lee, J., Bryan, N. J., Salamon, J., Jin, Z. and Nam, J. (August 2020), Metric Learning vs Classification for Disentangled Music Representation Learning, in *Proceedings of the 21st International Society for Music Information Retrieval Conference*. [Cited on Page 87.]

Libman, D. E. (March 2019), How To Shape Your Sound With Drumsticks, URL `https://drummagazine.com/how-to-shape-your-sound-with-drumsticks/`, [Online; accessed 21/5/2022]. [Cited on Pages 26 and 28.]

Ludwig, W. and Cook, R. (2001), *The Making of a Drum Company: The Autobiography of William F. Ludwig II*, Musical Instruments Series, Rebeats Publications, ISBN 9781888408058. [Cited on Page 15.]

Macaulay, E. J. (2003), Properties of Drum Shells and Bearing Edges, Technical report, Department of Physics, University of Illinois at Urbana-Champaign. [Cited on Page 14.]

MacEachran, A. (July 2019), Metal Snares...What's The Difference?!, URL `https://drummersonly.co.uk/metal-snares-whats-the-difference/`, [Online; accessed 23/12/2022]. [Cited on Page 12.]

Major, M. (2014), *Recording Drums: The Complete Guide*, Course Technology PTR, ISBN 9781133788935. [Cited on Pages 16, 23, 25, 29, 30, 31, 32, 33, 35, 36, 54, 68, 85, and 92.]

Man, B. D. and Reiss, J. (April 2014), APE: Audio Perceptual Evaluation Toolbox for MATLAB, in *136th Convention of the Audio Engineering Society*. [Cited on Pages 41 and 76.]

Martínez-Ramírez, M. A., Stoller, D. and Moffat, d. (March 2021a), A Deep Learning Approach to Intelligent Drum Mixing with the Wave-u-net, in *Journal of the Audio Engineering Society*, volume 69, no. 3, pages 142–151. [Cited on Page 86.]

Martínez-Ramírez, M. A., Wang, O., Smaragdis, P. and Bryan, N. J. (May 2021b), Differentiable Signal Processing with Black-Box Audio Effects, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. [Cited on Pages 86, 87, 92, and 100.]

Massenburg, G. (May 1972), parametric equalization, in *journal of the audio engineering society*. [Cited on Pages 64 and 65.]

McAllister, M. (August 2022), Top 10 Transient Shaper Plugins, URL `https://producelikeapro.com/blog/top-10-transient-shaper-plugins/`, [Online; accessed 26/12/2022]. [Cited on Page 92.]

McKinnie, D. (May 2006), Measuring the Perceived Differences between Similar High-quality Microphones, in *120th Audio Engineering Society Convention*. [Cited on Page 63.]

Mellor, D. (March 1995), EQ: How & When To Use It, URL `https://www.soundonsound.com/techniques/eq-how-when-use-it`, [Online; accessed 7/2/2022]. [Cited on Page 66.]

Mellor, D. (Novemeber 2018), What is a graphic equalizer (graphic EQ)?, URL `https://www.audiomasterclass.com/blog/what-is-a-graphic-equalizer-graphic-eq`, [Online; accessed 1/8/2022]. [Cited on Page 64.]

Messitte, N. (October 2021), Principles of Equalization: How to EQ Your Mix, URL `https://www.izotope.com/en/learn/principles-of-equalization.html`, [Online; accessed 11/1/2022]. [Cited on Page 64.]

Messitte, N. (May 2022), What Is the Difference Between Mixing and Mastering?, URL `https://www.izotope.com/en/learn/what-is-the-difference-between-mixing-and-mastering.html`, [Online; accessed 14/7/2022]. [Cited on Page 63.]

Meyer, B. (January 2014), What You Need to Know About...Snare Drums, URL `https://www.moderndrummer.com/2014/12/need-know-snare-drums/`, [Online; accessed 17/11/2022]. [Cited on Pages 6, 11, 16, 17, 19, 20, and 21.]

Mitzner, R. (November 2021), How To Tune A Snare Drum, URL `https://www.drumeo.com/beat/how-to-tune-a-snare-drum/`, [Online; accessed 27/8/2022]. [Cited on Page 10.]

Modern Drummer (September 2013), What You Need to Know About...Bearing Edges, URL `https://www.moderndrummer.com/2014/12/need-know-bearing-edges/`, [Online; accessed 21/8/2022]. [Cited on Pages 13 and 14.]

Moffat, D. and Sandler, M. (October 2019), Machine Learning Multitrack Gain Mixing of Drums, in *Audio Engineering Society Convention 147*, Audio Engineering Society. [Cited on Page 86.]

Mynett, M. (July 2011), Sound at Source: The creative practice of re-heading, dampening and drum tuning for the contemporary metal genre, in *The Journal on the Art of Record Production (JARP) - Issue 05*. [Cited on Page 22.]

Newbold, M. (September 2022), What Wood Sounds Good? Your Guide to Shell Materials, URL `https://www.gear4music.com/blog/what-wood-sounds-good-your-guide-to-shell-materials/`, [Online; accessed 25/10/2022]. [Cited on Pages 11 and 12.]

Newell, P. (2012), *Recording Studio Design*, Audio Engineering Society Presents Series, Focal, ISBN 9780240522401. [Cited on Page 29.]

Nicholls, G. (2008), *The Drum Book: A History of the Rock Drum Kit*, G - Reference, Information and Interdisciplinary Subjects Series, Backbeat, ISBN 9780879309404. [Cited on Pages 7, 11, 15, 17, 18, 20, and 21.]

Nichols, G. (April 2009), Tempus Fibreglass and Carbon-Fibre Snare drums review, URL `https://www.musicradar.com/reviews/drums/tempus-fibreglass-and-carbon-fibre-snare-drums-203805`, [Online; accessed 23/10/2022]. [Cited on Page 11.]

Nichols, G. (August 2013), Truth Custom Drums Maple/Walnut/Acrylic Snare review, URL `https://www.musicradar.com/reviews/drums/truth-custom-drums-maple-walnut-acrylic-snare-581901`, [Online; accessed 22/5/2022]. [Cited on Page 11.]

Oliver, R. J. and Jot, J.-M. (October 2015), Efficient Multi-Band Digital Audio Graphic Equalizer with Accurate Frequency Response Control, in *139th Convention of the Audio Engineering Society*. [Cited on Page 79.]

Owsinski, B. (2005), *The Recording Engineer's Handbook*, Gale virtual reference library, Thomson Course Technology, ISBN 9781932929003. [Cited on Pages 13, 25, 30, 32, 54, and 63.]

Owsinski, B. (2017), *The Mixing Engineer's Handbook*, The Mixing Engineer's Handbook, BOMG Publishing, ISBN 9780998503349. [Cited on Pages 52, 66, 68, and 92.]

Owsinski, B. and Moody, D. (February 2009), *The Drum Recording Handbook*, MusicPro guides, Hal Leonard Books, ISBN 9781423443438. [Cited on Pages 10, 12, 13, 16, 21, 22, 25, 26, 30, 31, 32, 33, 36, 54, 69, 85, and 92.]

Parsons, M. and Van Horn, R. (1996), *The Drummer's Studio Survival Guide: How to Get the Best Possible Drum Tracks on Any Recording Project*, Musical Instruments Series, Modern Drummer Publications, ISBN 9780793572229. [Cited on Pages 30 and 85.]

Pearce, A., Brookes, T. and Dewhirst, M. (October 2015), Validation of Experimental Methods to Record Stimuli for Microphone Comparisons, in *139th Convention of the Audio Engineering Society*. [Cited on Page 72.]

Pearce, A., Brookes, T., Dewhirst, M. and Mason, R. (2016), Eliciting the most prominent perceived differences between microphones, in *139th Journal of the Acoustical Society of America*. [Cited on Pages 63 and 72.]

Pedersen, K. and Grimshaw-Aagaard, M. (2018), *The Recording, Mixing, and Mastering Reference Handbook*, Oxford University Press, ISBN 9780190686659. [Cited on Pages 25, 29, 30, 31, 33, 34, 36, and 68.]

Pestana, P. D., Reiss, J. D. and Barbosa, A. (May 2013), Loudness Measurement of Multitrack Audio Content Using Modifications of ITU-R BS.1770, in *Audio Engineering Society Convention 134*. [Cited on Page 51.]

Quiroga, A. F., Garcia, J. D. and Páez, D. (May 2015), Tom–Tom Drumheads Miking Analysis, in *Audio Engineering Society Convention 138*. [Cited on Pages 37, 54, 72, and 93.]

Richardson, P. G. (2010), Acoustic analysis and tuning of cylindrical membranophones. [Cited on Page 51.]

Richardson, P. G. M., Robert Toulson, E. and Nunn, D. J. E. (January 2012), Analysis and manipulation of modal ratios of cylindrical drums, in *The Journal of the Acoustical Society of America*, volume 131, no. 1, pages 907–913, ISSN 0001-4966. [Cited on Page 24.]

Ritz, D. (September 2016), Understanding Skins – Resonant Drumhead Explained, URL `https://www.thenewdrummer.com/resonant-head-explained`, [Online; accessed 20/2/2022]. [Cited on Pages 16 and 17.]

Rossing, T. D., Bork, I., Zhao, H. and Fystrom, D. O. (July 1992), Acoustics of snare drums, in *The Journal of the Acoustical Society of America*, volume 92, no. 1, pages 84–94, ISSN 0001-4966. [Cited on Page 10.]

Saitis, C. and Weinzierl, S. (May 2019), *The Semantics of Timbre*, pages 119–149, ISBN 978-3-642-54671-6. [Cited on Page 8.]

Savage, S. (2011), *The Art of Digital Audio Recording: A Practical Guide for Home and Studio*, Oxford University Press, USA, ISBN 9780195394092. [Cited on Pages 25, 26, and 30.]

Schroedl, S. (2002), *Drum Tuning: The Ultimate Guide*, Instrument Instruction, H. Leonard, ISBN 9780634039256. [Cited on Pages 21, 22, and 23.]

Schroth, M. (March 2020), The Heart of Your Drums: A Guide to the Sound of Different Shell Materials, URL https://www.long-mcquade.com/blog/576/, [Online; accessed 8/3/2022]. [Cited on Page 11.]

Senior, M. (June 2008), Kick & Snare Recording Techniques: Tried & Tested Techniques From 50 Top Producers, URL https://www.soundonsound.com/techniques/kick-snare-recording-techniques, [Online; accessed 8/7/2022]. [Cited on Pages 32, 33, 36, 54, and 72.]

Senior, M. (2018), *Mixing Secrets for the Small Studio, 2nd Edition*, Routledge. [Cited on Pages 30, 68, and 93.]

Senior, M. and White, P. (August 2001), Using EQ, Understanding Instrument & Voice Frequency Ranges, URL https://www.soundonsound.com/techniques/using-eq, [Online; accessed 2/8/2022]. [Cited on Pages 66, 68, and 69.]

Seymour, M. (August 2010), Engineer's Guide To Tuning and Damping Drums, URL https://www.soundonsound.com/techniques/engineers-guide-tuning-and-damping-drums, [Online; accessed 20/6/2022]. [Cited on Pages 56 and 85.]

Solid State Logic (October 2010), *AWS 924-948 Owner's Manual*, SSL, Begbroke, Oxford, England. [Cited on Page 65.]

Southall, C., Wu, C.-W., Lerch, A. and Hockman, J. (October 2017), MDB Drums: An Annotated Subset of MedleyDB for Automatic Drum Transcription. [Cited on Page 93.]

Stewart, I. (September 2021), How to Use Dynamic EQ in Mastering, URL https://www.izotope.com/en/learn/how-to-use-dynamic-eq-in-mastering.html, [Online; accessed 12/3/2022]. [Cited on Page 88.]

Talbot-Smith, M. (2017), *Sound Engineer's Pocket Book*, Taylor & Francis Group, ISBN 9781138412767. [Cited on Page 34.]

Tashev, I. (2009), *Sound Capture and Processing: Practical Approaches*, Wiley, ISBN 9780470994436. [Cited on Page 34.]

Thibodeaux, W. M. (May 2022), What Are the Different Types of Snare Drum Heads?, URL https://www.musicalexpert.org/what-are-the-different-types-of-snare-drum-heads, [Online; accessed 15/7/2022]. [Cited on Page 15.]

Tindale, A. (March 2004), Classification of snare drum sounds using neural networks, in *McGill University*. [Cited on Page 51.]

Tindale, A., Kapur, A. and Fujinaga, I. (January 2004), Towards Timbre Recognition of Percussive Sounds, in *ICMC*. [Cited on Page 93.]

Tingen, P. (July 2005), Recording Queens Of The Stone Age, URL https://www.soundonsound.com/techniques/recording-queens-stone-age, [Online; accessed 26/10/2022]. [Cited on Page 32.]

Tomczak, M., Drysdale, J. and Hockman, J. (September 2019), Drum Translation for Timbral and Rhythmic Transformation, in *Proceedings of the International Conference on Digital Audio Effects*. [Cited on Page 100.]

Toulson, R. (October 2009), The Perception and Importance of Drum Tuning in Live Performance and Music Production, in *Journal on the Art of Record Production*. [Cited on Page 85.]

Toulson, R. (2021), *Drum Sound and Drum Tuning: Bridging Science and Creativity*, Audio Engineering Society presents, Routledge, Taylor & Francis Group, ISBN 9780367611194. [Cited on Pages 9, 13, 15, 16, 21, 23, 24, 25, 67, and 92.]

VandeStadt, A. (June 2022), Drum Hoops Explained, URL `https://www.sweetwater.com/insync/drum-hoops-explained/`, [Online; accessed 6/7/2022]. [Cited on Page 18.]

Vinson, W. L. (May 2012), Snare Wires from Yesteryear, URL `http://blog.bostondrumbuilders.com/2012/05/snare-wires-from-yesteryear.html`, [Online; accessed 24/7/2022]. [Cited on Pages 19 and 20.]

Wachtel, D. (Novemeber 2022), Do Snare Wires Make A Difference To Your Sound?, URL `https://aboutdrumming.com/do-snare-wires-make-a-difference-to-your-sound`, [Online; accessed 21/12/2022]. [Cited on Page 20.]

Wagner, A. (2006), *Analysis of Drumbeats: Interaction Between Drummer, Drumstick and Instrument*, Kunglia Tekniska Högskolan. [Cited on Page 22.]

Wagner, A., Dahl, S. S. and Askenfelt, A. (2005), Analysis of Drumbeats – Interaction between Drummer, Drumstick and Instrument. [Cited on Page 51.]

Waxman, G. (August 2022a), 4 Best Snare Wires To Enhance Your Snare Sound, URL `https://drumhelper.com/accessories/best-snare-wires/`, [Online; accessed 25/10/2022]. [Cited on Page 20.]

Waxman, G. (July 2022b), Triple Flange Hoops vs Die Cast Hoops: Ultimate Guide, URL `https://drumhelper.com/drum-hardware/triple-flange-hoops-vs-die-cast-hoops/`, [Online; accessed 26/8/2022]. [Cited on Page 18.]

Waxman, G. (October 2022c), What Is the Best Wood for Drum Shells?, URL `https://drumhelper.com/drums/best-wood-for-drum-shells/`, [Online; accessed 1/12/2022]. [Cited on Page 12.]

Weekhout, H. (2019), *Music Production: Learn how to Record, Mix, and Master Music*, Routledge, ISBN 9780429860898. [Cited on Pages 68 and 69.]

Westera, B. (August 2018), Does Drum Material Actually Matter?, URL `https://drumheadauthority.com/articles/forget-the-best-drum-wood/`, [Online; accessed 16/8/2022]. [Cited on Page 12.]

White, P. (October 2006a), Mixing Essentials: The Fundamentals Of Mixing, URL `https://www.soundonsound.com/techniques/mixing-essentials`, [Online; accessed 15/12/2022]. [Cited on Page 63.]

White, P. (September 2006b), SOS Guide To Choosing & Using Studio Microphones, URL `https://www.soundonsound.com/sound-advice/sos-guide-choosing-using-studio-microphones`, [Online; accessed 16/9/2022]. [Cited on Page 63.]

Wise, D. K. (March 2009), Concept, Design, and Implementation of a General Dynamic Parametric Equalizer, in *Journal of the Audio Engineering Society*, pages 16–28. [Cited on Page 91.]

World Of Music (January 2015), Snares: What's the Difference?, URL https://www.worldofmusic.com.au/blog/snares-whats-the-difference-in-snare-drum-materials/, [Online; accessed 13/11/2022]. [Cited on Pages 10, 13, and 14.]

Worrell, B. (Decemeber 2015), A Beginners' Guide to Drum Mic'ing, URL https://reverb.com/uk/news/a-beginners-guide-to-drum-micing, [Online; accessed 21/7/2022]. [Cited on Page 54.]

Yoshii, K., Goto, M. and Okuno, H. (January 2005), INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums, in *Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, pages 205–212. [Cited on Page 86.]

Zwicker, E. (1961), Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen), in *The Journal of the Acoustical Society of America*, volume 33, page 248. [Cited on Page 57.]

Appendix A

Publications

Audio Engineering Society

Convention Paper 10040

Presented at the 145th Convention
2018 October 17 – 20, New York, NY, USA

Microphone Comparison for Snare Drum Recording

Matthew Cheshire, Jason Hockman, and Ryan Stables

Digital Media Technology Lab (DMT Lab), Birmingham City University

Correspondence should be addressed to Matthew Cheshire (`Matthew.Cheshire@bcu.ac.uk`)

ABSTRACT

We present two experiments to test listener preference for snare microphones within real-world recording scenarios. In the first experiment, listeners evaluated isolated recordings captured with 25 microphones. In the second experiment, listeners performed the same task with the addition of a kick drum and hi-hat as part of a performed drum sequence. Results indicate a prominent contrast between the highest and lowest rated microphones and that condensers were rated higher than other subsets tested. The preference for three microphones significantly changed between the two listening test conditions. A post-test survey revealed that most listeners compared high-frequency characteristics, which were measured using spectral features. A positive correlation was observed between test scores of cardioid microphones and the brightness feature.

1 Introduction

The Shure SM57 has long been the go to choice of recording engineers when close miking the snare drum [1], [2], [3], [4]. Close microphones or "spot" mics are used in conjunction with overhead microphones (often a matched stereo pair of microphones placed above the drums to capture every element of the kit) in order to enhance specific parts of the kit, most notably the snare and kick drum [5], [6]. This technique gives the mixing engineer more control over how they manipulate the recordings with EQ and compression whilst not affecting the tonality of the other drums. Manufacturers, such as Audix and Beyerdynamic, produce microphones specifically designed for recording snare drums.

In this paper, a comparison of different microphones is undertaken to investigate whether microphone selection plays an important role in preference of snare drum recording. The purpose of carrying out subjective comparisons is to help users of such equipment make informed choices. One study of tom drum miking [7] compared three different microphones, two dynamic microphones and one condenser microphone, this study also incorporated multiple microphone positions and different drum head materials. Comparisons between microphones are also often made informally for magazine and online articles, with the intention of providing readers insight into a range of microphones for a specific application. Fourteen microphones for vocal recording were compared based on sonic characteristics of four singers [8]. Another informal online article [9] used two condenser and nine dynamic microphones for snare drum recording. While these results provide useful information to potential users, they are not scientific comparisons and the listening tests are only carried out on a short list of microphones. In general, listening tests may be conducted and samples were presented for listeners to make their own judgements.

The remainder of this paper is structured as follows:

Section 2 presents our methodology for determining preference for snare drum microphone selection and the results of the experiments are presented in Section 3. A discussion is then provided in Section 4, and conclusions presented in Section 5.

2 Method

2.1 Experimental Design

In order to determine whether preference plays a role in the selection of snare drum microphones, two experiments were carried out: the snare drum played on its own (`Single Hits`) and the snare drum played as part of a beat involving a hi-hat and kick drum (`Hits With Bleed`). The latter scenario being more relevant to real world applications of snare drum recording. There may however be situations where the snare is recorded on its own, for example, sample recording. It is also common for engineers in live or studio settings to request the drummer to play each element of the drum kit separately to assess the quality and timbre. Comparison of these two tests would show if microphone preference for snare drum changes with the presence of extraneous sounds from the other elements of the drum kit, often referred to as “spill” or “bleed”.

2.2 Microphones

Table 1 presents a list of the microphones used in the recording experiments. The total number of microphones used was 25, comprised of 15 dynamic (D) microphones and 10 condenser (C) microphones. Out of these microphones 14 had a cardioid polar pattern, eight had supercardioid and three had hypercardioid. Specification were taken from the manufacturers websites. All microphones used were commercially available at the time of publication. Microphones were selected based on availability and appropriateness, only small diaphragm condensers were used as large diaphragm condensers are often difficult to position between the hi-hat and the rack tom without obstructing the drummer. Several of the microphones had built in filters, such as the Sennheiser MD421 which has a five position bass roll off switch however the recordings were carried out with all filters switched off and no additional processing.

Brand	Model	Type	Polar Pattern
AKG	C451B	C	Cardioid
Audix	ADX51	C	Cardioid
Audix	D2	D	Hypercardioid
Audix	D4	D	Hypercardioid
Audix	i5	D	Cardioid
Beyerdynamic	M201	D	Hypercardioid
DPA	4099	C	Supercardioid
Electro Voice	PL80	D	Supercardioid
Electro Voice	RE20	D	Cardioid
Neuman	KM184	C	Cardioid
Rode	M2	C	Supercardioid
Rode	M3	C	Cardioid
Rode	NT5	C	Cardioid
Rode	NT55	C	Cardioid
Sennheiser	e609	D	Supercardioid
Sennheiser	e614	C	Supercardioid
Sennheiser	MD421	D	Cardioid
Sennheiser	Md441	D	Supercardioid
Shure	Beta57a	D	Supercardioid
Shure	SM57	D	Cardioid
Shure	SM7B	D	Cardioid
T.Bone	CC100	C	Cardioid
T.Bone	CD55	D	Cardioid
T.Bone	MB75	D	Cardioid
Telefunken	M80	D	Supercardioid

Table 1: List of microphones used in both recording experiments.

2.3 The Snare Drum

The snare drum used in this study and other variables that affect the timbral qualities were carefully considered to produce a sound representative of a typical snare drums characteristics. Common drum heads were used as well as tuning the heads appropriately for a wide range of genres. This provided a generalisable and realistic scenario of drum recording. A Mapex Black Panther Velvetone 14” x 5.5” snare drum was used. It had an 8.1mm shell consisting of a 3 mm exterior burl maple outer layer, enclosing a 3.4 mm walnut wood middle layer and a 1.7 mm maple interior layer. The drum had 10 tension rods for the batter head and 10 for the resonant head, as well as 20 strand snare wires. An Evans B14HBG Hydraulic drumhead was used on the batter side and a Remo Ambassador Black Suede Snare side for the resonant head.

A digital DrumDial was used to tune both the batter and resonant heads. This device ensured that the tension of the heads were uniform around the drum and allowed for accurate, repeatable tuning. The tension was set to 90 for the batter head at every tension rod position and set to 80 for the resonant head. These tension settings were suggested by the DrumDial tuning chart based on the dimensions of the snare drum and the types of heads used. Once the drum was tuned a BigFatSnare-Drum dampening disk was placed on the batter head, this was to reduce excessive overtones of the drum. This device was chosen over other products such as MoonGel Dampening pads as the placement of the Big-FatSnareDrum takes up the whole drum head, ensuring placement repeatability unlike devices that could be placed anywhere on the head. The drum sticks used were Vic Firth 5B Nova Hickory wood tip sticks.

2.4 Recordings

All recordings[1] were undertaken in an acoustically treated studio control room, with an ambient noise level of ~20 dBA. This was opted for over a larger studio live room for its shorter reverberation time and flatter frequency response, which minimized the impact of the room on the character of the recordings. The microphone was positioned at 60° with the diaphragm of each microphone placed 10 cm above the rim of the drum, pointing directly at the the centre of the drum head. This position was chosen for consistency as it was easy to replicate with every microphone. The utmost care was taken to ensure each microphones position was matched as accurately as possible. A triangular jig was used to aid the alignment of the microphones (Fig. 1), This measured 10 cm x 17.78 cm x 20.4 cm, where applicable the distance of the diaphragms location in relation to the external grill was compensated for.

The drummer (with over 9 years experience) attempted to maintain consistency of velocity and striking position throughout. A recording was made by each microphone consisting of four hits of the snare drum. In addition to this, without moving the microphone away from the snare drum, four hits of the kick drum and four hi-hat hits were also recorded, as well as four hits of kick drum and hi-hat played simultaneously. A four-bar phrase was played to a 110 BPM metronome recorded using each microphone, maintaining the same position

[1] 16-bit .wav recordings can be downloaded from dmtlab.bcu.ac.uk/matthewcheshire/audio/aes145

Fig. 1: Position of M201 with triangle jig.

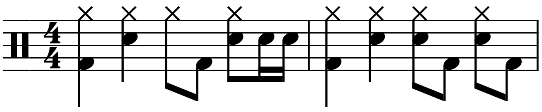

Fig. 2: Score for drum beat used in `Hits With Bleed` recording experiment.

for individual snare hits. No additional microphones were used for the kick drum or hi-hat, and these were captured solely through the snare microphone. The musical score for the drum beat is shown in Fig. 2. The recordings were captured using a Metric Halo ULN-2 2D analogue to digital converter into Pro Tools 12 running at 32 bit-float and 44.1 kHz. The level of the microphone preamplifier was set so that no clipping occurred for any recording. The tuning of the drums was checked with the DrumDial after every third recording.

2.5 Audio Pre-Processing

The four-bar phrase was manually edited to ensure quantization of all drum hits to the beat. This removed any of the player's timing variation from the recordings. The whole phrase was then normalised to -23 LUFS which removed any loudness variation between samples [10]. For the individual hits, each hit was separately normalised to -23 LUFS, to ensure that the perceived loudness between hits and different microphones was as consistent as possible for the listening test.

2.6 Listening Test

The listening test was carried out in the same studio that the recordings were captured. The speakers used were PMC IB1S, with a Bryston 2B-SST2 amplifier and an RME Fireface 802 audio interface. The test was conducted using the Web Audio Evaluation Tool (WAET) [11] with the APE interface [12]. A multi-stimuli approach was used over an A/B comparison to minimise the duration of the tests as in a previous study [13] results from multi-stimuli and AB test were found to produce comparable results. Two listening tests were carried out, one evaluating the `Single Hits` recordings and one evaluating the `Hits With Bleed`, these two tests were presented to participants in a random order. It required participants to position 25 markers corresponding to each audio sample along a slider, leftmost representing least preferred and rightmost for most preferred. Participants were instructed to "rate the audio samples based on the quality of just the snare drum. Using the full range of the scale". Selecting a marker would play the sample on a loop, and the loop position was maintained when switching between sample for uninterrupted playback. Participants could not complete the test until every sample had been played at least once and the marker had been moved from its original position. Starting position and marker number were randomised for every participant and for each test.

2.7 Participants

Twelve participants took part in the listening test, all of which had previous experience using both condenser and dynamic microphones in studio or live sound applications. The range of the subjects age was 22 to 48 with a mean of 27. The participants were asked how many years experience they had in sound recording/audio production related fields (range: 3–27 years, mean: 10 years). The participants took on average 8 mins 54 s to complete the `Hits With Bleed` test, and 9 mins 14 s to for the `Single Hits` test.

3 Results

A one-way analysis of variance (ANOVA) was carried out on the results of the listening test to determine if the differences between the mean scores of each microphone was significant. `Single Hits` *p*-value = 5.581e-8 and `Hits With Bleed` *p*-value = 0.0045.

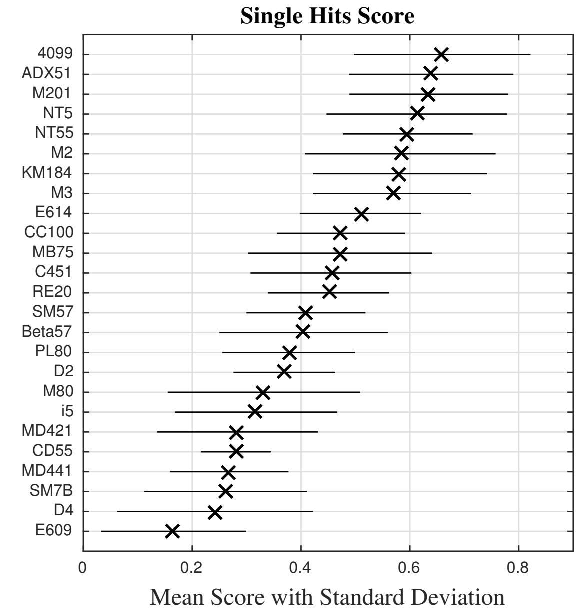

Fig. 3: Mean score (x) and standard deviation (horizontal lines) for all microphones tested in the `Single Hits` listening test.

This shows that the mean scores of the microphones are significantly different.

In Fig. 3 and Fig. 4 results are presented in order of their mean score, where the horizontal line shows the standard deviation for each microphone and the cross shows the mean across participants. A paired T-Test was used to compare the results from the `Single Hits` and the `Hits With Bleed` test. This was used to determine if scores for any microphone were significantly reduced or improved between the two conditions. As can be seen in Table 2, three microphones had significant differences; the DPA 4099 and the Audix ADX51, which had the two highest mean scores for the `Single Hits` test both received significantly lower score for the `Hits With Bleed` test. However the Audix D4 significantly improved with the addition of bleed.

4 Discussion

4.1 Ranking the Data

Although participants were asked to use the full range of the scale (i.e., placing their least preferred at 0 and their most preferred at 1), only one participant used the maximum rating, and two participants used

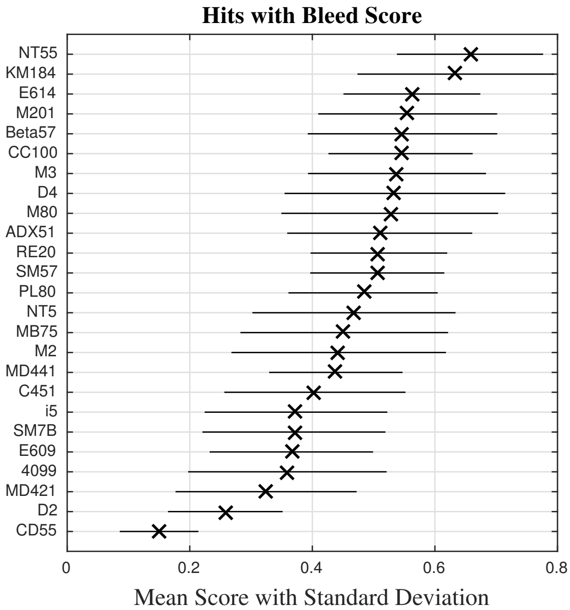

Fig. 4: Mean score (x) and standard deviation (horizontal lines) for all microphones tested in the `Hits With Bleed` listening test.

Microphone	Single hits mean score	Hits with bleed mean score	Paired T-Test *p* value
ADX51	0.63	0.51	0.04
Audix D4	0.24	0.53	0.02
DPA 4099	0.65	0.35	0.01

Table 2: Paired T-test results, showing microphones with significant *p*-values.

the minimum rating. Normalisation of data is a common procedure used to compare results between participants. In this case however, normalisation to the entire scale might misrepresent the intention of the participant (e.g., moving similarly scored microphones farther away from each other). Alternatively, we chose to assess the rank order as it is a robust against normalisation, and thereby a more comparable measure between participants than the raw data. Once the data was ranked, a clear preference for condenser microphones over dynamics was observed. The top eight out of ten ranked microphones in the `Single Hits` test were condensers, with the average rank being five places above the average rank of the dynamic microphones. For the `Hits With Bleed` test, condensers only made up five of the top ten ranked microphones and the average rank was two places above the average rank of dynamics.

4.2 Brightness

Once a participant had completed the test they were asked "What qualities of the samples were you comparing?" the answers included: "Resonant frequencies on the snare, depth, clarity", "brightness, fullness", "punch, crispness", "sharp transient, bright", "how hard the top-end sounded", "the attack and frequency content", "punch, warmth, highs", and "the tone of the snare, the snap of the impact". From the responses, frequency content—in particular the high-frequency energy—seemed to be an attribute to which participants were listening. Two spectral features, spectral centroid and brightness are used to measure the high frequency characteristics mentioned by participants in the post-test survey. The spectral centroid refers to the center of gravity of the frequency spectrum. This can be a good indication of how bright a sound is perceived [14], as a higher spectral centroid contains more energy within higher frequencies than in lower frequencies. We use the [15] definition of brightness, which measures the amount of spectral energy above 1.5 kHz, the result is expressed as a number between 0 and 1. The mean spectral centroid for `Single Hits` was 2.6 kHz, and for the `Hits With Bleed` was 5 kHz. The spectral centroid for each microphone was higher for the `Hits With Bleed`. The brightness for every microphone was higher for `Hits With Bleed` in comparison to the `Single Hits`, with a mean increase across all mics of 0.13. This increase is likely caused by the addition of the hi-hat cymbal placed to the side of the snare. It was noted that the mean increase in brightness for the cardioid microphones was 0.14, while the mean for hypercardioid and supercardioid mics together was a 0.12 increase. This demonstrates the effect of off-axis rejection of the directional pickup patterns (i.e., hypercardioid, supercardioid) on the amplitude of the hi-hat bleed from the side of the microphones.

Spearman correlation, with a significant *p*-value ($p >$ 0.05) was used to observe both the relationship between the spectral centroid of a sample and it's mean rank, and between brightness and mean rank.

A positive correlation was found between the centroid and the mean rank ($R = 0.51$, $p = 0.008$) for the `Single Hits` (Fig. 5). There was no correlation between the

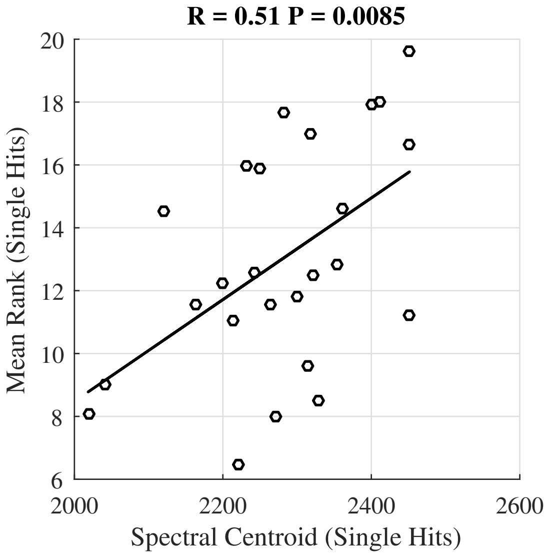

Fig. 5: Spectral centroid and mean rank for `Single Hits`, with regression line.

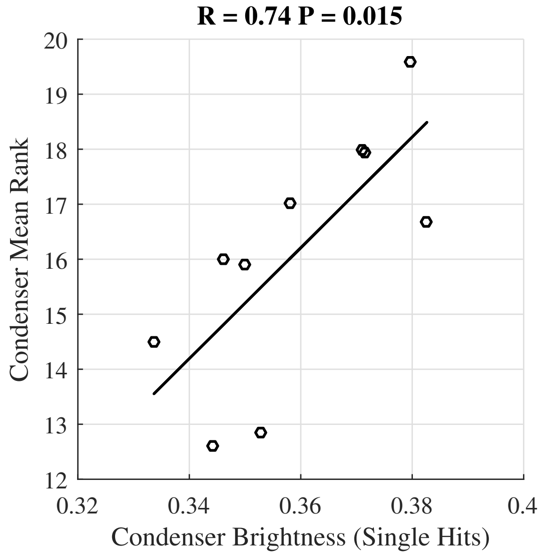

Fig. 6: Condenser brightness and condenser mean rank for `Single Hits`, with regression line.

`Hits With Bleed` and the mean rank (R = 0.28, p = 0.17) and no significant correlation between any subsets (e.g., polarity, type) of the microphones.

When taking all the samples together no correlation was found between brightness and mean rank for either `Single Hits`, or `Hits With Bleed`. However when the condenser and dynamic microphones were analysed separately, positive correlation was found for condenser microphone brightness for `Single Hits`, and the mean rank (R = 0.74, p = 0.015) (Fig. 6). This indicates that for the condenser used for `Single Hits` the “brighter” microphones received higher rank scores.

When taking only the hypercardioid and supercardioid microphones into account, a negative correlation (R = -0.65) was observed between the change in mean rank and the change in brightness across the two different tests (Fig. 7). The change in brightness can be described as the influence of the hi-hat on the microphones brightness. A relatively small increase in brightness from the `Single Hits` to `Hits With Bleed` means the high frequency bleed from the hi-hat was not as prevalent than had there been a much larger relative increase in brightness. The change in mean rank could be described as how much better or worse the rank was with the addition of the hi-hat and kick drum. Positive values show higher mean rank for the `Hits With Bleed`, whereas values below zero show where microphones mean rank reduced, and zero would show where the microphones rank was the same for both tests. The negative correlation (R = -0.65, p = 0.03) in Fig. 7 shows that as the change in brightness increases, the change in mean rank decreases. This would indicate that hypercardioid and supercardioid microphones that have a higher rejection of high frequencies are more likely to be ranked higher when used for `Hits With Bleed` over `Single Hits`.

4.3 Signal-to-Bleed Ratio (SBR)

The most common use of a snare microphone is when the rest of the drum kit is also being played. For this reason, it is important to examine the microphones behaviour when used in a real world application. The amount by which the microphone captures these other drums as well as the snare, is likely to affect listener preference to some degree. As previously mentioned, as well as isolated snare hits, the kick drum and hi-hat were also recorded through the snare microphone without the position being changed. These recordings were used to quantify the amount of bleed picked up by every microphone. A ratio was taken between the RMS of the snare hits and the RMS of the bleed. To calculate the signal-to-bleed ratio (SBR), we calculate the sum RMS of each drum hit (Sn) of a given microphone, and for the corresponding bleed recording (Bn) of the same microphone.

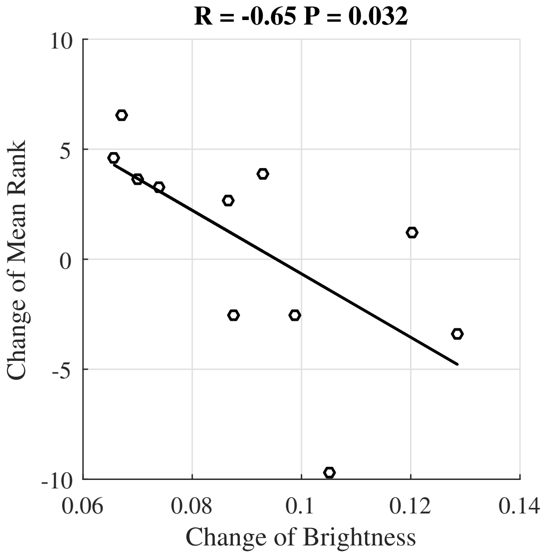

Fig. 7: Change in brightness and change of mean rank, with regression line.

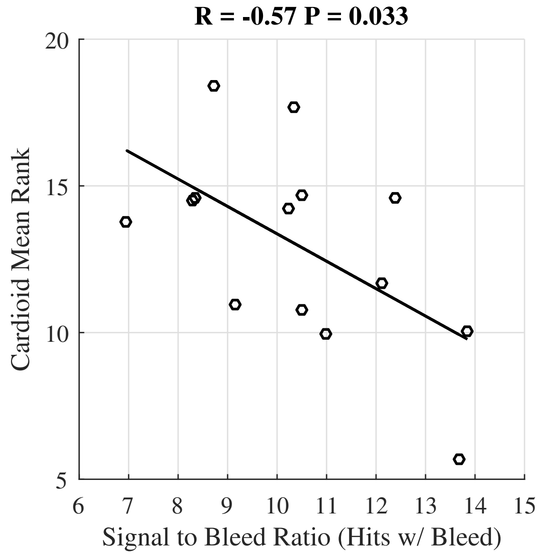

Fig. 8: Cardioid signal to bleed ratio and mean rank, with regression line.

$$SBR = \frac{\sum_{n=0}^{N-1} RMS(Sn)}{\sum_{n=0}^{N-1} RMS(Bn)} \tag{1}$$

A high ratio indicated that the signal captured in front of the microphone is stronger than the signal off axis, while a low ratio shows the bleed is close to or as strong as the snare signal. The mean SBR for all hypercardioid and supercardioid microphones was 12, whereas the mean for all cardioid microphones was 10.4. Spearman correlation was used to examine the relationship between the SBR and the mean rank of the microphones. Interestingly, a negative correlation was observed between the mean rank and the SBR of the cardioid microphones used in the `Hits With Bleed` test (Fig. 8). This shows that as the SBR decreases, the mean rank increases, or the better ranked cardioid microphones are those with worse off-axis rejection. A potential explanation for this could be that the listening test participants preferred a more complete sounding drum beat that included the hi-hat and kick drum, which they are more accustomed to hearing in produced music. Therefore the stronger the presence of bleed, the more they preferred those microphones. This effect could be investigated further through the inclusion of omnidirectional microphones to capture even greater bleed from the other drums.

5 Conclusions

In order to determine if microphone selection plays a role in the preference of snare drum recording, we performed two experiments with 25 different microphones. These experiments were designed to mimic real world recording scenarios. The results of the listening test demonstrate a clear disparity in score between the highest and lowest rated microphones. However, due to the broad standard deviation of the scores, providing conclusions regarding the preference of microphones with close mean scores is not possible. Other factors outside the scope of this study, such as microphone positioning, may also dictate microphone performance. The paired T-test revealed a significant change in score for three microphones between the two recording experiments, with the Audix D4 microphone receiving a better score for `Hits With Bleed`, and the Audix ADX51 and the DPA4099 getting the lowest scores. This highlights the importance of choosing the right microphone for a specific application. Of the subsets assessed (i.e., polar pattern, type), the condenser microphones demonstrated the strongest correlation with the mean rank. This effect could be further investigated by artificially enhancing the brightness of the recordings through equalisation to determine the degree to which brightness improves the score.

References

[1] Owsinski, B., *The Recording Engineer's Handbook*, Gale virtual reference library, Thomson Course Technology, 2005, ISBN 9781932929003.

[2] Huber, D. and Runstein, R., *Modern Recording Techniques*, Audio Engineering Society Presents, Focal Press/Elsevier, 2010, ISBN 9780240810690.

[3] Senior, M., "Kick & Snare Recording Techniques," *www.soundonsound.com*, 2008.

[4] Tingen, P., "Recording Queens Of The Stone Age," *www.soundonsound.com*, 2005.

[5] Owsinski, B. and Moody, D., *The Drum Recording Handbook*, MusicPro guides, Hal Leonard Books, 2009, ISBN 9781423443438.

[6] Major, M., *Recording Drums: The Complete Guide*, Course Technology PTR, 2014, ISBN 9781133788935.

[7] Quiroga, A. F., Garcia, J. D., and Páez, D., "Tom–Tom Drumheads Miking Analysis," in *Audio Engineering Society Convention 138*, 2015.

[8] Houghton, M., "Vocal Mics: Matching Mics and Voices," *www.soundonsound.com*, 2010.

[9] Fuston, L., "Best Mics for Recording Snare Drums," *www.sweetwater.com*, 2017.

[10] EBU-R-128, *Loudness Normalisation and Permitted maximum Level of Audio Signals*, 2014.

[11] Jillings, N., Man, B. D., Moffat, D., and Reiss, J. D., "Web Audio Evaluation Tool: A Browser-based Listening Test Environment," in *12th Sound and Music Computing Conference*, 2015.

[12] Man, B. D. and Reiss, J., "APE: Audio Perceptual Evaluation Toolbox for MATLAB," in *136th Convention of the Audio Engineering Society*, 2014.

[13] Man, B. D. and Reiss, J., "A Pairwise and Multiple Stimuli Approach to Perceptual Evaluation of Microphone Types," in *134th Convention of the Audio Engineering Society, Rome*, 2013.

[14] Grey, J. and Gordon, J., "Perceptual effects of spectral modifications on musical timbres," 63, pp. 1493–1500, 1978.

[15] N. Juslin, P., "Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception," 26, pp. 1797–813, 2000.

Audio Engineering Society

Convention Paper 10263

Presented at the 147th Convention
2019 October 16 – 19, New York

Microphone comparison: Spectral feature mapping for snare drum recording

Matthew Cheshire, Ryan Stables, and Jason Hockman

Digital Media Technology Lab (DMT Lab), Birmingham City University

Correspondence should be addressed to Matthew Cheshire (`Matthew.Cheshire@bcu.ac.uk`)

ABSTRACT

Microphones are known to exhibit sonic differences and microphone selection is integral in achieving desired tonal qualities of recordings. In this paper, an initial multi-stimuli listening test is used to categorise microphones based on user preference when recording snare drums. A spectral modification technique is then applied to recordings made with a microphone from the least preferred category, such that they take on the frequency characteristics of recordings from the most preferred category. To assess the success of the audio transformation, a second experiment is undertaken with expert listeners to gauge pre- and post-transformation preferences. Results indicate spectral transformation dramatically improves listener preference for recordings from the least preferred category, placing them on par with those of the most preferred.

1 Introduction

Microphone choice plays a key role in achieving the desired sonic characteristic of an audio recording [1, 2]. Particular microphones are known to work well with certain instruments or for certain styles of music [3]. Differences between microphones occur based on their physical construction, which affects properties such as frequency response and polar pattern [4, 5, 6]. With a plethora of microphones for the recording engineer to choose from, microphone selection is often based on personal experience and preference [7, 8].

In a previous study [9], an evaluation was undertaken that ranked recordings of a single snare drum from multiple microphones. Positive correlation was observed between listening test preference scores and spectral energy above 1.5 kHz. This indicates that the frequency response of the microphones is in part responsible for preference. Another study [10] found that the perceptual attributes of *brightness, harshness* and *clarity* contributed the most to describing inter-microphone differences—descriptors closely related to frequency content.

One important process during mixing is the enhancement and correction of the frequency content of a recorded track [11], performed using an equaliser (EQ)—an analogue or digital effect that boosts or attenuates specific user defined frequencies. McKinnie [12] suggests that when microphones of similar build type and polar-pattern are equalised to have near identical on-axis frequency response they exhibit varied timbral qualities, yet this claim was not investigated. The study instead aimed to identify the most salient perceived differences between the nine condenser microphones under evaluation. However, the study found that listeners could not distinguish between many of the stimuli.

Hebrock et al. [13] developed a method for measuring time domain responses of 25 microphones to understand why microphones with similar performance features are perceived differently by listeners. The results proved inconclusive due to the large amount of variables. It was noted that the frequency responses of microphones under evaluation played a role in the characterisation of the sound, affecting listener opinion.

The frequency response of microphones referenced in previous studies caused the greatest perceptual variance. This suggests that modifying the spectral characteristics of a recording through equalisation could improve subjective responses. In order to determine this, a comprehensive multi-stimulus listening test is conducted with 12 microphones across four distinct snare drums, to identify a ranked categorisation of microphones. To evaluate the effects of spectral modification, the least-preferred microphone is transformed to take on the frequency response of the most highly-preferred microphones. A second listening experiment is then conducted to determine the extent to which the preference of the least-preferred microphone has been improved.

The remainder of this paper is structured as follows. Section 2 presents the methodology used for creating the recordings used for analysis. Section 3 outlines the methods used for evaluating listener preference of the recordings and presents the findings. Section 4 details how the spectral features of the microphones are analysed and transformed. Section 5 presents the methods and results from a pre- and post-transformation A/B listening test. Conclusions and suggestions for future work are presented in Section 6.

2 Methods

The listening tests in this study utilise professional quality recordings of consistent snare drum performances. In all, 12 microphones are selected across a range of variables including cost and manufacturer (Section 2.1). To ensure that listener preference is not an effect of snare drum selection, multiple snare drums are selected with varying configurations and specifications (Section 2.2). Great care is taken to ensure that the recording equipment and procedure are of a professional level (Section 2.3) and that snare drum excitation is as consistent as possible (Section 2.4).

Brand	Model	Type	Polar Pattern
AKG	`D5`	D	Supercardioid
Beyerdynamic	`M201`	D	Hypercardioid
DPA	`4011A`	C	Cardioid
Neumann	`KM184`	C	Cardioid
RØDE	`M5`	C	Cardioid
RØDE	`NT55`	C	Cardioid
SE	`V7X`	D	Supercardioid
Sennheiser	`e614`	C	Supercardioid
Sennheiser	`MKH40`	C	Cardioid
Shure	`Beta57a`	D	Supercardioid
Shure	`SM57`	D	Cardioid
Telefunken	`M80`	D	Supercardioid

Table 1: Makes, models, types and polar patterns of dynamic (D) and condenser (C) microphones used in experiments.

2.1 Microphone Selection

In this study, six small diaphragm condenser microphones and six dynamic microphones were selected. Table 1 shows the full list of microphones used. Microphones are selected from a range of available manufacturers, as well as recommendations from recording engineers and online articles (e.g., [14, 15]). Only microphones that were commercially available at the time of this study and deemed appropriate for snare recording were selected. Microphones that could not be positioned without obstructing a drummer or specialist microphones (e.g., kick drum microphones) were excluded. A previous study [9] that compared solo snare drum recordings, and snare drum recordings including a kick drum and hi-hat found that listener preference for the majority of microphones did not significantly change between the two recording scenarios. However, as the preference for three microphones changed significantly, these three microphones were excluded from this study as only recordings of solo snare drums were investigated. These microphones were: Audix `ADX51`, Audix `D4`, and DPA `4099`.

2.2 Snare Drums

Table 2 presents the four snare drums selected for this study, including two steel shell drums and two maple shell drums. All four drums were 14" in diameter, used

Snare	Batter Head	Shell	Depth
Black Panther Machete	Evans Hydraulic	Steel	6.5"
Black Panther Velvetone	Evans Hydraulic	Maple	5.5"
Premier Artist Maple	Evans HD Dry	Maple	5.5"
Tama Rockstar	Tama Power Craft II	Steel	6.5"

Table 2: Configurations and specifications of four snare drums used in experiments.

a Remo Weatherking Ambassador Hazy Snare Side[1] for the resonant head, and were fitted with 20 strand snare wires. The resonant and batter heads of the snare drums were tuned with the aid of a digital DrumDial[2] to ensure uniform tension at every lug position, allowing for repeatable tuning. A 1" Evans E-ring[3] was placed on the batter head to dampen excessive overtone.

2.3 Recording

The recordings were carried out in a sound-treated isolation recording booth measuring H2.5 x W3.0 x L4.5 metres, with an ambient noise level of ~40dBA. Each snare drum was recorded separately using a Metric Halo ULN-2 into Apple Logic Pro X, at 32-bit resolution and 44.1kHz sample rate. The gain of the preamplifier was set to avoid clipping during any recording.

The option to record with all microphones concurrently as in [2, 10] was considered. All microphones could not be recorded simultaneously as microphone positioning is linked to variance in timbral characteristics [16, 17, 18]. As suggested in [19], an ideal microphone comparison test should locate microphones under observation at the exact point in space, maintaining an identical pressure-gradient or soundfield. Recordings were therefore made serially, at a near identical position, as opposed to positioning the 12 microphones around the rim of the snare drum. Great care was taken to ensure the position was matched as accurately as possible. This was achieved by aligning the microphones to a triangular jig (removed from the drum prior to any recording), measuring H10 x W17 x L20 cm, see Figure 1.

Fig. 1: Recording setup demonstrating robotic drum arm, triangular jig and Neumann `KM184` microphone.

To avoid listener fatigue in the subsequent evaluations, a 4-bar rhythmic pattern consisting of 16 snare hits was played at 120 beats per minute (BPM). This aimed to produce a more engaging stimuli than isochronous events for listeners and to provide a more realistic application of snare drum recording.

2.4 Robotic Drum Arm

As the position and velocity at which a drum skin is struck strongly impacts the tonality of the resultant sound, a robotic drum arm (RDA) is used to provide consistent excitation in the serial recording with each microphone (see Figure 1). The RDA is controlled through a MIDI interface with events sequenced in Logic Pro X. An Arduino Uno is used to convert MIDI messages into voltages, thereby switching a relay connected to an actuator that triggered the RDA to move. An elastic band is used to initialise the actuator position after each hit. The striking distance of the drumstick was calibrated to ensure that the stick would not dampen any resonance after it had excited the drum head and would not prevent the drumstick tip from reaching the drum head. A striking distance of 5 cm above the centre of the drum head was chosen to meet this criteria, producing hits of approximately 90dB SPL. The excitation consistency of the RDA was assessed by measuring the MIDI velocities achieved from 500 hits on a MIDI drum pad (*mean*: 118, *std*: 2).

[1] `www.remo.com`
[2] `www.drumdial.com`
[3] `www.evansdrumheads.com`

3 Multi-Stimuli Listening Test

A multi-stimuli listening test was used to evaluate listener preference of recordings generated through the use of different microphones. While previous studies have evaluated listener perception based on semantic descriptors (e.g., *warm*, *bright*) [20], the listening test in the present study was undertaken to produce a categorisation of highly-preferred and least-preferred microphones for snare drum recording prior to spectral modification.

3.1 Methodology

The listening test was performed in an acoustically-treated mastering studio using a pair of PMC IB1S speakers with a Bryston 2B-SST2 amplifier and an RME Fireface 802 digital-to-analogue converter. The tests were conducted using the Web Audio Evaluation Tool (WAET) [21] with the APE interface [22]. The samples were loudness normalised to -23 LUFS using the EBU specification [23] to remove any perceived loudness disparity between samples. As a previous study [2] found results from a multi-stimuli and an A/B test produced comparable findings, a multi-stimuli approach was implemented to minimise test duration. In total, 42 participants took part in the test (*range*: 19 to 49 years, *mean*: 23.7 years, *std*: 6.2 years) with an average of 6.9 years of music production/recording/mixing experience (*range*: 2 to 25 years, *std*: 5.1 years).

The recordings from the 12 microphones were presented one snare drum at a time, resulting in four separate listening tests presented in a random order. Participants were instructed to rank recordings from least-preferred to most-preferred, from left to right. The 12 samples were randomised and represented by large green boxes labeled alphabetically as seen in Figure 2. These were ranked by the listener based on personal preference and moved using a select and drag method. When switching between samples, loop position was maintained for uninterrupted playback. Participants were not able to complete the test until every sample had been played at least once. The average test duration for all four listening tests was 13 min 36 s, with participants spending 3 min 24 s on average on each test.

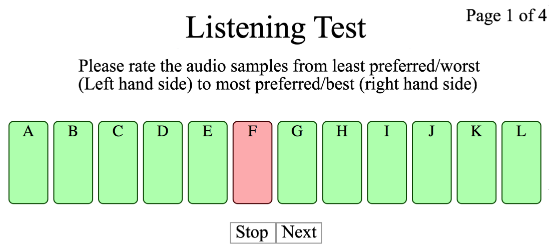

Fig. 2: Web Audio Evaluation Tool (WAET) interface used for multi-stimuli listening test.

3.2 Results

As the participants ranked the stimuli from 1 to 12, the resulting data was ordinal. The Kruskal-Wallis test is used to test the null hypothesis—that is, listeners did not collectively show a preference between stimuli recorded from different microphones. The results from the Kruskal-Wallis test ($p<0.05$) reject this null hypothesis, indicating that listeners do indeed show preference between the different microphone stimuli. This is essential to ascertain prior to any further investigation on the preference of different stimuli.

A certain microphone might be better suited to a particular sonic characteristic inherent in the snare drum by emphasising desired tonal features. By recording multiple snare drums with the same 12 microphones and recording procedure, the impact of snare drum selection on microphone preference can be established. The Kruskal-Wallis test was also used to investigate if the rank of microphones was significantly different for the four drums. For 11 of the 12 microphones there is no statistical significance between rankings across snare drums ($p>0.05$). The only microphone exhibiting snare drum dependent results is the Sennheiser `e614`, ($p<0.05$). Although the test can not be used to differentiate which specific snare rankings are significantly different from each other, post hoc analysis indicates that the mean rank of the Machete snare was significantly lower than the other three snare drums ($p<0.05$). As a result, this microphone is not used in the remainder of this study, leaving a total of 11 microphones (i.e., five condenser microphones and six dynamic microphones).

3.3 Pairwise Comparison

To determine if the subjective rank of each microphone exhibits a significant difference to that of the other

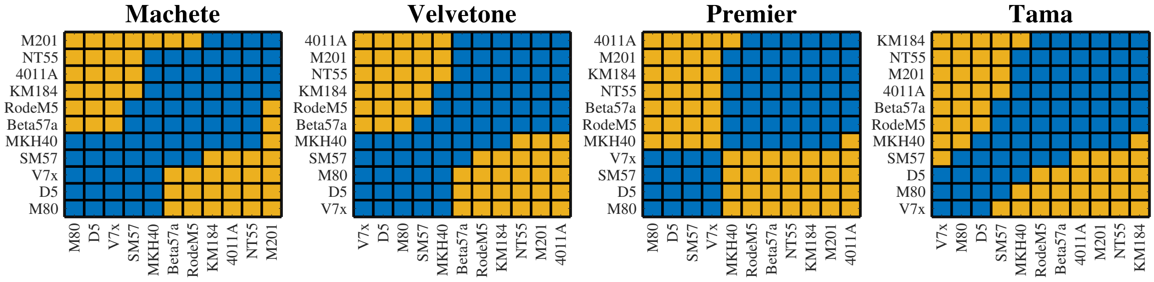

Fig. 3: Results from pairwise comparison test with microphones ordered by mean rank. Yellow squares indicate pairs of microphones that are significantly different from one other (p<0.05) and blue squares indicate pairs that are not significantly different (p>0.05).

microphones, a non-parametric pairwise multiple comparisons test was carried out using Dunn's test [24]. Figure 3 presents the results of the pairwise comparison test, with microphones ordered by the mean rank across participants. Here, the yellow squares depict pairs of microphones which are significantly different, based on the result of Dunn's test (p<0.05). The matrices show that the top four ranked microphones consistently exhibit significant differences to the lowest four ranked microphones. The top four microphones can not be considered to have significantly different ranks from each other, so no single microphone out of these may be considered optimal. Additionally, the lowest four microphones can not be considered to be ranked differently from each other, so no single microphones should be interpreted as the least-preferred, with one exception. For the Tama snare drum, the `SM57` can be considered more highly-preferred than the `V7X`. The remaining three middle-ranked microphones did not have consistent results across all four snares.

Using the results of the multiple pairwise comparison test, the microphones can be classified into three categories as seen in Table 3. The Category-1 classification denotes highly-preferred microphones that ranked in the top four positions for all snares. These were significantly different from the microphones in Category-2, which are the least-preferred microphones ranked in the bottom four positions for all snares. Category-3 includes the remaining microphones which did not fall into either category.

Category-1	Category-2	Category-3
4011A	D5	Beta 57a
KM184	M80	MKH40
NT55	SM57	M5
M201	V7X	

Table 3: Microphone categorisation achieved through multi-stimuli listening test.

4 Spectral Modification

Towards improvement of listener ratings for the least-preferred microphones, an exploratory audio effect is introduced to modify the frequency response of a Category-2 recording such that it mimics those of the Category-1 microphones. The adopted approach is to perform spectral analysis (Section 4.1), followed by a timbral transformation through an graphic EQ stage (Section 4.2). A graphic EQ is preferred over a parametric EQ due to the inclusion of the American National Standards Institute (ANSI) standardisation [25]. The Category-2 microphone chosen for transformation is the `D5`, which is the least expensive Category-2 microphone, allowing for the greatest monetary disparity between the microphone categories.

4.1 Frequency response analysis

The discrete Fourier transform (DFT) is used to extract the frequency response from the snare drum recordings under evaluation. To minimise differences between the drum excitations, the average of three individual excitations is used, once these are aligned in the time-domain by cross correlation. While this is not necessary with recordings made with the RDA, it provides consistency in performances with more variability.

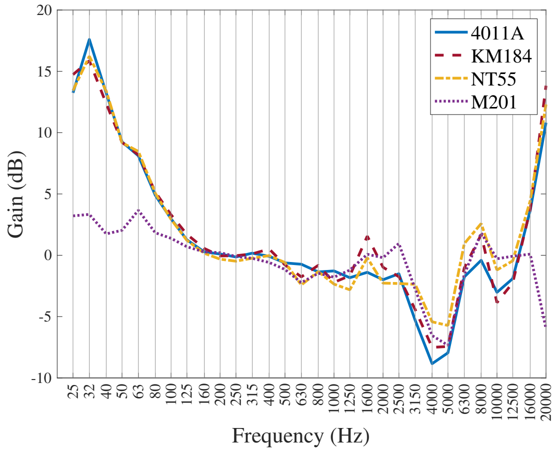

Fig. 4: Gain values for 30-band EQ applied to D5 recording to minimise spectral difference with those of the Category-1 microphones.

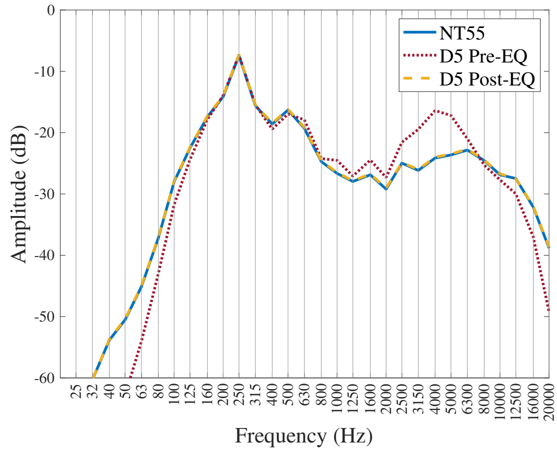

Fig. 5: Spectral difference between NT55 and D5 recordings of the Velvetone snare pre- and post-EQ.

The DFT bins are then mapped to gain values associated with a 30-band graphic EQ. The centre frequencies for the 30 bands were selected based on the ANSI standard for fractional-octave-band digital filters—the same bands used by the graphic EQ. $\frac{1}{3}$-octave bands were used with nominal mid-band frequencies ranging from 25–20,000 Hz. The DFT frequency bins associated with the bands of the $\frac{1}{3}$-octave EQ were summed and converted to decibels (dB).

4.2 Equalisation

The difference between all 30 frequency bands of a Category-1 microphone recording and that of the D5 is calculated. These 30 values are used to set the gains of a 30-band digital graphic EQ [26], using a 24th order cascaded design. The high-filter order is required to minimise the difference between microphones. Figure 4 displays the gain used for each band of the EQ in the modification of the D5 recording. This process is repeated for all Category-1 microphones and applied to the D5 recording. Figure 5 shows the difference between the D5 and the NT55 recordings of the Velvetone snare as well as showing the D5 recording post-EQ, having been equalised to match the NT55 recording. All original and modified recordings from this study are available online.[4]

[4] dmtlab.bcu.ac.uk/matthewcheshire/audio/aes147

5 Pre- and Post-EQ A/B Tests

To assess the success of the spectral modification in reducing the bias towards Category-1 microphones, a comparison is made between pre- and post-EQ A/B listening tests with expert listeners with over seven years of music production experience.

5.1 Methodology

A pre-EQ A/B test compares the *unmodified* D5 recording and those of the four Category-1 microphones, while a post-EQ A/B test compares the *modified* D5 recording to those of the four Category-1 microphones. These tests use the same listening environment and equipment as in Section 3.1 and are conducted with the WAET software in an A/B comparison configuration. In both tests, the sample order is randomised and repeated in 10 trials for each microphone pair, resulting in a total of 40 A/B comparisons. As it was found that microphones did not exhibit snare-dependent results (Section 3.2), only recordings from the Velvetone snare drum are used.

The pre-EQ A/B test was completed by 10 participants (range: 24 to 55 years, mean: 35 years, std: 10.4 years) with at least 8 years experience (*range*: 8 to 29 years, *mean*: 16.7 years, *std*: 9 years). The post-EQ A/B test was completed by 10 participants (*range*: 20 to 49 years, *mean*: 29.4 years, *std*: 9.5 years) with at least 7 years experience (*range*: 7 to 27 years, *mean*: 13 years, *std*: 6.7 years).

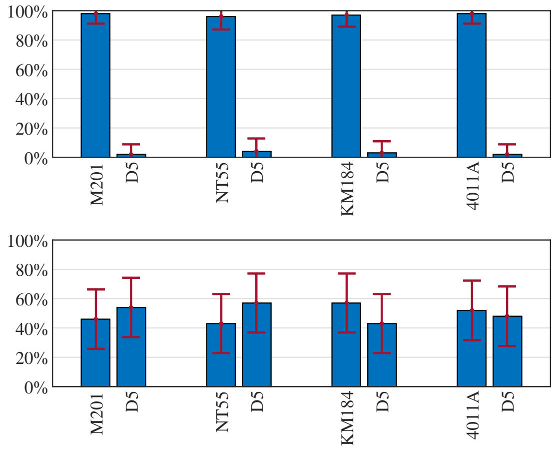

Fig. 6: Results from A/B test with 95% confidence interval. Top plot shows a comparison between the *unmodified* `D5` and Category-1 microphone recordings; Bottom plot shows *modified* `D5` and Category-1 microphone recordings.

5.2 Results

Observations of consistency in selection for both A/B listening tests were possible due to participants repeating each pairwise comparison in 10 trials. Due to the small trial size, only participants who preferred the same sample greater than eight times can be considered to be statically significant. For the pre-EQ A/B test, all participants were consistent in their preference for Category-1 microphones; however, for the post-EQ A/B test, participants were rarely consistent. No participant was consistent for the `4011A`, and two participants had a consistent preference for the `KM184`. In the `M201` comparison, one participant had a consistent preference for the *modified* `D5`, and one participant had a consistent preference for the `NT55`.

A binomial test was used to observe preference, with the null hypothesis being that two categories are equally likely to occur. The upper plot in Figure 6 depicts the results of the pre-EQ A/B test with 95% confidence intervals, indicating that preference is significant for the Category-1 microphones. The lower plot of Figure 6 shows that preference in the post-EQ A/B test is not significant and is closer to chance.

The results from the post-EQ A/B test indicate that participants showed no preference between Category-1 microphones and the *modified* `D5`. This suggests that matching frequency responses of recordings through equalisation is an effective approach to improve the preference of recordings made with a less-preferred microphone. The preference for the recordings of the *unmodified* `D5` over the Category-1 microphones is: `M201` (2.0%), `NT55` (4.0%), `KM184` (3.0%), `4011A` (2.0%). The preference for the *modified* `D5` over the Category-1 microphones is: `M201` (54.0%), `NT55` (57.0%), `KM184` (43.0%), `4011A` (48.0%).

6 Conclusions

In this study a multi-stimuli listening test was carried out to categorise highly-preferred and least-preferred microphones for recording snare drums. The results of this test revealed that for 11 of the 12 microphones, listener preference did not significantly change between the four different snare drums used. Once this was determined, recordings from one of the least-preferred microphones was equalised to have the same frequency characteristics as that of the top four highly-preferred microphones. Two A/B listening tests were carried out to compare preference pre- and post-transformation. Listener consistency was observed across repeated comparisons in both tests, which revealed that very few participants could choose the same sample consistently in the post-EQ A/B test. This demonstrated a trend for no preference between the recordings of the highly-preferred microphones and those of the modified least-preferred microphone.

In future work, the importance of phase response and off-axis microphone characteristics will be examined. In addition, non-linear characteristics on preference could be investigated, as recording a loud snare drum closely will induce harmonics in active circuitry and transformers that contribute to tonality. A range of striking velocities could also be applied to the drum for a more realistic emulation of human players. Towards an improvement of the spectral transformation, finer adjustment of spectral characteristics could be achieved through a higher-resolution EQ, and formalised testing would validate the suitability of the modification process across a range of microphones and stimuli.

References

[1] White, P., "SOS Guide To Choosing & Using Studio Microphones," *www.soundonsound.com*, 2006.

[2] De Man, B. and Reiss, J., "A Pairwise and Multiple Stimuli Approach to Perceptual Evaluation of Microphone Types," in *134th Convention of the Audio Engineering Society*, 2013.

[3] Bartlett, B., "Choosing the Right Microphone by Understanding Design Tradeoffs," *35th Journal of the Audio Engineering Society*, 1987.

[4] Bartlett, B., "Microphones: Small is Beautiful," in *81st Convention of the Audio Engineering Society*, 1986.

[5] Butler, J. L., "Microphone Selection," in *86th Conventions of the Audio Engineering Society Convention*, 1989.

[6] Eargle, J., *The Microphone Book*, Focal Press, 2004.

[7] Owsinski, B., *The Recording Engineer's Handbook*, Thomson Course Technology, 2005.

[8] Houghton, M., "Vocal Mics, Matching Mics & Voices," *www.soundonsound.com*, 2010.

[9] Cheshire, M., Hockman, J., and Stables, R., "Microphone Comparison for Snare Drum Recording," in *145th Conventions of the Audio Engineering Society*, 2018.

[10] Pearce, A., Brookes, T., Dewhirst, M., and Mason, R., "Eliciting the most prominent perceived differences between microphones," *139th Journal of the Acoustical Society of America*, 2016.

[11] Owsinski, B. and O'Brien, M., *The Mixing Engineer's Handbook*, Mix pro audio series, Mix Books, 1999.

[12] McKinnie, D., "Measuring the Perceived Differences between Similar High-quality Microphones," in *120th Audio Engineering Society Convention*, 2006.

[13] Hebrock, J., Statham, K., and Kraft, C., "Psychoacoustic Effects of Ringing in Microphones," in *103rd Conventions of the Audio Engineering Society Convention*, 1997.

[14] Elliott, B., "Mic Comparison: A Snare Drum Microphone Shootout," *www.drummercafe.com*, 2014.

[15] Fuston, L., "Best Mics for Recording Snare Drums," *www.sweetwater.com*, 2017.

[16] Bartlett, B. A., "Tonal Effects of Close Microphone Placement," *29th Journal of the Audio Engineering Society*, 1981.

[17] Quiroga, A. F., Garcia, J. D., and Páez, D., "Tom–Tom Drumheads Miking Analysis," in *138th Convention of the Audio Engineering Society*, 2015.

[18] Senior, M., "Kick & Snare Recording Techniques," *www.soundonsound.com*, 2008.

[19] Pearce, A., Brookes, T., and Dewhirst, M., "Validation of Experimental Methods to Record Stimuli for Microphone Comparisons," in *139th Convention of the Audio Engineering Society*, 2015.

[20] C. Disley, A., Howard, D., and Hunt, A., "Timbral description of musical instruments," 2006.

[21] Jillings, N., De Man, B., Moffat, D., and Reiss, J. D., "Web Audio Evaluation Tool: A Browser-based Listening Test Environment," in *12th Sound and Music Computing Conference*, 2015.

[22] De Man, B. and Reiss, J., "APE: Audio Perceptual Evaluation Toolbox for MATLAB," in *136th Convention of the Audio Engineering Society*, 2014.

[23] EBU-R-128, *Loudness Normalisation and Permitted maximum Level of Audio Signals*, 2014.

[24] Dinno, A., "Nonparametric pairwise multiple comparisons in independent groups using Dunn's test," *Stata Journal*, 2015.

[25] "American National Standard Specification for Octave-band and Fractional-Octave-band Analog and Digital Filters: ANSI S1.11-2004," *Acoustical Society of America*, 2009.

[26] Oliver, R. J. and Jot, J.-M., "Efficient Multi-Band Digital Audio Graphic Equalizer with Accurate Frequency Response Control," in *139th Convention of the Audio Engineering Society*, 2015.

Audio Engineering Society
Convention Paper
Presented at the 148th Convention
2020 May 25 – 28, Vienna, Austria

Investigating timbral differences of varied velocity snare drum strikes

Matthew Cheshire, Ryan Stables, and Jason Hockman

Digital Media Technology Lab (DMT Lab), Birmingham City University

Correspondence should be addressed to Matthew Cheshire (`matthew.cheshire@bcu.ac.uk`)

ABSTRACT

Adjusting striking excitation velocity for percussion instruments changes characteristics of the sound output, most notably in loudness and timbre. In this study, a listening test is carried out to assess participant abilities in distinguishing between varied velocity snare strikes when the loudness disparity had been removed from recordings made with four common studio microphones. Results indicate that all participants are able to identify different velocities based on timbral differences alone. Temporal and spectral features were then extracted from the recordings to gain insight into which quantifiable differences are present between varied velocity recordings. Analysis revealed various features such as attack and decay time, fundamental frequency, and brightness to have significant differences for the varied velocity snare strikes.

1 Introduction

Variation in the striking velocity for a percussion instrument results in modification of the sound output, with the main effects being related to the volume envelope and timbre. An obvious example of this phenomenon is the difference between a high velocity full strike and a substantially less energetic stick bounce. While loudness models are used to define the relationship between sound pressure level and perception of complex sounds such as snare drums in the case of [1], it has yet to be determined if timbral differences alone are enough for participants to easily identify velocity variation.

Timbral differences between snare drum recordings are derived from the response of the instrument to different striking velocity dependant on its physical construction [2, 3, 4], as well as non-linearities of the microphone used for recording. Many studio microphones have a total dynamic range of 125–130dBA [5]. A snare drum is capable of producing a sound pressure level (SPL) of up to 140dB [6], exceeding the maximum SPL-handling capability of the microphone, inducing harmonic distortion. The amount of distortion produced will be dependant on the specification of the microphone and the amount by which its maximum SPL tolerance is exceeded. The recording engineer must therefore consider the amount of dynamic fluctuation during a performance when selecting an appropriate microphone.

While loudness and timbre are intrinsically linked, it is important to know how the timbral character of a snare sound changes with varying velocity and the effect this has on perception. This information would help to detangle the mixing preferences of engineers that utilise dynamic compression to minimise relative volume differences between high and low velocity strikes. This paper investigates the role of timbral differences in distinguishing between high and low velocity strikes, in the extreme case where stimuli have been loudness normalised. Timbral differences are then assessed through signal analysis to characterise the high and low velocity strikes.

Linking timbral differences to features has many potential applications: In the context of audio production, such a process could afford new tools for subtle timbre modification of recorded drums, even out highly dynamic performances, or add a humanisation effect to sample-based production without a trade-off in volume. In the information retrieval domain, this would allow for sorting and searching of sample libraries by perceived velocity. Other uses might include a novel mixing task in which high and low velocity strikes are processed independently.

The remainder of this paper is structured as follows: Section 2 presents the listening test methodology and results. Section 3 presents signal analysis methods for characterising the timbral differences between strikes of varied velocity. Section 4 provides a discussion on the implications of the results from the previous two sections. Section 5 presents conclusions, highlighting some of the key findings of the study and Section 6 provides recommendations for future work.

2 Listening Test

An A/B listening test was conducted to evaluate whether participants with sound engineering training could identity loudness normalised snare strikes of varied velocities. In removing volume cues, participants would be required to evaluate differences between strikes based on inherent attributes in the recordings other than loudness.

2.1 Methodology

An A/B listening test was used to evaluate if participants could distinguish between high and low velocity strikes from each microphone. All recordings were loudness adjusted to -23 LUFS [7], which removed any loudness variation between samples to ensure that perceived loudness between strikes and different microphones was as consistent as possible. By removing the cue of loudness, participants would have to evaluate differences based on any temporal and spectral variations between velocities. To create more engaging stimuli for the listening test, high and low velocity strikes from each microphone were sequenced into a two-bar drum

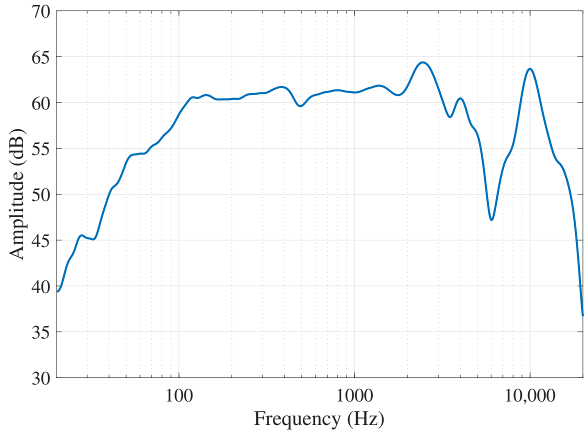

Fig. 1: Average frequency response of left and right headphone channels shown with $\frac{1}{6}$−octave smoothing.

phrase. For each microphone, participants were presented with high and low velocity phrases 10 times each in a random order. For 5 pairs, participants were asked to select the phrase that had lower velocity and for the other 5, participants were asked to select the phrase with higher velocity, resulting in a total of 40 comparisons. The participants could not proceed to the next evaluation until they had played both phrases and made a selection. 15 participants aged 21–50 years (mean: 26.8 years) took part in the listening test, and their experience in audio related fields was 3–30 years (mean: 8.5 years). AKG K240 studio headphones were used for playback, and participants were encouraged to adjust the volume to a comfortable level. Frequency response of the headphones was measured with an Earthworks M30 omnidirection measurement microphone while placed on a Sennheiser dummy head. Figure 1 shows the average frequency response as the left and right channels were nearly identical.

2.2 Recordings

In order to evaluate snare strikes of different velocities, a set of recordings that reflect professional standards were required (e.g., microphone position, microphone selection). In addition, objective measurements of velocity were required for categorisation of high and low strikes. Four common studio microphones (i.e., two dynamic microphones and two condenser microphones in Table 1) were selected for recording samples to investigate if microphone selection altered perception of strike velocity. All stimuli were recorded in mono with a 44.1kHz sample rate and 16-bit depth.

As microphone position has an impact on timbral characteristics [8, 9, 10], a consistent and generalisable placement was essential. For this reason a *close micing* technique was used. In the context of recording a drum kit, close micing refers to the placement of a microphone in near proximity to an individual instrument to reduce the effects of room acoustics and to minimise leakage from the rest of the kit. [11, 12, 13, 14]. Close micing is most notably used on the snare, toms and kick drum [15, 16]. This technique was used to achieve professional grade recordings of quasi-isolated snare strikes.

Recordings of the snare drum were captured at the same time SPL measurements were taken. Lower velocities corresponded to lower SPL measurements, and higher velocity with high SPL to produce varied velocity recordings with a measurable output.

As seen in Figure 3, a Cirrus CK:162C Optimus Red sound level meter[1] was placed at the same position as the microphones, in a typical close microphone position 10cm above the drum head and 5cm over the rim, pointing directly at the centre of the drum [17, 18, 19, 20].

The SPL was measured in LZFMAX, which measures the Z-weighted, fast-response maximum sound level. Z-weighted refers to no weighting across frequency response between 10Hz and 20kHz ±1.5dB. The fast response has a 125ms rise and decay time. In order to minimise the effects of room acoustics, the recordings were carried out in an acoustically treated isolation booth, measuring 2.5m x 3m x 4.5m with a RT60 of 112ms (mean of $\frac{1}{3}$-octave measurements). The snare drum was struck in the centre of the head, such that it produced SPL of 100dBZ and 125dBZ. These values were selected as they were consistently playable by the drummer, when asked to play high and low velocity strikes. Multiple recordings were captured until a strike was within ±0.5dBZ of the target SPL, as depicted in Figure 2.

To produce a generalisable and realistic scenario of drum recording, the snare drum and tuning method was the same used in [21]. For snare recording it is typical to dampen the batter head [22, 23]. This is done to reduce unwanted overtones and shorten the

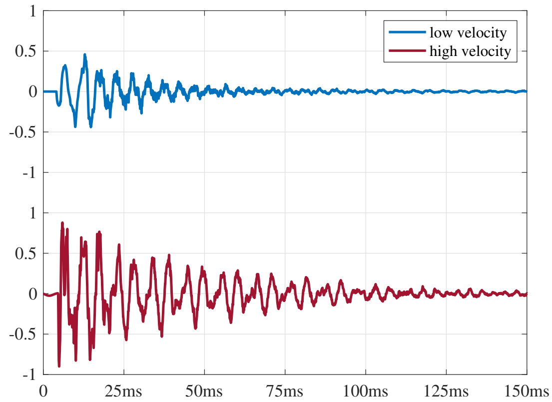

Fig. 2: Recording of low velocity strike (100dBZ) and high velocity strike (125dBz).

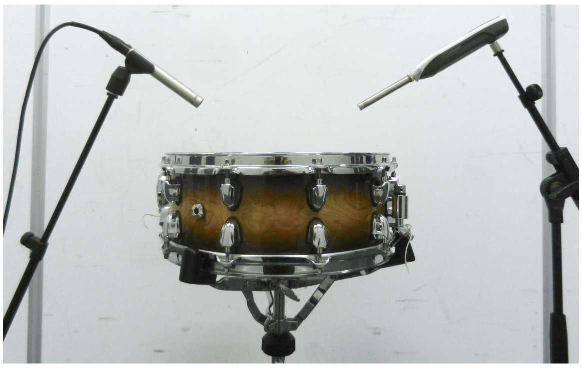

Fig. 3: Snare drum with RØDE `NT55` microphone (left) and Cirrus CK:162C SPL meter (right).

sustain or *ringing* of the snare, producing different tonal characteristics with less high frequencies. For this study MoonGel,[2] a self-adhesive gel rectangle (3.5cm x 2.5cm) was placed directly on the batter head 2.5cm from the edge of the rim.

2.3 Results

A binominal test was used to determine if the percentage of correct responses was significant with the null hypothesis being that participants could not distinguish between varied velocity strikes when perceptual loudness differences were removed. The null hypothesis would be accepted if the percentage of correct answers was below the significance level of 58.67% based on

[1] `www.cirrusresearch.co.uk`

[2] `www.rtom.com/moongel-damper-pad`

Brand	Model	Type	Polar Pattern	Frequency Range	Sensitivity	Impedance	Maximum SPL
Neumann	KM184	Condenser	Cardioid	20Hz to 20kHz	15mV	50 Ω	138 dB
RØDE	NT55	Condenser	Cardioid	20Hz to 20kHz	12.6mV	100 Ω	136 dB
Shure	Beta57a	Dynamic	Supercardioid	50Hz to 16kHz	2.8mV	290 Ω	not specified
Shure	SM57	Dynamic	Cardioid	40Hz to 15kHz	1.6mV	310 Ω	not specified

Table 1: Specifications of microphones used for recording, taken from manufacture websites.

150 trials (i.e., 10 trials for each participant per microphone; 15 total participants). Table 2 shows the percentages of correct responses for each microphone (correct), with the standard deviation (std) and the p-value from the binomial test (p). Small p-values (<0.05) for each microphone suggest that the percentage of correct responses was significant, thus the null hypothesis is rejected. This shows that the participants were able to distinguish between, and successfully identify varied velocity recordings.

When striking the snare drum at different velocities there is a loudness disparity, which was demonstrated by the SPL measurements taken at the time of recording. When perceptual loudness was normalised between recordings of varied velocity strikes, experienced participants with sound engineering training were able to successfully identify which velocities corresponded to the different recordings. This highlights that there are cues other than loudness variation that participants use to distinguish between striking velocities.

3 Feature Extraction and Analysis

In order to characterise the perceptual differences experienced between the varied velocity stimuli in Section **??**, signal analysis methods for feature extraction are applied to the recordings. Statistical analysis is then performed on the extracted features to identify if features from the high and low velocity strikes are statistically different.

Model	Correct (%)	std	p
Beta57	88.67	1.73	8.6×10^{-24}
KM184	89.26	1.36	1.9×10^{-24}
NT55	93.29	1.16	1.6×10^{-30}
SM57	91.95	0.99	2.4×10^{-28}

Table 2: Correct responses (%) across all participants with standard deviation (std) and p-values.

3.1 Methodology

To conduct timbral analysis of varied velocity strikes, a second set of recordings were created, which comprised 22 high velocity strikes (125dBZ $\pm$2dBZ) and 22 low velocity strikes (100dBZ $\pm$2dBZ). For each of the four microphones, recordings were captured in the same manner as in section **??**. Prior to any feature extraction, all samples were peak normalised, truncated to 1 second and synchronised using cross correlation.

Spectral and temporal analysis was undertaken to examine the different properties of the recordings that made identification of velocity possible by participants when the cue of loudness was removed. A variety of features from the MIRtoolbox [24] were selected to reflect features relevant to the spectral and temporal domain of a snare drum (Table 3), including *attack time* and *decay times* to define temporal envelope characteristics; *fundamental frequency* ($f0$); *entropy*, *flatness*, and *kurtosis* to describe the peakiness of a spectrum; *spectral rolloff* and *brightness* to estimate high frequency.

The frequency spectrum of the recordings was divided into 24 Bark scale critical bands as in Figure 4. These perceptual subdivisions of the spectrum are based on the natural division of the audible range by the human ear, and are known to correlate closely to cochlear mechanics [25].

Comparing perceptually relevant frequency bands allows observations of significantly different bands, thus aiding in explaining which characteristics contribute to perception of timbre-related velocity variation.

3.2 Statistical Tests

In order to test whether the features extracted from the varied velocity recordings were significantly different, a two-sample Kolmogorov-Smirnov (KS) test was used, as the data was from a non-normal continuous distribution. The Anderson-Darling test was used to check for normality. The null hypothesis for the KS test is

	`Beta57a`		`KM184`		`NT55`		`SM57`		**All Mics**	
Features	Low	High	Low	High	Low	High	Low	High	Low	High
Attack (ms)	30.99	33.26	30.28	35.80	30.41	32.75	30.72	33.08	30.60	33.73
	0.29	0.28	0.42	2.27	0.36	0.32	0.23	0.21	0.43	1.67
Decay (ms)	99.49	129.13	99.61	126.06	100.55	128.53	98.47	121.67	99.54	126.35
	2.85	4.08	2.59	3.12	2.31	3.98	2.41	2.99	2.62	4.59
$f0$ (Hz)	226.38	206.79	226.47	206.49	226.42	206.92	226.49	206.72	226.45	206.72
	0.72	0.59	0.69	0.41	0.70	0.58	0.70	0.56	0.69	0.56
Centroid (Hz)	2694.70	2410.70	2691.00	2471.10	2818.00	2498.00	269.4.80	2410.70	2641.00	2360.70
	65.35	53.30	53.01	92.12	59.74	61.28	65.35	53.27	180.59	187.86
Spread (Hz)	3215.60	2881.50	3891.50	3654.30	3982.40	3652.70	3339.80	3093.80	3607.30	3320.60
	33.63	41.43	34.85	71.99	35.83	44.57	32.21	46.18	337.67	347.09
Rolloff (Hz)	5548.60	4674.00	6504.70	5705.00	6933.00	5959.90	6358.50	5636.10	6336.20	5491.20
	118.85	155.28	130.88	277.92	142.96	193.80	110.02	151.93	519.26	528.84
Entropy	0.82	0.81	0.83	0.83	0.84	0.83	0.84	0.83	0.81	0.83
	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01
Flatness	0.14	0.11	0.23	0.20	0.23	0.20	0.15	0.14	0.18	0.16
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.05	0.04
Irregularity	1.09	0.58	1.04	1.13	1.08	1.02	1.17	1.09	1.09	0.95
	0.14	0.25	0.17	0.07	0.13	0.07	0.11	0.05	0.15	0.26
Kurtosis	6.89	8.87	6.91	8.16	6.15	7.53	5.31	7.29	6.33	7.96
	0.18	0.40	0.19	0.48	0.16	0.32	0.15	0.30	0.66	0.72
Roughness	50.21	396.93	31.86	765.98	45.33	398.92	34.19	339.94	42.39	475.44
	30.04	174.31	23.71	99.59	27.13	161.30	28.58	112.22	27.65	219.26
Skewness	1.94	2.28	2.00	2.24	1.86	2.13	1.58	1.94	1.85	2.15
	0.04	0.06	0.04	0.09	0.04	0.06	0.04	0.06	0.17	0.15
Brightness	0.39	0.35	0.38	0.36	0.32	0.36	0.44	0.41	0.40	0.37
	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.03	0.03

Table 3: Mean (upper value) and standard deviation (lower value) features extracted from low and high velocity recordings for each microphone. All Mics presents analysis of all 88 recordings.

that the data in two vectors are from the same continuous distribution. The two-sample KS test was used to evaluate every feature pair in Table 3 from the 22 low and 22 high velocity recordings. All microphones were evaluated separately as well as pooling all 88 low and 88 high velocity recordings. The test revealed that all features for low velocity recordings were significantly different from that of the high velocity recordings ($p < 0.05$). The only features which showed no significant difference were for the `KM184` microphone; these were entropy ($p = 0.56$) and irregularity ($p = 0.08$).

The Bark scale critical bands from the high and low velocity recordings for each microphone were evaluated to identify significant differences. The two-sample t-test was used to compare the distributions of each critical band, and the Anderson-Darling test was used to check normality of the distributions. The null hypothesis of the two-sample t-test is that two normal distributions have equal means and equal but unknown variances and the alternative hypothesis is that the distributions comes from populations with unequal means. Table 4 shows the rank of the top 5 critical bands which are significantly different based on the t-test. Table 5 shows which Bark bands had no significant difference between the high and low velocity strikes. All

Rank	`Beta57`	`NT55`	`KM184`	`SM57`
1	700Hz	250Hz	250Hz	150Hz
2	350Hz	700Hz	350Hz	840Hz
3	840Hz	840Hz	700Hz	700Hz
4	150Hz	570Hz	840Hz	1370Hz
5	570Hz	350Hz	570Hz	570Hz

Table 4: Top five ranked statically-different critical bands for each microphone.

Bark bands for the `KM184` recordings were significantly different between velocities.

4 Discussion

Analysis of the high and low velocity recordings show various timbral differences. These differences made it possible for participants to distinguish between velocities when the cue of loudness was removed. All but two features extracted from the high and low velocity recordings were significantly different to each other. A notable feature which displayed significant difference was decay time, which was slower for high velocity strikes with an average time of 27ms. This

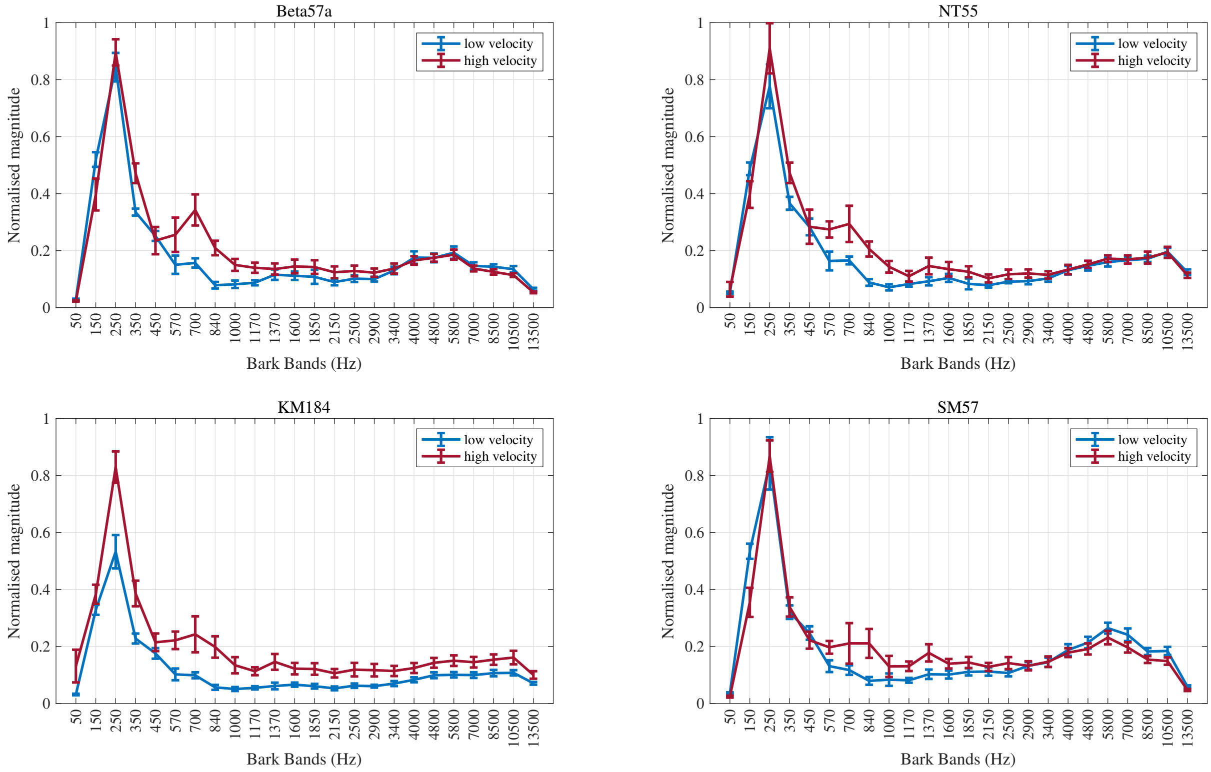

Fig. 4: Means and standard deviations of Bark band magnitudes for high and low velocity strikes.

confirms intuition, as the higher velocity strikes excite the drum skin with more energy and thus more time is required for energy to dissipate. Interestingly, attack time was found to be approximately 3ms quicker for the lower velocity strikes. Although statistically significant, the degree to which this was perceivable by participants is out of the scope of this study. The fundamental frequency was significantly lower when the snare was struck harder by roughly 20Hz. Centroid, rolloff and brightness were all lower for high velocity strikes, indicating that the low velocity strikes have proportionally more high frequency energy. The analysis of the critical bands revealed high velocity strikes have increased energy between the 570 and 840Hz critical bands (as seen in Figure 4). This suggests that although the low velocity strikes have proportionally more high frequency energy, this is actually attributed to the additional low frequency energy created by higher velocities. Although all critical bands were significantly different for the `KM184` than the other three microphones, the critical band centred at 4KHz was not significantly different. Not all the microphones had the same statistically different critical bands, which indicates that the varied frequency responses of the microphones had a non-linear effect.

`Beta57`	`NT55`	`SM57`
450Hz	450Hz	250Hz
3.4kHz	4.0kHz	2.9kHz
4.0kHz	4.8kHz	3.4kHz
4.8kHz	7.0kHz	4.0kHz
5.8kHz	8.5kHz	
	10.5kHz	

Table 5: Critical bands found to be not statistically different.

5 Conclusions

In this study a listening test was carried out to assess if participants could distinguish between high and low velocity snare strikes when loudness disparity had been removed from recordings made with four common studio microphones. It was discovered that all participants could identify the velocities with the absence of loudness cues. This indicated that participants were using temporal and spectral differences to accurately select the different velocity recordings.

Nearly all features extracted from the recordings were significantly different between high and low velocity strikes. This revealed that attack time was shorter for the low velocity strikes, whilst decay time was longer for the high velocity strikes. Fundamental frequency also varied with change in velocity, with high velocity strikes being 20Hz lower. Statistical analysis of the Bark scale critical bands using a two sample t-test showed the biggest disparity between bands centred at 570Hz to 840Hz.

6 Future Work

In this study, the output of only one snare drum was assessed. In order to gain a more comprehensive understanding of tonal differences between high and low velocity strikes, additional strikes from a range of snare drums would need to be analysed. This may include snare drums of different material, such as different woods and metals, as well as drum head type and tuning. The tension of the snare wires and number of snare strands may result in velocity dependant spectral variation. Drum stick material (e.g., nylon, wood) may even play a role in timbral differences. Player technique and location of strike are also likely to produce measurable variations, and this could certainly be included in any future investigations. Various techniques and products are utilised to dampen the batter head of snare drums; this process is deliberately designed to alter both temporal and spectral features. An in depth analysis should be carried out to determine how different dampening affects the tonal differences. Finally, while this study has concentrated on differences produced by high and low velocity strikes, a range of strikes could be assessed to reveal exactly how feature variation correlates with velocity change.

7 Appendix

Table 6 shows the percentage of correct responses for each of the microphones for the 15 individual participants.

ID	Beta57	KM184	NT55	SM57	Total
1	100	100	100	100	100
2	70	70	60	80	70
3	100	100	90	100	97.5
4	100	80	80	80	85
5	100	100	100	80	95
6	100	80	100	90	92.5
7	100	100	100	100	100
8	70	100	100	90	97.5
9	80	70	100	100	85
10	90	70	90	90	82.5
11	90	100	100	90	95
12	90	100	100	100	97.5
13	40	70	80	70	65
14	100	90	100	100	97.5
15	90	100	90	100	95

Table 6: Correct responses (%) for each participant (ID) from the 10 repeated trials.

References

[1] Pestana, P. D., Reiss, J. D., and Barbosa, A., "Loudness Measurement of Multitrack Audio Content Using Modifications of ITU-R BS.1770," in *Audio Engineering Society Convention 134*, 2013.

[2] Wagner, A., Dahl, S. S., and Askenfelt, A., "Analysis of Drumbeats – Interaction between Drummer, Drumstick and Instrument," 2005.

[3] Tindale, A., "Classification of snare drum sounds using neural networks," 2004.

[4] Richardson, P. G., "Acoustic analysis and tuning of cylindrical membranophones," 2010.

[5] Eargle, J., *The Microphone Book*, Elsevier/Focal Press, 2004, ISBN 9780240519616.

[6] DPA Microphones, "How To Mic A Snare Drum," *www.dpamicrophones.com*, 2015.

[7] EBU-R-128, *Loudness Normalisation and Permitted maximum Level of Audio Signals*, 2014.

[8] Bartlett, B. A., "Tonal Effects of Close Microphone Placement," *29th Journal of the Audio Engineering Society*, 1981.

[9] Quiroga, A. F., Garcia, J. D., and Páez, D., "Tom–Tom Drumheads Miking Analysis," in *138th Convention of the Audio Engineering Society*, 2015.

[10] Senior, M., "Kick & Snare Recording Techniques," *www.soundonsound.com*, 2008.

[11] Owsinski, B., *The Recording Engineer's Handbook*, Thomson Course Technology, 2005, ISBN 9781932929003.

[12] Huber, D. and Runstein, R., *Modern Recording Techniques*, Focal Press/Elsevier, 2010, ISBN 9780240810690.

[13] Ballou, G., *Handbook for Sound Engineers*, Focal, 2008, ISBN 9780240809694.

[14] Kokkinis, E. K., Georganti, E., and Mourjopoulos, J., "Statistical Properties of the Close-Microphone Responses," in *Audio Engineering Society Convention 132*, 2012.

[15] Owsinski, B. and Moody, D., *The Drum Recording Handbook*, Hal Leonard Books, 2009, ISBN 9781423443438.

[16] Major, M., *Recording Drums: The Complete Guide*, Course Technology PTR, 2014, ISBN 9781133788935.

[17] Henshall, M., "How to Mic a Snare Drum with a Shure SM57," *www.shure.com*, 2014.

[18] Fuston, L., "How to Mic a Snare Drum," *www.sweetwater.com*, 2017.

[19] Gonzalez, D., "Home recording tips - 9 mic placements for recording a snare drum," *www.discmakers.com*, 2014.

[20] Worrell, B., "A Beginners' Guide to Drum Mic'ing," *www.reverb.com*, 2015.

[21] Cheshire, M., Hockman, J., and Stables, R., "Microphone Comparison for Snare Drum Recording," in *Audio Engineering Society Convention 145*, 2018.

[22] Seymour, M., "Engineer's Guide To Tuning and Damping Drums," *www.soundonsound.com*, 2010.

[23] D'Virgilio, N., "How to Control Drum Sustain with Dampening," *www.sweetwater.com*, 2014.

[24] Lartillot, O. and Toiviainen, P., "MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio." *Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France*, 2007.

[25] Zwicker, E., "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *The Journal of the Acoustical Society of America*, 33, p. 248, 1961.

Audio Engineering Society

Convention e-Brief 626

Presented at the 149th Convention
Online, 2020 October 27-30

Snare Drum Data Set (SDDS): More snare drums than you can shake a stick at

Matthew Cheshire[1]

[1] *Digital Media Technology Lab (DMT Lab), Birmingham City University*

Correspondence should be addressed to Matthew Cheshire (`Matthew.Cheshire@bcu.ac.uk`)

ABSTRACT

A comprehensive acoustic Snare Drum Data Set (SDDS) is recorded, providing extreme tonal variation of the snare drum. SDDS affords researchers a collection of recordings that mimic a wide range of real world studio practices, suitable for diverse research tasks. This is achieved by recording varied velocity snare strikes using 53 microphones, 10 tonally distinct snare drums and several common dampening methods, recorded in an acoustically treated studio utilising both close and far micing positions and 3 mic preamps. The resulting data set is 2,544 WAV audio files captured from 48 performances.

1 Introduction

For various machine learning tasks, such as automatic drum transcription (ADT) and instrument or playing style classification, large data sets containing multiple varied examples of the problem can help avoid overfitting. These multiple examples will be task specific, reflecting variations that will likely be encountered when attempting to solve real world problems. A data set designed for snare drum specific tasks should include a broad range of examples as a snare's timbral characteristics are heavily genre dependant. The timbral properties of a snare are affected by several factors, including; shell material and dimension, head type and tuning, amount and method of dampening, amount of snare wires, hoop material, and playing method (i.e., sticks, rods, or brushes). Distinct timbral alterations can also be achieved through modification of the recording process, including adjustments of the acoustic environment, microphone make, model, and placement [1, 2, 3].

1.1 Background

In [4] researchers used a feed-forward backpropogated neural network for realtime classification of different snare drum playing techniques based on their spectral properties. In total 20 examples of 5 different playing techniques were recorded using a single Neumann U-87. Although this addressed a specific classification problem, similar research requiring timbraly varied acoustic drum samples would benefit from SDDS which includes multiple mics and several timbraly distinct snare drums.

Several data sets exist containing acoustic drum samples, however very few offer extensive timbral variations. MDB Drums [5], a subset of MedleyDB [6] and constructed for ADT, consists of drum annotations and audio files for 23 tracks from various genres. IDMT-SMT-Drums [7] also used for ADT and source separation, is comprised of 104 polyphonic drum loops containing kick, snare, and hi-hat. Alongside acoustic drums it includes synthesised drums and loops created from sampled drums, eliminating velocity fluctuations and the recording process entirely. The ENST drum dataset [8] is an audio-visual database for signal processing and ADT. It contains annotated drum recordings from 3 drummers with 3 different drum kits. Although recorded on 8 audio channels, only one close mic was used to capture the snare. Due to the intended use of these data sets, using a range of snare recording techniques was not prioritised. Limitations for timbral analysis include sparse or absent metadata of mics, positions and snare drums, as well as a lack of multiple recordings of snares captured by several mics simultaneously, preventing comparisons of identical strikes.

Commercially available drum sample software such as FXpansion's [1] BFD3 and Toontrack's [2] Superior Drummer 3, offer varied velocity recordings of acoustic drum kits. Designed to emulate a live studio drummer, they are not ideal for research application due to a traditional mic set-up that utilises only a few close mics, as well as their optimised and enhanced timbral properties. A sample library recorded with the extensive selection of mics used by SDDS has yet to be found.

2 Description of data set

2.1 Recording

The recordings were captured in an acoustically treated live room at 44.1kHz, 16 bit. Out of all 53 microphones 24 were recorded through a SSL AWS 924 mixing console, 15 through a MIDAS M32R console, and a further 14 channels were connected to the MIDAS via a Behringer S16 digital stage box, see Table 1. All equaliser and dynamic processing effects on the mixing consoles were disengaged. Gain was set so that the loudest strike produced would not cause clipping.

2.2 Microphones

The recordings were undertaken using 53 mics (32 condenser, 18 dynamic, and 3 ribbon), see Table 1, collated from various sources [9, 10, 11, 12]. Built in high pass filters or frequency emphasis selectors were switched off on any mics that had them. Mics with variable polar pattern were set to cardioid.

[1] www.fxpansion.com
[2] www.toontrack.com

Fig. 1: Top, bottom and shell microphone positions.

2.3 Position

For drum recording it is common to place several mics around the entire kit to selectively emphasis certain drums, most notably the snare and kick drum [13, 14]. This is referred to as *close micing*, and minimises the effect of room acoustics and leakage from other drums. Conversely an overview of the drum kit as a whole may be captured using overhead or room mics [15]. During recording 35 close, 14 Overhead (OH), and 4 Room positions were utilised. Close positions are subdivide into 17 Top, 9 Shell, and 9 Bottom. The 3 close positions can be seen in Figure 1, every mic's position is listed in Table 1. Multiple positions were chosen to emulate several real world recording techniques. The data set can be easily augmented by combining multiple mic positions together, a mixing technique used to emphasise desired timbral attributes that can not be achieved through a single mic [16].

2.4 Snare drums

In total 10 snares of varied dimensions, shell material, and head type were chosen to represent a broad range of timbral properties. The batter heads were tuned to different fundamental frequencies to further emphasise timbral variation, whilst maintaining traditional characteristics (i.e., not extremely low or high). A digital drum dial was used ensuring even tension across all lugs. Table 2 shows specifications of the snares used. The snare wires for all snare drums were fitted and tightened such to minimise rattling and buzzing issues.

SSL AWS 924							
Brand	**Model**	**Type**	**Position**	**Brand**	**Model**	**Type**	**Position**
AKG	C414	C	OH	Electro Voice	RE20	D	Top
AKG	C414	C	Shell	Neumann	KM184	C	OH
AKG	C414	C	Top	Neumann	KM184	C	Top
AKG	C451B	C	OH	RØDE	NT55	C	Top
AKG	C451B	C	Top	RØDE	NTG2	C	OH
Audix	i5	D	Top	Royer	R-121	R	OH
Beyerdynamic	M201	D	Top	Royer	R-121	R	Shell
Coles	4038	R	OH	Sennheiser	MD421	D	Top
DPA	4090	C	OH	Sennheiser	e614	D	Top
DPA	4090	C	Top	Shure	Beta57A	D	Top
DPA	4099	C	Top	Shure	SM57	D	Top
Earthworks	M30	C	OH	Shure	SM7B	D	Top
Midas M32R				**Behringer S16**			
Brand	**Model**	**Type**	**Position**	**Brand**	**Model**	**Type**	**Position**
AKG	C414	C	Bottom	AKG	C451B	C	Bottom
AKG	C414	C	Room	AKG	C451B	C	Shell
Audix	D2	D	Top	AKG	D5	D	Bottom
Audix	ADX51	C	OH	Audix	ADX51	C	Shell
Brauner	Phantera	C	Shell	Audix	i5	D	Bottom
Lauten Audio	FC-357	C	Room	Audix	i5	D	Shell
Neumann	TLM 103	C	OH	DPA	4090	C	Bottom
RØDE	M3	C	OH	DPA	4099	C	Bottom
RØDE	M3	C	Top	DPA	4099	C	Shell
RØDE	NT1-A	C	OH	Electro Voice	RE20	D	Bottom
RØDE	NT5	C	OH	Sennheiser	MD421	D	Bottom
SE	2200	C	OH	Sennheiser	MD421	D	Shell
Sennheiser	e901	C	Room	Shure	SM57	D	Bottom
Shure	Beta91A	C	Room	Shure	SM57	D	Shell
Telefunken	M80	D	Top				

Table 1: All 53 microphones used, subdivided by preamp. Condenser (C), dynamic (D), ribbon (R).

2.5 Dampening

When recording snare drums it is typical to dampen the batter head [17, 18]. For SDDS 4 different dampening products were employed to reduced unwanted overtones and shorten sustain by varying amount. Firstly *Moon Gel*, a self-adhesive gel rectangle (3.5cm x 2.5cm), dampened the least of the 4 methods. Secondly an *Evans E-ring 1"* dampened more so than *Moon Gel*. Thirdly an *Evans E-ring 2"* covered more area and reduced overtones to a greater degree than the 1" version and lastly a *Big Fat Snare Drum* (BFSD) covered the entirety of the batter head, producing the greatest dampening effect. The 13" snare was only dampened using *Moon Gel* and a 13" version of the *Evans E-ring 1"* whilst all other snares were dampened with the 4 products. Additionally all snares were recorded without any dampening applied, resulting in 5 recording scenarios for 9 of the snares and 3 for the 13" drum.

2.6 Performance

One performance took place for each of the distinct recording configurations, (i.e., one snare drum with one dampening technique) resulting in 48 performances. The drummer was instructed to strike using the full range of velocity intensities they were capable of, to repeat the same velocity several times, and to allow each strike to ring out before playing the next. They were not limited to where on the drum head they could strike and could play with both hands. The length of the performances ranged from 171secs to 246secs (mean: 208secs), with a range of 69 to 120 (mean: 86) varied velocity strikes being played per performance. The amount of strikes for each performance can be seen in Table 3. The whole data set features 4,015 unique strikes captured with 53 mics, equating to 212,795 strikes from all audio files. A limitation of the methodology prevents direct comparison of velocities from different snares or dampening methods, due to the sound pressure level (SPL) being dependant on those factors, which vary between scenarios. However the relational difference of velocity can be examined between strike from the same performance. The range of velocities during each performance will be identical in all 53 microphones due to simultaneous recording, therefore allowing analyse of timbral variation that occurs as a result of velocity fluctuations.

Brand	Material	Dimensions	Lug Amt.	Batter Head	Resonant Head
Gretsch	Birch	14 x 5.5"	8	Remo Emperor Smooth White	Remo Weatherking Ambassador
Mapex	Maple	13 x 6.0"	8	Mapex Remo UX Coated	Mapex Remo UX Resonant
Mapex	Steel	14 x 6.5"	10	Evans Hydraulic	Remo Weatherking Ambassador
Mapex	Walnut	14 x 5.5"	10	Evans Hydraulic	Remo Weatherking Ambassador
Premier	Maple	14 x 5.5"	10	Evans Level 360 HD Dry	Remo Weatherking Ambassador
Tama	Steel	14 x 5.5"	10	Evans Hydraulic	Remo Ambassador Black Suede
Tama	Steel	14 x 6.5"	8	Remo Ambassador X	Remo Weatherking Ambassador
Yamaha	Birch	14 x 5.5"	8	Remo Emperor Smooth White	Tama 200 Hazy Snare Side
Yamaha	Maple	14 x 6.5	8	Remo Emperor Smooth White	Evans Level 360 Snare Side 300
n/a	Mixed	15 x 8.0"	8	Remo Emperor Smooth White	Remo Weatherking Ambassador

Table 2: Specifications of all snare drums used.

Snare	Un-dampened	Moon Gel	E-Ring 1"	E-Ring 2"	BFSD	Total
Gretsch	73	104	85	93	97	452
Mapex Maple	69	91	72	N/A	N/A	232
Mapex Steel	71	95	102	75	72	415
Mapex Walnut	105	79	120	91	84	479
Premier	76	97	98	98	100	469
Tama 5.5	85	110	93	92	85	465
Tama 6.5	62	87	93	79	85	406
Yamaha Birch	58	73	75	82	87	375
Yamaha Maple	50	74	87	96	93	400
n/a Mixed Wood	49	68	74	52	79	322
Total	698	878	899	758	782	4,015

Table 3: Amount of strikes for each performance, showing total for both dampening method and snare.

2.7 Data set file structure

The audio files are first separated by snare drum and then into dampening method folders containing 53 WAV files. The files are named microphone brand_model_position, for example `Premier/moongel/AKG_414_OH.wav` is the audio for the AKG C414 in the overhead position of the premier snare drum dampened using *Moon gel*. The position names have been abbreviated, Top (TP), Shell (SHL), Bottom (BTM), Room (RM), and Overhead (OH).

3 Discussion

SDDS allows researchers to group the recordings by snare drum, dampening method, mic make, mic model, type, or placement. The aim was to create diverse recordings that emulate real world recording scenarios and techniques, whilst also allowing for a level of scientific control. The absence of acoustic bleed facilitates investigation of the snare drum's timbral properties in isolation. As mics can be an expensive commodity, the plethora of mics were selected to uncover the expanse of characteristics associated with industry standard and specialised mics for recording snare drums. Although it is not possible to directly compare mic to mic due to the difference in mic position, which will also affect timbre, the audio files presented here may serve as a reference for future research.

4 Usage

There are several potential research topics SDDS may be useful for. The data set could aid in classification tasks for automatic mixing scenarios. By first classifying the mic position, EQ or compression settings could be applied automatically to suit the distinct timbral differences associated with top, bottom and overhead mic location. Classifying snare mics into condenser or dynamic would provide useful information to a mixing engineer when little or no metadata is provided, informing their mix decisions. Classifying based on velocity would facilitate sorting and searching sample libraries based on perceived velocity or allow novel mixing tasks where high and low velocity strikes are processed independently. Timbral variation linked to velocity could be extracted and mapped onto programmed drum samples, creating a humanisation effect without a compromise in volume. Furthermore, timbral components associated with mic placement could be modelled to facilitate spectral transformations from one position to emulate that of another. A similar approach could be used to model the characteristics produced by the different dampening methods, allowing the engineer to retroactively "dampen" a snare drum recording.

SDDS will provide recording engineers a resource to hear how various mics perform across different snares, positions, and velocities. This will serve as a reference for mic characteristics when access to specific mics is limited. The methodology presented here will allow others to emulate the recording configuration in order to compare timbral properties of other mics, positions, or snares to those in this data set.

5 Distribution

SDDS can be downloaded from `dmtlab.bcu.ac.uk/matthewcheshire/audio/sdds` The audio files are published under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License 🅭🅯🄏🄎.

6 Summary

A varied strike velocity snare drum data set comprised of 2,522 WAV files was recorded and the methodology described. The data set's intended use is for machine learning and music information retrieval tasks as well as providing insight into timbral changes that occur when snare type, dampening method, mic model and placement are altered. 10 tonally distinct snare drums were recorded over 48 performances, using 53 mics simultaneously. Real world scenarios were mimicked whilst still maintaining consistency over several variables, such as the recording space, location of the mics, and using the same drummer for all performances.

References

[1] Bartlett, B. A., "Tonal Effects of Close Microphone Placement," *29th Journal of the Audio Engineering Society*, 1981.

[2] Quiroga, A. F., Garcia, J. D., and Páez, D., "Tom–Tom Drumheads Miking Analysis," in *138th Convention of the Audio Engineering Society*, 2015.

[3] Senior, M., "Kick & Snare Recording Techniques," *www.soundonsound.com*, 2008.

[4] Tindale, A., Kapur, A., and Fujinaga, I., "Towards Timbre Recognition of Percussive Sounds," 2004.

[5] Southall, C., Wu, C.-W., Lerch, A., and Hockman, J., "MDB Drums: An Annotated Subset of MedleyDB for Automatic Drum Transcription," 2017.

[6] Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J., "MedleyDB: A Multitrack Dataset for Annotation-intensive MIR Research," 2014.

[7] Dittmar, C. and Gärtner, D., "Real-Time Transcription and Separation of Drum Recordings Based on NMF Decomposition," in *DAFx*, 2014.

[8] Gillet, O. and Richard, G., "ENST-Drums: An Extensive Audio-visual Database For Drum Signals Processing," in *ISMIR*, 2006.

[9] Elliott, B., "Mic Comparison: A Snare Drum Microphone Shootout," *www.drummercafe.com*, 2014.

[10] Fuston, L., "Best Mics for Recording Snare Drums," *www.sweetwater.com*, 2017.

[11] Cheshire, M., Hockman, J., and Stables, R., "Microphone Comparison for Snare Drum Recording," in *Audio Engineering Society Convention 145*, 2018.

[12] Cheshire, M., Stables, R., and Hockman, J., "Microphone Comparison: Spectral Feature Mapping for Snare Drum Recording," in *Audio Engineering Society Convention 147*, 2019.

[13] Owsinski, B. and Moody, D., *The Drum Recording Handbook*, Hal Leonard Books, 2009.

[14] Major, M., *Recording Drums: The Complete Guide*, Course Technology PTR, 2014.

[15] Senior, M., "Kick & Snare Recording Techniques: Tried & Tested Techniques From 50 Top Producers," *www.soundonsound.com*, 2008.

[16] Albano, J., "Mixing Snare Drums By Blending Multiple Microphones," *ask.audio*, 2016.

[17] Seymour, M., "Engineer's Guide To Tuning and Damping Drums," *www.soundonsound.com*, 2010.

[18] D'Virgilio, N., "How to Control Drum Sustain with Dampening," *www.sweetwater.com*, 2014.

OPEN ACCESS Freely available online

M. Cheshire, J. Drysdale, S. Enderby, M. Tomczak and J. Hockman,
"Deep Audio Effects for Snare Drum Recording Transformations"
J. Audio Eng. Soc., vol. 70, no. 9, pp. 742–752, (2022 September).
DOI: https://doi.org/10.17743/jaes.2022.0041

Deep Audio Effects for Snare Drum Recording Transformations

MATTHEW CHESHIRE, JAKE DRYSDALE, SEAN ENDERBY,
(matthew.cheshire@bcu.ac.uk) (jake.drysdale@bcu.ac.uk) (sean.enderby@bcu.ac.uk)

MACIEJ TOMCZAK, AND JASON HOCKMAN
(maciej.tomczak@bcu.ac.uk) (jason.hockman@bcu.ac.uk)

Sound and Music Analysis Group (SoMA), Digital Media Technology Lab (DMT Lab), School of Computing and Digital Technology, Birmingham City University, Birmingham, United Kingdom

The ability to perceptually modify drum recording parameters in a post-recording process would be of great benefit to engineers limited by time or equipment. In this work, a data-driven approach to post-recording modification of the dampening and microphone positioning parameters commonly associated with snare drum capture is proposed. The system consists of a deep encoder that analyzes audio input and predicts optimal parameters of one or more third-party audio effects, which are then used to process the audio and produce the desired transformed output audio. Furthermore, two novel audio effects are specifically developed to take advantage of the multiple parameter learning abilities of the system. Perceptual quality of transformations is assessed through a subjective listening test, and an object evaluation is used to measure system performance. Results demonstrate a capacity to emulate snare dampening; however, attempts were not successful for emulating microphone position changes.

0 INTRODUCTION

The positioning and recording of a standard acoustic drum kit—comprising of kick, snare, toms, and an assortment of hi-hats and other cymbals—is a technical and time-consuming endeavor. Recording drums may account for as much as 25% of the whole recording project [1]. During a typical session, an engineer must modify a large number of recording parameters to achieve a desired result. Key considerations include the selection of drums, drumheads, tuning, and dampening and the selection, arrangement, and positioning of microphones. These decisions impact the overall timbral quality of a recording, with certain modifications producing greater effects than others [2, 3].

Time permitting, an engineer may test different parameter options to identify an appropriate configuration for a song before committing to the final recording; however, with many variables, this can easily become a lengthy process. As such, the ability to perceptually modify these recording parameters in a post-recording process would be of great benefit to engineers limited by time or equipment, especially during sessions in which compromises may need to be made. In this work, a system is proposed for post-recording modification of the dampening and microphone positioning parameters associated with snare drum capture.

0.1 Background

Several methods for the automatic mixing of drums have been proposed [4–6]. Although these look at emulating processes of the digital mixing stage, the proposed system attempts to emulate techniques that are carried out prior to the recording stage. Two notable techniques an engineer can use to modify snare drum timbre include treating the drum heads directly through dampening or varying the position of the microphones around the drum in order to emphasize or subdue certain timbral characteristics.

Snare batter head dampening is a common timbre manipulation practice in drum recording [7, 8], which involves adding mass to the drumhead to remove unwanted overtones and shorten decay time to produce a perceptually tighter, more controlled sound [9, 10]. Engineers place various materials (e.g., cloth, duct tape, wallet) directly onto the drumhead to achieve subtle to extreme dampening effects. Many commercial products such as Big Fat Snare Drum, Snare Weight, and Moongel allow for the adjustment of dampening amount [11]. The recording engineer may use several of these techniques to create the intended drum sound [12]. Once dampening has been applied, those timbral properties are then committed to the recording, and one loses the ability to apply additional dampening if later required or to remove any if too much was used.

Microphone selection also impacts the timbre of recordings [13, 14]. The authors of [15] modified the spectral characteristics of a snare drum recording to mimic those of another through the use of a 30-band graphic equalizer (EQ); however, a limitation of this work was that access to recordings with target characteristics were required.

Audio effects are an integral part of the music production workflow that can be used to modify sound characteristics, such as dynamics, frequency, and timbre. Utilizing audio effects for a predefined audio transformation can be a laborious task that often requires mastery over a large number of parameters. As a result, there has been an increasing focus on audio effects modeling and intelligent audio effects within the field of music information retrieval.

In recent years, deep learning has demonstrated excellent performance in tasks such as emulating audio effects through end-to-end transformation methods [16–18], estimating audio effect parameters [19], mapping semantic descriptors to the parameter space of audio effects [20], and generating audio through differentiable digital signal processing [21]. More recently, Martinez et al. [22] emulated three common audio production tasks (i.e., mastering, breath/plosive removal, and tube amplification) through the use of a deep encoder, which performs parameterization of third-party audio effects within layers of the network.

0.2 Motivation

The system in [22] facilities training of audio plugin parameters or a chain of plugins for any desired transformation, given the appropriate training data. In this paper, the ability to modify the timbre of an undampened snare recording in order to elicit a perceptual change that corresponds to that of a dampened snare, referred to as Undampened-to-Dampened (U2D), will be explored through the use of multiple audio effects by utilizing the tools presented in [22]. The inverse transformation is also examined, whereby a dampened snare recording is modified to perceptually emulate qualities of an undampened snare recording, referred to as Dampened-to-Undampened (D2U). In addition to these dampening transformations, two positional recording parameter changes are explored: bottom-to-top (B2T) and top-to-bottom (T2B) microphone position.

The remainder of this paper is structured as follows: SEC. 1 outlines the proposed system. SEC. 2 describes the evaluation methodology for subjective objective comparisons. SEC. 3 presents the results from the evaluation, and SEC. 4 provides a discussion. Conclusions and suggestions for future work are presented in SEC. 5.

1 METHODOLOGY

An overview of the system configuration for transforming an undampened snare drum into a dampened snare is provided in Fig. 1. In order to automatically carry out different perceptual transformations, DeepAFx [22] is utilized for its powerful parameter learning and audio processing capabilities. DeepAFx consists of a deep encoder that first analyzes the input audio and then predicts the optimum parameters of one or more effect, which then processes the audio, producing the desired transformed output audio. The system makes use of the LV2 audio plugin open standard and incorporates third-party audio effects as a black box layer within a deep neural network. The authors provide the code used in their experiments.[1]

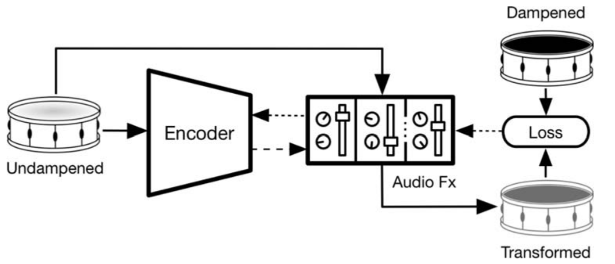

Fig. 1. System overview for snare dampening with DeepAFx with third-party audio effect. Solid lines depict flow of audio, the longer dashed line represents the predicted parameter values, and shorter dashed lines depict gradient flow.

1.1 Network Architecture

Following [22], an inception-based encoder network [23] is implemented to predict the audio effect parameter values required for a desired snare drum transformation. The input to the network is a log-scaled Mel-spectrogram represented as a 4D tensor $t \in \mathbb{R}^{b\times w\times h\times c}$, with batch size b, number of frames w, number of frequency bins h, and channels c. The model consists of 64 convolutional filters with a 5×5 sized kernel followed by 2×2 strided max-pooling. This is followed by six inception blocks with mixed kernel sizes, each comprised of a naive module with a stride of 2 and a dimension reduction module [24]. Rectified linear unit activations are used for all layers apart from the network's last layer, which is a fully connected output layer consisting of r output nodes and a sigmoid activation function, in which r is the number of parameters associated with a particular audio effect. The network outputs estimate audio effect parameter values for each snare drum transformation under observation.

1.2 Audio Effects

For this study, two novel LV2 audio effects are specifically developed to take advantage of DeepAFx's multiple parameter learning abilities; both effects have high parameter counts that would make it tedious and time-consuming for a human engineer to fine tune each control. Typically audio production tools are designed with the audio engineer in mind, graphic user interfaces (GUIs) are implemented, and variables such are parameter amount, layout, size, and color are considered in order to enhance the experience of the user. Allowing DeepAFx to learn the parameters a GUI is not required, nor are any considerations to the impracticality to a human user.

Both effects are investigated for their timbre-transforming abilities: a 10-band dynamic EQ (DEQ10) and

[1] https://github.com/SoMA-group/snarefx.

30-band dynamic EQ (DEQ30). Typically, dynamic EQs will consist of four to seven parametric frequency bands [25, 26], allowing the user to specify center frequency, Q-factor, and shelf or bell filter types [27, 28]. However, unlike traditional dynamic EQ, DEQ10 and DEQ30 are implemented as fixed-band graphic equalizers, with fixed center frequencies based on the specification for octave bands and fractional-octave bands described in [29], allowing for complete dynamic control over the full spectrum.

Dynamic EQ was specifically chosen in order to provide both spectral and temporal manipulation within one audio effect [30], often used in mastering applications [31]. The ability to control specified frequency bands over time lends itself to transformations in which some frequencies may be similar and others are disparate, such as in the case of dampening a snare, and in which both high frequencies are attenuated and their associated envelopes shortened, whereas the lower frequencies remain mostly unaffected. This would be difficult to achieve through the use of a standard full spectrum compressor; thus, dynamic EQ has the potential to perform better than a standard EQ and compressor combined for particular production tasks.

Both DEQ10 and DEQ30 have the same architecture, the signal path consisting of cascaded bi-quad peaking filters. Each frequency band comprises of two such filters; the gain of the first is controlled dynamically and that of the second is controlled through the *make-up gain* parameter for the band. Dynamic control of each band is achieved through a standard feed-forward compressor architecture. Within the side chain for each band, the signal first passes through a bi-quad band-pass filter, with center frequency and bandwidth matching that of the corresponding peaking filter in the signal path. Level detection and ballistics are carried out within the gain computer of the compressor's side chain. The output of this filter undergoes peak amplitude detection and then feeds a gain computer with the following parameters: *threshold*, *attack*, *release*, *ratio*, and *knee*. Each effect has an *output gain* parameter at the end of the signal path. A graphical representation of this architecture is given in Fig. 2. The principle difference between DEQ10 and DEQ30 is that the first uses an octave band layout, whereas the second uses third-octave increments. With six parameters per band and output gain, this gives 61 trainable parameters for DEQ10 and 181 for DEQ30.

In addition to the two novel effects, two open-source plugins were used.[2] Firstly an eight-band parametric EQ (PEQ), was chosen for its frequency sculpting ability and for the ubiquitous nature of parametric EQs in audio engineering. Secondly, because applying dampening to a snare drum alters its envelope characteristic, a transient designer (TD) was chosen as a possible candidate for a tool that might perform well at emulating this feature. A transient designer provides level-independent processing of the signal's envelope by using envelope followers to control output dynamics; this allows transients to be accelerated or slowed down and sustain to be prolonged or shortened [32].

[2]http://calf-studio-gear.org/.

DeepAFx also has the ability to train multiple plugins in a series; chaining multiple effects together is a common practice among mixing engineers [33], so for this reason, this aspect was also investigated. The PEQ and TD were used in conjunction with one another to determine whether they were able to perform better together, providing both spectral and temporal manipulations. The order of PEQ and TD were tested in both configurations, placing TD before and after PEQ. This was found to have very little audible difference on the processed audio; for this reason, only the PEQ+TD configuration was chosen for investigation.

1.3 Loss Function

The objective of the proposed model is to minimize the multi-scale spectrogram loss (MSL) between target snare drums and predicted snare transformations. MSL allows the network to extract information at multiple spectro-temporal resolutions and is calculated as the sum of the L2 difference between magnitude and log magnitude spectrograms computed with different fast Fourier transform resolutions: $r = \{2048, 1024, 512, 256, 128, 64\}$. The spectral loss for each resolution is defined as

$$\mathrm{MSL}_{\mathrm{stft}}(S, \hat{S}) = \sum_{r_i} \big[||S_{r_i} - \hat{S}_{r_i}||_2 + || \log S_{r_i} - \log \hat{S}_{r_i} ||_2 \big], \tag{1}$$

where magnitude spectrograms S and $\hat{S}$ are computed with a given fast Fourier transform resolution r_i from the target snare drums and predicted snare transformation audio.

1.4 Network Training

The deep encoder takes data x as input and parameters λ. Audio is pre-processed through resampling and conversion to a spectrogram representation. Following [22], snare drum recordings are resampled to 44.1 kHz, and the short-time Fourier transform (STFT) of each snare is calculated using a Hanning window with a size of 1,024 samples and a hop size of 256 samples to facilitate the desired temporal resolution of the network input. The magnitudes of STFT are transformed to log-scaled Mel-spectrograms with 128 Mel-frequency bands.

The model is trained using the Adam optimizer [34] with a learning rate 1e–4, where each iteration takes a mini-batch of 100 examples. Network weights are initialized using He's constant [35] to promote equalized learning. Once model performance ceases to improve over 25 epochs, early stopping is applied to complete training, and the epoch that achieves the best accuracy on the validation set is used for testing. Training was carried out on a Nvidia TESLA M40.

2 EVALUATION METHODOLOGY

The system presented in SEC. 1 is assessed through two evaluations to determine 1) perceptual quality of the transformations through a subjective listening test and 2) similarity of the transformed audio compared to the target audio through an objective evaluation using various comparative metrics. For each type of transformation under investiga-

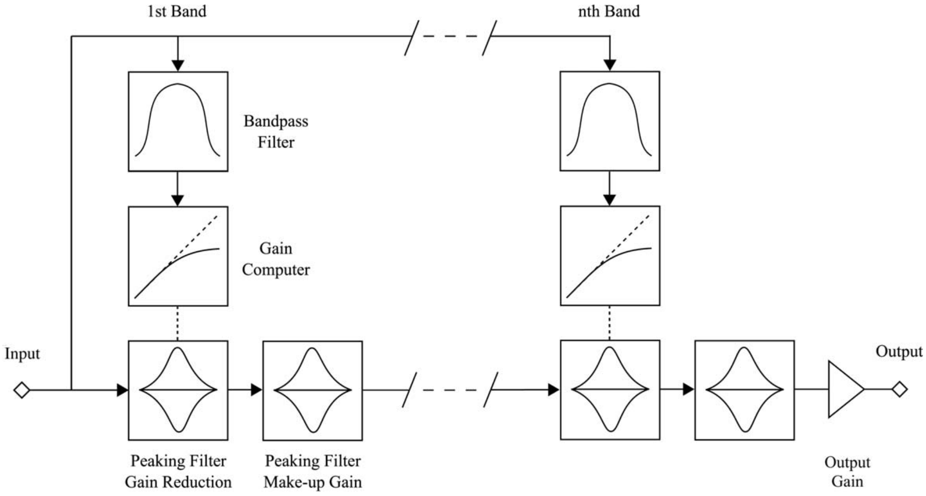

Fig. 2. Architecture of 10-band dynamic equalizer (DEQ10) and a 30-band dynamic equalizer (DEQ30) audio effects.

tion, the unprocessed snare drums from the test dataset of input-target pairs are transformed using the proposed audio effect configurations, where parameter values for each audio effect are inferred from the trained encoder network.

2.1 Dataset

In order to train DeepAFx to learn the most suitable parameters for any given audio processing task it requires input-target paired audio as supervision. The training data for each of the four transformation tasks is comprised of specific subsets from the Snare Drum Data Set (SDDS) [36].[3] From the four subsets, 3,000 input-target pairs were randomly selected to create the test set. SDDS is a comprehensive acoustic snare drum dataset, featuring multi-velocity recordings of ten different snare drums, each captured with 53 studio microphones, using various commercial dampening techniques.

One of the dampening methods used in SDDS was a BigFatSnareDrum (BFSD), a specialized device designed to dampen a snare or tom, placed directly on top of the batter head. This allows for exact repeatability because it covers the entirety of the drumhead and could only be placed in one position unlike other products. Although SDDS included other dampening methods such as MoonGel, BSFD was chosen to be used for the dampening transformations because it produces a distinct timbral change. The BFSD is also used for the D2U transformation. For each U2D and D2U input-target pair, the snare drum, microphone, and mic position were all identical, with the only variable being the dampening, either undampened or dampened with a BFSD.

Individual strikes from each pair were matched based on closest peak amplitude levels and time-aligned using cross correlation. For the positional transformation of T2B and B2T, only eight of the same microphones were used in both top and bottom positions. These pairs were used on all 10 snare drums and for all dampening methods; the paired strikes were identical performances because the top and bottom microphones were recorded simultaneously. For each subset, 80% was used for training, 10% for validation, and the remaining 10% for test data for later evaluation. Once processed by the trained models, the evaluation data was used for the comparative metrics and provided stimuli for the subjective listening tests.

[3]http://dmtlab.bcu.ac.uk/matthewcheshire/audio/sdds.

2.2 Subjective Evaluation

A subjective listening test was carried out using a multiple stimulus approach in order to determine whether participants would perceive the transformed audio as comparable to the real-world recording parameter adjustments it was emulating. The test was implemented using the Web Audio Evaluation Tool [37] and was carried out by 25 participants between the ages of 20–42 (mean: 27), and their experience in audio-related fields ranged from 1 to 25 years (mean: 9). Participants were instructed to use the highest-quality playback system available to them. They were required to provide the specification of equipment used, and all systems reported were deemed to be suitably professional.

The four transformations were evaluated on separated pages of the listening test. On each page, participants were presented with seven sliders, each corresponding to a different audio sample. The page and slider order were randomized, and slider starting position was as well. The seven audio stimuli were comprised of the unprocessed input used as a baseline for similarity, with the target acting as a *hidden reference*, and the five samples of the input processed by the five different plugin chains. Participants were instructed to arrange stimuli based on their similarity to the *reference* and use the full range of the scale, placing the most similar at the top and least similar at the bottom. The *hidden reference* was used to ensure participants could accurately identify the identical sample to the *reference*. No low anchor was used, in order to allow participants to rate the perceptually least similar stimuli lowest on the rating scale. Stimuli were loudness normalized to –23 LUFs.

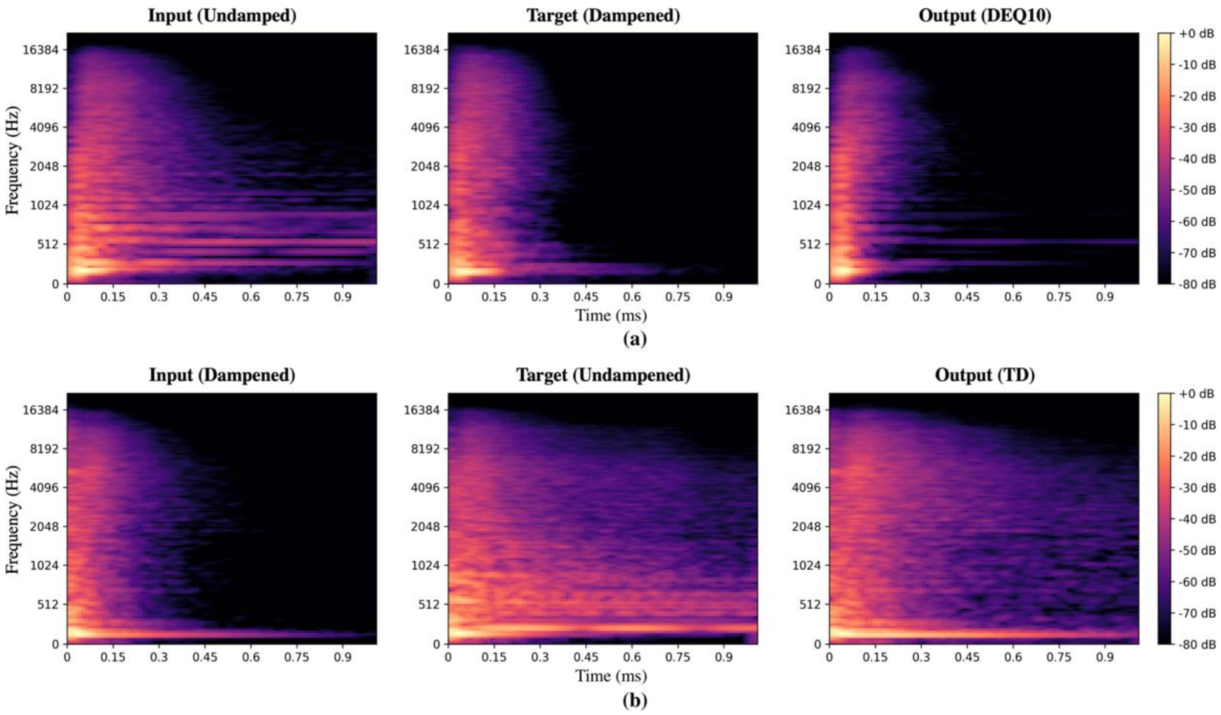

Fig. 3. Mel-scaled log frequency spectrograms for (a) Undampened-to-Dampened with 10-band dynamic equalizer (DEQ10) and (b) Dampened-to-Undampened with a transient designer (TD). Input snare drums (left), target (center), output transformations (right).

Listening test stimuli are available for audition.[4] The input-target pairs for each transformation were randomly selected from the test data subset (SEC. 2.1). Participants could not move on to the next page until all stimuli were played at least once and all sliders were moved. Fig. 3 presents an example of (a) U2D snare transformation using DEQ10 and (b) D2U snare transformation using TD.

2.3 Reconstruction Metrics

In order to evaluate the ability of the model to produce desired transformations of snare recordings, how accurately the transformed examples $\hat{S}$ match the target examples S Recording pairs in the test set introduced in SEC. 2.1 are evaluated using reconstruction metrics in two experiments comparing timbre and pitch characteristics of the transformed snare drums. Each transformation type is grouped into two tasks: 1) dampening (i.e., U2D and D2U) and 2) positional (i.e., T2B and B2T) and is evaluated with a range of spectral representations and metrics focused on timbral (see SEC. 2.3.1) and pitch (see SEC. 2.3.2) reconstruction capabilities of the model.

To extract the selected comparative metrics, a magnitude spectrogram S_{stft} is computed using the STFT for each audio file using an n-length Hann window ($n = 2{,}048$) with a hop size of $\frac{n}{4}$. S_{stft} is additionally mapped onto the Mel-scale or converted to Mel-frequency cepstral coefficients, resulting in S_{Mel} and S_{mfcc}, respectively.

[4] https://dmtlab.bcu.ac.uk/matthewcheshire/audio/jaes_samples/.

2.3.1 Timbral Reconstruction

Timbral reconstruction metrics in the first experiment include MSL (see SEC. 1.3) and spectral cosine distance (SCD) metrics as used in [22], along with log-spectral distance (LSD) [38] and Pearson correlation (PC) coefficients, which were previously employed in evaluations of deep generative models for music signals as an objective measure of audio quality [39, 40]. Additionally, the cosine similarity (CS) metric based on spectral difference functions (SDFs) used in research on automatic event detection [41] and automatic music remixing [42] are used. The implementation by [22] is followed for the computation of MSL and SCD metrics, in which the former uses STFT magnitudes and latter uses 13 Mel-frequency cepstral coefficients (excluding the first coefficient). The LSD is calculated using Mel-spectrograms as follows:

$$LSD_{Mel}(S, \hat{S}) = \sqrt{\sum [10 \log_{10}(|S|/|\hat{S}|)]^2}. \quad (2)$$

Following [41], spectral difference envelopes E are computed as

$$E_S(t) = \sum_{k=0}^{K-1} \Big\{ H(|S_k(t+1)| - |S_k(t)|) \Big\}, \quad (3)$$

where S represents a Mel-spectrogram with K bins. The $H(x) = (x + |x|)/2$ is a half-wave rectifier, which returns zero for negative arguments. The calculations of the E_S envelopes is the same for $E_{\hat{S}}$. Following [43], envelope reconstruction of the transformations is evaluated with cosine similarity calculated between envelopes extracted from target and transformed recordings as follows:

$$CS_{sdf}(S, \hat{S}) = \frac{E_S \cdot E_{\hat{S}}}{\|E_S\| \|E_{\hat{S}}\|}, \quad (4)$$

where $\cdot$ represents a dot product between E. CS_{sdf} will be close to unity for very similar drum envelopes and nearer to zero for dissimilar ones. Spectral difference functions are then calculated as the sum of the first-order difference between each spectrogram (e.g., [44]). The resulting envelopes are then normalized between [0, 1].

All reported timbral reconstruction experiments are presented as means calculated over the test set (see SEC. 2.1) except the MSL_{stft} metric, which is represented as the sum of L2 differences [see Eq. (1)]. Although the computation of PCs is described in the following section, here they are reported as mean PC coefficients averaged over the frequency axis.

2.3.2 Pitch Reconstruction

In the second experiment, a pitch-based reconstruction metric, which was previously used to evaluate the audio quality of pitched instruments generated with an adversarial autoencoder [39], was implemented. This approach is modified to suit snare drum frequency ranges. The use of Mel-spectrograms is opted for, as opposed to constant-Q transform spectrograms used in [39], because a logarithmically-spaced frequency range provides a more even representation over the fundamental frequencies of snare signals than frequency representations spaced over musical octaves (e.g., constant-Q transforms).

3 RESULTS

3.1 Subjective Results

3.1.1 Dampening

Fig. 4 presents normalized violin plots showing the dampening transformation results for the subjective listening test (means are depicted by asterisks, and medians are denoted by black horizontal lines). A one-way analysis of variance (ANOVA) was used to determine whether distributions of the responses have a common mean—that is, whether the plugin chains under evaluation had a different effect on the subjective scores of similarity. U2D (p = 3.12e–14) and D2U (p = 4.81e–14) both had $p < 0.05$. The small p values allow for rejection of the hypothesis that all group means are equal and indicate that the different ratings are not the same as each other.

A post-hoc multiple pairwise comparison was used to establish which of the ratings were significant based on the results from the ANOVA test. As per the recommendations in [45], Tukey's Honestly Significant Difference (HSD) test was used for this comparison. The U2D subjective listening test showed promising results. It can be seen in Fig. 4 that DEQ10 (mean: 0.66) and DEQ30 (mean: 0.58) are rated more similarly to the hidden reference (mean: 1) than the input (mean: 0.3). All participants correctly identified the hidden reference, placing it at the top of the rating scale.

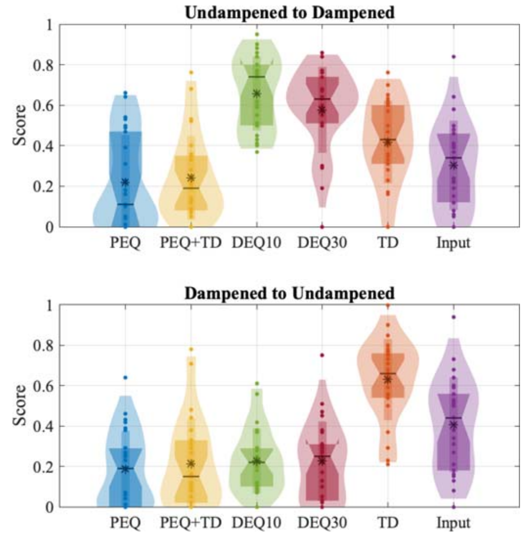

Fig. 4. Dampening results from listening test. DEQ10 = 10-band dynamic equalizer; DEQ30 = 30-band dynamic equalizer; PEQ = parametric equalizer; TD = transient designer.

The ratings for DEQ10 and DEQ30 were both statistically higher than the input (p = 2.07e–08 and p = 9.84e–06, respectively) using HSD. This suggests that both of these effects moved the processed input perceptually closer in similarity to the reference, which in this instance was a snare drum recording dampened with a BFSD.

Although not able to completely emulate the real dampening effect, these results indicate that the transformation is indeed creating a more dampened sound compared with the undampened recording. It should be noted that all participants were able to correctly identify the hidden reference and placed it at the top of the scale for all four test pages. Although TD (mean: 0.42) was rated higher than the input overall, the ratings were not significantly higher (p = 0.054). Likewise, although PEQ and PEQ+TD do have lower overall ratings than the input, they are not statistically different. For D2U, the only effect that had a significantly higher rating (p = 0.0012) than the input (mean: 0.4) was TD (mean: 0.63) based on HSD, which can be seen in Fig. 4.

3.1.2 Positional

The listening test results for the positional transformation are presented in Fig. 5. All participants correctly identified the hidden reference (mean: 1). An ANOVA was used again to determine whether any of the ratings were significantly different; for B2T transformations, it was found that there were no statistical differences between any of the scores (p = 0.42). This can be seen by the relatively close means and overlapping ranges of the different ratings. Although DEQ10 has a higher rating (mean: 0.49) than the input (mean: 0.36), these ratings were not statistically different from each other when the HSD test was conducted.

For T2B, some significant differences were shown based on the results from an ANOVA (p = 1.54e–11). The HSD test revealed that the performance of both PEQ and PEQ+TD (mean: 0.34 and 0.16, respectively) were statisti-

Table 1. Dampening task results using Mel-spectrograms: mean multi-scale loss (MSL), spectral cosine distance (SCD), log-spectral distance (LSD), mean Pearson correlation (PC), and envelope cosine similarity (CS). Lower values indicate greater similarity, except for the PC and CS metrics, for which higher values do.

	MSL$_{stft}$		SCD$_{mfcc}$		LSD$_{Mel}$		PC$_{Mel}$		CS$_{sdf}$	
Name	U2D	D2U	U2D	D2U	U2D	D2U	U2D	D2U	U2D	D2U
PEQ	8.31	65.57	0.75	0.90	2.53	3.09	0.68	0.52	0.86	0.69
TD	6.92	12.90	0.73	**0.85**	2.78	**2.72**	0.64	0.60	0.70	**0.91**
PEQ+TD	8.91	39.96	0.64	0.87	2.45	3.49	0.62	0.45	0.61	0.52
DEQ10	**4.77**	11.83	**0.55**	0.80	**2.13**	4.32	**0.70**	**0.68**	**0.89**	0.90
DEQ30	5.46	**8.01**	0.63	0.87	2.25	4.71	0.69	**0.68**	0.86	0.90

Bold values indicate best score (highest or lowest based on metric). DEQ10 = 10-band dynamic equalizer; DEQ30 = 30-band dynamic equalizer; D2U = Dampened-to-Undampened; mfcc = Mel-frequency cepstral coefficient; PEQ = parametric equalizer; sdf = spectral difference function; stft = short-time Fourier transform; TD = transient designer; U2D = Undampened-to-Dampened.

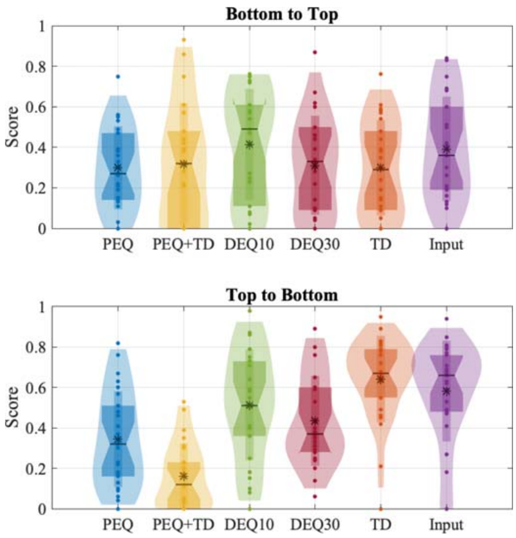

Fig. 5. Positional results from listening test. DEQ10 = 10-band dynamic equalizer; DEQ30 = 30-band dynamic equalizer; PEQ = parametric equalizer; TD = transient designer.

cally lower than the input (mean 0.58), with PEQ+TD being rated least similar to the target. TD had slightly higher ratings (mean 0.64) than the input, but again, these ratings were not statistically different from each other. This showed that for T2B positional changes, no method was successful at moving the input perceptually closer to the target, with both PEQ and PEQ+TD statistically worsening similarity. For B2T, no significant effects were seen, either positively or negatively, by any of the transformations.

3.2 Objective Results

Several of the objective metrics for U2D shown in Table 1 display similar trends to the subjective evaluations. For U2D all metrics showed DEQ10 to be most similar to the target. For D2U, TD rated most similar in the subjective evaluation and measured most similar when using SCD, LSD, and CS; however, unlike the subjective ratings when using MSL and PC, DEQ30 performed the best.

The objective metrics for the positional tasks can be seen in Table 2, DEQ10 had the highest similarity for T2B and B2T when measured with MSL and PC, respectively. PEQ also showed favorable results for T2B when using LSD and B2T when using both MSL and CS. TD was another effect that performed well across different metrics because it displayed the highest similarity with both SCD and PC for the T2B transformation. PEQ+TD was the only effect that presented strong similarity for one metric alone, with it scoring most similarly when using SCD for B2T.

4 DISCUSSION

The results from the listening test indicate that D2U may be a harder transformation to emulate than U2D, with both DEQ10 and DEQ30 being rated statistically more similar

Table 2. Positional task results, metrics are the same as those used in Table 1. Lower values indicate greater similarity, except for the PC and CS metrics, for which higher values do.

	MSL$_{stft}$		SCD$_{mfcc}$		LSD$_{Mel}$		PC$_{Mel}$		CS$_{sdf}$	
Name	B2T	T2B	B2T	T2B	B2T	T2B	B2T	T2B	B2T	T2B
PEQ	**7.86**	10.63	0.39	0.43	2.09	**2.11**	0.64	0.53	**0.91**	0.87
TD	10.16	7.35	0.40	**0.39**	2.34	2.07	0.61	**0.64**	0.89	**0.92**
PEQ+TD	17.86	23.09	**0.35**	0.42	**1.81**	2.45	0.52	0.38	0.48	0.57
DEQ10	8.17	**5.83**	0.54	0.50	2.39	2.54	**0.66**	0.54	0.83	0.87
DEQ30	8.27	6.33	0.68	0.62	2.61	3.01	0.65	0.54	0.81	0.88

Bold values indicate best score (highest or lowest based on metric). B2T = bottom-to-top microphone position; CS = cosine similarity; DEQ10 = 10-band dynamic equalizer; DEQ30 = 30-band dynamic equalizer; LSD = log-spectral distance; mfcc= Mel-frequency cepstral coefficient; MSL = multi-scale loss; PC = Pearson correlation; PEQ = parametric equalizer; SCD = spectral cosine distance; sdf = spectral difference function; stft = short-time Fourier transform; T2B = top-to-bottom microphone position; TD = transient designer.

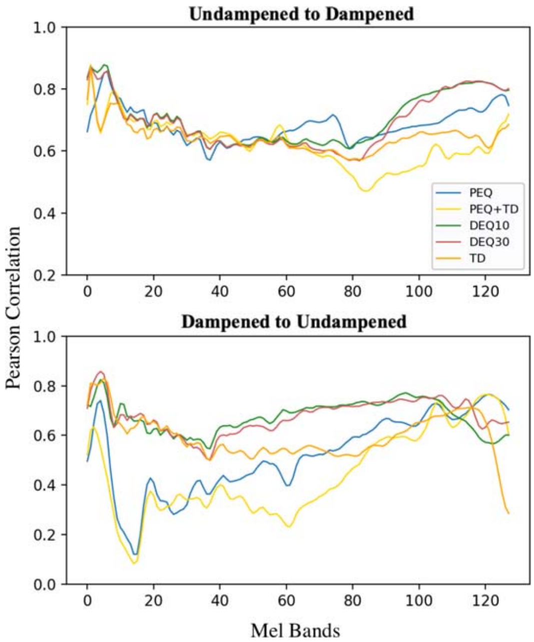

Fig. 6. Mean smoothed Pearson correlation results computed with Mel-spectrograms for the dampening task. DEQ10 = 10-band dynamic equalizer; DEQ30 = 30-band dynamic equalizer; PEQ = parametric equalizer; TD = transient designer.

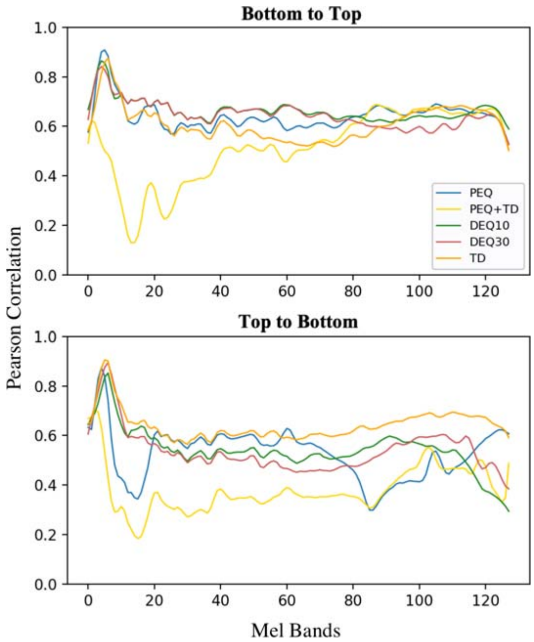

Fig. 7. Mean smoothed Pearson correlation results computed with Mel-spectrograms for the positional task. DEQ10 = 10-band dynamic equalizer; DEQ30 = 30-band dynamic equalizer; PEQ = parametric equalizer; TD = transient designer.

to the target for U2D but had ratings that were not significantly different to the input when used for D2U. Dampening a snare drum removes high-frequency energy, whereas removing dampening increases higher frequencies. When dealing with a heavily dampened snare recording, the high-frequency content has already been removed, and it shows that DeepAFx was not able to learn optimal parameters for the effects to enhance the missing information.

TD was most successful for the D2U transformation, likely because of TD's *release boost* parameter, shaping the envelope of the drum recording to better emulate an undampened strike. One possible alternation to DEQ10 and DEQ30 that may have facilitated better results for D2U would be to change the *ratio* parameter to allow values below 1. This would create an expansion effect instead of a compression effect for each frequency band, which could possibly be used to create a similar effect to that of the TD. Fig. 6 displays the mean smoothed PC results for the dampening tasks. High degrees of similarity to the target can be observed by both DEQ10 and DEQ30 only for the higher-frequency ranges for U2D. Little difference is seen between any of the plugins for the lower-frequency bands. Because high frequencies are most affected by dampening, the high measure of similarity in these important bands is likely responsible for the significantly higher ratings in the subjective evaluation.

For D2U, DEQ10 and DEQ30 have the highest measures of similarity in the mid-range and upper-mid–range frequency bands; however, this similarity is not reflected in the subjective tests. Although TD was subjectively the most similar to the target, the PC in Fig. 6 shows that it does not outperform DEQ10 or DEQ30, suggesting that envelope similarity is more important for D2U than spectral similarity. The subjective evaluation for B2T did not show any effect chain to statistically produce different ratings. In the case of T2B, PEQ and PEQ+TD produced ratings that were statistically lower than the input. A possible cause for this may be that the input is rated similar to the target. With little timbral disparity between input and target, it may be more difficult for DeepAFx to use the provided plugins to make the necessary improvement. PEQ and PEQ+TD also showed very low similarity for the mean smoothed PC results for T2B seen in Fig. 7, with the most notable dissimilarity being in the lower-frequency ranges and upper-mid range. PEQ+TD also showed very poor similarity in the lower frequencies for B2T; however, this was not reflected in the subjective evaluations.

The stimuli selected for the listening tests may not best exemplify the ideal transformation because the input-target pairs were chosen randomly from the available evaluation data. Thus, more representative samples that were not able to be assessed during the subjective evaluations may exist. Other variables, such as timbral differences associated with velocity disparity, could also play a part in the subjective perception of similarity. The effects of dampening or microphone placement may be less pronounced when the snare is played very lightly. Certain microphones may be more adept at capturing timbral subtleties making it easier for a listener to distinguish changes, or particular snare drums may emphasis the effects of parameter changes more so than others. The relationship between subjective ratings and objective metrics cannot be strongly linked because the objective measures made use of all samples from the evaluation data. In most cases DEQ10 outperformed DEQ30, which indicates that octave-band control (i.e., DEQ10) had sufficient timbral shaping abilities and third-octave band (i.e., DEQ30) had no additional benefits.

5 CONCLUSION

In this study, a deep learning system for automatic modification of snare drum recording parameters has been investigated. Two novel audio effects, an octave-band and third-octave–band dynamic EQ with fixed center frequency bands and trainable parameters, were created specifically for use within this system. Results from a subjective evaluation demonstrated that with particular effects, the system was able to move perceptually closer to the real-world targets for dampening tasks but was unsuccessful in positional transformations. Objective metrics also revealed a tendency toward improvements in similarity for certain transformations. Most notably, DEQ10 performed best at Undampened-to-Dampened in all measures.

A possible direction for future research in this area would be to assess the benefits of additional computational power, larger datasets, and alternative architectures to improve the quality of the transformations. The authors would also like to explore more aspects of the recording process, for example, transformations between different drum shell materials and investigation of other audio effects, such as distortions or reverbs for their timbral shaping capabilities. Additionally, subsequent studies could investigate methods for navigating the network's latent space. Navigation controls could be provided as a GUI to creatively interpolate between transformations or refine the estimated parameters.

6 REFERENCES

[1] R. Toulson, "The Perception and Importance of Drum Tuning in Live Performance and Music Production," in *Proceedings of the 4th Art of Record Production Conference* (Lowell, MA) (2008 Nov.).

[2] B. Owsinski and D. Moody, *The Drum Recording Handbook* (Hal Leonard, Milwaukee, WI, 2009).

[3] B. Bartlett and J. Bartlett, *Practical Recording Techniques: The Step-by-Step Approach to Professional Audio Recording* (Focal Press, Waltham, MA, 2008), 5th ed.

[4] K. Yoshii, M. Goto, and H. Okuno, "INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums," in *Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, pp. 205–212 (London, UK) (2005 Nov.). https://doi.org/ds85t4.

[5] D. Moffat and M. B. Sandler, "Machine Learning Multitrack Gain Mixing of Drums," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), e-Brief 527.

[6] M. A. Martínez Ramírez, D. Stoller, and D. Moffat, "A Deep Learning Approach to Intelligent Drum Mixing With the Wave-U-Net," *J. Audio Eng. Soc.*, vol. 69, no. 3, pp. 142–151 (2021 Mar.). https://doi.org/h4zh.

[7] M. Seymour, "Engineer's Guide To Tuning and Damping Drums," https://www.soundonsound.com/techniques/engineers-guide-tuning-and-damping-drums (2010 Aug.).

[8] N. D'Virgilio, "How to Control Drum Sustain With Dampening," https://www.sweetwater.com/insync/how-to-control-drum-sustain-with-dampening/#:~:text=Tear%20a%20small%20piece%20of,to%20really%20dampen%20the%20head.&text=Moongel%20is%20a%20great%20product%20to%20keep%20in%20your%20stickbag (2014 Sep.).

[9] M. Major, *Recording Drums: The Complete Guide* (Course Technology PTR, Boston, MA, 2014).

[10] M. H. Parsons, *The Drummer's Studio Survival Guide: How to Get the Best Possible Drum Tracks on Any Recording Project* (Modern Drummer Publications, Cedar Grove, NJ, 1996).

[11] B. Gibson, *Sound Advice on Recording & Mixing Drums* (Alfred Music, Los Angeles, CA, 2004).

[12] B. Owsinski, *The Recording Engineer's Handbook* (Bobby Owsinski Media Group, Burbank, CA, 2017), 4th ed.

[13] M. Cheshire, J. Hockman, and R. Stables, "Microphone Comparison for Snare Drum Recording," presented at the *145th Convention of the Audio Engineering Society* (2018 Oct.), paper 10040.

[14] K. Pedersen and M. Grimshaw-Aagaard, *The Recording, Mixing, and Mastering Reference Handbook* (Oxford University Press, New York, NY, 2019).

[15] M. Cheshire, R. Stables, and J. Hockman, "Microphone Comparison: Spectral Feature Mapping for Snare Drum Recording," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), paper 10263.

[16] M. A. Martínez Ramírez, E. Benetos, and J. D. Reiss, "Deep Learning for Black-Box Modeling of Audio Effects," *Appl. Sci.*, vol. 10, no. 2, paper 638 (2020 Jan.). https://doi.org/gppmr6.

[17] A. Wright, E.-P. Damskägg, L. Juvela, and V. Välimäki, "Real-Time Guitar Amplifier Emulation With Deep Learning," *Appl. Sci.*, vol. 10, no. 3, paper 766 (2020 Jan.). https://doi.org/ggvv8t.

[18] S. Hawley, B. Colburn, and S. I. Mimilakis, "Profiling Audio Compressors With Deep Neural Networks," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), paper 10222.

[19] J. Rämö and V. Välimäki, "Neural Third-Octave Graphic Equalizer," in *Proceedings of the 22nd International Conference on Digital Audio Effects*, paper 21 (Birmingham, UK) (2019 Sep.).

[20] S. Stasis, R. Stables, and J. Hockman, "Semantically Controlled Adaptive Equalisation in Reduced Dimensionality Parameter Space," *Appl. Sci.*, vol. 6, no. 4, paper 116 (2016 Apr.). https://doi.org/h4zj.

[21] J. Engel, L. H. Hantrakul, C. Gu, and A. Roberts, "DDSP: Differentiable Digital Signal Processing," in *Proceedings of the International Conference on Learning Representations* (Addis Ababa, Ethiopia) (2020 Apr.).

[22] M. A. Martínez Ramírez, O. Wang, P. Smaragdis, and N. J. Bryan, "Differentiable Signal Processing With Black-Box Audio Effects," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Sig-*

nal Processing, pp. 66–70 (Toronto, Canada) (2021 May). https://doi.org/h4zk.

[23] J. Lee, N. J. Bryan, J. Salamon, Z. Jin, and J. Nam, "Metric Learning vs Classification for Disentangled Music Representation Learning," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, pp. 439–445 (Montreal, Canada) (2020 Oct.).

[24] C. Szegedy, W. Liu, Y. Jia, et al., "Going Deeper With Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (Boston, MA) (2015Jun.).

[25] A. Fox, "Top 10 Best Dynamic EQ Plugins for Your DAW in 2022," https://mynewmicrophone.com/top-best-dynamic-eq-plugins-for-your-daw/ (2022).

[26] M. Hahn, "How to Use Dynamic EQ for A Better Mix," https://blog.landr.com/dynamic-eq/ (2020 Jul.).

[27] I. Stewart, "How to Use Dynamic EQ in Mastering," https://www.izotope.com/en/learn/how-to-use-dynamic-eq-in-mastering.html (2021 Sep.).

[28] G. Brown, "Multiband Compressors vs. Dynamic EQs: Differences and Uses," https://www.izotope.com/en/learn/multiband-compressors-vs-dynamic-eqs.html (2020 Apr.).

[29] ANSI, "Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters," *Standard ANSI S1.11-2004* (2009 Jun.).

[30] D. K. Wise, "Concept, Design, and Implementation of a General Dynamic Parametric Equalizer," *J. Audio Eng. Soc.*, vol. 57, no. 1/2, pp. 16–28 (2009 Jan.).

[31] R. Izhaki, *Mixing Audio: Concepts, Practices, and Tools* (Taylor & Francis, New York, NY, 2017), 3rd ed. https://doi.org/h4zn.

[32] H. Gier and P. White, "Transient Designer - Dual-Channel, Model 9946," https://spl.audio/wp-content/uploads/transient_designer_2_9946_manual.pdf (1999 Jul.).

[33] B. Owsinski, *The Mixing Engineer's Handbook* (Bobby Owsinski Media Group, Burbank, CA, 2016), 4th ed.

[34] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the International Conference on Learning Representations* (San Diego, CA) (2015 May).

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep Into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034 (Santiago, Chile) (2015 Dec.). https://doi.org/gfw55h.

[36] M. Cheshire, "Snare Drum Data Set (SDDS): More Snare Drums Than You Can Shake a Stick At," presented at the *149th Convention of the Audio Engineering Society* (2020 Oct.), e-Brief 626.

[37] N. Jillings, D. Moffat, B. De Man, and J. D. Reiss, "Web Audio Evaluation Tool: A Browser-Based Listening Test Environment," in *Proceedings of the 12th Sound and Music Computing Conference*, pp. 147-152 (Kildare, Ireland) (2015 Jul.).

[38] A. Bitton, P. Esling, A. Caillon, and M. Fouilleul, "Assisted Sound Sample Generation With Musical Conditioning in Adversarial Auto-Encoders," in *Proceedings of the International Conference on Digital Audio Effects*, paper 24 (Birmingham, UK) (2019 Sep.).

[39] J. W. Kim, R. Bittner, A. Kumar, and J. P. Bello, "Neural Music Synthesis for Flexible Timbre Control," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 176–180 (Brighton, UK) (2019 May). https://doi.org/h4zq.

[40] M. Tomczak, J. Drysdale, and J. Hockman, "Drum Translation for Timbral and Rhythmic Transformation," in *Proceedings of the International Conference on Digital Audio Effects*, paper 25 (Birmingham, UK) (2019 Sep.).

[41] J. P. Bello, L. Daudet, S. Abdallah, et al., "A Tutorial on Onset Detection in Music Signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047 (2005 Sep.). https://doi.org/dw7kcc.

[42] M. E. P. Davies, P. Hamel, K. Yoshii, and M. Goto, "AutoMashUpper: Automatic Creation of Multi-Song Music Mashups," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 12, pp. 1726–1737 (2014 Dec.). https://doi.org/h4zr.

[43] M. Tomczak, M. Goto, and J. Hockman, "Drum Synthesis and Rhythmic Transformation With Adversarial Autoencoders," in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2427–2435 (Seattle, WA) (2020 Oct.). https://doi.org/h4zs.

[44] S. Dixon, F. Gouyon, and G. Widmer, "Towards Characterisation of Music via Rhythmic Patterns," in *Proceedings of the International Conference on Music Information Retrieval*, pp. 509 (Barcelona, Spain) (2004 Jan.).

[45] ITU-R, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," *Recommendation ITU-R BS.1534-3* (2015 Oct.).

THE AUTHORS


Matthew Cheshire


Jake Drysdale


Sean Enderby


Maciek Tomczak


Jason Hockman

Matthew Cheshire is a Ph.D. student in the SoMA Group at Birmingham City University. His main interests are timbral modifications, recording techniques, and perceptual evaluations. He has been a member of the AES since 2018, and he was the recipient of the Saul Walker Award (2019) and the John Eargle Award (2020 and 2021).

•

Jake Drysdale is a Ph.D. student in the SoMA Group, where he specializes in deep learning methods to assist in sample-based music production. His main interests are neural audio synthesis and music sample retrieval.

•

Sean Enderby completed his Ph.D. in Intelligent Music Production at Birmingham City University in 2017. He is currently a Lecturer and Researcher in the SoMA Group. His main areas of interest are music digital signal processing, virtual analog, and "intelligent" parameterization of effects.

•

Maciek Tomczak is a Ph.D. student in the SoMA Group. His main areas of interest are machine learning, computational rhythm analysis, and audio style transfer.

•

Jason Hockman received a doctorate in Music Research from McGill University in 2014. He is currently Associate Professor of Audio Engineering at Birmingham City University, where he leads the SoMA Group. His main areas of interest are in music informatics and computational rhythm and meter analysis.