MSc in Audio Production

School of Computing, Science and Engineering

University of Salford MANCHESTER

MSc Dissertation

EVALUATION OF PITCH MODULATION AS A CREATIVE EFFECT

Author: Jay Harrison

Supervisor: Dr Philip Duncan

2019

Abstract

Pitch modulation is employed by vocalists and musicians as a musical device in the form of frequency vibrato. Wow & Flutter is pitch modulation caused by mechanical inconsistencies in the recording and playback of analogue media. The pitch modulation introduced by Wow & Flutter has traditionally been considered objectionable by listeners and industry. However, in the last 20 years Wow has become a favourable creative effect that is often used in the production of several Hip Hop, Lofi & Alternative subgenres. This study investigates the preferential perception of Wow as a creative effect. Existing research is reviewed and provides a foundation for the study. Listening test stimuli is produced by a modified cassette deck, enabling discrete control over the pitch modulation of program material. Two subjective listening tests consisting of 20 participants each are carried out to assess the influence of modulation rate and extent on preferential perception. The results suggest statistical significance between modulation rate and participant preference and no statistical significance between modulation extent and participant preference. It is recommended that further research is carried out to investigate the influence of other modulation parameters on preferential perception.

Acknowledgements

I would like to thank Dr Philip Duncan for his expert guidance and supervision of the project; Niels Larsen of The Reel Deal Denmark for providing calibrated cassette tapes used in the production of the listening test stimuli; Ben Smith of PandaSound for assisting with the servicing and calibration of the modified cassette deck; Henric Mattsson for PAT testing the modified cassette deck; Professor Chris Coleman and the development team of Maxuino for providing intuitive interfacing of Max/MSP & Arduino and the Institute of Sound Recording at the University of Surrey for their MUSHRA Max/MSP patch.

Contents

List of Figures

List of Tables

1. Introduction

The modulation of pitch has been and continues to be a popular musical device. It is achieved using different techniques on a wide range of instruments and is found in most musical cultures and genres around the world. Traditionally pitch modulation is a device that is brought about as a result of a musician's technique during performance, usually in the form of vibrato. Slower pitch modulation is also generated as a symptom of poorly performing tape machines and record players. This type of modulation is referred to as Wow & Flutter and has historically been considered an undesirable, unpleasant effect.

More recently however developments in several musical subgenres have come to feature this slower pitch modulation as a prominent intentional creative effect. This thesis will explore a review of literature concerning pitch modulation. The procedure and findings of subjective listening tests designed to examine the perception of Wow as an intentional creative effect will then be documented.

The motivation for this study primarily stems from the observation of the prominence of exaggerated Wow as a constituent part of several disparate musical genres. It is of particular interest that this development has occurred from the midst of Wow & Flutter traditionally being widely considered so profoundly undesirable.

1.1. Aims & Objectives

The following aims have been identified as being suitable focal points for this study:

- Conduct a comprehensive review of literature relevant to pitch modulation.

- Execute new research exploring the perception of Wow as a creative effect.

In order to realise these aims several objectives have been identified as being necessary for the successful completion of the study. They are as follows:

- Review current literature including research on the psychoacoustic perception of pitch modulation & a review of the use of Wow in commercially released music.

- Develop a bespoke tape machine capable of producing pitch modulation effects with control over rate, depth and waveform.

- Produce a number of listening stimuli to be processed by the tape machine.

- Design and carry out controlled subjective tests that will allow the comparison of different pitch modulation parameters.

- Conduct a statistical analysis of the results of the subjective tests.

- Draw conclusions on the influence of different pitch modulation parameters on listener perception.

2. Background & Literature Review

The focus of the experiments carried out in this study is to investigate the critical perception of Wow as an intentional creative effect. At the time of writing there have been no published studies looking at Wow as a creative effect, this is most likely due to it only recently developing within several musical subcultures over the past 25 years. In the context of this study the term pitch modulation refers to the broadly cyclical fluctuation of frequency observed in an audio signal. It could be argued that the most common example of intentional pitch modulation is frequency vibrato. Whilst frequency vibrato is primarily a musical device used by musicians during performance, its two main parameters of rate and extent are shared with Wow.

Research and experiments have been carried out on vibrato, including studies looking at the average modulation rates and extents exhibited by performers, in addition to the perception of these parameters by the average listener. As these parameters are shared by both frequency vibrato and Wow it was considered appropriate to study frequency vibrato to examine its critical perception relative to Wow and the results of this study. All available literature considered relevant to the study of Wow is also presented to provide a contextual background and foundational basis for the research carried out in this study.

2.1. Vibrato Definition

Vibrato is a performance technique used by vocalists and instrumentalists to rapidly modulate the frequency of notes around a central pitch. It is formally defined by the American National Standards Institute (1960) as-

> *"…a family of tonal effects in music that depend on periodic variations of one or more characteristics in the sound wave*."

Within this definition several further categories exist including frequency vibrato, amplitude vibrato and phase vibrato. The resonances of the instrument or voice and an acoustic performance space means frequency vibrato and amplitude vibrato are almost always heard together (Rossing et al., 2002, pp. 141-142). As this study is concerned primarily with pitch modulation it has been decided that only frequency vibrato will be discussed in detail.

Frequency vibrato has two main parameters, these are rate and extent. The rate describes the number of fluctuations in a second, whilst the extent indicates the depth of the frequency fluctuation. The extent is most often expressed in cents, in some instances however it is illustrated as a percentage of the average frequency (Sundberg, 1999, pp. 195-197). A cent is a unit of pitch used to describe the intervals within a whole pitch, 100 cents are equal to one semitone. Seashore (1967, p. 141) states that a good vibrato is-

> *"…of such extent and rate as to give a pleasing flexibility, tenderness, and richness to the tone."*

2.1.1. Frequency Vibrato Rate

Rossing et al. (2002, pp. 141-142) state that the rate of a vibrato varies between performers and that the average for vocalists and instrumentalists is 7 Hz. The rate of a natural vibrato has also been shown to vary slightly over the course of a tone. Studies conducted by Seashore (1931, pp. 623-624) found that vibrato produced by singers and instrumentalists both averaged approximately 6 or 7 Hz, with 50 percent of participants exhibiting vibrato rates between 6 and 7 Hz. Risset & Wessel (1999, p. 130) define the frequency vibrato rate of a violinist to be around 6 Hz, similarly Backus (1977, p. 208) describes the average frequency vibrato rate for violin to be 5 to 6 Hz. Prame (1994, pp. 616-621) conducted a study of 10 singers measuring the rate of their frequency vibrato. The mean rate was 6 Hz. The maximum varied between 5.8 and 7.3 Hz, whilst the minimum was between 5.0 and 6.0. Ferrante (2011, p. 1687) conducted a review of vibrato rate and extent in soprano voice over the last century. It was found that there was a clear decrease in mean rate by 1.8 ± 0.3 Hz. Vibrato rates from the beginning of the 19^{th} century tended to be around 7 Hz, whereas vibrato from the end of the century were predominately in the region of 5 Hz.

2.1.2. Frequency Vibrato Extent

Studies carried out by Seashore (1931, p. 623) determined that the average vibrato extent produced by singers possessed a deviation of approximately half a tone, whilst the average vibrato extent for string instruments was approximately a quarter of a tone. Backus (1977, p. 208) defines the average extent of frequency vibrato exhibited by violinists as being between 30 and 45 cents. Ferrante's (2011, p. 1687) study on vibrato rate and extent in soprano voice over the previous century found a clear increase in mean extent of 56.4 ± 0.3 cents. Mean Vibrato extent at the start of the 19th century was in the region of 0.25 cents, whereas a century later extent was found to be around 0.8.

2.1.3. Perception of Frequency Vibrato Rate

Rossing et al. (2002, pp. 141-142) report that it is possible to recognise the periodicity of pitch change in vibrato between the rates of 1 & 5 Hz, with it being most clear around 4 Hz. Between 6 and 12 Hz the tone is perceived as a single average pitch, whilst vibratos with rates of over 12 Hz are interpreted as confusing and undesirable combinations of multiple tones. Zwicker & Fastl (1999, pp. 185-186) also states that the human auditory system is better able to recognise abrupt changes in frequency as opposed to continuous modulations. Corso & Lewis (1950, pp. 206-112) conducted studies on the preferred rate and extent of frequency vibrato, the results established that non-expert listeners considered rates of 6, 6.5 & 7 Hz as being equally preferable, whilst expert listeners preferred 6 & 6.5 Hz equally.

Fletcher (2010, p. 4) studied the physics and psychophysics of vibrato and stated that in the case of operatic singing, vibrato adds an emotional character to the musical line. It is suggested that the emotional perception of vibrato potentially indicates related psychological effects. This is supported by Fletcher's assertion that vibrato frequencies are typically between 5 & 10 Hz, equating to 0.1 to 0.2 seconds, which is equal to the onset time of numerous muscular and neural responses in humans. Fletcher also refers to the contiguity between the typical range of vibrato rates and the 8 to 12 Hz range of the 'alpha rhythm' of brain cell oscillation, as measured by electro encephalography. It is noted that these observations are as yet unproven suppositions. However, they were included to highlight the various academic hypotheses of potential influence of psychological effects on pitch modulation perception.

2.1.4. Perception of Frequency Vibrato Extent

Seashore (1931, p. 625) states experiments have shown that extents of a quarter of a tone are perceived as the most pleasing for singers and instrumentalists. Seashore (1967, pp. 45-46) also asserts that vibrato extent is perceived by the listener as being very much smaller than the actual extent observed in the physical tone. He states that this is part of a wider auditory illusion whereby the level of perceived extent underestimation is increased by larger extents, faster rates and the richness of tone. Research conducted by Corso & Lewis (1950, pp. 206-212) found that non-expert listeners considered extents of 25 cents to be most preferable, in contrast expert listeners considered extents of 10 cents to be most favoured.

Benade (1990, p. 375) shows that there is an inherent natural frequency vibrato present in the voice. Without the addition of intentional performed frequency vibrato, the human voice possesses an irregular random fluctuation with a typical extent of 0.5 percent with a series of random rates. This is attributed to unavoidable muscle tremor and unstable aerodynamic flow present in the voice. Some of the irregularity exhibited in performed frequency vibrato can be attributed to this phenomenon.

2.1.5. Artificial Frequency Vibrato

Typically, vibrato that is created by electronic instruments is perceived as static and artificial. Taylor (1992, pp. 81-83) states that humans are highly skilled at recognising the irregularities that are present in natural vibrato performed by singers or instrumentalists. Electronic vibrato following a perfectly sinusoidal oscillation is therefore easily identified as artificial. Pierce (1999, p. 8) supports this point of view, stating that the shared rise, fall and wavering supplied by frequency vibrato is an essential characteristic of traditional instruments, the lack of which can cause synthesized tones to be perceived as unnatural.

The Moog Theremin is a monophonic electronic synthesizer, the frequency of the note it produces is controlled by moving your hand closer or further away from a sensor. It is difficult to maintain accurate intonation playing in this way, therefore Theremins include an optional frequency vibrato effect. Seashore (1967, p. 267) states that without this vibrato the intonation of the notes is intolerable, however its addition has the effect of smoothing the intonation pitch perception, allowing for a greater margin of error in performance accuracy.

2.2. Wow & Flutter Definitions

The following definitions are provided for Wow and Flutter by Bartlett (1975)-

> *"Flutter (wow): In recording or reproducing, flutter is the deviation in frequency or pitch which results from minor periodic or random changes in the motion of the medium. (Note: The term "flutter" usually refers to cyclic deviations occurring at a relatively high rate, as, for example, 10 cycles per second. The term "wow" usually refers to cyclic deviations occurring at a relatively low rate as, for example, a once-per- revolution speed variation of a turntable.)"*

Drift is defined by the International Electrotechnical Commission (1972) as the-

> *"Slow variation of the velocity of the recording medium during recording and reproduction."*

The majority of academic discussion regarding Wow and Flutter was published between the 1960's and 1980's when tape machines and vinyl records were the dominate formats in recording and record distribution. These definitions were selected as they provided insightful descriptions of the effects of Wow, Flutter and Drift. The modulation rates specified with each definition have since been revised and updated by more modern standards. The most recent standard to provide definitions is provided by the Audio Engineering Society (2013), this is detailed later in the study.

Burstein (1975, pp. 138-139) states that Wow & Flutter should be inaudible in a high-quality tape machine and that professional standards specify that Wow extent should be no more than 0.2 percent of the total signal. Godsill & Rayner (1998, p. 6) describes Wow and Flutter as pitch variation defects that result in a very disturbing modulation of all frequency components. These definitions serve to illustrate the extent to which Wow & Flutter have been universally considered highly unpleasant and undesirable effects by the academic community.

Wow is normally caused by one or more inherent mechanical faults in the playback systems of tape machines, record players or media. Consequently, the influence of a Wow modulation tends to persist for the duration that its mechanical faults remain present. In contrast however, Flutter is often brought about by various intermittent faults in the transport system or tape media, tending to result in sporadic and unpredictable signal distortion.

2.2.1. Measurement Standards of Wow & Flutter

The most recent standard outlining best practice for measuring speed fluctuations in analogue recording and playback equipment is AES6-2008 (s2013) (Audio Engineering Society, 2013). This is stated as being technically identical to IEC 60386-am1 (IEC, 1988). It also asserts that measurements made in accordance with the AES6-2008 (s2013) are identical to those made adhering to IEEE Std-193 (IEEE, 1971), IEC 60386 (IEC, 1972), CCIR 409-2 (CCIR, 1970) and DIN 45507 (Deutsches Institut für Normung, 1966).

All of these standards provide specifications for the performance of Wow and Flutter meters. Wow and Flutter meters work by measuring the speed fluctuation in a piece of equipment and calculating a nominal percentage of Wow. Specially created media including vinyl records, cassettes & reel to reel tape is produced containing highly accurate, calibrated test tones. The rendered tones on this media are produced using extremely high-quality equipment and feature exceptionally low Wow. The most common tone specified by current and past standards is 3,150 Hz (McKnight, 1972, p. 78), however 3,000 is also supported with most hardware and software meters.

The media is played back on the piece of equipment under test and the signal is fed to the Wow and Flutter meter. The meter will measure any fluctuation away from the nominal 3,150 or 3,000 Hz and produce the Wow percentage figure. Modern measurement standards factor in a modulating frequency weighting to account for the subjective perception of human hearing. It has been shown that the human ear is most sensitive to pitch modulation between the modulating frequencies of 0.2 and 200 Hz, peaking at around 4 Hz (Hongwei, 2019, p. 4). Figure 2.1 shows the modulating frequency weighting curve as specified by (IEC, 1972). Wow meter readings can be expressed in either Peak or RMS as a percentage. Despite meters measuring peak to peak values the peak reading is actually expressed as one half of the peak to peak value.

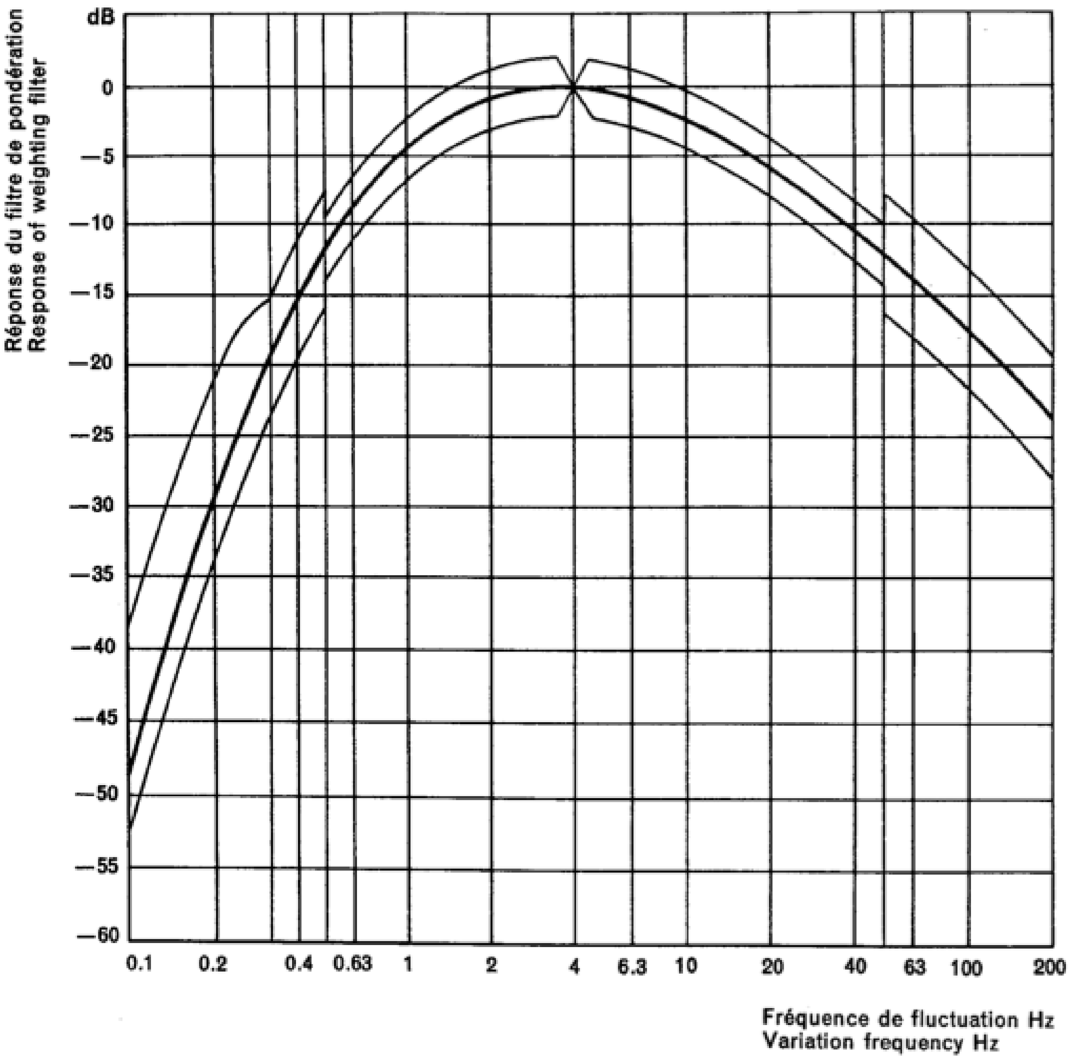

Figure 2.1 Modulating frequency weighting curve (IEC, 1972).

2.2.2. Development of Wow & Flutter Meter Weighting Curve Standards

Current standards concerning the measurement of Wow & Flutter specify that metering hardware and software should include a weighting curve to account for the way that the human ear perceives pitch modulation. The following research has been selected to demonstrate how the pitch weighting curves found in current Wow measurements standards were developed.

The development of the weighting curve originally stemmed from research by Shower & Biddulph (1931, pp. 275-287) on the “Differential Pitch Sensitivity of the Ear”. This study explored the minimum percentage change in frequency and sensation that is detectable by the human ear. Detectability was measured between a frequency range of 32 to 11,600 Hz using test tones supplied binaurally via headphones and separately through bone conductions. From 1000 Hz and above the minimum detectable frequency change remained approximately constant at 0.003%. In the range of 500 to 1000 Hz frequency percentage remained constant across all sensation levels. Below 500 Hz the ear is found to be more sensitive to frequency variation.

It is important to make the distinction that this study and the related differential pitch sensitivity research that led to the development of the weighting curve found in current Wow & Flutter measurement standards is concerned only with how sensitive the ear is to variations in pitch. It does not address the question of how Wow is perceived preferentially by listeners in either a critical or favourable context.

Stott & Axon (1955, pp. 643-656) carried out further experiments on the discrimination of pitch in the context of recording systems. This research used complex signal-based test stimuli to more accurately represent the musical program material commonly afflicted by Wow from recording systems observed in the field. The study was also devised with the expressed intention of using the results in the development of a universal pitch weighting curve specifically for Wow & Flutter measuring equipment.

Comerci & Oliveros (1956, pp. 419-425) conducted experiments to test the perceived accuracy of an existing flutter index weighting curve. The study concluded that no relationship was observed between peak-to-peak flutter measurements, however an appreciable relationship was found between the established flutter index and preferential ranking. This remained the case with the exception of flutter rates from 10 to 25 Hz, where a non-linearity was observed between the subjective preferential ranking and the flutter index. In order to address this two flutter rate weighting networks were used to form a modified flutter index meter that employed automatic switching between the alternate weighting networks, this was controlled by the unweighted flutter amplitudes observed at rates exceeding 10 Hz. The study also found that the flutter index weighting curve originally derived from experiments conducted using test tones could be reliably extrapolated to music and speech program material.

In Comerci (1962, pp. 1-8) the concept of the flutter index is presented formally in the Journal of the Society of Motion Picture and Television Engineers. The concept is proposed as forming the basis of a new IRE standard designed to measure the subjective effect of flutter on program material. The design criteria characteristics of a new flutter index meter are also presented including operation using a 3,000 Hz test signal, the ability to disregard amplitude modulation from the test signal, the facilitation of modulation flutter frequencies ranging between 0.25 and 100 Hz and metering sensitivity derived from flutter index weighting curves presented in the research.

Belger (1972, pp. 79-80) presents an updated translation of a paper originally published in Belger (1958, pp. 168-169) that summarized the work of several studies on flutter measurement frequency weighting curves. Figure 2.2 is a graph from the paper showing a collection of several frequency weighting curves presented in prominent studies published by Zwicker (1952, pp. 239-246), Zwicker & Kaiser (1952, pp. 125-133), CCIR (1953), Zwicker (1953, 342-346), Stott & Axon (1955, pp. 642-656), CCIR (1956) and Comerci & Oliveros (1956, pp. 419-425). Some of the larger variations between curves are attributed to some of the studies using music and speech as the test program test material rather than pure tones. It is however observed that they all follow a similar trend. As such the weighting curve shown in

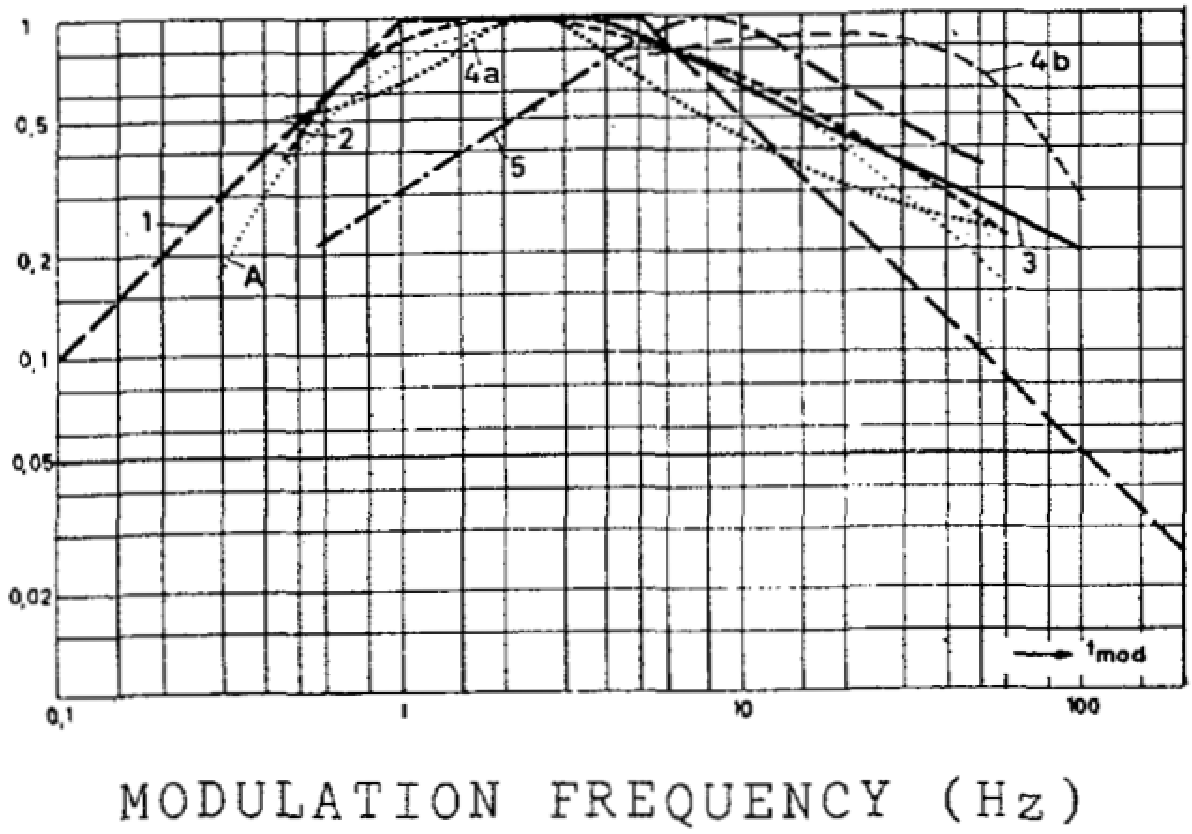

Figure 2.2 Several Modulating frequency weighting curve (Berger, 1972).

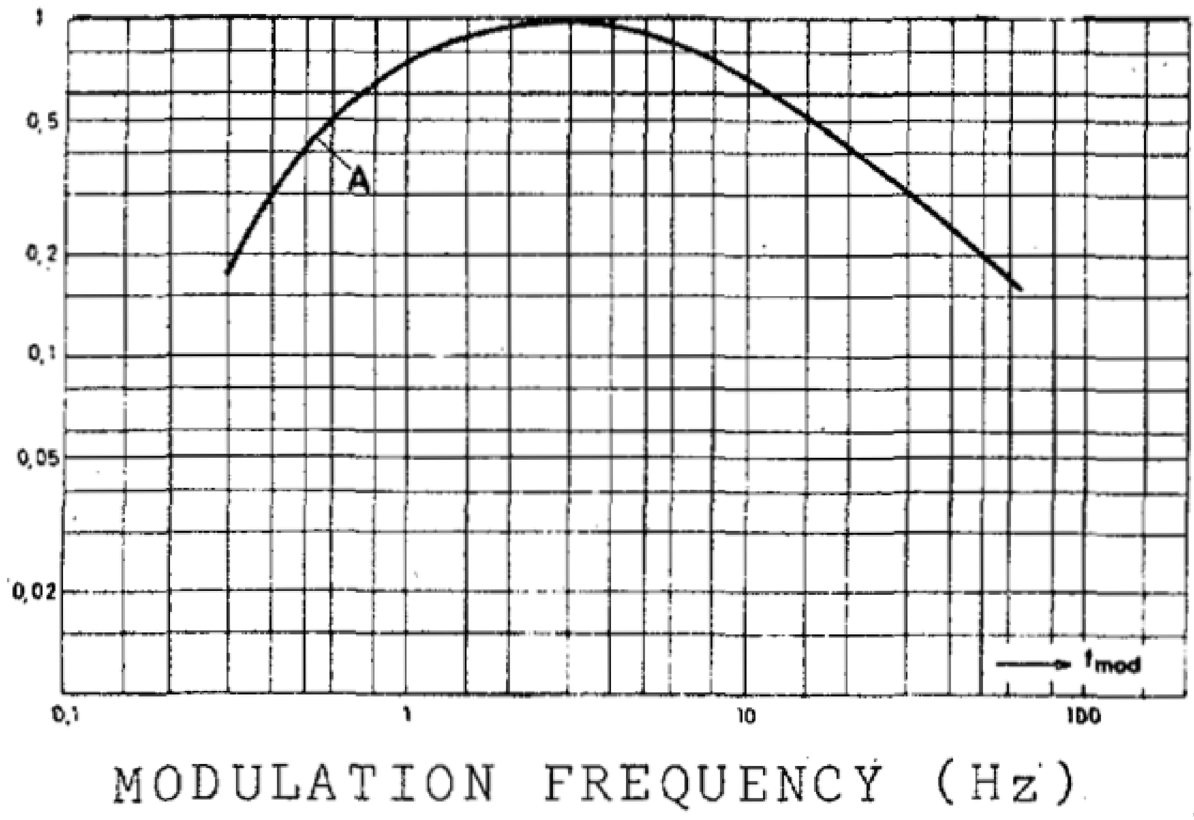

Figure 2.3 Modulating frequency weighting compromise (Berger, 1972).

Figure 2.3 was proposed as a compromise between the data presented by the aforementioned literature. This weighting specifies the curve peaking at a modulating frequency of 4 Hz, meaning that the subjective effects of Wow are considered at their greatest around this point.

This weighting curve was consequently adopted by the German DIN 45507 standard in 1962, later revised to DIN 45507-10 (Deutsches Institut für Normung, 1966). This standard would remain in service until it was superseded by DIN IEC 60386:1995-04 (Deutsches Institut für Normung, 1995) which still specified the use of the same frequency weighting curve based on the proposal by Belger (1958). It was also adopted in the following standards (National Association of Broadcasters, 1965), (CCIR, 1966), (CCIR, 1970) and (IEEE, 1971) effectively making it the universally agreed weighting curve for subjective modulation frequency perception.

2.2.3. Causes of Wow in a Tape Machine

There are several mechanical issues that can cause a tape machine to impart Wow, the most common faults being found within the tape transport system. The speed that tape passes the tape machine playback head is directly related to the pitch at which the material on the tape is played back. Faster playback speeds will result in higher pitched material playback. As such any fluctuation in the speed of the tape will translate directly to fluctuations in pitch, potentially producing Wow.

Different faults will impart specific fluctuations to the rate and extent. Dry bearings will most likely not impart the same pitch fluctuation as a worn capstan roller. The superposition of several mechanical faults and their associated nominal rate and extent parameters often results in random subtle irregularities and unique Wow pitch modulations. The various constituent modulations caused by the faults will fluctuate with independent phase, this also adds to the unpredictable irregularity of the overall pitch modulation once summed. Due to the unpredictable nature of these mechanical faults and their specificity to the exact tape machine model, it is difficult to quantify these random irregularities.

Mechanical faults that can influence the tape speed include warped spool cheeks, dry bearings, sticky or worn rubber rollers, a layer of oil or oxide on the drive spindle, worn drive belts, a damaged idler wheel and an imbalanced flywheel (Capel, 1972, pp. 34-35). Godsill & Rayner (1998, p. 171) asserts that it is also possible for the tape media itself to become warped or stretched during storage or playback which would result in pitch modulation during playback. This is a problem often encountered by archivists working to digitise collections of tape and film artefacts.

Tape machines are designed to eliminate as much Wow as possible and certain mechanical design features such as the flywheel are able to partially mitigate some of the faults that can lead to fluctuations in tape speed. Flywheels are used in the drive mechanism of tape machines to resist changes in speed using their inertia. Flywheels that rotate at faster velocities and possess greater mass will tend to provide improved resistance to fluctuations in speed (Capel, 1972, pp. 34-35). Hood (1999, p. 25) describes higher quality machines as often using a dual-capstan system featuring two capstan drives that ensure the tape is kept at a constant tension whilst being pulled across the playback head. They will also tend to make use of speed-controlled motors that can guarantee a greater consistency of speed through the use of internal feedback systems and electromechanical governors.

It is possible to artificially introduce mechanical Wow into a tape machine transport system by adding simple modifications to certain components. Capel (1972, pp. 134-135) describes how this can be achieved by applying a piece of adhesive tape around the drive capstan, ensuring that the ends overlap to create an uneven bulge. The profile of the modified drive capstan will behave as an offset cam, resulting in an irregular speed and pitch modulation of the magnetic tape.

2.2.4. Causes of Flutter in a Tape Machine

In tape machines Flutter is usually either caused by a hard pressure pad or the friction causing low quality unpolished tape to catch on the playback head. This can result in the tape moving past the head in a series of abrupt, jerky motions which is perceived aurally as Flutter with the occasional addition of noise and distortion (Capel, 1972, p. 170).

2.2.5. Causes of Wow in a Record Turntable

Wow can also be present as a result of mechanical faults in record turntables and vinyl records. Godsill & Rayner (1998, p. 171) states that Wow can be imparted as a result of the hole in a record not being punched exactly in the centre. This results in an eccentricity in the rotational speed of the record which translates into an audible Wow pitch modulation. Wow can also be introduced if the turntable is susceptible to any low frequency vibrations or rumble, these low frequency vibrations can then cause subtle speed variations in the motor (Backus, 1977, p. 322). It is also possible that the vinyl records themselves may become warped. Usually this is due to prolonged storage in an environment that experiences regular and pronounced fluctuations in ambient temperature, causing the vinyl to warp from constant expansion and contraction.

2.2.6. Wow & Flutter Rates

Burstein (1975, pp. 138-139) defines Wow as modulating with a frequency of up to 10 Hz with modulations beyond 10 Hz being classified as Flutter. Eargle (2003, p. 161) describes Flutter as modulations in the range of 10 to 30 Hz. IEC 60386 (1972) defines Wow as ranging from 0.1 to 10Hz and Flutter from 10 Hz and above. Prandolini & Moody (1995, p. 241) assert that Wow is attributed to oscillations below 6 Hz. The most recently published standard concerning the measurement of Wow & Flutter in sound recording and reproducing equipment (Audio Engineering Society, 2013) provides the following definitions for different rates of frequency modulation.

> *"Drift: Frequency modulation of the signal in the range below approximately 0,5 Hz resulting in distortion which may be perceived as a slow changing of the average pitch.*
>
> *Wow: Frequency modulation of the signal in the range of approximately 0,5 Hz to 6 Hz resulting in distortion which may be perceived as a fluctuation of pitch of a tone or program.*

Flutter: Frequency modulation of the signal in the range of approximately 6 Hz to 100 Hz resulting in distortion which may be perceived as a roughening of the sound quality of a tone or program."

2.2.7. Wow & Flutter Extents

Sakai (1970, p. 291) conducted testing on 50 different tape recorders comprising of 10 professional console machines, 16 professional field units, 12 reporter field devices and 12 consumer machines intended for home use. The Wow percentage of each machine was measured and were found to range from 0.17 to 3.8%. The console type ranged from 0.17 to 0.6%, professional field type from 0.4 to 1.1%, report field type from 1.4 to 3.8% and consumer type from 0.6 to 1.2%. It should be noted that this research was originally conducted a decade prior and as such the Wow & Flutter specifications on equipment of this era was largely inferior to that of more modern decks produced from the 1980's onwards.

2.2.8. Perception of Wow & Flutter

Sakai (1970, pp. 295-298) also carried out subjective listening tests with 12 participants exploring both the mean threshold of perceptibility and objectionability of Wow. The listening test program material comprised of piano music, 3 separate symphony excerpts and speech. The Wow percentage was expressed in weighted RMS, Table 2.1 shows the results.

Type	Threshold of Perceptibility %	Threshold of Objectionability %
Piano music	0.14	0.31
Symphony 1	0.35	0.69
Symphony 2	0.43	0.67
Symphony 3	0.42	0.62
Speech	0.86	1.89

Table 2.1 Threshold of perceptibility & objectionability data (Sakai, 1970).

Sakai also established that perception is dependent on program material. Pitch modulation imparted by a tape transport is more easily perceptible in acoustic piano than speech and orchestral music, this is due to acoustic pianos not producing vibrato. Comerci & Oliveros (1956, p. 419) assert that the actual pitch perceived by listeners is a function of the perceived signal over the course of approximately 140 milliseconds and therefore the ear is unable to perceive a variation in pitch when the fluctuation rate exceeds around 5 to 7 Hz.

2.2.9. Notable Examples of Wow as an Intentional Creative Effect

It has been observed that the intentional use of Wow as a creative effect has recently become a popular and desirable production tool for artists working in certain musical genres. Three primary examples have been identified which exemplify the development of intentional Wow as a core component of specific music subcultures. There are many more genres that feature intentional Wow prominently than it would be possible to discuss in this study. These examples have been selected as they broadly represent three distinctly disparate musical genres and are each born out of different pitch modulation production techniques.

The first example to consider is the band ‘Boards of Canada” and their contemporaries within the ambient and electronica scenes. Much of the band’s discography has a distinctive, warped, pitch modulated and degraded quality. They have stated that this is in part achieved by processing audio through old media, going so far as analysing the specific medium they plan to work towards for each track (The Guardian, 2013). ‘Dawn Chorus’ from the album ‘Geogaddi’ features this effect prominently throughout the track (Boards of Canada, 2002). This use of old media and the pitch modulation and other degradation it provides is highly evocative and helps to foster their nostalgic aesthetic, particularly when heard in the context of their signature use of stock video. In addition to the Wow heard on Boards of Canada records it is also likely that they have used the built-in vibrato that is available in some of the hardware synthesizers they have been known to use.

The second area of focus is the advent of digital hardware sampling in the late 1980’s and the consequent work by artists such as J Dilla, Madlib & Flying Lotus.

This involved sourcing samples from existing tracks, sampling them to a digital sampler and cutting them up to form parts of new tracks. Record players would frequently be used to sample vinyl records directly into the digital samplers. Sampling from vinyl in this way often resulted in the presence of Wow also being recorded and this quickly became a popular sound associated with Hip Hop sampling culture. The Boss SP-303 digital sampler was released in 2001 and included 26 effects that could be applied to samples. One of these was named 'Vinyl Sim' and provided a digital simulation of vinyl sampling including exaggerated pitch modulation, compression and needle noise. The Wow pitch modulation Vinyl Sim provided became a highly popular effect for Hip Hop producers. Coupled with the pitch modulation inherently introduced by physical vinyl sampling it would go on to define a new direction in Hip Hop and influence a whole generation of producers (Stenson, 2016). Vinyl Sim from the SP-303 can be heard in a video produced by Cagee (2016).

The final examples to consider are the collective Lofi genres of music originally born out of the DIY and cassette culture scenes in the 1980's. Music classified as Lofi was originally produced on cassette predominately using consumer level 4 track tape machines. The use of cassettes as a recording medium and the typical low fidelity production aesthetics meant that Wow pitch modulation, if not necessarily by design, was a common feature of Lofi music to begin with. Some artists would deliberately introduce exaggerated Wow into their recordings by various means, the most common of which was to manually adjust the tape playback speed control on 4 track recorders. Mac DeMarco used this technique prominently on his first 4 albums, this is demonstrated by DeMarco in Bentler, (2014). The effect can also be heard in the instrumental backing of the track 'My kind of woman' from the album '2' (DeMarco, 2012). The use of Wow within Lofi has now fed through into a multitude of new subsidiary genres that feature pitch modulation as a core production component.

3. Experimental Design

As the primary focus of this study is to investigate the use of Wow as an intentional creative effect, it was decided to devise and conduct subjective listening tests to explore its critical perception. The literature review found that there has been no published research or experiments carried out in this area. As such the decision was made to devise the experimental design of the tests with reference to existing studies focusing on the perception of vibrato.

In order to conduct the listening tests, it would be necessary produce stimuli for participants to assess. As the focus of the study is the perception of Wow as a creative effect it was felt necessary to produce stimuli using authentic production techniques, whilst also ensuring full control over any modulation parameters applied and confidence in the inherent technical performance of any equipment used. It became clear this brief would require the development of new hardware capable of satisfying this specification.

3.1. Design of Modified Tape Deck

After assessing the merits of several options, it was decided the best solution would be to modify an existing tape deck to enable precise control of the speed of the capstan drive motor. Controlling the drive speed in this way would produce pitch modulation exactly as it occurs in a tape machine transport system in the real world. This includes the sonic characteristics associated with the tape medium and changes in speed perfectly syncing with pitch modulations. The modified tape deck could then be used to play back excerpts of program material, applying specific pre-programmed pitch modulations to create a collection of listening test stimuli. The design, development and production of the modified tape deck is detailed in Appendix A.

3.2. Subjective Listening Test Design

As there were no studies published on the perception of pitch modulation as a creative effect it is not possible to base the listening test design for this study on any existing research. It was also acknowledged that it would be difficult to source an established methodological standard that was completely suited to assessing Wow as an intentional effect, especially considering the well documented historical aversion to it.

The two parameters most frequently associated with studies on the perception of Wow and vibrato are rate and extent. The other major pitch modulation component in both Wow and vibrato is the shape of the modulating waveform and any irregularities that are present in it, however very few studies on the perception of vibrato and none concerning Wow have considered the shape or regularity. As such there is very little existing research in this area that could be used to inform further tests in this study. It was therefore decided that Wow rate and extent should be the subject of two separate listening tests carried out independently from one another. Conducting the tests independently would prevent the introduction of familiarity bias by ensuring participants were limited to only taking part in either of the tests.

It was decided that the test design should be identical for both rate and extent with the only differences being the stimuli supplied to participants. The tests would be based around the BS.1534-3 standard (International Telecommunication Union, 2015). This describes a standardised method for the subjective assessment of intermediate quality level of audio systems and is intended to be used to assess audio systems that impart significant impairments to the signal. The main reason for selecting this standard was its use of the method referred to as MUSHRA – Multi Stimulus Hidden Reference and Anchor. This provides a methodological test apparatus capable of comparing a wide range of audio stimuli including both impairments and intentional effects. This is considered important as it cannot be guaranteed that the participants involved in the tests will perceive the pitch modulation as a preferable creative effect as opposed to an impairment.

Certain design criteria for the BS.1534-3 are highly specific to the specific purpose of the standard and as such are not considered relevant to this research. Any revisions in the design criteria that have been applied and implemented to this research will be acknowledged and justified. The primary example of this is the use of reference, hidden reference and anchor stimuli. In the BS.1534-3 standard the reference and anchors are intended to both aid and assess participants ability to rate the severity of impairments in audio codecs. In this context the reference is a fully uncompressed, optimal stimulus that serves as the gold standard against which participants can judge the impaired stimuli.

In the context of this study the concept of a reference and anchors is not considered relevant. With respect to the creative and preferential focus of this study, it is not considered possible to define a universal gold standard of pitch modulation to act as a reference against which other stimuli could be rated. For this reason, it was decided to remove the reference and anchor stimuli from the listening tests used in this research, instead asking participants to assess the modulated stimuli based on their subjective preference.

3.3. Subjective Listening Test Hypotheses

The following hypotheses were outlined for the two subjective listening tests. For the test comparing different rates of Wow the null hypothesis was that no statistically significant relationship will be found between the rate of Wow and participant preference. The alternate hypothesis being that a statistically significant relationship is observed. Similarly, for the second test concerning extent the null hypothesis was that no statistically significant relationship would be found between Wow extent and participant preference, the alternate being again that a relationship would be observed.

3.4. Subjective Listening Test Stimuli

As discussed in the literature review the circumstances in which Wow is perceived and used intentionally as a favourable creative effect are highly specific to certain musical sub genres and production methods. As the focus of the listening tests conducted as part of this study is to investigate how parameters relating to Wow may influence preference amongst listeners, it was decided that the choices of stimuli used in the listening tests should reflect this by selecting complex signal-based program material closely related to the music identified in the literature review. It was also considered important to select material from multiple genres in order to gather data on listener preference of Wow between different musical genres, the importance of this observation was expressed by Sakai (1970, pp. 295-298). In contrast it would be very unlikely to expect listeners to have a favourable perception of Wow applied to test tones in isolation. This scenario is completely at odds to how intentional Wow is experienced in practice.

It was crucial however that the stimuli chosen should have no inherent Wow present in the production to begin with. If this were not the case, then it would be impossible to include stimuli with negligible and lower extents of Wow in the tests. After extensive exploratory listening the following 3 tracks were selected to serve as a basis for the production of the stimuli; Dayvan Cowboy by Boards of Canada (Boards of Canada, 2005), Two Can Win by J Dilla (J Dilla, 2006) and Still Beating by Mac Demarco (Mac DeMarco, 2017). 44.1 kHz, 16-bit digital WAV copies of each track were sourced to produce the stimuli.

The BS.1534-3 standard recommends that stimuli should be-

> *"…approximately 10 s, preferably not exceeding 12 s."* (IEC, 2015).

The justification being that excessive exposure to stimuli can begin to fatigue listeners and impact the validity of the results. Consequently 12 second sections were selected from each track. Due to there not being a natural break in the program material around the 12 second mark for all tracks, the decision was taken to extend the excerpts to 16 seconds with a uniform 4 second fade applied to each track for continuity. Once the edits were applied using Logic Pro X the loudness of each excerpt was then measured using the Youlean Loudness Meter 2 Pro plugin (Youlean, 2019). The gain of each track was then attenuated until they all measured -12 dB LUFS integrated. This was carried out to ensure that there could be no bias attributed to perceived differences in amplitude when comparing listener preference between genres. The stimuli were then bounced out to a single WAV file in order to be rendered to cassette tape.

It was considered important that the stimuli were transferred to tape with as little inherent Wow as possible. The Reel Deal Denmark carried out the rendering of the stimuli onto cassette tape using the same Studer A710 that was used to produce the test tones calibration cassette. The Studer features Wow & Flutter of < 0.04% (DIN) (The Reel Deal Denmark, 2019). Upon receiving the rendered test track cassette, it was placed in the modified JVC deck and played back with the various modulation parameters applied. The output of the deck was recorded into Logic Pro X using an Audient ID44 interface.

At this juncture the decision was taken to fully adhere to the BS.1534-3 stimuli length recommendation and edit the stimuli down to 12 seconds in length with 50 millisecond fades being applied to the beginning and end. The stimuli then underwent further loudness metering to ensure continuity post cassette modulation. It was found that the stimuli all varied in loudness within a range of approximately 3 dB LUFS. This was attributed to inconsistencies in the JVC deck playback level and anomalies caused by the modulation phase alignment. The gain of each stimuli was consequently adjusted to -18 dB LUFS to ensure the removal of external bias, in accordance with the recommendations in BS.1534-3 and BS.1116-3.

3.5. Selection of Test Stimuli Modulation Parameter Levels

The BS.1534-3 standard specifies that no more than 12 stimuli should be included in a single multi stimulus test, whilst the advisory documentation provided with the MUSHRA Max/MSP patch (IoSR Surrey, 2017) states that as many as 10 stimuli can be difficult for listeners to assess. As such it was decided that a maximum of 9 stimuli would be included in assessments for both listening tests.

The breadth of literature available on the study of the favourable perception of Vibrato and the literature concerning the development of the pitch fluctuation weighting curves used in Wow & Flutter meters were identified as the two main areas of existing research that held the greatest influence over the design of the listening tests in the study. Consequently, the focus of the first listening test was identified as the preferential perception of modulation rate in Wow. In accordance with the Audio Engineering Society (2013) definition of Wow, the following rates of modulation stimuli were selected for the test: 0.3, 0.5, 1, 2, 3, 4, 5, 6 & 7 Hz. This selection covers the whole range of rates found in Wow and also expands into the Drift and Flutter regions to explore the preferential perception beyond the boundaries of Wow.

The fixed extent percentage selected for use in this first test was derived from the data on the thresholds of perceptibility and objectionability presented by Sakai (1970, pp. 295-298) and shown in Table 2.1 of the literature review. The data derived from the 3 symphony extracts used in the tests was considered applicable to this study, as the orchestral symphony music was most representative of the complex signals exhibited by the program music selected in this study. The mean threshold of perceptibility calculated from the 3 symphony extracts is 0.4%, whilst the mean threshold of objectionability is 0.66%. It was therefore decided that the fixed extent for the first test should be at the equilibrium between the mean of perceptibility and objectionability thresholds, which resulted in a weighted RMS Wow rounded to 0.5%. The rationale underpinning the selection of this figure was to identify an average extent considered both easily perceptible and unobjectionable in equal measure, so as to facilitate impartial testing of the Wow rates.

The focus of the second listening test was identified as a study of the preferential perception of Wow percentage extents. The selection of extents for this study was derived from the perceptibility and objectionability thresholds presented in Sakai (1970, pp. 295-298) and Table 2.1 of the literature review. It was decided that data from the piano music test would be used in addition to the 3 symphony tests in order to investigate a wide range of extents. The lowest threshold of perceptibility is the 0.14% of the piano music, whilst the highest threshold of objectionability is the 0.69% of the first symphony extract. It was therefore decided that the extents featured in the second test should comprise of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 & 0.9. This would facilitate the investigation of extents within and either side of the stated limits of perceptibility and objectionability.

Internal listening tests were conducted to determine the fixed rate for each stimulus in the extent listening test. 3.25 Hz is the midway point between the lower and upper bounds of the Wow rate as defined by AES (2013). Therefore, rates from 2 to 4 Hz were assessed for their suitability. Following these tests 2 Hz was selected as the fixed rate, this was considered an average example of Wow rate that was both clearly perceptible but not objectionable, enabling unbiased testing of the extent percentage to be carried out. The stimuli used in the modulation rate and extent listening tests have been uploaded to SoundCloud and are available to listen to at the following two links respectively:

https://soundcloud.com/jay-harrison-2/sets/modulation-rate-listening-test/s-W9DPw

https://soundcloud.com/jay-harrison-2/sets/modulation-extent-listening-test/s-INIpI

3.6. Software Test Interface

The MUSHRA test method was designed to be presented to participants in the form of an interactive software interface. Participants should have the ability to compare audio stimuli in isolation and then be able to rate their preference using a sliding scale for each stimulus. The Institute of Sound Recording at the University of Surrey developed a Max/MSP based MUSHRA test interface that fully adheres to the BS.1534-3 standard (IoSR Surrey, 2017). It presents test material in a randomised order and allows participants to conduct A/B comparisons between each of the stimuli before providing ratings. It does this by looping all of the material in unison and then muting all stimuli apart from the one selected for audition by the participant. Upon successful completion of contextual testing the IoSR Max patch implementation of the MUSHRA interface was selected for use in the studies listening tests.

The patch was modified to accommodate 9 stimuli as opposed to the default 7, the reference stimuli button was removed, the patch was duplicated for rate and extent tests and the stimuli files were written in for playback. Both rate and extent modified MUSHRA listening test patches are available to download at the following link:

https://github.com/JayHarrison1/Evaluation-of-Pitch-Modulation-as-a-Creative-Effect-Max-MSP-Patches.git

3.7. Listening Test Participant Training

As recommended by the BS.1543-3 standard, a training procedure was included in the test design to ensure that participants were fully aware of how to interact with the Max/MSP MUSHRA interface and what was expected of them during the tests. This comprised of a general information document, shown in Figure B.1 of Appendix B, to be read by participants prior to beginning the tests, followed by a further verbal induction into the operation of the test interface lasting approximately 1 minute. This was intended to prevent any potential external bias caused by participants misunderstanding the practicalities of the test infrastructure.

3.8. Criteria for Selecting Participants

The BS.1534-3 standard states that listening tests consisting of small impairments should always require the use of expert listeners, who are experienced in the critical auditioning of audio, to ensure the validity of the results. It also states that the importance of the use of expert as opposed to non-expert listeners lowers as the impairments become larger and consequently more easily perceptible. The BS.1534-3 is itself intended to assess intermediate impairments. In a critical context, akin to the traditional, unfavourable perception of Wow & Flutter, the 'impairments' introduced by the pitch modulation to program material would be considered medium to large. It should also be noted that due to the recently established prevalence of the use of intentional Wow in many musical subcultures, it can be understood that it is favoured by large numbers of musical consumers, a majority of which would statistically not be considered expert listeners in the traditional sense. Upon consideration of these observations it was decided that a mixture of both expert and non-expert listeners should be sourced as participants for the listening tests, so as to achieve a truly representative listener population.

A pre-screening participant questionnaire, shown in Figure C.1 of Appendix C, was produced and supplied to participants to establish if they had any known hearing impairments and to determine if they would be considered expert listeners. Participants were also required to complete a written consent form, shown in Figure D.1 of Appendix D.

3.9. Subjective Listening Test Conditions

The test stimuli were delivered to participants using a pair of Beyerdynamic DT770 PRO 250 Ohm headphones. The specifications of this model are shown in Figure E.1 of Appendix E. The listening tests were conducted in several venues including, libraries, classrooms and residential addresses. All test venues were deemed to provide quiet listening environments and were directly supervised by the researcher.

4. Testing & Results

This section will present and analyse the results obtained from both subjective listening tests. Results from all 3 genres are grouped into separate rate and extent data sets in order to conduct extensive global analysis of both modulation parameters. Summaries of statistical testing carried out on the results from each of the individual rate and extent musical genre tests are also presented in Appendix F. IBM SPSS statistics is used throughout this section for all data analysis, statistical testing and graphing.

4.1. Modulation Rate Subjective Listening Test

4.1.1. Analysis of Rate Pre-Screening Questionnaires

Of the 20 participants that completed the modulation rate subjective listening test 15 were considered sufficiently experienced to qualify as expert listeners. Two participants stated that they had known minor hearing impairments and another 2 were over the age where it is common that minor natural hearing performance losses have begun. However, none of the hearing impairments observed in these 4 participants were considered severe enough to warrant their exclusion from the results.

4.1.2. Global Modulation Rate Data Analysis

In this analysis the results from all 3 listening test pages exploring the perception of modulation rate for the Boards of Canada, J Dilla & Mac DeMarco program material were grouped and then sorted according to their modulation parameters. Kolmogorov-Smirnov and Shapiro-Wilk tests of normality were carried out. This showed that there were 4 outliers identified in the data, as determined by the analysis of the boxplot for results greater than 1.5 box lengths from the edge of the box. This boxplot is shown in Figure 4.1, the central markers display the median, the boxes indicate the central 50% region of the data and the upper and lower whisker shows the 5th and 95th percentile data points. This format is identical for all boxplots

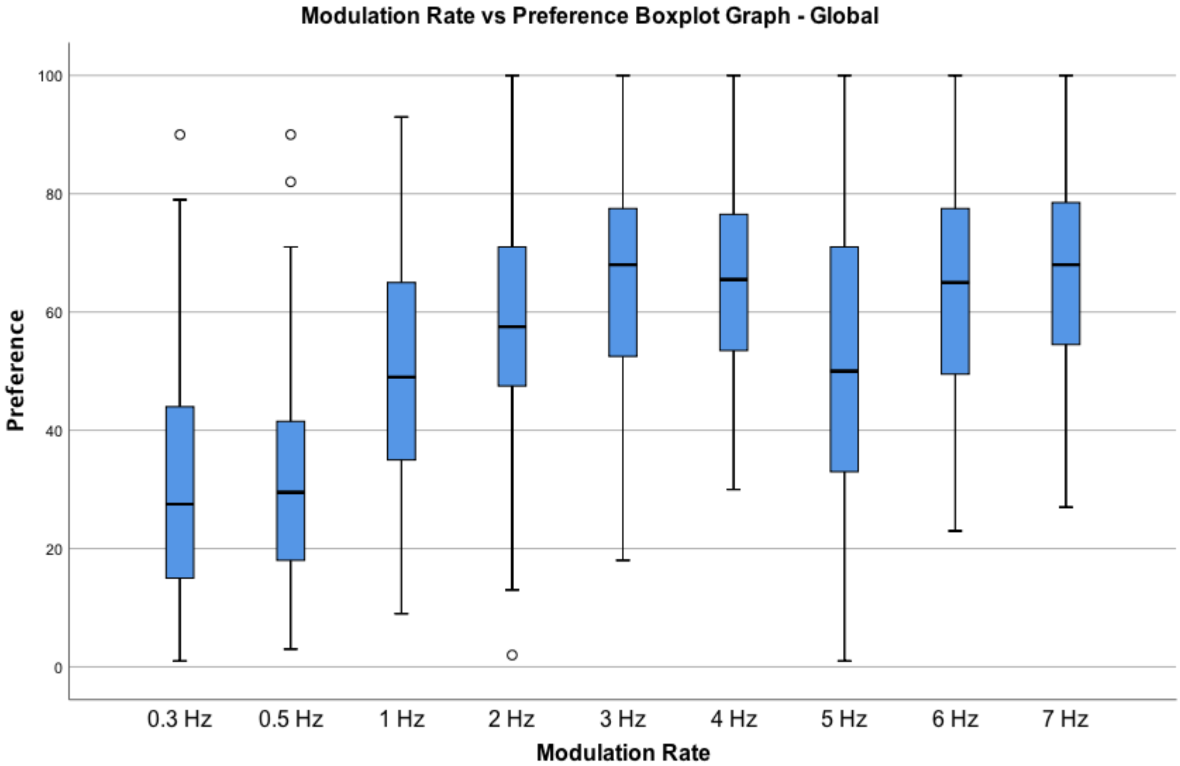

Figure 4.1 Global modulation rate vs mean preference boxplot.

presented in the study. It was decided to continue with the statistical testing irrespective of the outliers until a more comprehensive data analysis was achieved. The outliers would be addressed should statistical significance not initially be found.

The Shapiro-Wilk test for normality was then used to assess whether the data was normally distributed. This test was selected as it is recommended for sample sizes of less than 50 participants. The results of the Shapiro-Wilk normality tests for each of the modulation rate parameters are shown in Table 4.1. They show that participant preference was normally distributed at all modulation rates with the exception of 0.3 and 0.5 Hz, this was assessed at $p > .05$. It was decided to continue with the statistical testing irrespective of the observed non-normality until a more comprehensive data analysis was achieved. The issue of non-normality would be addressed should statistical significance not initially be found.

Modulation Rate	Statistic	df	Significance
0.3 Hz	0.954	60	0.025
0.5 Hz	0.942	60	0.007
1 Hz	0.987	60	0.753
2 Hz	0.980	60	0.426
3 Hz	0.974	60	0.229
4 Hz	0.985	60	0.673
5 Hz	0.981	60	0.453
6 Hz	0.986	60	0.726
7 Hz	0.980	60	0.433

Table 4.1 Global modulation rate Shapiro-Wilk normality test results.

A one-way repeated measures ANOVA with post hoc test was then carried out on the data. The mean, standard deviation, upper and lower bounds for each modulation parameter are shown in Table 4.2.

Modulation Rate	Mean	Standard Deviation	95% Confidence Interval	
			Lower Bound	Upper Bound
0.3 Hz	30.350	20.585	25.032	35.668
0.5 Hz	32.750	18.531	27.963	37.537
1 Hz	49.933	19.419	44.917	54.950
2 Hz	57.400	20.401	52.130	62.670
3 Hz	64.883	18.398	60.131	69.636
4 Hz	66.100	15.805	62.017	70.183
5 Hz	51.600	25.474	45.019	58.181
6 Hz	64.200	16.754	59.872	68.528
7 Hz	67.517	17.024	63.119	71.914

Table 4.2 Global modulation rate ANOVA test results.

Mauchly’s test of sphericity was then conducted and was found to be statistically significant, indicating that the assumption of sphericity had been violated, Chi-squared = 6.270, df = 35 & p = 7.8868E-8. Consequently, the results were then interpreted using the Greenhouse-Geisser correction.

The Epsilon (ε) was 0.666, as calculated in accordance with Greenhouse & Geisser (1959, pp. 95-112) and was used to amend the one-way repeated measures ANOVA. Participant preference was found to be statistically significantly different at the different modulation rates presented during the listening test, $F(5.327, 314.321) = 53.358$, $p = 9.6503E-42$. As such the null hypothesis can be rejected and the alternate hypothesis is accepted. The final mean participant preference results are graphed with 95% error bars in Figure 4.2.

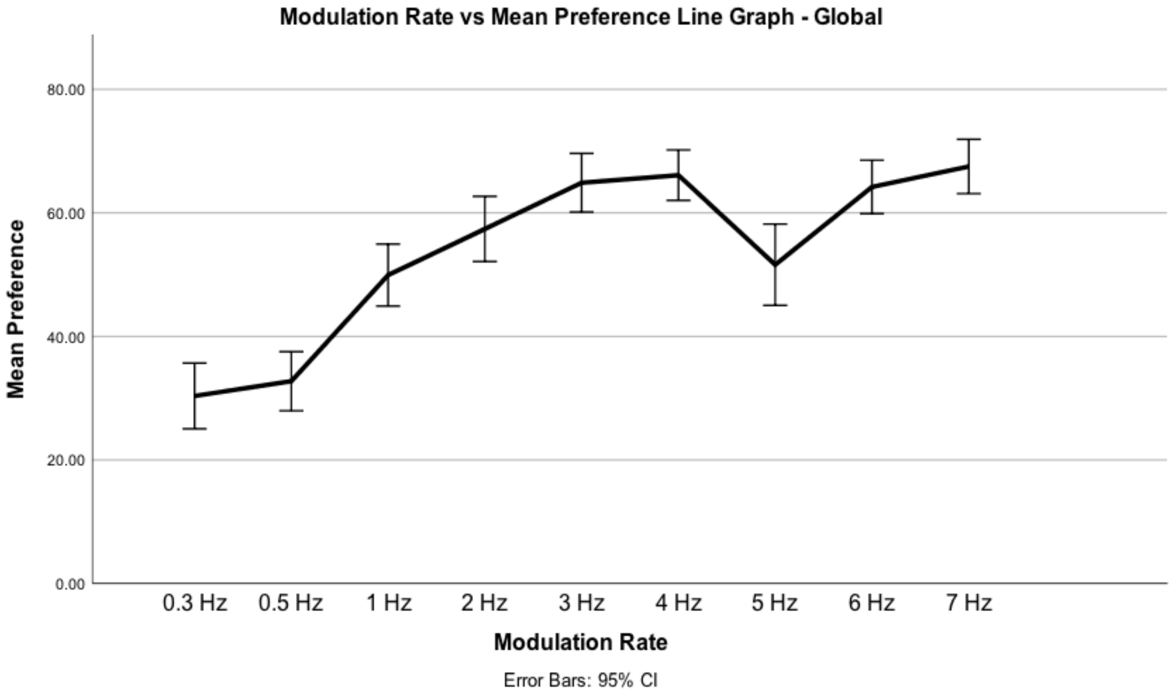

Figure 4.2 Global modulation rate vs mean preference line graph.

Modulation Rate	Mean Difference	Standard Error	Significance
0.3 – 0.5 Hz	2.400	2.030	1.000
0.5 – 1 Hz	17.183	2.360	0.000
1 – 2 Hz	7.467	2.691	0.266
2 – 3 Hz	7.483	2.332	0.078
3 – 4 Hz	1.217	1.982	1.000
4 – 5 Hz	-14.500	3.198	0.001
5 – 6 Hz	12.600	3.060	0.004
6 – 7 Hz	3.317	2.130	1.000

Table 4.3 Global modulation rate Bonferroni post hoc test results.

The Bonferroni post hoc test was then used to perform pairwise comparisons between consecutive modulation rate levels. The mean difference, standard error and significance results for this test are shown in Table 4.3. The results show that statistical significance was found in the mean differences between 0.5 - 1 Hz, 4 - 5 Hz and 5 - 6 Hz modulation rates.

4.1.3. Global Rate Data Analysis with Omission of 5 Hz

Upon the initial assessment of the global rate results, the mean preference for 5 Hz was considered of particular interest. It was identified as an unusual point in the data given that mean preference remained stable in the mid 60’s region for modulation rates of 3, 4, 5 & 6 Hz either side of it. It also featured the widest data set of all the modulation rates as demonstrated by the boxplot in Figure 4.1.

As this would be of significant further interest in the discussion section, the decision was taken to carry out a separate one-way repeated measures ANOVA with the 5 Hz data excluded from the analysis. If the result was found to be the same, then it could be understood that the 5 Hz data did not influence the overall results of the test.

This ANOVA analysis again found that Mauchly’s test of sphericity had again been violated, Chi-squared = 72.230, df = 27 & p = 0.000006. The Greenhouse-Geisser correction was then used (ε = 0.685). Participant preference was again found to be statistically different at the different modulation rates presented, with the omission of the 5 Hz data, $F(4.794, 282.843) = 68.084$, p = 2.5929E-45. Therefore, with the omission of the 5 Hz data the null hypothesis can still be rejected, and the alternate hypothesis accepted.

4.2. Modulation Extent Subjective Listening Test

4.2.1. Analysis of Extent Pre-Screening Questionnaire

Of the 20 participants that completed the modulation extent subjective listening test 13 were considered sufficiently experienced to qualify as expert listeners. One participant stated that they had a known minor hearing impairment, and another was over the age where it is common that minor natural hearing performance losses have begun. However, neither of the hearing impairments observed in these participants were considered severe enough to warrant their exclusion from the results.

4.2.2. Global Modulation Extent Data Analysis

This analysis concerns the results from all 3 listening test pages researching the perception of modulation extent. Results from the Boards of Canada, J Dilla & Mac DeMarco program material were grouped and then sorted in accordance with their modulation parameters. Kolmogorov-Smirnov and Shapiro-Wilk tests of normality were carried out. This identified 12 outliers in the data as shown by the boxplot in

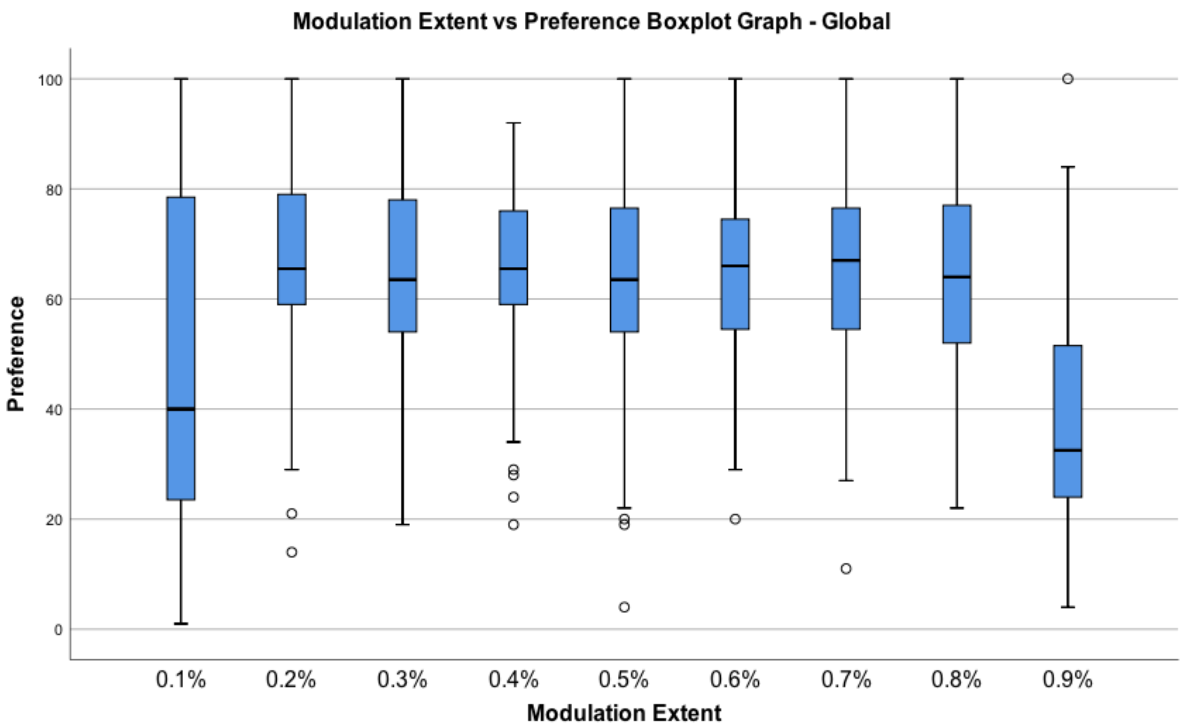

Figure 4.3 Global modulation extent vs mean preference boxplot.

Figure 4.3. It was decided to continue with the statistical testing irrespective of the outliers until a more comprehensive data analysis was carried out. The results of the Shapiro-Wilk normality tests for each of the modulation rate parameters are shown in Table 4.4. They show that participant preference was normally distributed at all modulation extents with the exception of 0.1, 0.4, 0.5 & 0.9%. It was decided to continue with the statistical testing irrespective of the observed non-normality until a more comprehensive data analysis was achieved. The results of a one-way repeated measure ANOVA test carried out on the data is shown in Table 4.5.

Modulation Extent	**Statistic**	**df**	**Significance**
0.1%	0.921	60	0.001
0.2%	0.969	60	0.134
0.3%	0.983	60	0.546
0.4%	0.945	60	0.009
0.5%	0.952	60	0.019
0.6%	0.983	60	0.549
0.7%	0.981	60	0.486
0.8%	0.979	60	0.398
0.9%	0.940	60	0.005

Table 4.4 Global modulation extent Shapiro-Wilk normality test results.

Modulation Extent	**Mean**	**Standard Deviation**	**95% Confidence Interval**	
			Lower Bound	**Upper Bound**
0.1%	46.983	29.671	39.318	54.648
0.2%	66.417	18.152	61.727	71.106
0.3%	64.667	18.802	59.810	69.524
0.4%	63.883	17.396	59.390	68.377
0.5%	62.017	19.858	56.887	67.146
0.6%	64.617	14.967	60.750	68.483
0.7%	65.033	17.123	60.610	69.457
0.8%	64.333	18.248	59.619	69.047
0.9%	39.133	21.620	33.548	44.718

Table 4.5 Global modulation extent ANOVA test results.

Mauchly's test of sphericity was then conducted and found to be statistically significant, indicating that the assumption of sphericity has been violated, Chi-squared = 104.724, df = 35, p = 7.6595E-9. As such the Greenhouse-Geisser correction was used.

The Epsilon (ε) was 0.623, as calculated according to Greenhouse & Geisser (1959) and was used to amend the one-way repeated measures ANOVA. Participant preference was found to be statistically significantly different at the different modulation extents presented during the listening test, $F(4.981, 293.867) = 22.958$, p = 2.154E-19. Therefore, the null hypothesis can be rejected, and the alternate hypothesis is accepted. The final mean participant preference results are graphed in Figure 4.4.

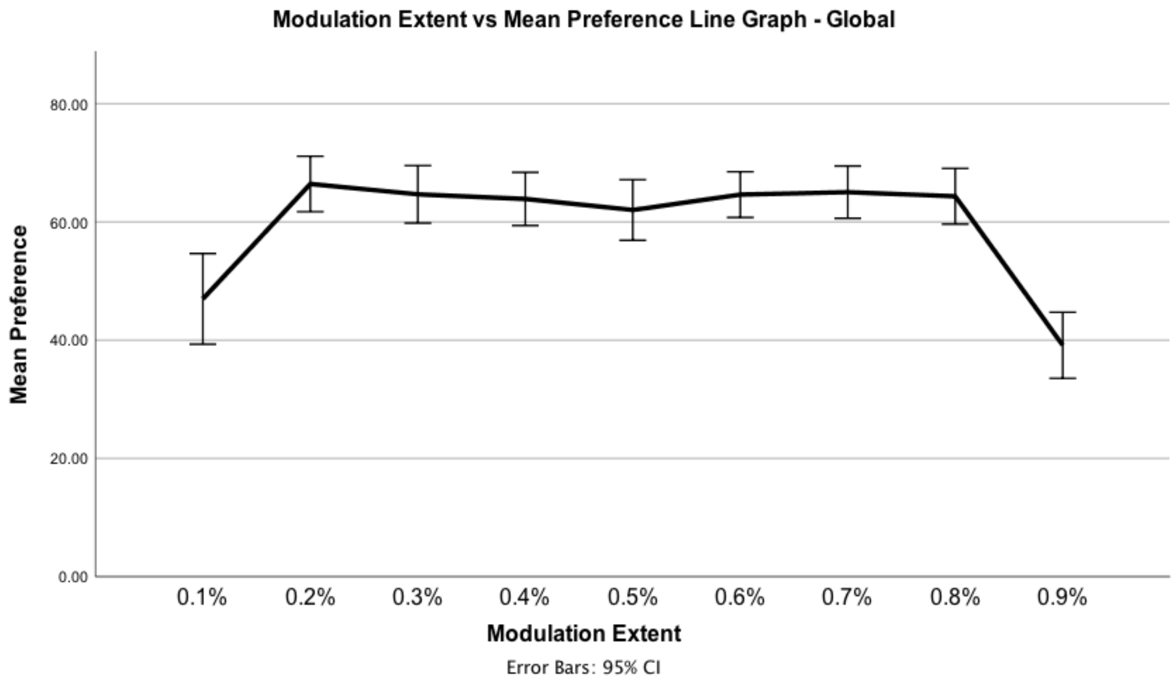

Figure 4.4 Global modulation extent vs mean preference line graph.

The Bonferroni post hoc test was used to execute pairwise comparisons between consecutive modulation extent levels, the results are shown in Table 4.6. Statistical significance was found in the mean differences between 0.1 – 0.2 % and 0.8 – 0.9 % modulations extents.

Modulation Extent	Mean Difference	Standard Error	Significance
0.1 – 0.2%	19.433	3.675	0.000
0.2 – 0.3%	-1.750	2.758	1.000
0.3 – 0.4%	-0.783	2.428	1.000
0.4 – 0.5%	-1.867	2.789	1.000
0.5 – 0.6%	2.600	2.319	1.000
0.6 – 0.7%	0.417	2.216	1.000
0.7 – 0.8%	-0.700	2.130	1.000
0.8 – 0.9%	-25.200	2.882	0.000

Table 4.6 Global modulation extent Bonferroni post hoc test results.

4.2.3. Global Extent Data Analysis with Omission of 0.1 & 0.9%

Upon the initial assessment of the global extent results, the mean preferences of 0.1 and 0.9% were considered of particular interest. They were identified as potentially unusual points in the data given that the mean preference remained stable in the mid 60's region for all of the other modulation extents in between.

As this would be of significant further interest in the discussion section if was decided to carry out a separate one-way repeated measures ANOVA with the 0.1 & 0.9% data excluded from the analysis. If the result was found to be the same, then it could be understood that the 0.1 & 0.9% data did not influence the overall significance results of the test.

This ANOVA analysis again found that Mauchly's test of sphericity had again been violated, Chi- squared = 36.169, df = 20 & p = 0.015. The Greenhouse-Geisser correction was then used (ε = 0.819). Participant preference was found not to be statistically different at the different modulation extents presented, $F(4.912, 289.905)$ = 0.581, p = 0.711. Therefore, with the omission of the 0.1 & 0.9% data the null hypothesis can be accepted, and the alternate hypothesis rejected.

5. Discussion

5.1. Omission of 5 Hz Rate Data

In the testing and results section, it was noted that the mean participant preference data obtained for the 5 Hz modulation rate in all 3 musical genre tests was considered potentially unusual. In all cases the post hoc Bonferroni pairwise comparison tests showed the mean participant preference for 5 Hz was statistically significantly lower than the 4 and 6 Hz means immediately either side of it. It was also observed that the 3, 4, 6 and 7 Hz means remained closely aligned within a maximum of 10 preference points from each other respectively and did not show statistical significance through pairwise comparison.

Consequently, separate ANOVA tests were carried out to assess the statistical significance of the results with the omission of the 5 Hz data. This showed that even with the omission of the 5 Hz data in all rate tests, participant preference was still found to be statistically different at the different modulation rates presented. This was primarily attributed to the statistically significant differences in means between 0.5 and 1 Hz as assessed by the post hoc Bonferroni pairwise comparison tests.

It was decided to investigate the 5 Hz test stimuli for any anomalies that may have contributed to the test results. Upon audition of the stimuli it became clear that the overall pitch of the 5 Hz stimuli for all 3 musical genres was approximately 1 semitone lower than the stimuli from the rest of the modulation rates. During the initial production of the stimuli using the modified cassette deck it was found that the 5 Hz stimuli featured a prominent audible inconsistency in the stereo image when compared to the rest of the samples. This was considered unacceptable for use in the listening tests as the anomaly with the stereo image would most likely influence participant preference, therefore introducing an element of external bias. Consequently the 5 Hz stimuli were recorded again at a later date to address the stereo image issue. The new 5 Hz stimuli were auditioned again, and it was found that the issues with the stereo image had been rectified.

With hindsight it is apparent that when the 5 Hz stimuli were rerecorded the overall speed of the modified cassette deck must have been slightly slower, thus causing the semitone drop in pitch. It is regrettable that this difference in overall pitch was not detected during the quality control auditioning immediately after production.

5.1.2. Omission of 0.1 & 0.9% Extent Data

The mean participant preference results for the 0.1 & 0.9% modulation extent test results were also highlighted as being potentially unusual in the context of the wider results. In all 3 genre tests the post hoc Bonferroni pairwise comparison tests established the mean participant preference for 0.1 & 0.9% was statistically significantly lower than the 0.2 and 0.8% means immediately adjacent to either respectively. It was also observed that all the modulation extent means from 0.2 to 0.8% were closely aligned within a maximum of 10 preference points from each other respectively and did not show statistical significance through pairwise comparison.

Separate ANOVA tests were carried out to assess the statistical significance of the results with the omission of the 0.1 & 0.9% data. This showed that with the omission of the 0.1 & 0.9% data in all extent tests, participant preference was found not to be statistically different at the different modulation extents. This is in stark contrast to the original ANOVA performed on all the extent data including 0.1 & 0.9% which found overall statistical significance.

It was decided to investigate the 0.1 & 0.9% test stimuli for any anomalies that that may have contributed to the test results. Upon audition of the stimuli it was found that as with the 5 Hz anomalies the overall pitch of the 0.1 & 0.9% stimuli were also approximately 1 semitone lower than the rest of the modulation extents. During the initial production of the stimuli using the modified cassette deck, the 0.1 & 0.9% recordings were found to be affected by the same unacceptable stereo imaging inconsistencies that affected the 5 Hz samples. They were then rerecorded alongside the 5 Hz stimuli and, upon audition, the stereo image issues were considered resolved.

With hindsight it is clear that as with the 5 Hz samples, the 0.1 & 0.9% stimuli were rerecorded at a slightly slower overall speed, resulting in the semitone drop in pitch. It is again regrettable that this difference in overall pitch was not detected during the auditioning immediately after production.

5.1.3. Omission of Data Conclusions

Whilst it is regrettable that these errors in quality control occurred, it was considered important to highlight them at the beginning of the discussion section in order to best communicate an accurate and comprehensive critical analysis of the testing and results carried out by this study. The nature of the potentially anomalous results highlighted here would suggest that they were, at least in part influenced by the participants perception of the lower overall pitch. However, as it is not possible to definitively prove this without further testing it was decided the results should be discussed both with and without the influence of the 5 Hz, 0.1 & 0.9% data. Despite this it is noted that the conclusions not derived from the omissible 0.5 Hz, 0.1 & 0.9% data are to be considered the most reliable.

5.2. Listening Test Participant Observations

As described in the experimental design section, upon the completion of their test participants were asked if they were able to provide any observations. Of the 40 participants that took part in the study across the two listening tests 21 provided observations. Full transcripts of all the comments provided by the participants are shown in Appendix G. of the appendices. Some of the observations that were considered particularly insightful have been extracted for further discussion in this section.

One participant provided a wider general observation that Wow pitch modulation could be a contributing factor in the perceived preference of vinyl records when compared to digital media. The implication being that the subtle pitch modulation introduced by the transport systems of record turntables is perceived favourably by some individuals when compared with a playback system that introduces no artificial pitch modulation. This is an interesting hypothesis which would require further independent research to investigate satisfactorily. The literature review in this study discussed the unique modulation waveform shapes or patterns that are often produced by different faults in tape and turntable transport systems, highlighting that Wow is rarely formed of a purely cyclical modulation. This would likely be a key area of interest in further research pertaining to the perception of Wow produced by record turntables.

One participant commented that they perceived the implementation of pitch modulation over the whole mix as opposed to individual instruments or stems to be unfavourable. They cited the pitch modulation applied to the strings in both the Boards of Canada and J Dilla tracks as being particularly unpreferable. The participant also commented that it was unusual for pitch modulation to be used on the whole mix in this way, going on to suggest that perhaps testing exploring the perception of pitch modulation on individual stems would be an area that warranted further study.

The participant’s assertion that pitch modulation is not usually applied to the whole mix in a creative context is considered accurate. This was reflected upon during the design of the experimental methodology, at the time it was decided that given the scope of the project and the timescale afforded to complete it, it would only be possible to carry out a limited number of narrowly focused listening tests and the perception pitch modulation as applied to the overall mix was considered a suitable area of initial investigation. Therefore, it is agreed that further study should seek to explore the perception of Wow on individual instruments and stems.

Three participants stated that they recognised the Mac DeMarco track. Two of them commented that they had tried to determine which stimuli was the original version. The third participant noted that they found the pitch modulation in Mac DeMarco track to be more preferable than the other two tracks. They stated that this was due to their prior knowledge and experience of Mac DeMarco’s tape speed manipulation production techniques.

5.3. Modulation Rate Listening Test Results

The results from the global modulation rate data analysis showed that participant preference was found to be statistically significantly different at the different modulation rates both with and without the omission of the 5 Hz data. Post hoc Bonferroni pairwise comparison tests established that statistical significance was found in the mean differences between 0.5 – 1 Hz, 4 – 5 Hz & 5 – 6 Hz. As shown in Table 4.3 mean participant preference rose slightly by 2.4 points from 0.3 to 0.5 Hz, then sharply increased 17.183 points from 0.5 to 1 Hz, it continued to increase by 7.467 from 1 to 2 Hz and 7.483 from 2 to 3 Hz. With the omission of the 5 Hz data, mean participant preference remained broadly stable in the mid 60's at 3, 4, 6 & 7 Hz. If the 5 Hz data is included in the analysis, mean participant preference is shown to drop by 14.500 from 4 to 5 Hz before rising back 12.600 points from 5 to 6 Hz.

Therefore, it can be concluded that both with and without the omission of the potentially anomalous 5 Hz data, the null hypothesis that no statistically significant relationship would be found between Wow rate and participant preference is rejected and the alternate hypothesis that a statistically significant relationship is found between rate and preference is accepted. The results indicate that pitch modulation is perceived as significantly less preferential at 0.3 & 0.5 Hz compared to all other rates included in the test. They also show a trend that preference steadily increases from 0.5 to 3 Hz, with the omission of the 5 Hz data a separate trend is observed that shows mean participant preference remaining largely the same from 3 to 7 Hz.

The Audio Engineering Society (2013) defined pitch modulations of 0.5 Hz and below as Drift, modulations rates from 0.5 to 6 Hz as Wow and modulations of 6 to 100 Hz as Flutter. Under these definitions the results from the modulation rate tests conducted in this study suggest that Drift is statistically significantly less preferential than Wow. The results also support the argument that within the range of Wow rates, a trend is observed for preference increasing significantly from 0.5 to 1 Hz and then continuing to rise steadily from 1 to 3 Hz. With the omission of the potentially anomalous 5 Hz data preference is then observed as remaining largely consistent from 3 to 7 Hz. The results do not show a significant difference between preference for 6 and 7 Hz as the border between Wow & Flutter is crossed. Further testing would be required to achieve a more comprehensive understanding of the effects of Flutter on perceptual preference.

It can also be observed that the modulation rate results share some similarities with the Wow meter perceptibility weighting curve graph shown in Figure 2.1, specified by IEC (1972). This graph shows that the human ear is most perceptible to fluctuations in pitch at 4 Hz. Below 4 Hz perceptibility is shown to fall in a logarithmic fashion. The nature of this perceptibility weighting curve below 4 Hz follows a very similar trend to the mean global participant results for modulation rates from 0.5 to 4 Hz in this study. This would support the supposition that between pitch modulation rates 0.5 & 4 Hz, a trend is observed between higher pitch fluctuation perceptibility and increased participant preference.

The data obtained for 0.3 Hz deviates somewhat from this trend and would be expected to fall at a greater rate if it were to also closely match the weighting curve. The logarithmic nature of the perceptibility weighting graph means that with the omission of the potentially anomalous 5 Hz data, the mean preference for 6 and 7 Hz modulation rates remains within the error bars of the curve. Further perceptual testing would be to establish if the trend between participant preference and the IEC (1972) perceptibility weighting curve continued above modulation rates of 7 Hz.

It could also be argued that the trend of global Wow rate results being considered most preferential from 3 to 7 Hz is comparable to the average singer and violinist vibrato rates of 5 to 7 Hz presented in the literature review by Rossing & Moore (2002, pp. 141-142), Seashore (1931, 623-624), Risset & Wissel (1999, p. 130), Prame (1994, pp. 616-621) & Ferrante (2011, p. 1687). Although no conclusive links can be drawn between the perception of Wow and Vibrato, it is thought that this observation presents a basis for further investigation in future studies.

5.4. Modulation Extent Listening Test Results

The results from the global modulation extent data analysis showed that participant preference was found to be statistically significantly different at the different modulation extents. However, with the omission of the potentially anomalous 0.1 &

0.9% data, the analysis showed that participant preference was not to be statistically significantly different at the different modulation extents. Post hoc Bonferroni pairwise comparison tests determined that statistical significance was found in the mean differences between 0.1 – 0.2% & 0.8 – 0.9% extents. As shown in Table 4.6 mean participant preference rose sharply by 19.433 points from 0.1 to 0.2%. It then remained consistently in the mid 60's from 0.2 to 0.8%, before falling acutely by 25.200 points between 0.8 & 0.9%.

Consequently, with the inclusion of all modulation extent data the null hypothesis can be rejected and the alternate hypothesis that a statistically significant relationship would be found between Wow extent and participant preference is accepted. However, with the omission of the potentially anomalous 0.1 & 0.9% data, the null hypothesis that no statistically significant relationship would be found between Wow extent and participant preference is accepted and the alternate hypothesis rejected. The only trend that can be observed from the extent data analysis is that participant preference remains consistently stable between 0.2 & 0.8%, suggesting that an extent of 0.8% Wow is considered equally as favourable as 0.2%. As such, with the omission of the 0.1 & 0.9% data it would appear than Wow extent has very little influence over participant preference for the extents included in the test.

With the omission of the 0.1 & 0.9% data the extent results neither supports or refutes the threshold of Wow perceptibility data presented by Sakai (1970, pp. 295-298). However, the extent results do offer alternate findings to the Wow threshold of objectionability data also presented by Sakai (1970, pp. 295-298). Sakai's results show the mean threshold of objectionability for a piano music extract as being 0.31% and for 3 pieces of symphony music as being 0.62, 0.67 & 0.69%. The threshold of objectionability in this study would be equivalent to a statistically significant drop in participant preference. The modulation extent data from 0.2 to 0.8% with the omission of 0.1 & 0.9% show no such significant drop in mean participant preference. A significant drop is observed from 0.8 to 0.9%, however as discussed previously it is unclear how much of this fall in preference is due to the participant preference of the potentially anomalous lower pitched stimuli. Therefore, it can only be confidently concluded that Sakai's threshold of objectionability results are not supported by the extent data ranging from 0.2 to 0.8% presented in this study.

5.5. Perception of Wow across different Musical Genres

The results showed that the general participant preference trends for both modulation rate and extent were remarkably similar across all individual genre specific listening tests. The overall participant preference between each genre test did show slightly more variation. However, it is not considered possible to draw any meaningful conclusions from this as the results are influenced by existing differences in participant's personal musical preferences.

5.6. Limitations of the Study

It should be noted that the results and consequent analysis presented in this study can only be considered accurate to certain tolerances. As discussed in the experimental design section, technical issues with the design of the modified cassette deck motor speed control meant that the Wow extent percentage WRMS figures reported can only be considered accurate to tolerances of ± 0.03%. It should also be acknowledged that the Wow rate frequencies cannot be guaranteed as being 100 percent accurate in all cases.

The implementation of pitch modulation using a cassette deck meant that the Wow it produced would also be at least partially influenced by inherent limitations in the mechanical performance of the tape transport system. As discussed in the experimental design section, the flywheel in cassette deck transport systems is designed to increase the stability of the capstan speed by reducing the drive mechanisms susceptibility to sudden fluctuations in speed. Whilst this was essential in achieving the low Wow & Flutter performance required for the production of test stimuli, it may also have resulted in inability of the transport system to accurately execute rapid fluctuations in motor speed required to produce the faster Wow rates. Although the absolute extent of this potential deficiency was not quantified, its influence was not considered sufficiently significant to compromise the veracity of the test results. Despite these technical limitations, it is still possible to establish significance and observe wider trends in the data

As discussed briefly in the literature review it is understood that Wow, as it occurs in the transport systems of tape machines and record turntables, is often not purely cyclical but instead is born of the summation of several phase independent cyclical speed modulations caused by various mechanical faults. The research presented in the literature review showed that this uneven, pseudo-cyclical waveform shape or pattern is a key component of Wow. Due to the specified timescale of this research it was not possible to include the waveform shape in the modulation parameter listening tests. However, the perceptual influence of waveform shape is still considered highly relevant to the study of the preferential effects of Wow pitch modulation. As such it is recommended it should be the focus of future areas of further study.

6. Conclusions & Further Work

It is noted that all of the aims and objectives outlined for the successful completion of this study were satisfied. The following project outcomes have been identified based on these aims and objectives:

- A detailed literature review was carried out that provided a comprehensive summary of the existing relevant research.

- A modified cassette tape deck was designed and produced that enabled the production of test stimuli. Whilst issues were encountered with the mechanical tolerances of the deck, it was capable of producing high quality stimuli that facilitated the listening tests, data analysis and consequent findings.

- The majority of stimuli were produced to a high standard, however an oversight in the stimuli production quality control did result in potential bias being introduced into 3 modulation parameter stimuli groups.

- Two subjective tests were designed and executed that culminated in the data analysis and conclusions of hypotheses, statistically significant differences and the identification of trends in the data.

These tests found statistically significant differences between mean participant preference observed across different pitch modulation rates. They also established that no such statistically significant differences were found between mean participant preference observed across different pitch modulation extents. Further analysis of the data showed a statistically significant increase in preference for Wow rates compared to Drift. Within Wow a trend was observed that found preference to steadily increase from 0.5 to 4 Hz, before levelling off from 4 to 7 Hz.

Upon the completion of the research the following points were identified as potential improvements to the methodology that could be implemented should the study be carried out again:

- Although it was originally decided not to use reference and hidden reference as detailed in the experimental design section, it is now thought that the inclusion of a hidden reference would have been advantageous in interpretation of the data analysis. The use of a hidden reference, free of any pitch modulation, in both listening tests would have potentially provided an insight into the participant's general perception of pitch modulation.

- As discussed previously issues were encountered with mechanical tolerances of the modified cassette deck that produced the test stimuli. It is recommended that future studies research and develop an improved stimuli production device in order to both reduce mechanical tolerances and increase the confidence in the data.

- If greater care and attention had been afforded to quality control during the production of the test material then the issues with the 0.5 Hz, 0.1 & 0.9% stimuli would have been identified earlier and rectified prior to the execution of the listening tests and data analysis. This would have provided a complete data set free of any potentially anomalous results.

- Whilst all listening tests were conducted in quiet environments using noise cancelling headphones, no single consistent venue was used, and in some cases, tests were carried out in public spaces such as university libraries. Although all tests were supervised by the researcher and it is not felt that participants were influenced by their surroundings, it must be acknowledged that some participants may have been subconsciously affected by slight visual distractions during their test. Conducting the tests in a controlled laboratory environment would have resolved this issue.

Based on the findings of this study the following avenues of research have been identified as recommended areas of future work:

- Further studies focusing on the relationship between participant preference and pitch modulation rates and extents, with stimuli formed of individual musical instrument stems to better represent the archetypal use of Wow as a creative effect. The methodological improvements discussed in the previous section would also be implemented.

- Studies on the preferential perception of pitch modulation at a wider range of rates and extents with a particular focus on the preferential perception of Flutter compared to Wow.

- Research investigating the psychological effects of Wow and the potential influence on preferential perception.

- Research into the influence of Wow on the perceived preference of the audio quality of vinyl records compared to digital media.

References

Audio Engineering Society. (2013). *Method for measurement of weighted peak flutter of analogue sound recording and reproducing equipment* (AES6-2008 (s2013)). Retrieved from http://www.aes.org/publications/standards/search.cfm?docID=15

American National Standards Institute. (1960). *American Standard Acoustical Terminology*. New York: American National Standards Institute.

Backus, J. (1977). *The Acoustical Foundations of Music* (2nd ed.). New York: W.W. Norton & Company, Inc.

Bartlett, G, W. (1975). *National Association of Broadcasters Engineering Handbook* (6th ed.). Retrieved from https://www.americanradiohistory.com/Archive-NAB-Engineering/NAB-6th-Edition/NAB-engineering-Handbook-6th-Edition.pdf

Belger, E. (1958). On Measuring Frequency Variations. *Rundfunktechnische Mitteihgen*, 2(4), 168-169.

Belger, E. (1972). On Measuring Frequency Variations. *IEEE Transactions on Audio and Electroacoustics*, 20(1), 79-80. doi 10.1109/TAU.1972.1162328

Benade, A, H. (1990). *Fundamentals of musical acoustics* (2nd ed.). Mineola: Dover Publications, Inc.

Bentler, R, J. (Executive Producer). (2014) Mac DeMarco – Pepperoni Playboy (Documentary) [Multimedia]. Retrieved from https://www.youtube.com/watch?v=JStLz_vkEm8&t=896s

Beyerdynamic. (2014). *DT 770 PRO Dynamic Headphone*. Retrieved from https://europe.beyerdynamic.com/dt-770-pro.html

Boards of Canada. (2002). Dawn Chorus [CD]. *Geogaddi.* London: Warp.

Boards of Canada. (2005). Dayvan Cowboy [CD]. *The Campfire Headphase*. London: Warp.

Burstein, H. (1975). *Questions and answers about tape recording.* Slough: Foulsham-Tab Limited.

Cagee, J. (Producer). (2016) Fiverr: Run your samples through SP-303 [Multimedia]. Retrieved from https://www.youtube.com/watch?v=v9ou4CwI4rQ

Capel, V. (1972). *Creative tape recording.* London: Fountain Press.

CCIR. (1953). *Document 187.* Paper presented at the VIIth Plenary Assembly of the CCIR, London, United Kingdom.

CCIR. (1956). *Document 364 Measurement of "Wow and "Flutter".* Paper presented at the VIIIth Plenary Assembly of the CCIR, Warsaw, Poland.

CCIR. (1966). *Recommendation 409-1 Measurement of wow and flutter in recording equipment and in sound reproduction.* Paper presented at the XIth Plenary Assembly of the CCIR, Oslo, Norway. Retrieved from http://search.itu.int/history/HistoryDigitalCollectionDocLibrary/4.277.43.en.1006.pdf

CCIR. (1970) *Recommendation 409-2 Measurement of wow and flutter in recording equipment and in sound reproduction.* Paper presented at the XIIth Plenary Assembly of the CCIR, New Delhi, India. Retrieved from http://search.itu.int/history/HistoryDigitalCollectionDocLibrary/4.278.43.en.1007.pdf

Comerci, F, A. (1962). Flutter Index Concept. *Journal of the SMPTE*, 71(1), 1-8.

Comerci, F, A., & Oliveros, E. (1956). An Audio Flutter Weighting Network. *Journal of the SMPTE*, 65(8), 419-425. doi: 10.5594/J13931

Corso, J, F., & Lewis, D. (1950). Preferred Rate and Extent of the Frequency Vibrato. *The Journal of Applied Psychology*, 34(3), 206-212.

DeMarco, M. (2012). My kind of woman [CD]. *2*. Brooklyn: Captured Tracks.

DeMarco, M. (2017). Still Beating [CD]. *This Old Dog*. Brooklyn: Captured Tracks.

Deutsches Institut für Normung. (1966). *Measuring Apparatus for Frequency Variations in Sound Recording Equipment* (DIN 45507:1966-10). Retrieved from https://www.beuth.de/en/standard/din-45507/1520543

Deutsches Institut für Normung. (1995). *Method of measurement of speed fluctuations in sound recording and reproducing equipment* (DIN IEC 60386:1995-04). Retrieved from https://www.beuth.de/en/standard/din-iec-60386/2497674

Eargle, J. (2003). *Handbook of recording engineering* (4th ed.). Boston: Kluwer Academic Publishers.

Ferrante, I. (2011). Vibrato rate and extent in soprano voice: A survey on one century of singing. *The Journal of the Acoustical Society of America*, 130(3), 1683-1688.

Fletcher, N, H. (2010). *Vibrato in music – physics and psychophysics.* Paper presented at the ISMA 2010, associated meeting of ICA 2010, Sydney and Katoomba, Australia. Retrieved from https://www.semanticscholar.org/paper/Vibrato-in-music-%E2%80%93-physics-and-psychophysics-Fletcher/245552bae146d77e058f6dc314999947bed486b9

Godsill, S, J., & Rayner, P, J, W. (1998). Digital audio restoration. London: Springer.

Greenhouse, S, W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika,* 24, 95–112.

Hongwei, W. (2019). *Wow and flutter measurement using multi-instrument*. Singapore: Virtins Technology.

Hood, J, L. (1999). *Audio electronics* (2nd ed.). Oxford: Newnes.

Institute of Electrical and Electronics Engineers. (1971). IEEE Standard Method for Measurement of Weighted Peak Flutter of Sound Recording and Reproducing Equipment (IEEE Std-193-1971). Retrieved from https://ieeexplore.ieee.org/document/7369902

International Electrotechnical Commission. (1972). *Method of measurement of speed fluctuations in sound recording and reproducing equipment* (IEC 60386). Retrieved from https://webstore.iec.ch/publication/2015

International Electrotechnical Commission. (1988). *Amendment 1 - Method of measurement of speed fluctuations in sound recording and reproducing equipment* (IEC 60386:1972/AMD1:1988). Retrieved from https://webstore.iec.ch/publication/2014

International Telecommunication Union. (2015). *Method for the subjective assessment of intermediate quality level of audio systems* (ITU-R BS.1534-3 (10/2015)). Retrieved from https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-3-201510-I!!PDF-E.pdf

IoSR Surrey. (2017). MUSHRA-MaxMSP. Retrieved from https://github.com/IoSR-Surrey/MUSHRA-MaxMSP

J Dilla. (2006). Two Can Win [CD]. *Donuts.* Los Angeles: Stones Throw.

McKnight, J, G. (1972). Development of a standard measurement to predict subjective flutter. *IEEE Transactions on Audio and Electroacoustics*, 20(1), 75-78.

National Association of Broadcasters. (1965). *Magnetic Tape Recording and Reproducing (Reel-to-Reel)*. Retrieved from http://www.richardhess.com/tape/history/NAB/NAB_Reel_Tape_Standard_1965_searchable.pdf

Pierce, J, R. (1999). The nature of musical sound. In. D. Deutsch (Eds.), *The Psychology of Music* (2nd ed.) (pp. 1-20). San Diego: Academic Press.

Prame, E. (1994). Measurements of the vibrato rate of ten singers. *The Journal of the Acoustical Society of America*, 96(4), 1979-1984.

Prandolini, R., & Moody, M. (1995). Brownian Nature of the Time-Base Error in Tape Recordings. *The Journal of the Audio Engineering Society*, 43(4), 241-247.

Risset, J, C., & Wessel, D, L. (1999). Exploration of timbre by analysis and synthesis. In. D. Deutsch (Eds.), *The Psychology of Music* (2nd ed.) (pp. 113-169). San Diego: Academic Press.

Rossing, T, D., Moore, F, R., & Wheeler, P, A. (2002). *The Science of Sound* (3rd eds.). San Francisco: Addison Wesley.

Sakai, H. (1970). Perceptibility of Wow and Flutter. *The Journal of the Audio Engineering Society*, 18(3), 290-298.

Seashore, C, E. (1931). The Natural History of the Vibrato. *Proceedings of the National Academy of Sciences,* 17(12), 623-626.

Seashore, C, E. (1967). *Psychology of Music.* Mineola: Dover Publications Inc.

Shower, E., & Biddulph, R. (1931). Differential Pitch Sensitivity of the Ear. *The Journal of the Acoustical Society of America, 3*(2A), 275-287.
doi: 10.1121/1.1915561

Stenson, R. (2015, 18 April). Vinyl Sim. [Weblog], Retrieved from https://tonal.goodhertz.co/vinyl-sim/

Stott, A., & Axon, P, E. (1955). The subjective discrimination of pitch and amplitude fluctuations in recording systems. *Proceedings of the IEE - Part B: Radio and Electronic Engineering*, 102(5), 642-656. doi: 10.1049/pi-b-1.1955.0125

Sundberg, J. (1999). The perception of singing. In. D. Deutsch (Eds.), *The Psychology of Music* (2nd ed.) (pp. 171-210). San Diego: Academic Press.

Taylor, C. (1992). *Exploring Music: The Science and Technology of Tones and Tunes.* Bristol: IOP Publishing.

The Guardian. (2013). Boards of Canada: 'We've become a lot more nihilistic over the years'. Retrieved from https://www.theguardian.com/music/2013/jun/06/boards-of-canada-become-more-nihilistic

The Reel Deal Denmark. (2019). The Reel Deal Denmark, Retrieved from https://www.ebay.co.uk/str/The-Reel-Deal-Denmark/cassette-test-tapes-/_i.html?_storecat=23076257016

Youlean. (2019). Youlean Loudness Meter. Retrieved from https://youlean.co/youlean-loudness-meter/

Zwicker, E. (1952). The limits of audibility of AM and FM of a tone. *Acustica*, 2(4), 239-246.

Zwicker, E. (1953). On the audibility of non-sinusoidal pitch variations. *Funk Ton*, 7, 342-346.

Zwicker, E., & Kaiser, W. (1952). The threshold of modulation within the auditory area. *Acustica*, 2(3), 125-133.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models*. Berlin: Springer.

Appendices

Appendix A – Development of Modified Tape Deck

A.1. Motor speed control modification

When exploring how to realise the motor speed control it was considered important to be able to have precise, quantifiable control over the nature of the modulations. To satisfy this it was decided that it should be possible to control the modulation using a computer, this would also provide a greater level of repeatability and consistency when applying modulation. The most accessible way to achieve this is through the use of a microcontroller. A microcontroller is a circuit board that enables software control over physical circuitry. Software programs can be written using code and loaded onto their built-in memory allowing both autonomous and real time control of analogue circuits.

An Arduino Uno microprocessor was selected for use in initial development and subsequent prototyping, it features both digital and analogue inputs and digital output terminals. This functionality allows you to take readings from external sensors and also to control components like LEDs or servos. Components that require larger voltages and draw more current can also be controlled with an Arduino by using a transistor. A transistor is an electrical device that acts as a switch in a circuit. Two of its three terminals are placed in a circuit and when voltage is applied to the third leg the transistor is switched on, completing the circuit. The use of a transistor with the digital output voltage of an Arduino in this way enables computer control of circuits featuring comparatively large electrical loads capable of power motors

Most cassette decks use DC motors to power their tape transport systems. With this in mind research was undertaken to identify circuits capable of controlling a DC motor with an Arduino. The circuit used in the ‘Motorized Pinwheel’ project featured in (Fitzgerald & Shiloh, 2012, pp. 94-101) was selected as a suitable basis from which to start, Figure A.1 shows this circuit diagram. The circuit includes a DC motor powered by a 9V battery. Digital pin D9 of the Arduino is being used as a digital

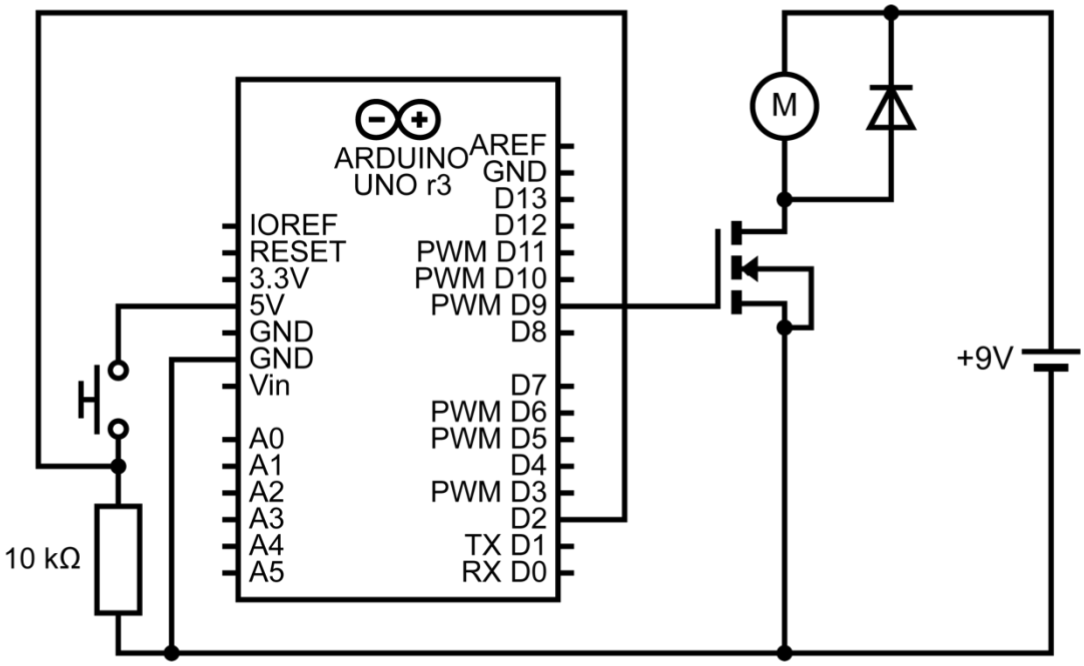

Figure A.1 Motorized Pinwheel circuit diagram.

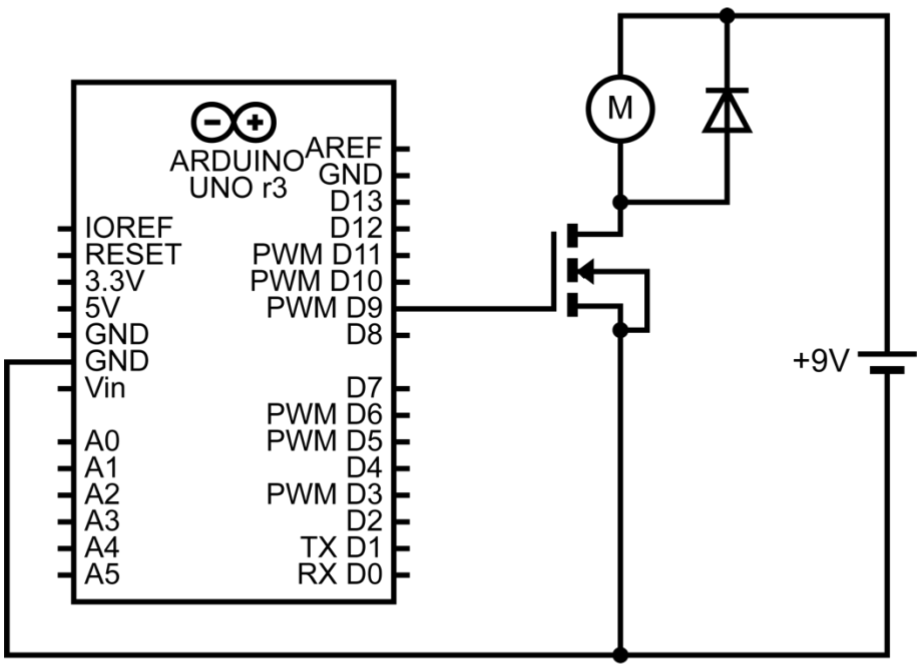

Figure A.2 Revised cassette deck circuit diagram.

output to open the N channel Mosfet transistor gate, which in turn provides power to the motor. Digital pin D2 is configured as an input to detect when the momentary push button is being pressed. The accompanying Arduino software sketch for this circuit is programmed to provide power to the motor whenever it senses the push button being pressed. The diode placed across the motor terminals protects the circuit by preventing large back voltages from damaging the Arduino.

This circuit was produced, tested and found to work well. Its design indicated that it would be possible to implement it into existing circuits in cassette decks by placing it between the DC motor of the deck and its immediate power and ground. As it was intended for the motor speed to be controlled directly by the computer it was decided that the push button was not a necessary feature. Consequently, the circuit design was revised to remove it, this revision is shown in Figure A.2.

A.2 Pulse width modulation

As previously stated, Arduinos do not feature analogue outputs. However, to address this it is possible to use a technique known as Pulse width modulation to enable a digital fixed voltage output to simulate a variable analogue output voltage. Pulse width modulation works on the principle that pulsing a fixed voltage by switching it on and off at very fast rates will result in an emulation of a lower voltage. The amount of time the PWM is turned on and off is referred to as the duty cycle. If a PWM has a duty cycle of 50% then this means the voltage is turned on for the same amount of time as it is turned off per cycle. For a 75% duty cycle the voltage is on 75% and off for 25% of the cycle, this is shown in Figure A.3. Pulse width modulation is commonly used to control servo motors. Sending a pulse width modulated signal to the gate terminal of the Mosfet transistor would enable control of the motor speed.

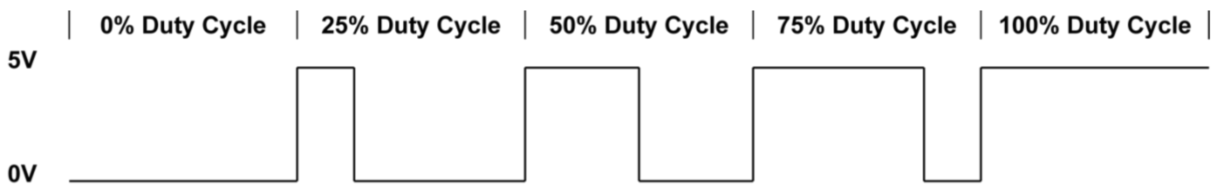

Figure A.3 Pulse width modulation duty cycle diagram.

Arduino sketch C++ code is able to facilitate control of PWM signals, however during development it was thought that it would be difficult and impractical to implement the necessary modulation control functionality required for the production of the listening test stimuli. The code-based Arduino sketch programming environment offers little opportunity for implementing modulation control as required, in this respect a graphical programming environment such as Max/MSP was considered much better suited.

A.3. Max/MSP

It is possible to interface Arduinos with Max/MSP through the use of an open source project called Maxuino. This is a collection of software comprising Max/MSP patches, external objects and an interfacing ‘Standard Firmata’ Arduino sketch that enables direct communication between Max and Arduino, including analogue and digital read and digital and PWM write functionality. Following initial testing it was decided to use Max to provide modulation control of the motor in the production of the listening test stimuli. The Maxuino patch providing PWM control via the Arduino to the Mosfet transistor and motor was found to provide precise and responsive control over the speed. This confirmed that sufficiently accurate speed and pitch modulation would be achievable using this method. The final Max patch used for the control of the modified cassette deck is available to download at the following link:

https://github.com/JayHarrison1/Evaluation-of-Pitch-Modulation-as-a-Creative-Effect-Max-MSP-Patches.git

A.4. Full Prototyping

The next stage was to implement the circuit in an existing tape machine to test how well it would perform using the external power and motor of the cassette deck. For this initial test a brand-new Bush CRS-132 Cassette Player/Recorder was sourced (Bush, 2016), this new machine was selected as it used a DC motor and provided optimal reliability, meaning any problems encountered with the performance were more likely to be due to the modification rather than the machine itself. The circuit was inserted in between the motor terminals and the immediate power and ground of the tape machine.

Upon testing it was found the circuit worked well and was capable of providing a wide range of precise control over the tape machine motor speed. Further tests were carried out with cassettes containing program material to determine the audio performance of the modification. It was found the motor speed modulation translated into the intended pitch modulation as expected and the modification implementation testing was considered successful.

The Bush CRS-132 is a low-quality machine by design and is only intended for rudimentary dictation work. The tape transport is very compact and lacks many of the features required for favourable performance, as such its base Wow and Flutter performance and frequency response is audibly very poor. As the implementation of controlled Wow would be the primary function of the modified deck, it was considered essential that its base Wow performance was of a sufficient standard so as not to unduly influence the perception of the controlled Wow in the listening tests. As a result, it was decided it would be necessary to source a higher quality cassette deck that would be capable of the performance levels required.

As discussed previously one of the methods to improve Wow & Flutter performance that is employed in cassette decks is the use of flywheels in the capstan drive system. Flywheels are especially effective at regulating against sudden, abrupt changes in drive speed. The more massive the flywheel, the harder it is for mechanical faults to influence the capstan speed. It was necessary for the final implementation of the modified tape machine to be able to reproduce a wide range of Wow & Flutter rates and extents for use in the listening tests. Therefore, it was decided that a cassette deck should be sourced that employed a comparatively modest flywheel whilst still possessing suitably low base Wow & Flutter performance. The consequent deck would therefore be capable of reproducing pitch modulation to sufficiently fast rates and large extents.

The JVC KD-V100 stereo cassette deck was selected and considered suitable for use in the study (JVC, 1983). A KD-V100 was sourced and the Arduino modification was installed inside the chassis of the deck. An Arduino Micro was used instead of the Uno that had previously been used in prototyping as the Micro is considerably smaller and therefore much easier to fit. The only other revision was the addition of a resistor and LED running from pin 12 to ground. The LED was installed in the rear chassis wall of the deck and was intended to serve as a visual feedback aid to check serial communication continuity between the Arduino and Max/MSP. Photographs of the Micro installed in the deck are shown in Figures A.4 & A.5.

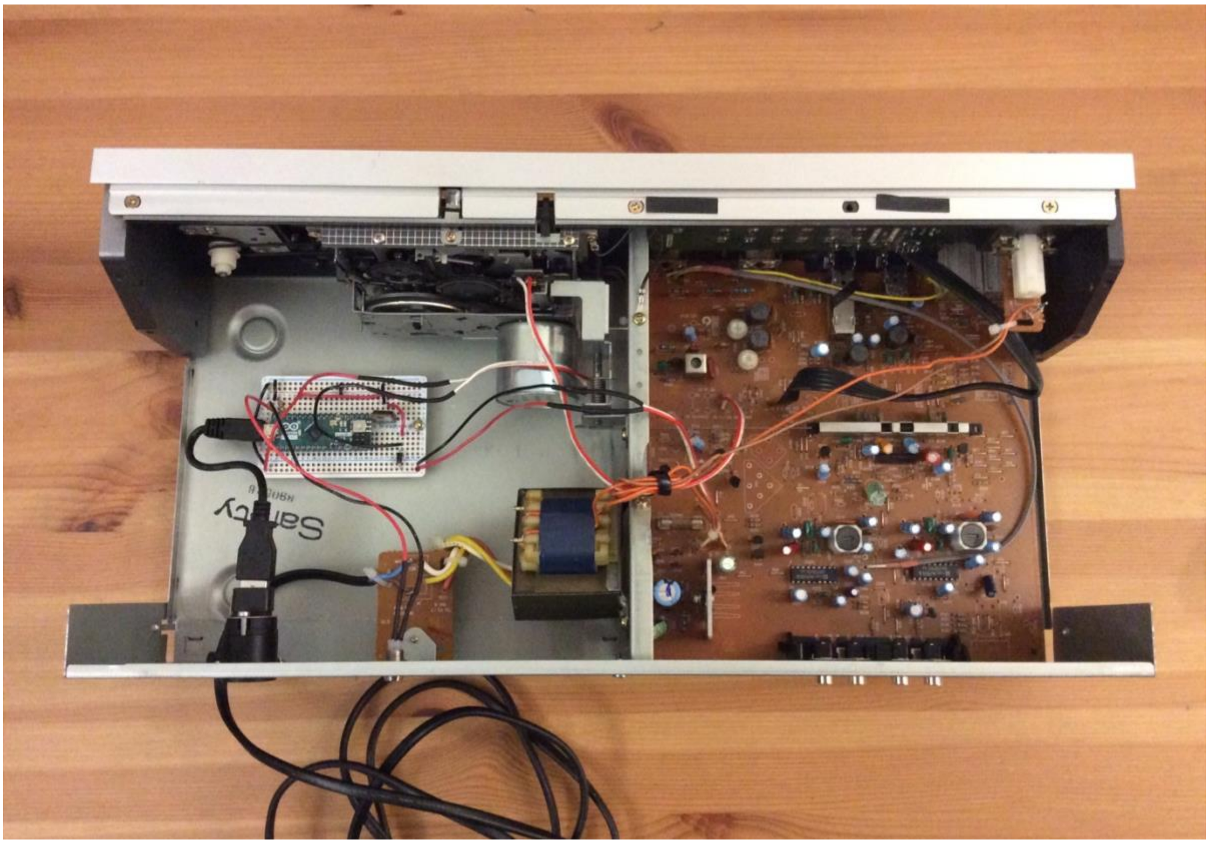

Figure A.4 Modified JVC KD-V100 cassette deck.

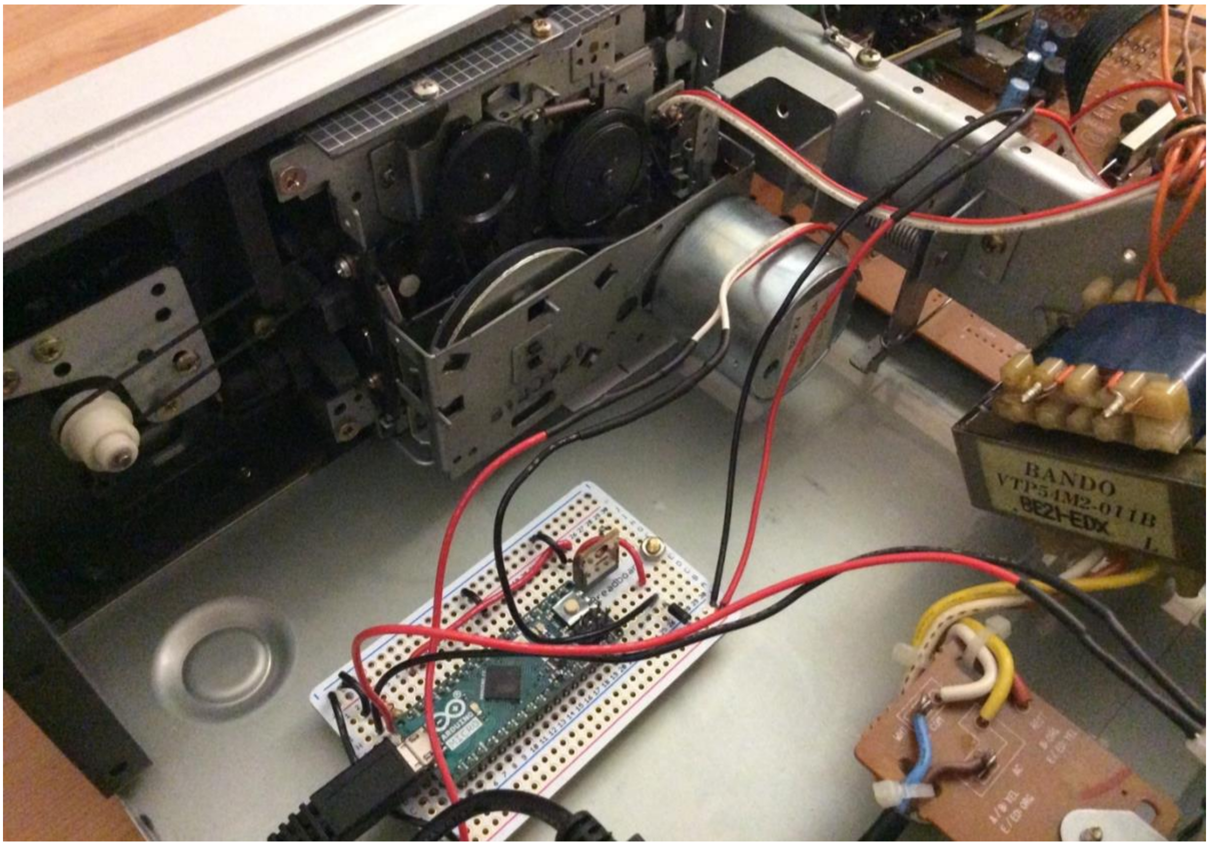

Figure A.5 Modified JVC KD-V100 cassette deck closeup.

A.5. Servicing and Calibration

Once the modification installation was complete it was decided the deck should undergo a full mechanical service. This would ensure all aspects of the deck were functioning correctly and optimise the base Wow & Flutter performance of the deck before any artificial modulation was applied. This work was carried out by the experienced tape service company PandaSound (PandaSound, 2012). The work consisted of head demagnetising and cleaning, azimuth adjustment, speed calibration and tape transport service. Prior to this work being carried out two professional calibration tapes were sourced from the calibration tape production company Reel Deal (The Reel Deal Denmark, 2019). These tapes contained several test tones required for accurate calibration of Wow & Flutter, speed, azimuth and level. Larsen (2018) details the stringent quality control involved in the production of these tapes.

It was essential for the deck to be able to modulate the pitch whilst also being able to deliver unmodulated playback at or close to the manufacture spec speed calibration. In doing so this would ensure that the deck would be able to playback any commercially produced cassette at the correct pitch speed. As the PWM modification was only capable of varying the speed from between 0 to 100% of the motor speed it was therefore necessary to raise the base speed to its maximum. On the KD-V100 this is achieved using the calibration pot located on the rear of the motor.

Once the motor speed was adjusted the 3000 Hz Wow & Flutter test tone tape and WFGUI frequency counter were used to establish the extent of the pitch increase brought about by the raised speed. The counter showed that the adjustment resulted in a frequency reading in the region of 3750 Hz. This difference in frequency was equivalent to over 2 semitones which would provide more than enough frequency headroom to accommodate the range of modulation extents planned for in the subjective listening test stimuli.

During final calibration testing was carried out to establish the Max/MSP PWM control value required to artificially revert the motor speed back to manufacturer specification. This involved gradually lowering the control value until the test tone tape was shown to be playing back at the correct 3000 Hz. The calibrated Max/MSP control value was found to be 0.294. Following servicing and calibration by PandaSound the Wow & Flutter performance was tested with the deck running at manufacturer spec speed under PWM control, it was measured at approximately 0.1% WRMS using the weighted DIN standard in WFGUI. There was a degree of error observed in the Wow and Flutter measurement of the modified deck, measurements taken by WFGUI were found to have a tolerance of ± 0.03%. Therefore, all Wow measurements stated in the listening test modulation stimuli are accurate to within ± 0.03%.

Appendix B – Subjective Listening Test Information Sheet

University of Salford MANCHESTER

Subjective Listening Test Participant Information Sheet

This study is being undertaken as part of an MSc Audio Production dissertation project at the University of Salford and is for educational purposes only.

Before you agree to take part in this study it is important that you aware of why the research is being undertaken and what it will involve. Please take time to read the following information carefully in order to decide whether or not you wish to take part in the study. Please feel free to ask the researcher to address any questions you may have regarding this information sheet or the wider study.

What is the purpose of the study?

To assess the perception of pitch modulation as a creative effect.

How long will my involvement in this study take?

Upon the completion of this sheet, the participant consent form and a verbal induction lasting approximately 1 minute, your involvement will be between 9 & 15 minutes.

What is the nature of the study and how will I be involved?

The study comprises of a brief questionnaire followed by a subjective listening test. The listening test is conducted using a pair of headphones and a laptop, both of which are provided by the researcher. You will be required to provide your responses using the laptop.

The listening test comprises of 3 pages. On each page you will listen to 9 audio samples and then rank them in order of preference on a scale of 0 to 100. You are encouraged to switch back and forth between the audio samples in order to aid your judgement.

You will be allowed a maximum of 5 minutes to assess the audio samples in each test. Click the ‘Next’ button when you are ready to move on you to the next page of the test.

What will happen to the responses provided as part of this study?

The responses you provide during this study will form a set of data that will be analysed as part of the aforementioned dissertation project. If you would like to see a copy of the dissertation, this will be available upon completion from the 1st of June 2019 and can be obtained by emailing the researcher using the details at the bottom of this document. Participants identities or personal data will not be included in the dissertation or shared.

Has this study been ethically reviewed?

Yes, this study has received ethical approval from the University of Salford.

Who should I contact if I have any further questions about the study?

Jay Harrison – University of Salford MSc Researcher - j.a.harrison1@edu.salford.ac.uk

Jay Harrison 1 University of Salford

Figure B.1. Subjective Listening Test Participant Information Sheet.

Appendix C – Subjective Listening Test Pre-Screening Questionnaire

University of Salford MANCHESTER

<u>Subjective Listening Test Pre-Screening Questionnaire</u>

The purpose of this questionnaire is to provide additional data required for the successful completion of the study. Please seek the assistance of the researcher if you have any questions regarding the completion of this questionnaire.

Surname: ……………………………………………………………………………

First Initial: ……………………………………………………………………………

Gender: ……………………………………………………………………………

Which of the following age brackets do you belong to? (Please circle)

18 – 25 26 – 35 36 – 45 46 – 55 56 – 65 66 – 75 75+

Do you have any known hearing impairments? If so, please provide details:

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Do you have any experience as a musician, in audio production or engaging in critical listening? If so, please provide details:

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Jay Harrison Ref: University of Salford

Figure C.1. Subjective Listening Test Pre-Screening Questionnaire.

Appendix D – Subjective Listening Test Consent Form

University of Salford MANCHESTER

Subjective listening test consent form

This form is for test subjects that agree to participate in a listening test and is to be completed prior to the beginning of the test.

I the participant agree to take part in a study on the following subject:

“Evaluation of Pitch Modulation as a Creative Effect”

I have read and understood the information sheet provided. I have been made aware of the premise, purpose and intended duration of the test and how I am expected to be involved as a participant. I have been afforded the opportunity to ask questions regarding the nature of the test and as a result I am confident in my understanding of the information sheet and consent form.

I agree to cooperate with any reasonable instructions provided to me during the test.

I understand that I am permitted to withdraw from the test at any juncture and will not be required to cite any justifications should I wish to do so.

I understand that any data held on a participant complies with the Data Protection Act (1998) and that participants identities or personal data will not be published or shared. I confirm that I will not request results obtained from this test to be restricted providing that my anonymity is maintained.

I understand I am participating in this listening test as a subject on purely a voluntary basis and that no fee is owed to me upon completion.

I confirm that I have read and understood the above in full and provide my consent to participate in the study.

Name of participant: (PRINTED) ....

Signed: ....

Date: ....

Name of witness: (PRINTED) ....

Signed: ....

Date: ....

Jay Harrison University of Salford

Figure D.1. Subjective Listening Test Consent Form.

Appendix E – Beyerdynamic DT770 PRO Specification

DT 770 PRO

Dynamic Headphone

Order # 459.046 (250 Ω)
Order # 474.746 (80 Ω)
Order # 483.664 (32 Ω)


32 Ω version


80 Ω version


250 Ω version

Features

- Closed diffuse field studio headphone
- Innovative bass reflex system
- Robust spring steel headband
- Single sided cable
- Soft inner headband
- Gold plated jack plug (3.5 mm) and adapter (6.35 mm)

Applications

The DT 770 PRO is a closed dynamic headphone of exceptional quality suitable for the most demanding professional and audiophile applications. The long term comfort and accurate performance make the DT 770 PRO the perfect monitoring headphone for recording studios, post production or broadcasting situations.
The low mass coil and diaphragm assembly produce a transient performance equalled only by electrostatic earphones, and, in combination with a carefully tailored frequency response offer a natural and balanced sound.
Soft earpads and adjustable, sliding, earpieces together with a single sided connecting cable ensure listening comfort during extended periods of use.
The DT 770 PRO features 32, 80 or 250 ohm drivers and a gold plated 3.5 mm stereo jack with 1/4" inch adapter, and is therefore suitable for use with almost all headphone amplifiers.

Technical Specifications

Transducer type	Dynamic
Operating principle	Closed
Nominal frequency response	5 - 35,000 Hz
Nominal impedance	32 Ω / 80 Ω / 250 Ω
Nominal SPL	96 dB SPL
Nominal T.H.D	< 0.2%
Power handling capacity	100 mW
Sound coupling to ear	Circumaural
Ambient noise isolation	
32 Ω version	approx. 20 dBA
80 Ω / 250 Ω version	approx. 18 dBA
Nominal headband pressure	approx. 3.5 N
Weight (without cable)	270 g
Length and type of cable	
32 Ω version	1.6 m / straight cable
80 Ω version	3 m / straight cable
250 Ω version	3 m / coiled cable
Connection	Gold plated stereo jack plug (3.5 mm) and 1/4" adapter (6.35 mm)

(all specifications according to EN 60 268-7)

Spare Parts

EDT 770 S	Ear pads, soft PVC, circumaural	Order # 904.783
EDT 770 V	Ear pads, velours, circumaural	Order # 926.660
EDT 770 VB	Ear pads, velours, circumaural, black	Order # 906.166
EDT 990 S	Ear pads, soft PVC, especially for 32 Ω version circumaural, black	Order # 904.791
BN 59-53/D	Headband pad	Order # 990.681

1 of 1

beyerdynamic GmbH & Co. KG
Theresienstr. 8 | 74072 Heilbronn – Germany
Tel. +49 (0) 71 31 / 617 - 0 | Fax +49 (0) 71 31 / 617 - 204
info@beyerdynamic.de | www.beyerdynamic.com

For further distributors worldwide, please go to www.beyerdynamic.com
Non-contractual illustrations. Contents subject to change without notice. E7/DT 770 PRO (07.14)

Figure E.1 Beyerdynamic DT770 Pro Spec. (Beyerdynamic, 2014).

Appendix F – Individual Genre Rate & Extent Data Analysis

F.1. Boards of Canada Modulation Rate Data Analysis

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in participant preference according to different levels of modulation rate. There was one outlier and the data was normally distributed, as assessed by the boxplot shown in Figure F.1 and Shapiro-Wilk test (p > .05), respectively. The assumption of sphericity was violated, as assessed by Mauchly's test of sphericity, $\chi^2(35) = 53.484$, $p = 0.028$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.564$).

Variation in modulation rate was found to bring about statistically significant changes in participant preference, $F(4.513, 85.744) = 18.888$, p = 6.1779E-12, partial $\eta^2 = 0.499$, with mean preference graphed in Figure F.2. Post hoc pairwise comparison analysis with a Bonferroni adjustment established that participant preference was not found to be statistically significant across any of the individual mean differences between modulation rates, as shown in Table F.1.

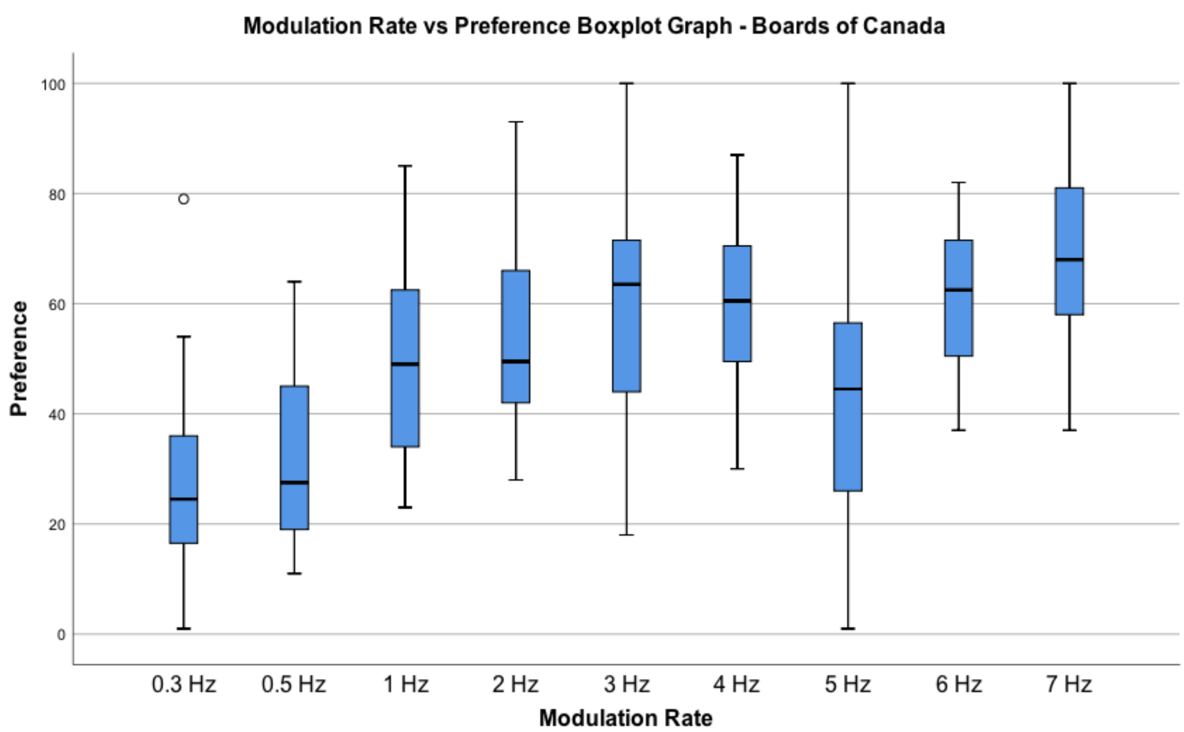

Figure F.1 Boards of Canada rate vs mean preference boxplot.

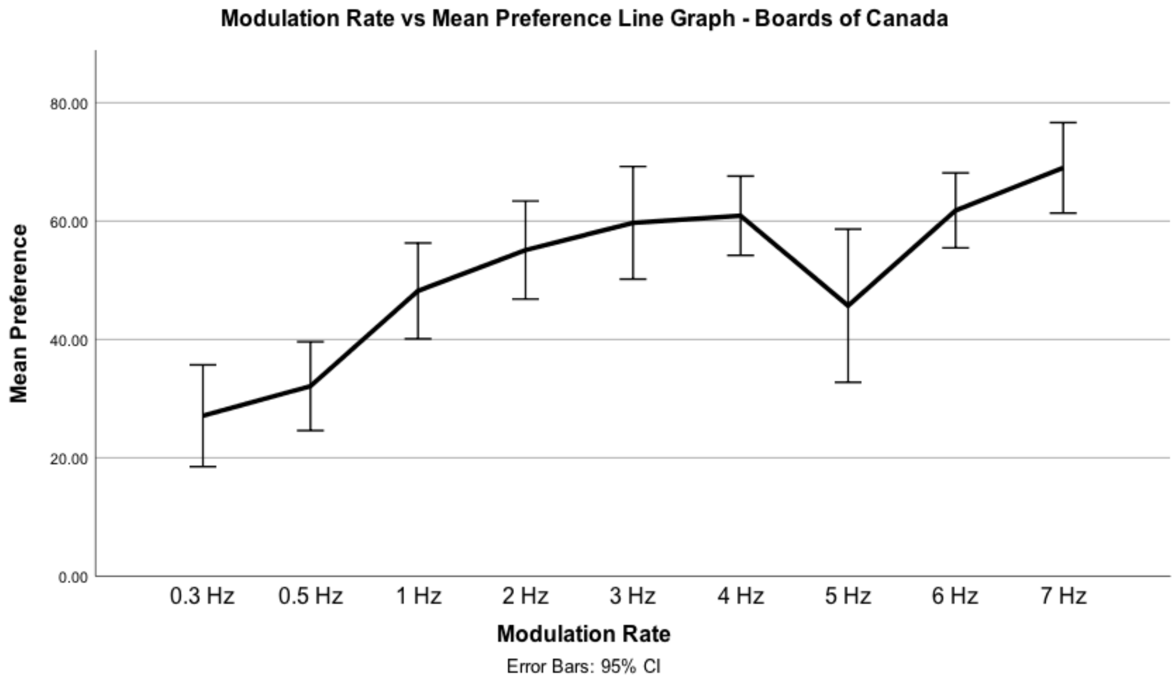

Figure F.2 Boards of Canada rate vs mean preference line graph.

As with the global modulation rate analysis a further one-way repeated measures ANOVA was carried out with the omission of the 5 Hz data. Mauchly's assumption of sphericity was again violated, $\chi^2(27) = 40.711$, $p = 0.049$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.612$). Variation in modulation rate was still found to bring about statistically significant changes in participant preference, $F(4.287, 81.452) = 24.735$, $p = 6.3779\text{E-}14$, partial $\eta^2 = 0.566$.

Modulation Rate	Mean Difference	Standard Error	Significance
0.3 – 0.5 Hz	5.000	3.126	1.000
0.5 – 1 Hz	16.100	4.642	0.093
1 – 2 Hz	6.900	3.317	1.000
2 – 3 Hz	4.600	4.005	1.000
3 – 4 Hz	1.200	3.910	1.000
4 – 5 Hz	-15.200	5.101	0.277
5 – 6 Hz	16.100	5.963	0.511
6 – 7 Hz	7.200	3.385	1.000

Table F.1 Boards of Canada modulation rate Bonferroni post hoc results.

F.2. J Dilla Modulation Rate Data Analysis

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in participant preference according to different levels of modulation rate. There were 4 outliers and the data was normally distributed, as assessed by the boxplot shown in Figure F.3 and Shapiro-Wilk test ($p > .05$), respectively. The assumption of sphericity was violated, as assessed by Mauchly's test of sphericity, $\chi^2(35) = 55.610$, $p = 0.018$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.550$).

Variation in modulation rate was found to bring about statistically significant changes in participant preference, $F(4.401, 83.614) = 10.113$, $p = 4.3698\text{E-}7$, partial $\eta^2 = 0.347$, with mean preference graphed in Figure F.4. Post hoc pairwise comparison analysis with a Bonferroni adjustment established that participant preference was found to be statistically significant across the difference in means between modulation rates 0.5 – 1 Hz, as shown in Table F.2.

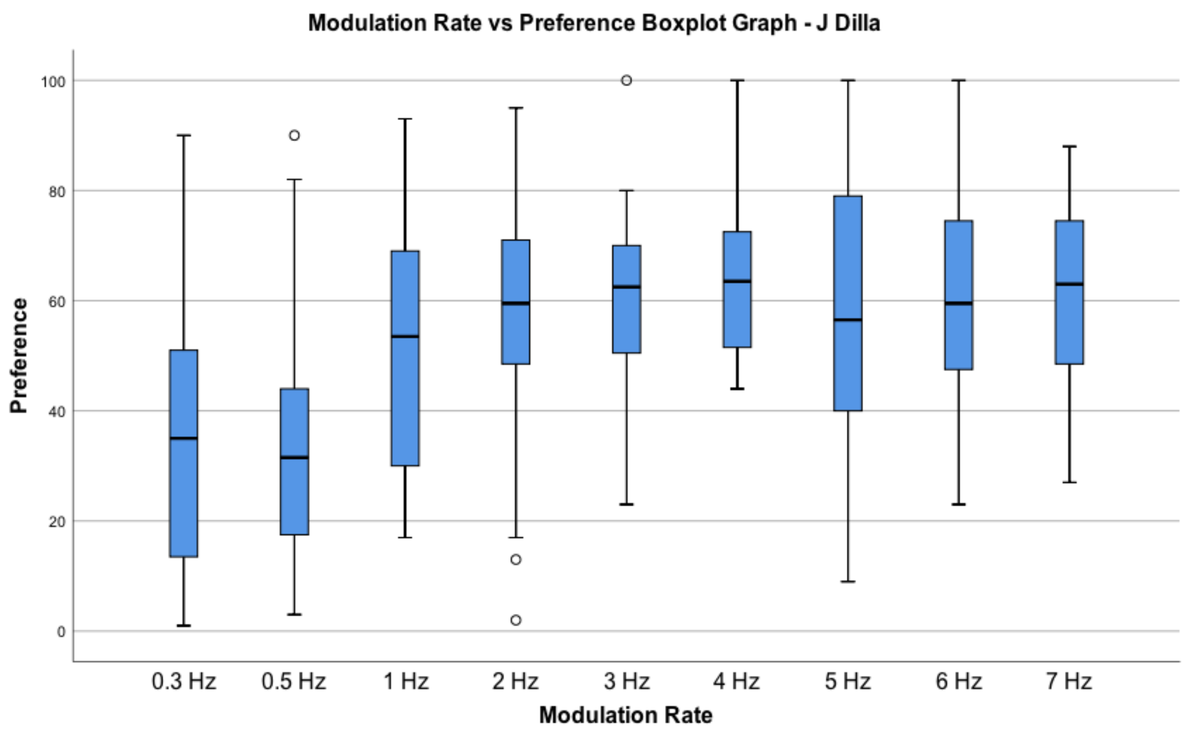

Figure F.3 J Dilla rate vs mean preference boxplot.

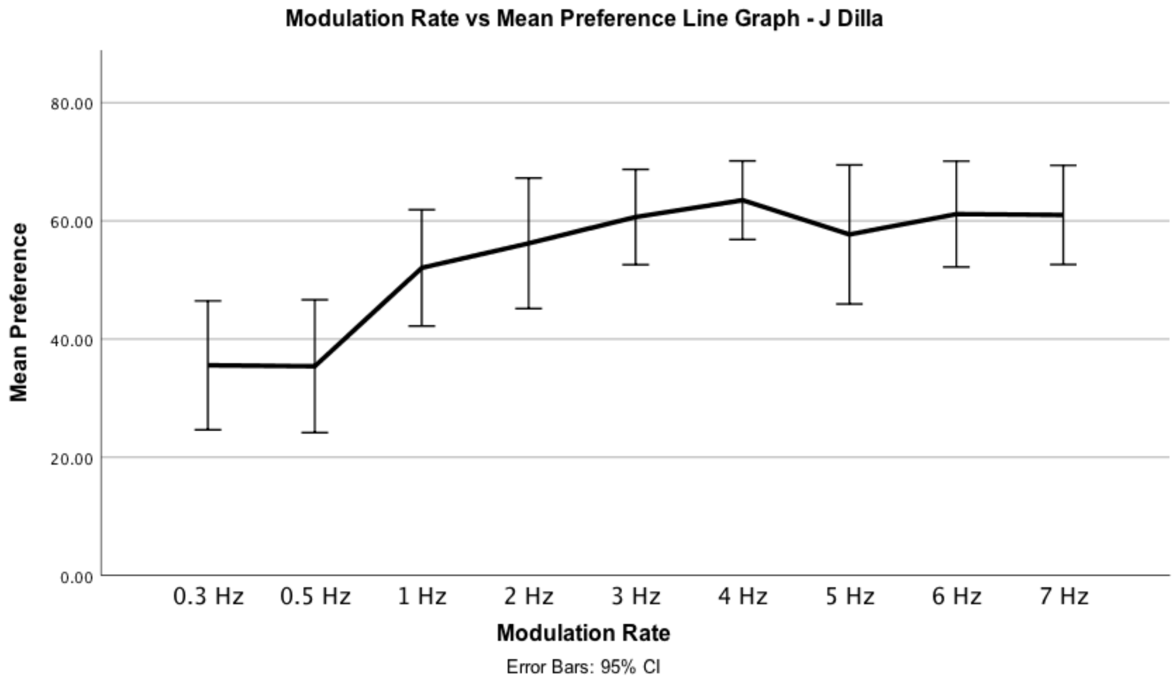

Figure F.4 J Dilla rate vs mean preference line graph.

As with the global modulation rate analysis a further one-way repeated measures ANOVA was carried out with the omission of the 5 Hz data. Mauchly's assumption of sphericity was again violated, $\chi^2(27) = 49.787$, $p = 0.006$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.501$). Variation in modulation rate was still found to bring about statistically significant changes in participant preference, $F(3.510, 66.686) = 12.596$, $p = 3.4854\text{E-}7$, partial $\eta^2 = 0.399$.

Modulation Rate	Mean Difference	Standard Error	Significance
0.3 – 0.5 Hz	-0.150	3.545	1.000
0.5 – 1 Hz	16.650	3.902	0.015
1 – 2 Hz	4.150	6.190	1.000
2 – 3 Hz	4.450	3.858	1.000
3 – 4 Hz	2.850	2.874	1.000
4 – 5 Hz	-5.800	5.356	1.000
5 – 6 Hz	3.450	5.500	1.000
6 – 7 Hz	-0.150	3.209	1.000

Table F.2 J Dilla modulation rate Bonferroni post hoc results.

F.3. Mac DeMarco Modulation Rate Data Analysis

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in participant preference according to different levels of modulation rate. There were no outliers and the data was normally distributed, as assessed by the boxplot shown in Figure F.5 and Shapiro-Wilk test (p > .05), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(35) = 41.055$, $p = 0.242$.

Variation in modulation rate was found to bring about statistically significant changes in participant preference, $F(8, 152) = 31.325$, $p = 1.3821\text{E-}28$, partial $\eta^2 = 0.622$, with mean preference graphed in Figure F.6. Post hoc pairwise comparison analysis with a Bonferroni adjustment established that participant preference was found to be statistically significant across the differences in means between modulation rates 0.5 – 1 Hz, 4 – 5 Hz & 5 – 6 Hz, as shown in Table F.3.

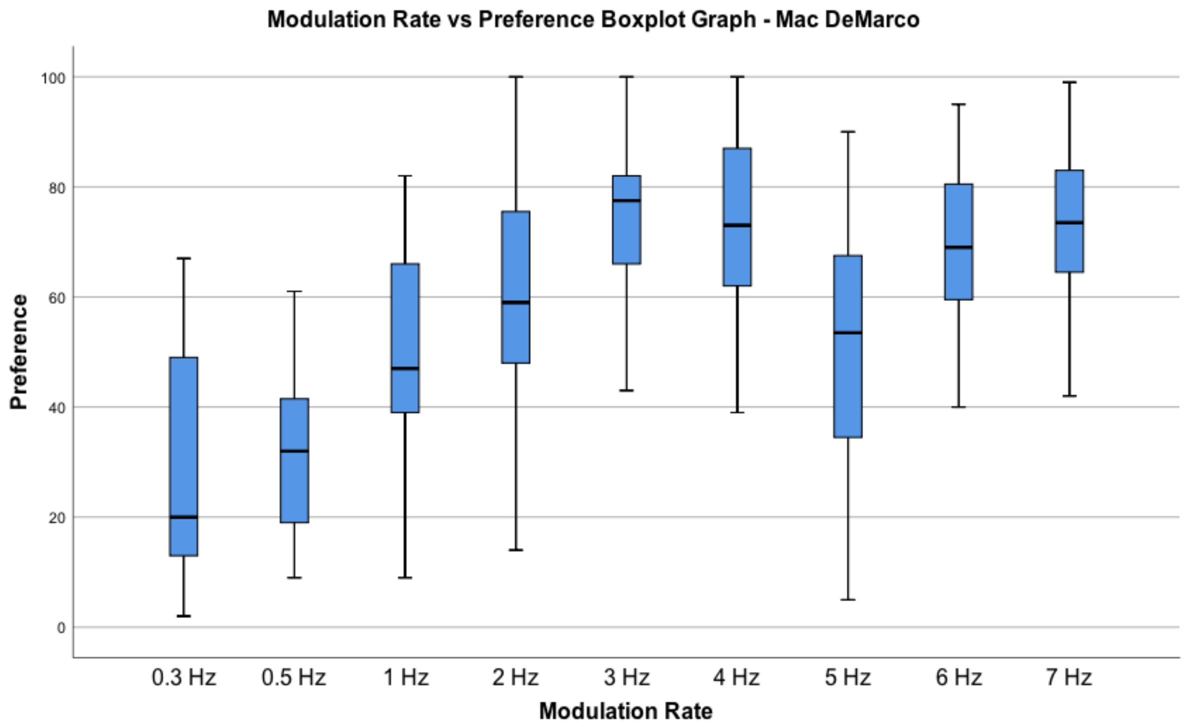

Figure F.5 Mac DeMarco rate vs mean preference boxplot.

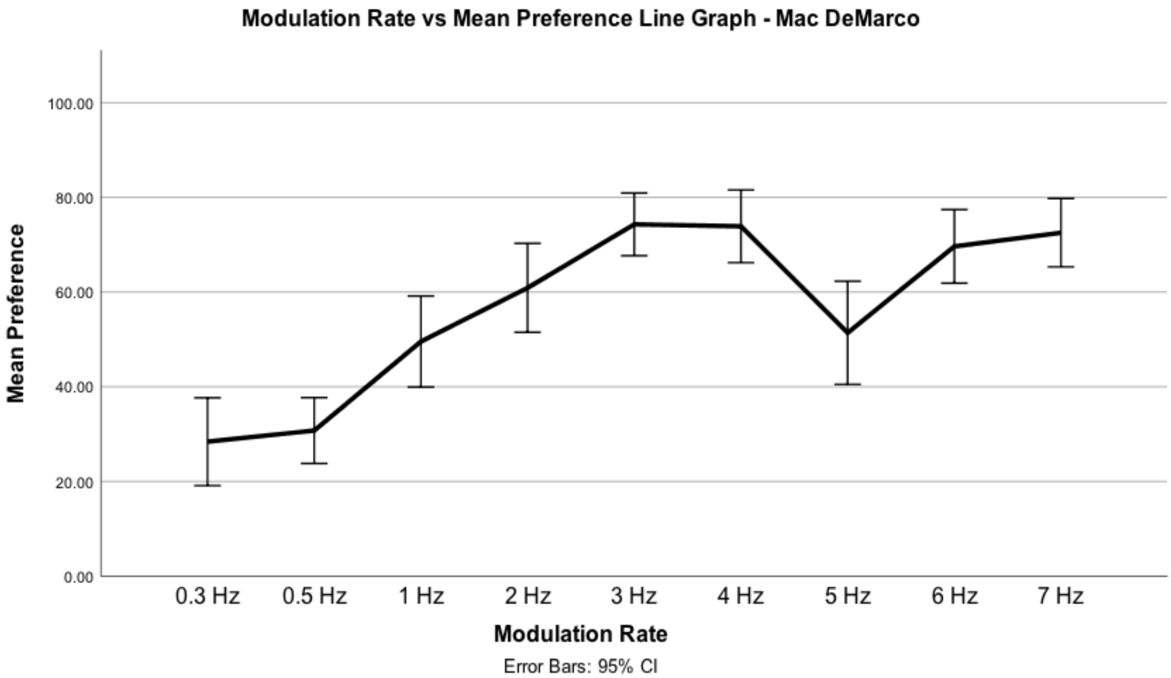

Figure F.6 Mac DeMarco rate vs mean preference line graph.

As with the global modulation rate analysis a further one-way repeated measures ANOVA was carried out with the omission of the 5 Hz data. Mauchly's assumption of sphericity was again met, $\chi^2(27) = 24.516$, $p = 0.615$. Variation in modulation rate was still found to bring about statistically significant changes in participant preference, $F(7, 133) = 38.605$, $p = 4.012\text{E-}29$, partial $\eta^2 = 0.670$.

Modulation Rate	Mean Difference	Standard Error	Significance
0.3 – 0.5 Hz	2.350	3.918	1.000
0.5 – 1 Hz	18.800	3.859	0.004
1 – 2 Hz	11.350	4.094	0.437
2 – 3 Hz	13.400	4.120	0.151
3 – 4 Hz	-0.400	3.574	1.000
4 – 5 Hz	-22.500	5.759	0.034
5 – 6 Hz	18.250	3.831	0.005
6 – 7 Hz	2.900	4.369	1.000

Table F.3 Mac DeMarco modulation rate Bonferroni post hoc results.

F.4. Boards of Canada Modulation Extent Data Analysis

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in participant preference according to different levels of modulation extent. There were 5 outliers and the data was normally distributed with the exception of 0.1, 0.5 & 0.9%, as assessed by the boxplot shown in Figure F.7 and Shapiro-Wilk test (p > .05), respectively. The assumption of sphericity was violated, as assessed by Mauchly's test of sphericity, $\chi^2(35) = 66.983$, $p = 0.001$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.506$).

Variation in modulation extent was found to bring about statistically significant changes in participant preference, $F(4.052, 76.987) = 11.018$, p = 3.6307E-7, partial $\eta^2 = 0.367$, with mean preference graphed in Figure F.8. Post hoc pairwise comparison analysis with a Bonferroni adjustment established that participant preference was found to be statistically significant across the difference in means between modulation rates 0.8 – 0.9%, as shown in Table F.4.

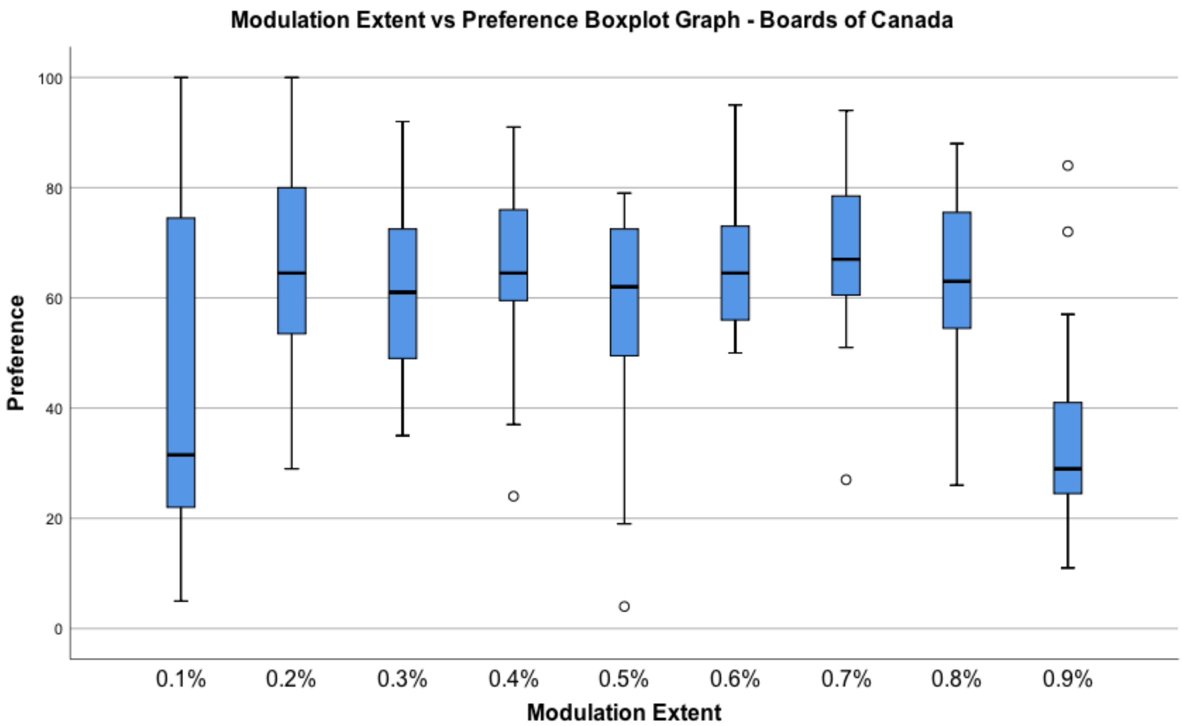

Figure F.7 Boards of Canada extent vs mean preference boxplot.

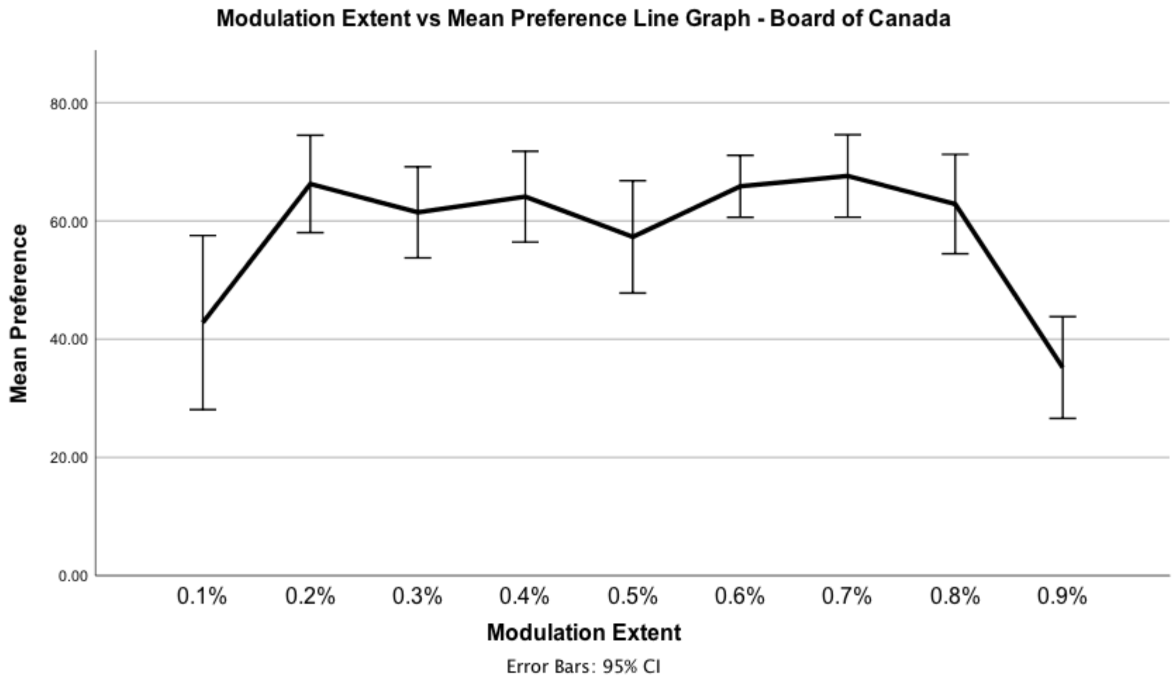

Figure F.8 Boards of Canada extent vs mean preference line graph.

As with the global modulation extent analysis a further one-way repeated measures ANOVA was carried out with the omission of the 0.1 & 0.9% data. Mauchly's assumption of sphericity was again violated, $\chi^2(20) = 33.456$, $p = 0.032$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.609$). Variation in modulation rate was not found to bring about statistically significant changes in participant preference, $F(3.657, 69.475) = 1.282$, $p = 0.286$, partial $\eta^2 = 0.063$.

Modulation Extent	Mean Difference	Standard Error	Significance
0.1 – 0.2%	23.450	6.835	0.101
0.2 – 0.3%	-4.800	5.350	1.000
0.3 – 0.4%	2.650	4.164	1.000
0.4 – 0.5%	-6.800	5.119	1.000
0.5 – 0.6%	8.550	3.686	1.000
0.6 – 0.7%	1.750	3.299	1.000
0.7 – 0.8%	-4.750	4.078	1.000
0.8 – 0.9%	-27.650	3.976	0.000

Table F.4 Boards of Canada modulation extent Bonferroni post hoc results.

F.5. J Dilla Modulation Extent Data Analysis

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in participant preference according to different levels of modulation extent. There was 1 extreme outlier and the data was normally distributed with the exception of 0.2%, as assessed by the boxplot shown in Figure F.9 and Shapiro-Wilk test (p > .05), respectively. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $\chi^2(35) = 43.481$, $p = 0.171$.

Variation in modulation extent was found to bring about statistically significant changes in participant preference, $F(8, 152) = 4.866$, $p = 0.000023$, partial $\eta^2 = 0.204$, with mean preference graphed in Figure F.10. Post hoc pairwise comparison analysis with a Bonferroni adjustment established that participant preference was found to be statistically significant across the difference in means between modulation rates 0.8 – 0.9%, as shown in Table F.5.

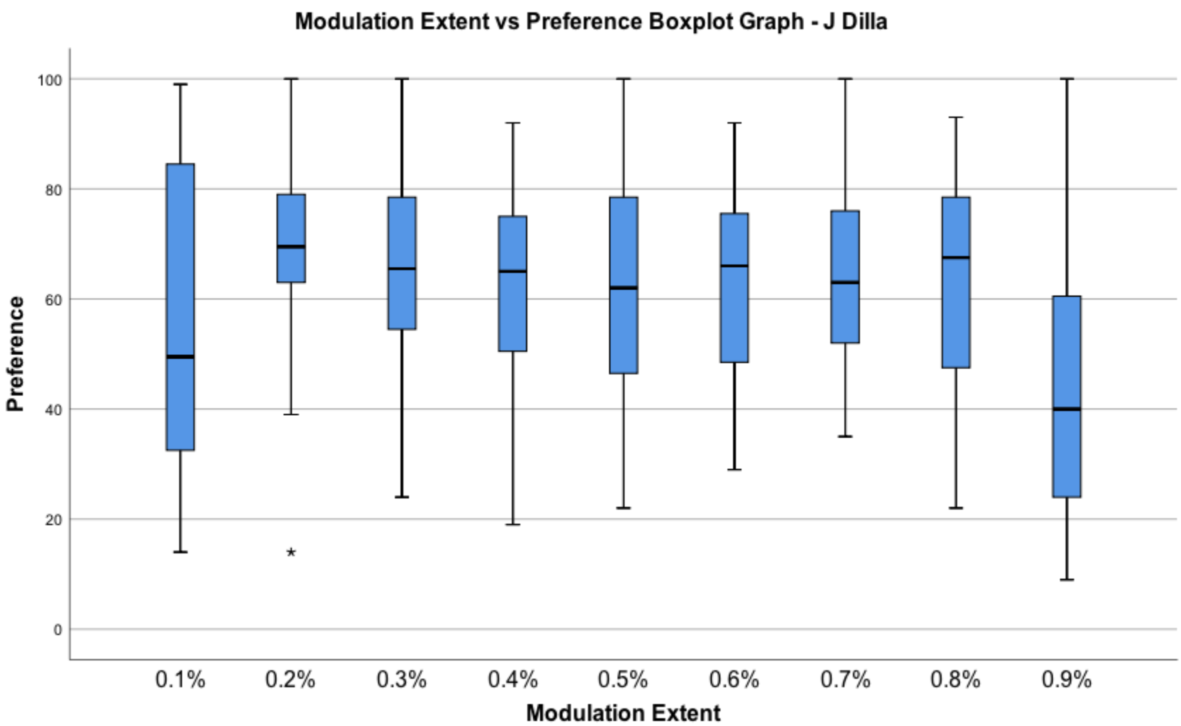

Figure F.9 J Dilla extent vs mean preference boxplot.

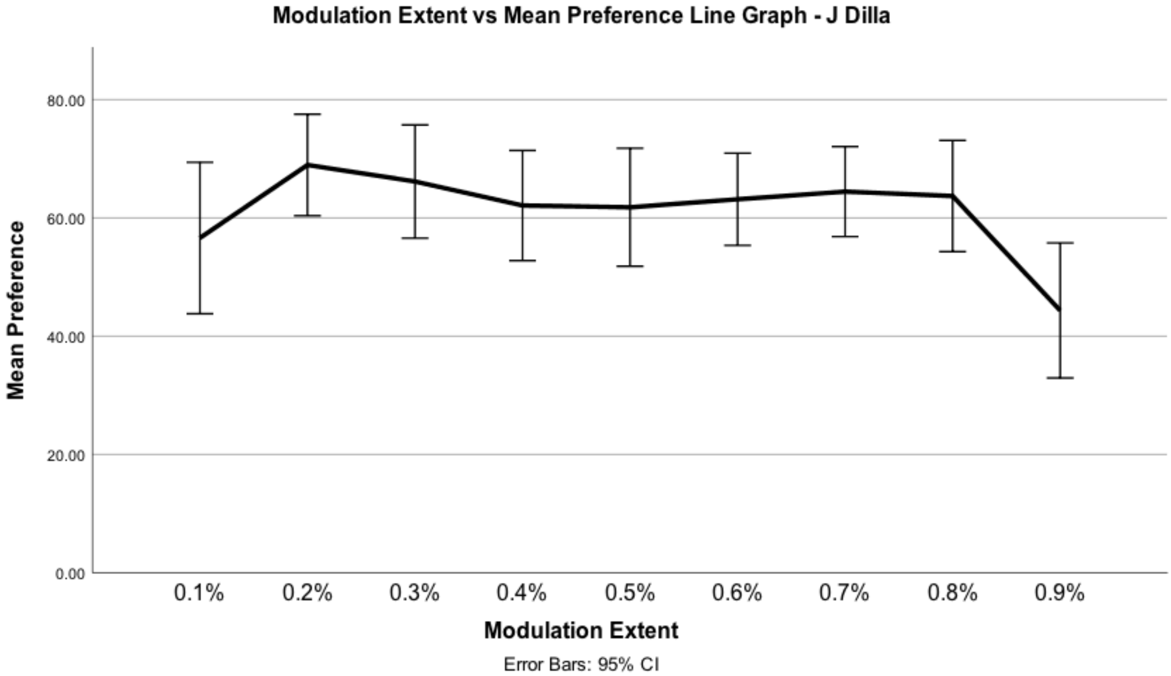

Figure F.10 J Dilla extent vs mean preference line graph.

As with the global modulation extent analysis a further one-way repeated measures ANOVA was carried out with the omission of the 0.1 & 0.9% data. Mauchly's assumption of sphericity was violated, $\chi^2(20) = 31.894$, $p = 0.048$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.630$). Variation in modulation rate was not found to bring about statistically significant changes in participant preference, $F(3.779, 71.798) = 0.710$, $p = 0.580$, partial $\eta^2 = 0.036$.

Modulation Extent	Mean Difference	Standard Error	Significance
0.1 – 0.2%	12.350	4.706	0.601
0.2 – 0.3%	-2.800	2.795	1.000
0.3 – 0.4%	-4.050	4.415	1.000
0.4 – 0.5%	-0.300	5.286	1.000
0.5 – 0.6%	1.350	3.980	1.000
0.6 – 0.7%	1.300	2.748	1.000
0.7 – 0.8%	-0.750	3.550	1.000
0.8 – 0.9%	-19.350	4.954	0.034

Table F.5 J Dilla modulation extent Bonferroni post hoc results.

F.6. Mac DeMarco Modulation Extent Data Analysis

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in participant preference according to different levels of modulation extent. There were 3 outliers and the data was normally distributed with the exception of 0.4%, as assessed by the boxplot shown in Figure F.11 and Shapiro-Wilk test ($p > .05$), respectively. The assumption of sphericity was violated, as assessed by Mauchly's test of sphericity, $\chi^2(35) = 83.834$, $p = 0.000011$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.444$).

Variation in modulation extent was found to bring about statistically significant changes in participant preference, $F(3.550, 67.442) = 9.218$, $p = 0.000012$, partial $\eta^2 = 0.327$, with mean preference graphed in Figure F.12. Post hoc pairwise comparison analysis with a Bonferroni adjustment established that participant preference was found to be statistically significant across the difference in means between modulation rates 0.8 – 0.9%, as shown in Table F.6.

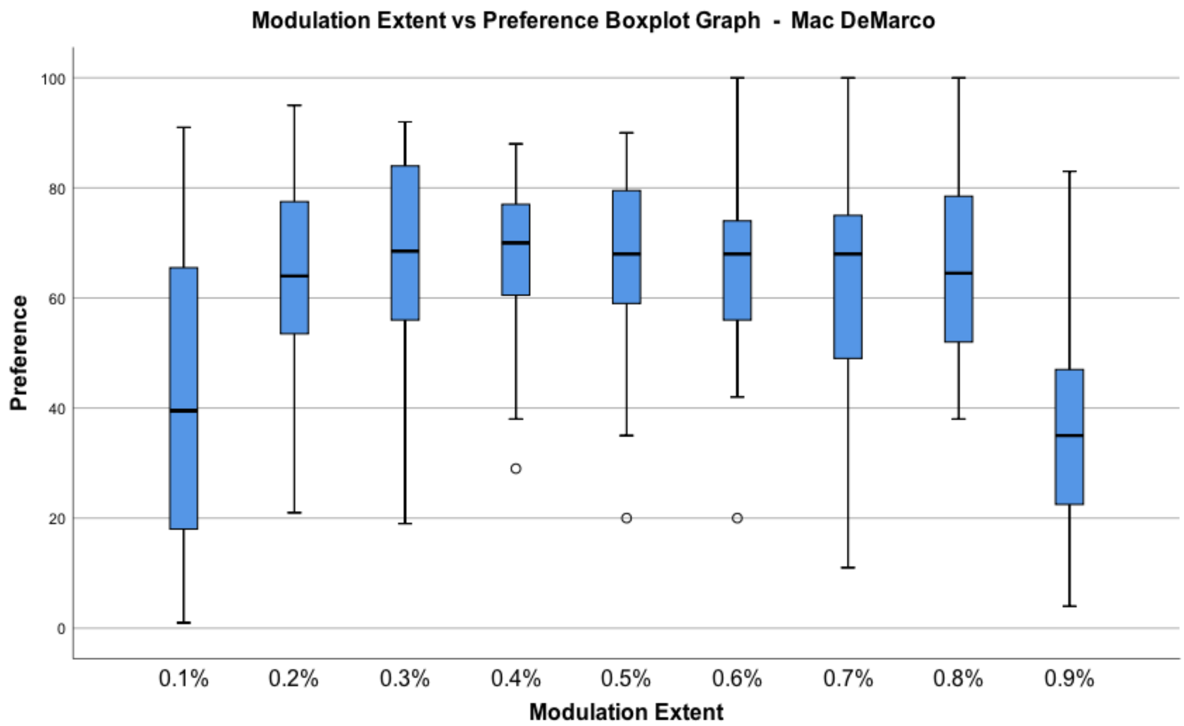

Figure F.11 Mac DeMarco extent vs mean preference boxplot.

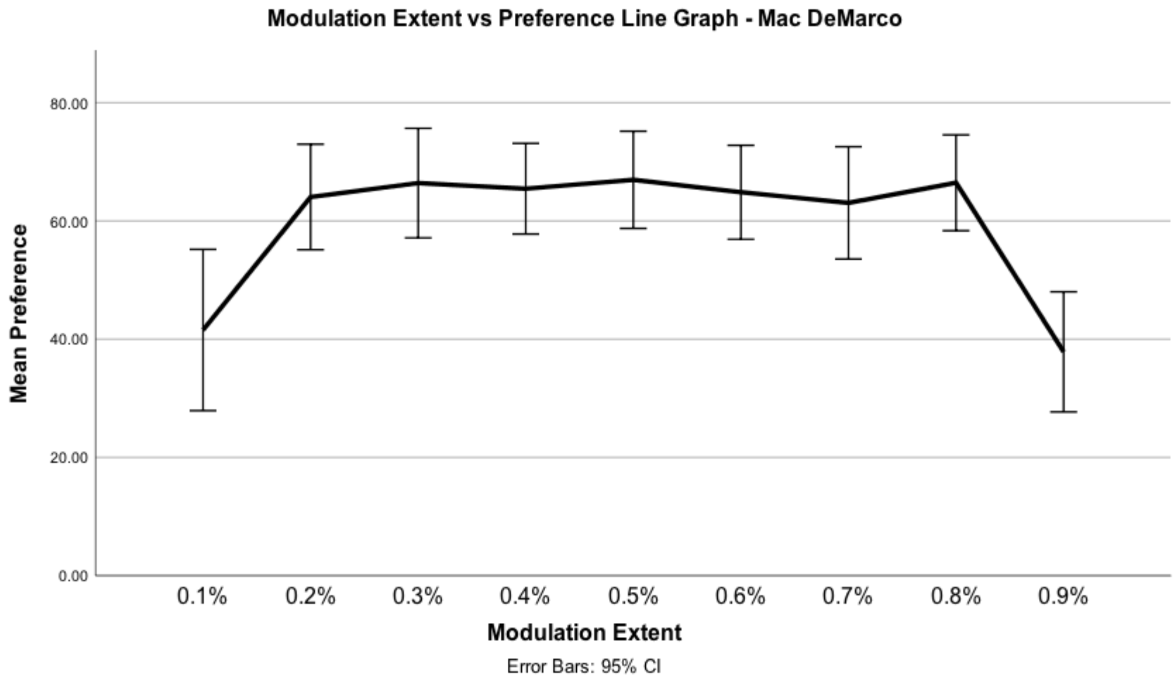

Figure F.12 Mac DeMarco extent vs mean preference line graph.

As with the global modulation extent analysis a further one-way repeated measures ANOVA was carried out with the omission of the 0.1 & 0.9% data. Mauchly's assumption of sphericity was again violated, $\chi^2(20) = 42.798$, $p = 0.002$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = 0.597$). Variation in modulation rate was not found to bring about statistically significant changes in participant preference, $F(3.581, 68.040) = 0.226$, $p = 0.907$, partial $\eta^2 = 0.012$.

Modulation Extent	Mean Difference	Standard Error	Significance
0.1 – 0.2%	22.500	7.277	0.216
0.2 – 0.3%	2.350	7.521	1.000
0.3 – 0.4%	-0.950	4.111	1.000
0.4 – 0.5%	1.500	4.041	1.000
0.5 – 0.6%	-2.100	4.193	1.000
0.6 – 0.7%	-1.800	5.188	1.000
0.7 – 0.8%	3.400	3.359	1.000
0.8 – 0.9%	-28.600	5.858	0.004

Table F.6 Mac DeMarco modulation extent Bonferroni post hoc results.

Appendix G - Subjective Listening Test Participant Observations

G.1. Rate

R1 commented that they weren't 100% sure what they were assessing, they stated they were listening for variations in pitch but also perceived some static interference in some of the stimuli.

R2 commented that they felt they were marking the stimuli overall quite low and that they perceived the differences as being quite minimal.

R3 commented that they felt the effect was more preferable on the Mac DeMarco track compared to the other two tracks. They cited Mac DeMarco's known use of tape speed manipulation as a reason for this.

R4 commented that they perceived small and large differences in the test and they found it difficult to assess different levels of modulation when the differences were very close.

R6 identified the Boards of Canada track as being more discordant so they didn't perceive as much difference in pitch. The pitch modulation in the Mac DeMarco tracks was most easily perceptible. They also commented that the modulation that 'was more in key' with the track was most preferable.

R8 commented that they felt they needed a reference to assist in assessing the stimuli. They found the pitch modulation was most preferable in the Mac DeMarco track, they cited the tempo as the main reason for this but also the instrumentation. Boards of Canada was not preferable because the modulation was applied to the whole mix, the modulation was particularly unfavourable when applied to the strings. The modulation applied to the strings in the J Dilla track took away from the overall preference of the track, they made it sound 'like an amateur orchestra'. They suggested a logical progression in research would be to test the modulation on stems.

R16 commented that they perceived some of the samples as having too much midrange and that the highly slowed down samples were highly non-preferential.

R18 commented that the pitch modulation in the J Dilla track was most perceptible. They also commented that they recognised the Mac DeMarco track and were conscious that they were trying to determine which stimuli was the original version.

R19 commented that they recognised the Mac DeMarco track and were conscious that they were trying to determine which stimuli was the original version.

R20 asked if some of the stimuli were the same.

G.2. Extent

E2 commented that they perceived the J Dilla track as having less pitch modulation.

E3 commented that the modulation that was closest to a whole pitch out (adhering to the western musical scale) was the most preferable. The modulation also reminded the participant of the rise in pitch sometimes heard at the end of a track, like a fanfare.

E8 commented that they felt that the middle ground of modulation was preferable, but wasn’t sure which samples represented the middle ground. They felt the modulation was most preferable on the Mac DeMarco track.

E11 commented that the ‘slower, pitched down’ samples in the J Dilla track were by far the most preferable. The participant said they gave an emotional response.

E12 commented that they preferred the pitch modulation on the Mac DeMarco track. They felt the pitch modulation was more easily perceptible on this track.

E15 commented that they much preferred the pitch modulation on the J Dilla track. They stated they liked the larger pitch modulation, specifically citing how the singers voice sounded preferable both high pitch ‘like a chipmunk’ and pitch down. They also stated that they felt the pitch modulation was most perceptible on the J Dilla track.

E16 commented that they felt the pitch modulation was easiest to perceive on the Boards of Canada track, whereas the pitch modulation on the J Dilla track was by far the most difficult to perceive.

E17 commented that the pitch modulation was most favourable on the J Dilla track, they cited the use common use of turntables in the production for this. They also commented that there was a ‘non- diatonic passing chord’ in the Mac DeMarco track that was off putting when trying to perceive the pitch modulation. They also said they perceived most of the modulation as being preferable and suggested that perhaps this was due to an existing personal preference for this kind of effect.

E18 commented that the extensive pitch modulation on the Mac DeMarco track was largely unfavourable, but what was perceived as the most severe modulation was actually favourable. The participant commented that it was ‘as though it was in tune with the track’.

E19 described the lesser modulation as being brighter and as such more preferable. They also stated that the pitch modulation in the Mac DeMarco and J Dilla tracks gave a ‘lofi chillout vibe’.

E20 commented that the pitch modulation was hardest to perceive in the J Dilla track, then the Board of Canada track and then the Mac DeMarco track. They also stated that perhaps the perception of pitch modulation was a contributing factor to why some individuals perceive vinyl records to be preferable to digital media. They also commented that they perceived 30 to 40% modulation as being the optimal extent within the stimuli presented.

Appendix H – Project Plan (Prior to amendments of project focus)

MSc Audio Production Dissertation Project Plan

Title: **Evaluation of pitch modulation as a creative effect**

Student: **Jay Harrison**

Supervisor: **Dr Phil Duncan**

Introduction

Vocalists and instrumentalists use Vibrato to produce cyclical pitch modulations that add to their performance. It is a popular technique that is widely used in many musical cultures. Wow and Flutter are more subtle pitch modulations caused by mechanical inconsistencies in tape machines and record players. It has become a favourable creative effect that is often used in the production of Hip Hop, Lofi & alternative sub genres. What aspects of these pitch modulations appeal to some listeners?

This project will involve the subjective evaluation of a number of pitch modulation effects by way of controlled listening tests. Different modulation parameters will be explored and tested including speed, depth and shape.

Objectives

- Review current literature including research on the psychoacoustic perception of pitch modulation
- Conduct historic and contemporary review of pitch modulation devices and plugins
- Develop a bespoke tape machine capable of producing pitch modulation effects with control over speed, depth and waveform shape.
- Produce a number of listening stimuli to be processed by the tape machine.
- Design a controlled subjective test that will allow the comparison of different pitch modulation parameters.
- Conduct a statistical analysis of the results of the subjective tests.
- Draw conclusions on the influence of different pitch modulation parameters on listener perception.

The following objectives may be attempted if the time allows:

- Carry out an online survey exploring specific pitch modulation effect preferences amongst professionals, enthusiasts and listeners
- Undertake objective measurements of the pitch modulation parameters of popular effects processors and implement them in the tests.

Figure H.1 Dissertation project plan.

Equipment requirements

No equipment requirements have been identified at this stage of the project.

Signature of student:________________________________Date_____________

Signature of supervisor:______________________________Date_____________

Week → **Activity ↓**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Prepare project plan														
Literature review														
Bespoke tape machine production														
Produce listening test stimuli														
Design and run Subjective test														
Analyse Results														
Write final report														
Optional Objectives														
Produce and distribute pitch modulation survey														
Take objective measurements of popular effects														

Figure H.1 Dissertation project plan.