An Investigation Into the Changes of Loudness Perception in Relation to Changes in Crest Factor for Octave Bands

By

Mark Wendl

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Masters of Science

The University of Huddersfield

January 2015

Abstract

Even though modern technology has the capability to make audio sound better than it ever has, there is still an ongoing argument about the compression levels with music. Heavy compression has been found in previous studies to be detrimental to quality but before a full understanding into how compression affects audio, a fundamental knowledge of the affects across different frequencies is important.

Using pink noise in a series of tests to determine the level of compression and how it changes the perceived loudness across different frequency bands and across different amplitude characteristics gave an insight into how compression affects complex audio.

It was found that lower octaves behaved differently to the expected outcome where it was found to be perceived as louder despite no extra treatment from other frequency bands. It was however also found that all frequency bands increased in loudness as a result of compression.

Table of Contents

List of Figures

Table of Tables

1 INTRODUCTION

1.1 Background

There are many aspects to a professional recording and each aspect includes a careful consideration into the equipment being used and the determined effect it has on the resultant audio. From the recording stage through to the mastering stage a level of understanding of the final release is required from the several engineers involved at differing stages. Each of these stages can impact how the audio is represented, however techniques applied during the mastering stage has caused significant controversy with the approach and use of a type of dynamic range compressor (DRC) called a limiter and the resulting effect it had on the audio.

The mastering stage of the production process can be interpreted as the bridging point from the audio sounding high quality in a studio sound system and ensuring it still sounds high quality in the less invested sound systems such as a regular home system [Meadows, 2008]. When the audio has finished being mixed it is presented to a mastering engineer to finalise the work. Depending on the format this can include many aspects for example a digital single would potentially require content edits, equalisation and DRC whereas the CD would require these several times over for different tracks and additionally potentially require level changes, appropriate fades and track ordering [Owsinski, 2008]. Another consideration for the mastering engineer will be the performance and genre. Genres such as classical and jazz will have a greater range of dynamics compared to a popular music production [Owsinski, 2008] and this is potentially accomplished by how much the audio is compressed via the use of a compressor or a limiter. The extent that the mastering engineer changes the audio however is dependant on the audio and the engineer. Katz [2007] noted that mastering is not about processing but about judging if the audio is of the quality to distribute. This could be interpreted as meaning that the mastering stage is less about artistic decisions and more about commercial decisions and so the mastering stage has been heavily influenced by trends of sales and business related factors.

The “loudness wars” were the result of this input of business lead competitiveness. The “loudness wars” are referred to usually as a problem of the compact disc (CD) era however it stems from the vinyl cutting era where record companies noticed that the louder the record played back, the more listeners would perceive it as sounding better and so it sold more [Owsinski, 2008]. This started with production techniques to increase the loudness such as adding more instrumentation [Vickers, 2010a] however it then moved to the application of DRC. This created the louder is better stigma which during the vinyl era was not too much of a problem because if a track was produced that was too loud then it would cause the needle to skip in the playback system and so would be inaudible [Owsinski, 2008].

When the CD was introduced a new capability of producing inherently loud audio came with it. The CD allowed for high fidelity listening due to the inherent low noise floor and large dynamic capabilities however the same trend occurred where the resultant master was for radio and so was to be as loud as possible as outlined by Nielsen & Lund [1999]. When the industry was introduced to digital audio workstations (DAWs), a sequencing piece of software that allowed for digital manipulation of audio, Nielson and Lund commented on the large amount of dynamic processing that became the normal practise. With these digital advancements the only limit that the audio had was that of 0 decibels relative to full scale (dBFS) meaning this was the digital maximum point that the audio could be pushed to.

The results of the competitiveness and the DAW work flow created an era of "hypercompressed" releases. The audio was being squeezed to accomplish louder playback. A comparison of audio produced in the pre CD era can be seen compared to that of a hypercompressed piece of audio produced during the DAW work flow era in **figure 1.1**.

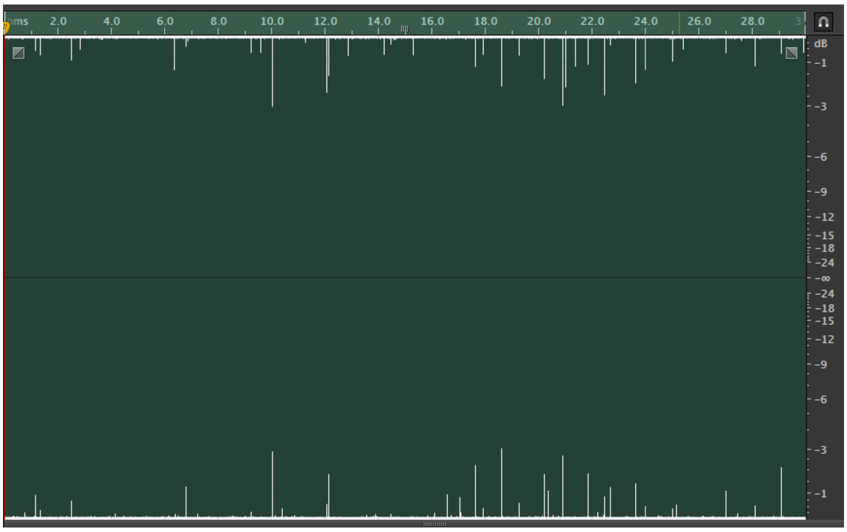


Figure 1.1 Example of hypercompression where the waveform has lost any natural peaks and decay to produce a flat and squashed waveform

The era of hypercompression lead to changes in playback systems especially as the era of streaming services and Internet radio became popular. Hypercompression took advantage of playback systems using peak normalisation where the audio played back was all referenced to 0 dBFS whereas Internet based services have loudness normalisation. This works on using an objective measure of loudness to assign a level for each track. Then when the tracks are played back to back using the Internet based services, they all sound the same loudness however the tracks that have been more compressed have less dynamics and so have less quality. Wendl & Lee [2014] showed this relationship between compression and quality with a removed loudness change when comparing the same track compressed to varying levels at a static playback level.

Objective measures for loudness have also been implemented in television broadcasting. The ITU-R 1770-3 (2012) created a new procedure for calculating perceived loudness and set a guideline level to reduce the loudness variations between channels and programs. The same procedure has also been demonstrated to be beneficial to monitoring inherent loudness changes within popular productions. Genres such as pop music or rock could benefit by using objective loudness measurements to ensure that “hypercompression” is avoided.

1.2 Motivation

The recent advancement into loudness normalisation using objective loudness calculations has lead to a new way that certain playback systems represent audio. Playback systems that use loudness normalisation allow for the avoidance of “hypercompression”. Objective measures of loudness remove the ear of the engineer and so the estimations of loudness have presented a need to further examine the processes applied during mastering and the perceived changes. Crogan *et al* [2012] conducted a series of tests that found that high levels of compression were either degrading to music quality or had little advantage suggesting that DRC can reduce the quality of audio.

Hjortkjaer & Walther-Hansen [2014] however questioned this finding by conducting a series of tests that compared original popular music releases with the their remastered and more compressed counterparts and found no significant preference was determined. The lack of a significant preference may suggest that no quality judgements could be made. This lack of quality assessment noted that a remaster will accompany compression with additional processing and concludes that further research is essential.

To understand how DRC effects loudness perception and quality perception, a study into the relationship of bands of frequencies and the effect of DRC would indicate if certain frequency bands have a greater effect on loudness perception and would further the knowledge about complex interactions between DRC and potential quality assessments. As suggested by Wendl & Lee [2014] loudness perception and quality perception may be correlated dependant on genre or on certain passages of audio with amplitude characteristics. With this ambiguity around the impact of genre or type of audio, an investigation into broad fundamental amplitude characteristics may further the understanding of how DRC effects audio in relation to loudness perception and therefore potential changes in quality perception.

1.3 Research Aims and Objectives

As can be seen in the previous discussion the impact that the mastering stage has on audio is apparent. Complex audio used in experiments has yielded mixed results and differing conclusions and so this has lead to a study into how fundamental noises and amplitude characteristics behave as a result of DRC. Another reason is to give an indication of the effect of DRC irrespective of genre and mixing preference. If a track has a heavy bass representation, for example, then the mastering considerations in regards to DRC could be taken from exploring how lower frequency bands behave under DRC. The findings of Fletcher & Munson [1933] suggested that frequencies are not perceived in an equal level and so this would suggest that differing bands of noise will be perceived differently in regards to the perceived loudness as a result of DRC. Within complex audio found in popular music productions there are usually a combination of continuous amplitude characteristic sounds and transient amplitude characteristic sounds. Though complex audio has an almost infinite variety of amplitude characteristics, a study that explores the difference in perceived loudness for continuous versus transient could explore the way that DRC effects both separately and give further lead into looking at how both are affected when combined. With all these considerations the aim of this research was to answer the following questions:

- Does each octave band have a change in perceived loudness when crest factor changes as a result of DRC?
- If so, at what point for each octave, in terms of crest factor, is a change in loudness perceived?

- Does each octave band have the same loudness perception changes when crest factor is changed?
- If not, which octave band is most sensitive, if any, to crest factor changes in regards to loudness perception changes?
- Are continuous amplitude characteristics and transient amplitude characteristics perceived the same in regards to changes in loudness perception as a result of DRC?
- If not, how do the different amplitude characteristics differ in regards to loudness perception as a result of DRC?

1.4 General Overview of Methodology

In order to answer the questions that inform this research, a careful methodology that isolated the variables was used. A detailed methodology will be given in corresponding chapters however a brief overview will outline further the aims of the research.

The research was driven at identifying how DRC affects the loudness perception that could in turn affect the quality perception and so using a method of monitoring the use of DRC became paramount. Several independant variables that could change how the perception of DRC affected audio were considered also to allow for a model for practical application. These variables included playback level, frequency bands and amplitude characteristics. The stimulus was created using each of these isolated variables and then had DRC applied to create a set of variable appropriate stimulus.

Through considering several listening tests subjects changed the playback level of differing stimulus to create a set of quantitative data that was analysed to support the understanding of the effects involved.

1.5 Structure of Thesis

Chapter two conducts a review into loudness perception in relation to many different factors such as the auditory system, equal loudness contours and temporal aspects. Chapter three focuses on dynamic range compression and how it relates to loudness perception and also past work on quality perception. Chapter four outlines the listening test designs using examples focused on in the previous two chapters. Chapter five is the actual listening test methodology for the equal loudness tests for octaves including a results and discussion section. Chapter six is the section which underlines the listening tests for the perceived changes in loudness in relation to changes in crest factor, explaining how crest factor changes with dynamic range compression. Chapter seven is the summary and conclusion where the findings are stated along with limitations and future work.

2 LOUDNESS

2.1 Introduction

Perceived loudness has many contributing factors. The complexity of understanding loudness lies in many forms of audio whether pure tone, complex tones, broadband noise, complex noise or combinations of all.

The following section will review the background of loudness perception exploring the basic functions of the ear including the response to differing levels, functions of equal loudness for pure tones and functions of masking for both pure tone and complex tone. Also previous methodologies for testing perceived loudness will be examined by reviewing the process of acquiring equal loudness for pure tone and complex tone whilst also focusing on octave band equal loudness. Finally a consideration into the temporal aspects of loudness perception will be reviewed to assess the impacts that changing amplitude characteristics has on loudness perception.

2.2 Background

2.2.1 The Ear

The function of the ear converts acoustical energy to mechanical energy [Pohlmann, 2005]. As sound propagates, via fluctuations of the atmospheric pressure, moving as a wave, the three sections of the human ear allow for the fluctuations to be converted into information that can be interpreted by the brain. The outer ear generally helps the location of sound and amplify certain frequencies whereas the middle ear transports the acoustical energy to the inner ear whilst allowing for the reaction to loud sounds over 75 dB sound pressure level (SPL) by stiffening the ossicular chain to protect the hearing system [Howard & Angus, 2009]. The inner ear consists of the cochlea that transduces the transported fluctuations into a neural code along the auditory nerve [Yates, 1995].

2.2.1.1 Inherent Auditory Filters

As sound resonates across the basilar membrane within the cochlea, different hair cells respond and generate a response in the fibres of the auditory nerve [Palmer, 1995]. This means that the response of each fibre is tuned to a certain frequency or at least has a greater sensitivity to a certain frequency and this is named the characteristic frequency. A representation of the filters that each fibre has is demonstrated in **figure 2.1**.

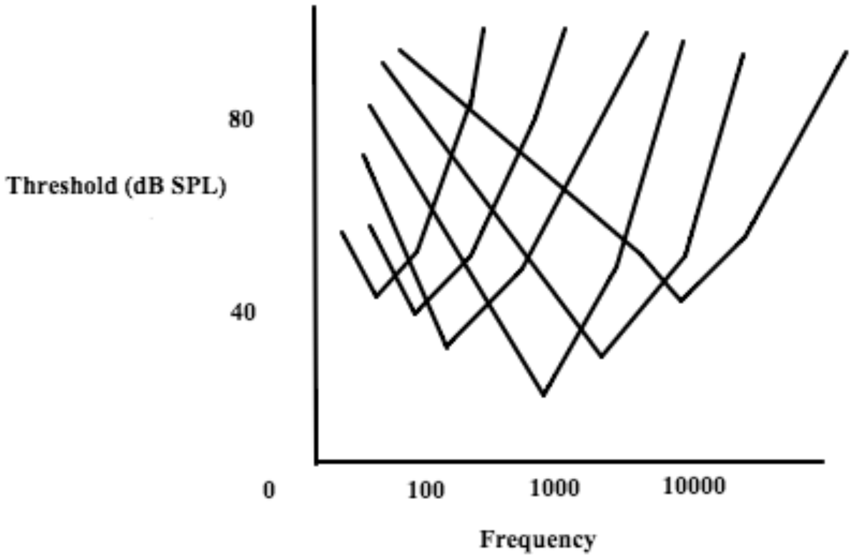


Figure 2.1 Simplified account of the tuning curves within the human auditory system based upon Doucet & Relkin [1997]

At lower sound pressure levels there are fewer fibres that respond and depending on the frequency content of the sound there may be only certain fibres that respond at all. The filters overlap and so this suggests that at times several fibres can respond. Palmer [1995] used the example of a 1 kilohertz (kHz) pure tone at different SPLs. When the pure tone is of a high SPL, this falls into many of the filters and level required for many fibres and so there is an increased activity across the auditory nerve.

The overlap of the filters can also create the phenomenon where certain frequencies are not heard because the amplitude of a pure tone or noise is too great to allow for the fibres to respond to the pure tone or noise of lesser amplitude. The relationships between the distance of the pure tones or noises and how they differ for different frequencies are explained as part of the critical band theory that is covered further in the section 2.3.1.

2.2.1.2 Auditory Nerve Fire Rate

The fire rate of the auditory nerve has been theorised as a contributor to loudness as early as 1933 by Fletcher and Munson [1933] and conclusions made by Lachs *et al* [1984] suggest that the fibre count mean relates to the sound intensity or loudness. The rationality behind this follows from the aforementioned filter crossovers and fibre response in the way that as the signal increases in intensity, the response from the ear increases and the interpretation of this increased activity is that the loudness perception increases. This may not however be entirely accurate as proposed by Relkin & Doucet [1997]. They suggested that the data found from auditory spikes (the sum of auditory nerve neuron fire) did not match loudness perception data especially when comparing the spike rate at lower frequencies and the lower spike rate of higher frequencies. Relkin and Doucet [1997] theorise that if the spike count was linearly related then a sound of higher frequency would sound quiet when produced alongside a sound of lower frequency and so the central auditory system has an impact on the loudness perception by implementing an equalisation. Though they offer this theory they acknowledge that the limitations are routed in a lack of knowledge of the central auditory system, which is defined as the part of the auditory chain that is not a part of the peripheral auditory system. Their work offers no explanation as to what this equalisation is or when it is applied and therefore can only highlight the need for further research in the area.

Moore *et al* [1997] also worked upon the principles of the auditory nerve fire rate by encompassing it within their work with excitation patterns. Simpson *et al* [2013] reviewed the work of Moore *et al* [1997] and explain the excitation model as being the neural representation of the basilar membrane's resonance. The excitation model however does not work upon the fire rate theory solely and works upon the summation of the excitation across the spectrum. If two frequencies play at the same time then the excitation is summed and therefore the loudness has increased. There are exceptions to this, however, when frequencies are too close and this is the basis of frequency masking that will be reviewed further on in this literature review.

2.2.2 Equal Loudness Contours for Pure Tones

Equal loudness contours give an indication into the non-linearity of the human ear in regards to frequency perception and demonstrate the changes at different loudness levels. There have been several key studies. A review of the methodologies will precursor the results and a summary of the implications will follow in this section.

2.2.2.1 Review of Methodology

Fletcher and Munson

A key early experiment that highlighted the non-linearity of the human ear was conducted by Fletcher and Munson [1933]. To accurately track the differences in frequency perception, a 1 kHz pure tone was used as a reference tone against other pure tones of predetermined frequencies. Fletcher and Munson gave reason to the use of the 1 kHz pure tone as the reference for five different reasons which are as follows:

1. *It is simple to define*
2. *It is sometimes used as a standard of reference for pitch*
3. *Its use makes the mathematical formulae more simple*

4. *Its range of auditory intensities (from the threshold of hearing to the threshold of feeling) is as large and usually larger than for any other type of sound*
5. *Its frequency is in the mid range of audible frequencies*

Normal hearing subjects compared the reference tone to other predetermined pure tones. Fletcher and Munson used normal hearing subjects to predict the typical human ear response. An interesting method for acquiring the differences in level was used where the reference tone was changed to match the loudness of different pure tones at predetermined different sound intensity levels. The tones were played from a monaural position in front of the subject so as to allow for the use of both ears.

Robinson and Dadson

Robinson and Dadson [1956] further explored the equal loudness to identify a better representation of low frequency equal loudness and to document a more methodical and sound research practice. To manage this, the playback system was adapted to create a better representation of lower frequencies so that the equal loudness contours remained fundamentally accurate and not biased by the system.

The use of SPL as the intensity reference was a noticeable difference and advancement. Because SPL is a measure of sound pressure with a reference to the threshold of human hearing, Robinson and Dadson helped to define the equal loudness contours in relation to the lowest point that a human can hear meaning that a fundamental foundation had been established.

Again a 1 kHz pure tone was the reference tone and again the subjects changed the reference tone to match the stimulus material of predetermined SPL level. A documented difference from the Fletcher and Munson [1933] tests can be found in the considerations into the time between tones and the onset and offset of tones. A time between 1 second and 3 seconds was used for the silence and the onset and offset times were 150 milliseconds (ms). Times between tones and onset and offset will be discussed later in the literature review and the importance that they hold in regards to methodologies.

BS ISO 226:2003

The equal loudness contours found by both Fletcher and Munson [1933] and by Robinson and Dadson [1956] gave great fundamental and methodological work regarding the non linearity of the human ear in regards to frequency and gave birth to many studies in the area. These many studies often focused upon different frequency ranges or subject age but often held the methodology in a similar format. The International Organization of Standards (ISO) [2003] reviewed the many pieces of literature to create a summarised and even more accurate account of Equal Loudness Contours.

2.2.2.2 Explanation of the Equal Loudness Contours for Pure Tones

Figure 2.2 shows the concluding equal loudness contours.

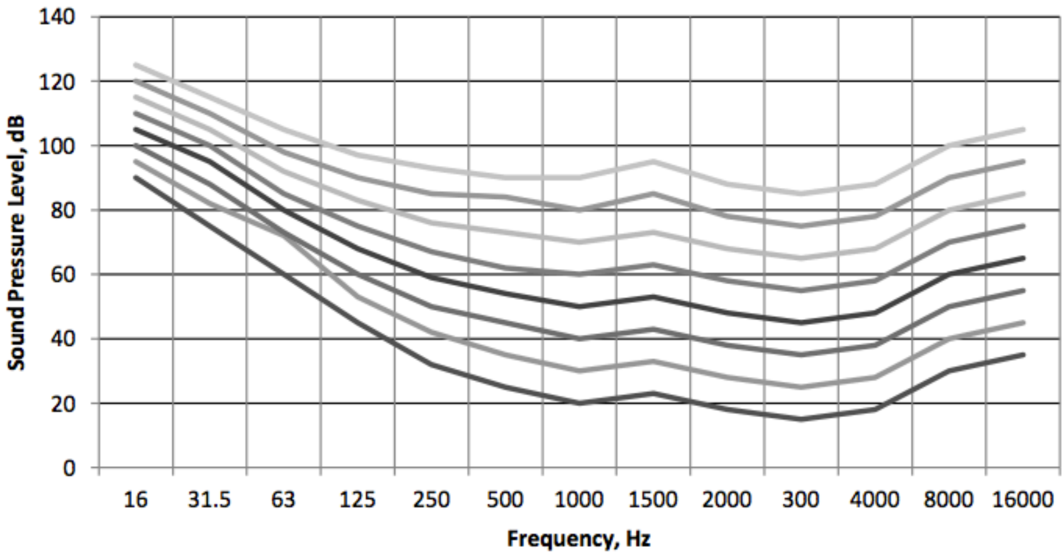


Figure 2.2 Equal Loudness Contours based upon BS ISO 226:2003 [2003]

As can be seen within **figure 2.2** the non-linearity of the human ear are apparent. By first looking at the equal loudness contour based upon a 20 dB SPL 1 kHz pure tone it can be clearly seen that the human ear has sensitivity within the 1 kHz to 5 kHz frequency range. The contours deviation below the 20 dB SPL marker shows that a frequency around 3 kHz to 4 kHz is the frequency that our auditory system responds to the most. During a study by Hammershøi & Møller [1996] into the impact of the propagation of the ear canal on localisation, microphones were inserted into the ear canal and took measurements on the natural amplification within the canal. By examining the results the ear canal naturally amplifies, via resonance, frequencies between 1 kHz to 5 kHz with a peak around 3 kHz. This suggests that our sensitivity to this region is due to the ear canal and not any kind of auditory filter or auditory equalisation. This sensitivity is comparable to the lack of sensitivity in the lower frequencies.

The advantage of the equal loudness contours is that they gave a representation not only for frequency selectivity but also for different loudness level and how the ear changes upon differing loudness levels. As can be seen within **figure 2.2**, at higher SPL levels of the reference tone there were less differences between perceived loudness for different frequencies. Comparing the curve of the 20 dB SPL line and the 100 dB SPL line demonstrates this. An explanation can given by considering the auditory systems response in regards to inherent filters as detailed before. As the SPL increased, more fibres became active as the level pushed the frequency over more of the filters' response threshold.

2.2.2.3 Implications of Equal Loudness Contours

This relationship between SPL level and equal loudness contours are one of the suggested reasons that the "louder is better trend" began. Vickers [2010a] noted that boosting the loudness of the audio results in a more linear response and this could be interpreted as a better defined balance.

2.2.3 Phon and Sone

The phon is a measure of loudness of pure tones in relation to a pure tone of 1 kHz. This follows on from the equal loudness contours for the contours show the phon scale. To match any of the tones to the reference level, the amount of required SPL is shown by the contours.

The sone is a measurement used to gauge loudness changes of pure tones in stead of SPL changes. Stevens [1936] named the unit and set the foundation for the sone measurements. A 1 sone measurement is equal to a 40 phon 1 kHz pure tone. The reasoning is given that this the level that pitch judgments can be executed. As the SPL or phon (can be interchanged for a 1 kHz pure tone) level changes for the pure tone the sone measurement will change proportionally to the loudness. As demonstrated by Stevens [1956] when a phon measurement increased by 10 phons, the sone measurement doubles.

2.2.4 Frequency Weightings

A product of the equal loudness contours were the different weighting that could be applied when working out the loudness of audio. **Figure 2.3** demonstrates the workflow using frequency weightings [Norcross & Soulodre, 2003].

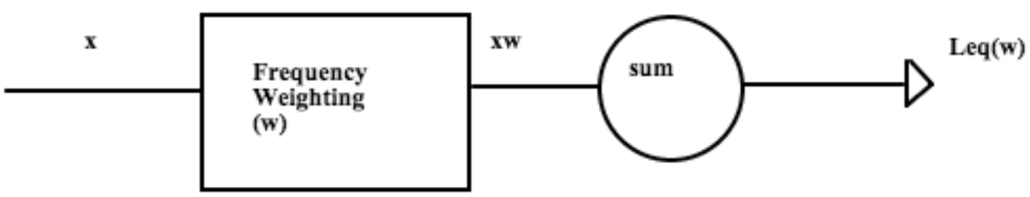


Figure 2.3 A block diagram to display the process of creating afrequency weighting based upon Norcross & Soulodre [2003]

Three well known frequency weightings are A, B and C. Each frequency weighting aligns to a different contour of the equal loudness contours for pure tones. A weighting was a reference to the 40 phon contour, B weighting was a reference to the 70 phon contour and C weighting was a reference to the 100 phon contour [Norcross & Soulodre, 2003]. As can be seen in **figure 2.4** the weightings loosely represent an inverse of the contours.

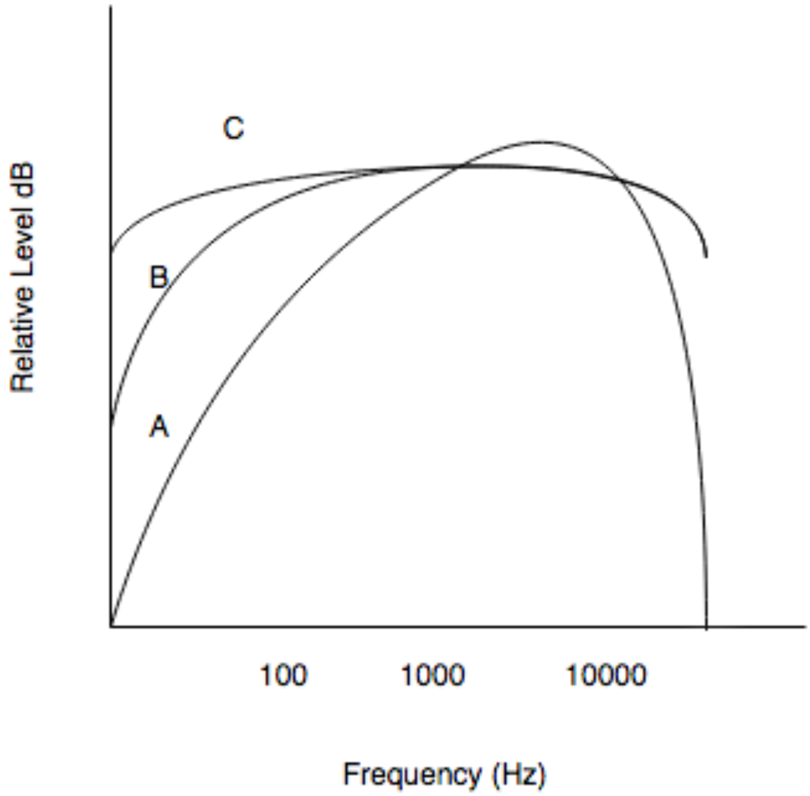


Figure 2.4 The curves showing the A, B and C weighting curves based upon Norcross & Soulodre [2003]

These weightings are useful in determining an appropriate SPL level when in context with the loudness level by applying the filter before the measurement.

2.3 CALCULATION OF LOUDNESS OF COMPLEX SOUNDS

Complex sounds vary in complexity ranging from a sound with more than a single pure tone to a fusion of complex tones where individual components could include a tone with many harmonics or several tones of different amplitudes. When calculating the loudness of such complex sounds there are considerations into how several tones are perceived at the same time or whether certain tones are not heard.

2.3.1 Critical bands

Loudness summation is the overall judgement of perceived loudness as a result of several tones contributing. The summation takes into account the loudness of each tone and how the loudness will increase as a result of adding the loudness perception together. For loudness summation there are however certain aspects that were discovered such as the spacing of the tones and the relation to loudness summation. Zwicker *et al* [1957] demonstrated how for sound with different centre frequencies the distance from the centre frequency of test tones had different distances before contributing to loudness. This is a foundation for frequency masking for the tones had no impact on perceived loudness until outside of a certain distance. Zwicker whilst working with Feldtkeller also tested the bandwidth of white noise and the effect it had on loudness perception by comparing different bandwidths against either a pure tone or noise. Both comparisons yielded the same bandwidth where loudness summation began. In **figure 2.5** this is demonstrated.

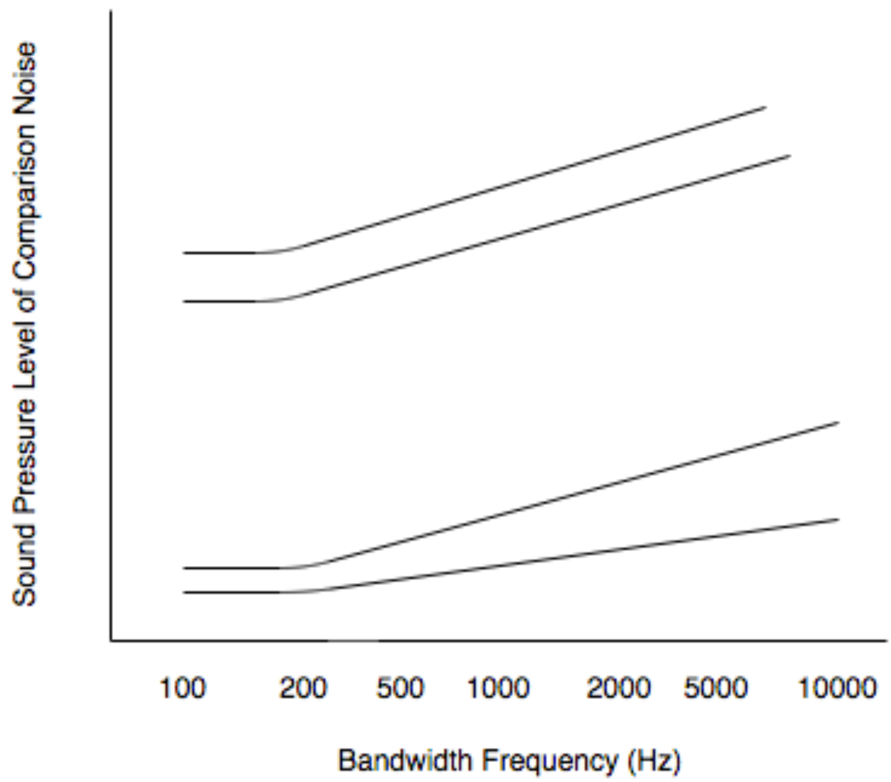


Figure 2.5 Development of loudness as the bandwidth increases of a noise with centre frequency at 1420 Hz when compared to two other noises with bandwidths of 210 Hz and 2300 Hz. Different playback levels were tested also. Based upon Zwicker *et al* [1957]

An interesting detail that can be seen within **figure 2.5** is how regardless of test SPL, the critical band was always the same. This is interesting when considering that different SPL levels have an effect of frequency perception, as show by the equal loudness contours, however for the loudness summation there was no impact. Zwicker [1961] published a list on critical bands in 1961 that demonstrated the change of critical bands across the frequency range.

Following on from this work, Moore & Glasberg [1983] detailed a theory of calculating the auditory filters widths that are used for summation by creating the equal rectangular bandwidth theory (ERB). ERB is the calculation of the upper and lower limits of the bandwidth of the auditory filter. The calculation follows a series of notched noise tests to determine these limits. The formulae to calculate an ERB for a frequency is a follows:

$$ERB = 24.7(4.37F + 1)$$

Where

$ERB\ is\ in\ Hertz$

$F = centre\ frequency\ in\ Hertz$

This calculation differs from the work of Zwicker by offering a better detailed critical band in lower frequencies. In theory what this provides is the ability to see if a noise has tones that are contributing to the loudness where they fall outside of the upper or lower boundaries of the critical band. An example would be to check an octave of noise with centre frequency of 1 kHz. If we apply the formulae the equation looks like so:

$$ERB = 24.7(4.37 * 1 + 1)$$

$$\text{ERB} = 132\ \text{Hertz (hZ)}$$

The upper and lower boundaries for an octave band with centre frequency of 1 kHz are 707 Hz and 1414 Hz. This means that there is a total of 322 Hz that are contributing towards the inherent loudness with the assumption that all tones are equally present. This demonstrates that the octave band would be perceived louder than the pure tone of 1 kHz. This difference in loudness gives a reason to find how different octave bands are perceived both in regards to each other. Also as the octave band increases, the amount the bandwidth that is outside the critical bandwidth increases. These changes are contributing factors into why the prediction of the affect of DRC on octave bands is not a simple prediction.

2.3.2 Equal Loudness of Complex Tones

Listening tests to find equal loudness for octaves were originally by Stevens [1956] to elaborate on the then pioneering work of Fletcher and Munson [1933]. Similar studies have been conducted (Jahn [1974] and Bauer & Torick [1966]) and a review of the methodologies and results will allow for considerations as later discussed in this thesis.

Before the review will take place it may be necessary to answer the question of why use octave bands? Zwicker & Fastl [1972] gave the reasoning that the human auditory filters are lower than an octave band, as discussed previously, in relation to critical bands. Therefore the use of pure tone data may not hold truly accurate for complex tones because complex tones can breach critical bands. Also the use of octaves allow for perceived loudness tests to be related to an almost musical element. Though octaves in regards to frequency and octaves in regards to pitch are different, the changing frequency bandwidth allows for a study into how changing bandwidths up the frequency spectrum change our perception. Multi-band compressors use frequency bandwidths that may not be too dissimilar from bandwidths found in octaves and so the study of these complex tones may offer suggestions in regards to final stage dynamic range compression as will be discussed in later chapters.

2.2.2.1 Review of Methodology

Stevens [1956]

A similar method was used as found in equal loudness determining tests for pure tones. Bands of white noise were filtered to a predetermined bandwidth. The listening test procedure had the subjects match the comparison frequency band to both another frequency band and to a 1 kHz pure tone. The subject wore earphones for this study and the frequency response of the earphones were taken in account.

The filtering of the white noise to obtain octave bands used filters with steep cut offs. The steeper cut offs gave less overlap in different octaves and therefore the results are less likely to have been influenced by outside frequencies. The octave bands however did not use centre frequencies that are often associated with normal octave bands. None of the octave bands had a centre frequency of 1 kHz which is often the basis of calculating different octave bands.

This first study gave a set of equal loudness contours for octave bandwidths however Stevens noted that these equal loudness contours could be biased to the earphones used and so a free field test was designed. The impact that the playback system is of high importance.

Using a reference noise that had a lower frequency of 500 Hz and an upper frequency of 2 kHz, the new listening test had the subject change the octaves to match this reference noise. The reference noise was set to a playback level of 73 phons.

Jahn [1974]

Though published in 1974, Jahn originally completed this study in 1958. Octave bands of non-specific noise were compared to a ⅓ octave bands of 1 kHz centre frequency. The calculations from the phon level to the perceived level of the ⅓ octave band was taken into account so the reference levels were set to 40 phons and 55 phons. The playback system was a free-field system like the previous mentioned method.

The octave band comparison tones were altered in incrementing steps and the subject controlled a dampening key to determine if the comparison was too loud or too soft. A problem with this methodology is that it does not give an accurate account of the loudness for the comparison tone may still be not equally represented in terms of loudness perception but be deemed as close enough.

The clarification on filters for the octave tones were not given and so this limits the ability to compare results. Definitions of the octave upper and lower ranges are important for, as discussed previously in regards to critical bands, the bandwidth of complex sound can impact the perceived loudness as a result of excitation patterns.

Bauer and Torick [1966]

Octave bands of pink noise were used to determine the loudness level judgements of complex sound. The use of pink noise is novel in this study and was justified as being distributed better across the spectrum for the following reason:

> White noise has equal average energy per cycle, whereas the average energy per cycle of pink noise is inversely proportional to frequency. Thus pink noise has equal energy distribution per octave band. [Bauer & Torick, 1966, p. 145]

The pink noise was filtered into nine different octave bands though the filter method was not documented. The octave bands were listed as ranging from 32 Hz to 16 kHz though the exact upper boundaries; lower boundaries and centre frequencies were also not documented. The comparison octave bands were compared to a reference ⅓ octave band with centre frequency of 1 kHz as similar to the Jahn [1974] methodology. The playback system was set to allow for multiple subjects to be tested and so a loudspeaker was positioned in the corner of a room with noted acoustical properties. The method of determining equal loudness followed a similar Jahn [1974] methodology again where the subjects had to judge if the comparison noise was louder or softer than the reference noise.

The study used several playback levels to map several equal loudness contours similar to the contours drawn by Fletcher and Munson [1933]. Tests conducted using 50, 60, 70 and 80 phons for the reference level. Due to the slight differences that the ⅓ octave band reference noise has in regards to SPL when compared to a pure tone, the SPL was slightly lower than it would be for a pure tone.

2.3.2.2 Equal Loudness Contours for Complex Noise

Figure 2.6 shows the equal loudness contours as presented by Stevens [1956].

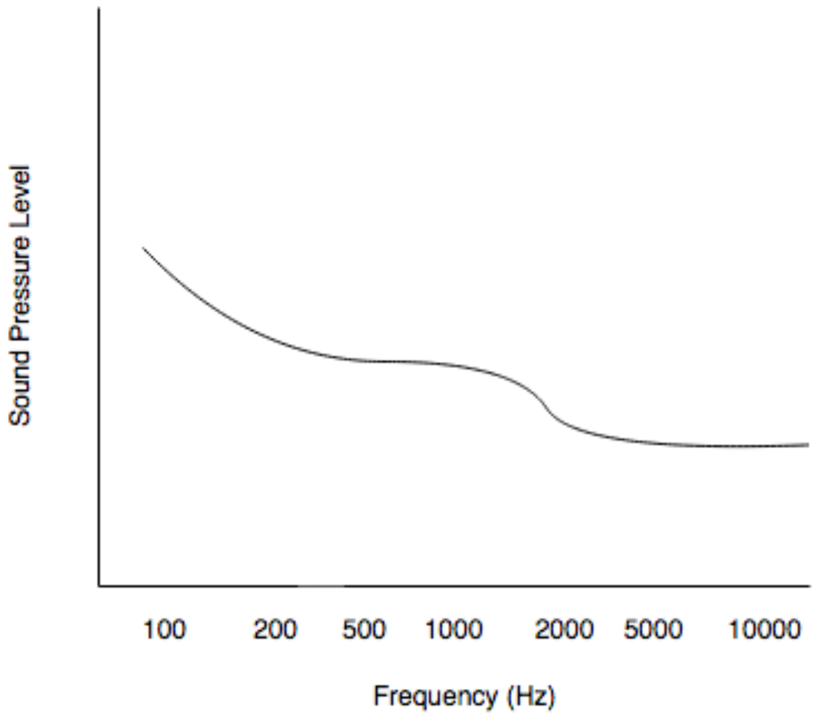


Figure 2.6 The Equal Loudness Contour for complex noise based upon Stevens [1956]

The equal loudness contour for octaves that Stevens provided can be compared to the equal loudness contour for pure tones. Stevens notes that there is an apparent difference in the high frequency range. Where the equal loudness contours for pure tones raise after 5 kHz meaning that sensitivity decreases, for octaves the contour showed a plateau suggesting that sensitivity for higher frequencies remain equally sensitive.

Figure 2.7 demonstrates the equal loudness contour found by Jahn [1974].

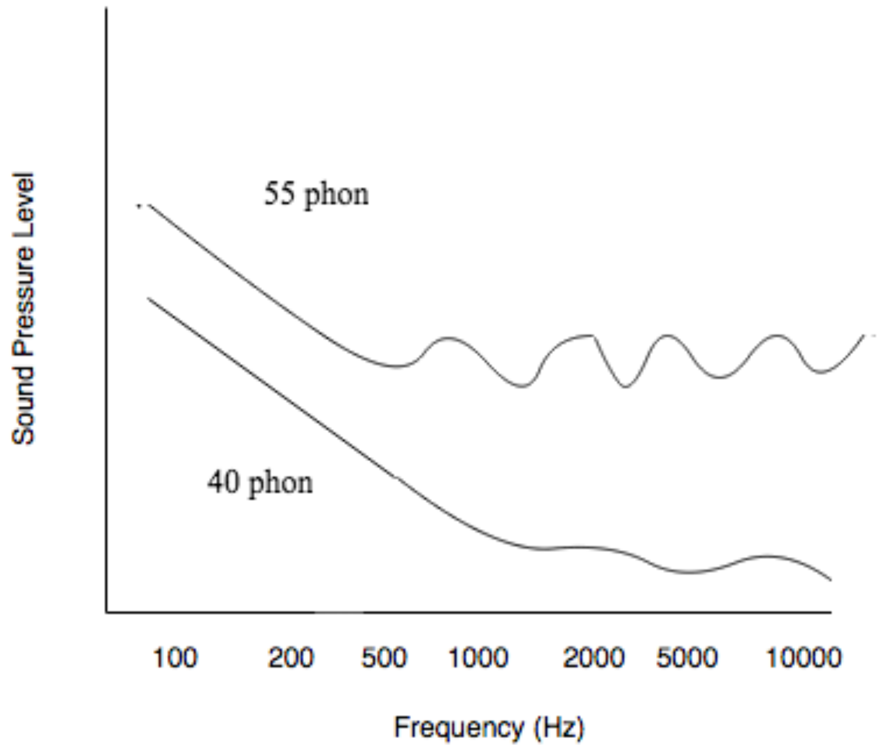


Figure 2.7 The equal loudness contour for octaves at 55 phon and for octaves at 40 phon when compared to 1/3rd octave band with centre frequency of 1 kHz based upon Jahn [1974]

The differences found as a result of the Jahn listening tests suggest a strange variation between sensitivities of octaves as demonstrated by the 55 phon contour. Upwards from octaves with centre frequencies of 800 Hz the octaves show that the sensitivity alternates between increasing and decreasing. This could be contribution of the methodology where the comparison noise was determined as only louder or quieter than the reference noise and so the comparison noise may never precisely match the reference noise. This ambiguity in the method may cause the alternating increasing and decreasing.

At 40 phons the trend differs slightly but shows that the most sensitive octave is that which has a centre frequency somewhere between 6.4 kHz and 12.8 kHz. This sensitivity contradicts findings made from pure tones as argued earlier in this thesis (Fletcher & Munson [1933], Robinson & Dadson [1956] and BS ISO 226:2003 [2003]).

Jahn [1974] noted that the difference between contours mapped from this study and those found in the aforementioned Stevens [1956] study suggest that at the time of the published document no correct contour could be decided.

Figure 2.8 shows the equal loudness contours for octaves as presented by Bauer & Torick [1966].

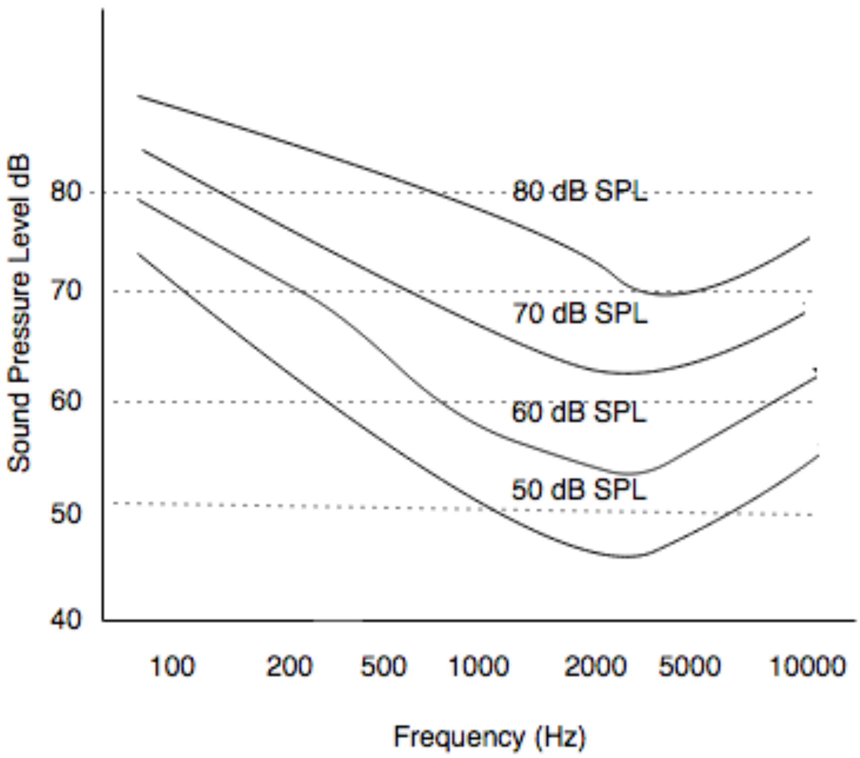


Figure 2.8 Equal Loudness Contours found for octave bands based upon Bauert & Torick [1966]

As can be seen in **figure 2.8** the similarities to the equal loudness contours of pure tones are a lot more present. Both sets of contours exhibit the sensitivity region in roughly the same region. There are differences found between the different phon contours, however, such as the highest frequency octave band, when compared to the loudest reference noise (80 phons), is perceived as louder. Comparing this to the 50 phon contour, the highest frequency octave band is perceived as quieter and by quite a large degree, almost six phons quieter. This change between perceived loudness, as a result of playback level, is not contributed by critical bands. Zwicker *et al* [1957] demonstrated, as discussed earlier in the thesis, how for varying SPL levels the critical band does not change. This relationship is not found in equal loudness contours for pure tones. Stevens [1961] made an observation, when working with airplane spectra, that high frequency content can be deemed as annoying and interfere with perceived loudness judgements. At low to moderate playback levels this annoyance factor may be lower due to naturally perceiving the higher frequencies as lower, however, when the playback level reaches a high level these high frequencies are perceived more and so the judgements become confused with annoyance.

2.3.3 Temporal Aspects

Loudness perception is not instantaneous. The aspect of the length of the noise or tone has an impact on loudness perception up to a certain time length as suggested by Zwicker [1977]. Zwicker [1977] used tones of varying length and used a masking white noise immediately after the test tones to determine the loudness as a result of the masking tone cutting off the development of loudness. The length of the tone found to be length where loudness no longer develops was 100ms. This suggests that any peak that is shorter than 100ms will have a different loudness level compared to a continuous tone.

Zwislocki [1969] also summarised the relationship between loudness level and sound duration as can be seen in **figure 2.9**.

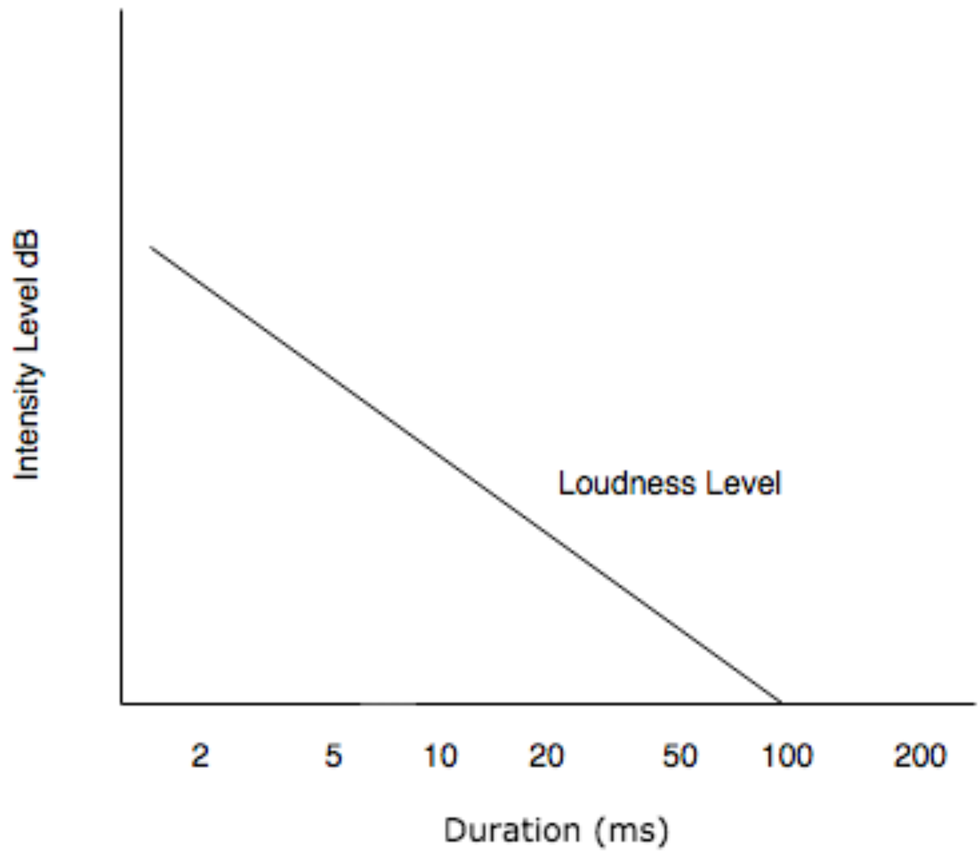


Figure 2.9 Effect of sound duration on the loudness level based upon Zwislocki [1969]

As can be seen in **figure 2.9** the amount of applied level to match a continuous noise declines up to the 100ms region. Though this length is accepted as the temporal integration of loudness, Florentine *et al* [1996] conducted further testing to see if playback level contributed to temporal integration or whether the relationship between short sounds and long sounds were uniform. Using both tones and noises based around 1 kHz, loudness differences were recorded for 5ms, 30ms and 200ms samples ranging in SPL level. It was found that the difference between 5ms and 200ms had changing perceived loudness levels for different SPL levels. The peak difference was around 18-19 dB Sensation Level (SL) for 56 dB SPL for pure tones and 76 dB SPL for noise. This altered from the 10 dB SL at threshold. These results suggest that the playback level changes the response from the auditory system. Another temporal aspect is the masking nature in regards to the time taken to decay a loudness contribution once the noise has ended. Though the decay time can vary depending on signal level, duration of signal and signal frequency, it has been suggested that up to 200ms can be the time that the ear takes to return to a sensitivity where masking no longer has an effect [Baumgarte, 2001].

2.3.4 Objective Measures for Loudness of Complex Noise

As mentioned previously, different weightings can be applied as a filter when determining the loudness of audio. For complex tones this was elaborated when considering that most complex tones are perceived differently to pure tones as also mentioned. A fairly recent advancement in broadcasting standards has lead to programme loudness assessment using a new form of objective measure. A previously used method within the industry used sample peak methods within DAWs as well as sample peak methods such as PPM or Quasi-peak level meters [Lund, 2007]. The newer models used are measured in Loudness Unites to Full Scale (LUFS) and the foundation of this objective loudness measure lies within the frequency weighting filter used as well as several temporal aspects and gates.

2.3.4.1 CBS Loudness Monitor

Using the equal loudness contours as mapped by Baeur & Torick [1966] that can be seen in figure (**FIGURE OF CBS**) and inverting the 70 phon contour, a basis for the filter for an objective loudness measure was developed. Similar to the methods of the previously mentioned frequency weighting for pure tones, this gave a new frequency weighting. Though Bauer *et al* [1967] concluded that the loudness monitor was successful and that they had made progress towards an accurate objective loudness measure, the lack of documented use suggests that it never became an industry standard.

2.3.4.2 ITU-R BS.1770-3

This documentation is the work associated with the LUFS measurements and has become the standard industry measurement for objective loudness and for ensuring that dynamics remain within broadcasting content and not hypercompressed. Hypercompression will be reviewed later on in this thesis.

Norcross and Soulode [2003] developed a precursor to the filter used as part of LUFS calculations. Leq(RLB) is an altered B weighting filter (seen in figure (**INSERT FIG**)) which stands for Revised Low-frequency B-curve for the lower frequency curve lies between a standard B-curve and a standard C-curve and has no upper frequency curve. This filter can be seen in figure (**INSERT FIG**).

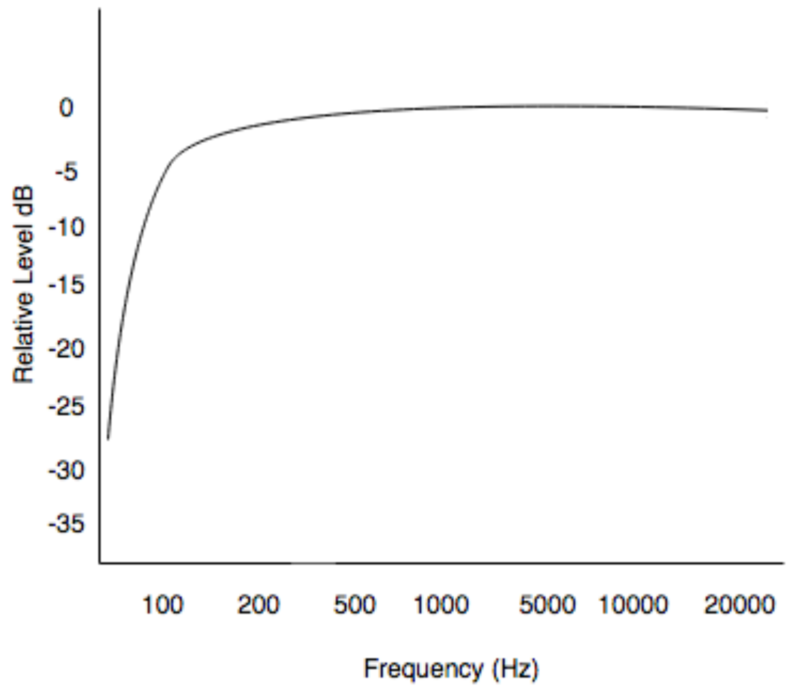


Figure 2.10 The frequency weighting using the RLB curve based upon Norcross and Soulodre [2003]

The application of frequency weighting to determine an objective loudness measurement is often a dB value in relation to Leq. For example a loudness measure using RLB would be

$$L_{eq}(RLB) = x \; dB$$

An Leq measurement is a type of Root Mean Square (RMS) so the objective model of loudness is based upon RMS measurements of the audio. The Leq formulae is as follows:

$$\mathrm{L_{eq}(W)} = 10 \log_{10}\left[\frac{1}{T}\int_0^T \frac{x_W^2}{x_{Ref}^2} dt\right], dB$$

Where

$w = Frequency\ weighting$

$T = time\ internal\ of\ the\ measurement$

$x_w =\ signal\ at\ the\ output$

$x_{ref} = reference\ signal$

12 subjects tested many frequency weighting as part of their study however the RLB curve was found to perform the best when comparing the objective measurements to the subjective measurements.

This RLB curve is similar to the K curve that is used within LUFS calculations as can be seen in **figure 2.11**

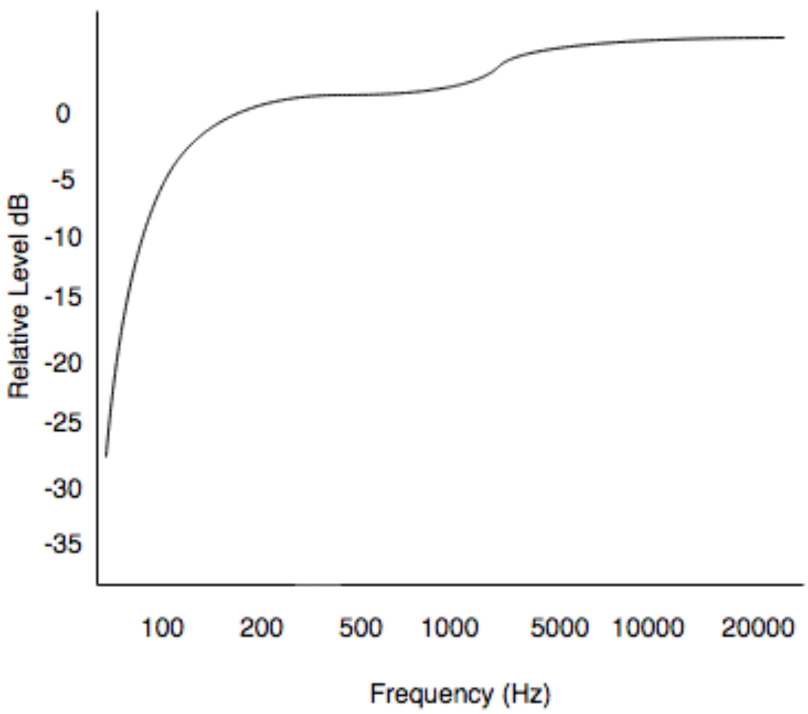


Figure 2.11 The frequency weighting using the K curve based upon International Telecommunications Union [2012]

The K-weighting has differences in the mod to high frequency treatment however which draw similarities to the equal loudness contour as drawn by Stevens [1961] in **figure 2.**6. The mid to high frequency boost in the filter is justified as required due to the acoustic properties on the human head [International Telecommunication Union, 2012]. As previously discussed the sensitive region, as a result of resonance within the ear canal, ranges between 1 kHz and 5 kHz and so this justification, paired with the loudness contour of Stevens, suggest that there is a rational reasoning to the curve. This curve does however differ from the Bauer & Torick [1966] curve and, because the filter is such an important part of the objective measure process, if the filter is wrong it could heavily impact a true objective loudness measure.

The LUFS measurement is intended for loudness normalisation across broadcasters and programs and so it also has three sets of time varying windows where the loudness is calculated and a level threshold gate. The time varying windows change the Leq calculation by changing the T variable. The time windows consist of 400 ms, 3 s and total sample [European Broadcasting Union, 2011]. The three gates are summed together to give a loudness value. For programming, where there is a constant amplitude change of the audio, these three windows make sure that there are no extremely large peaks or sections of extreme quietness and then extreme loudness to create the average of the overall sample. If the source drops 10 dB FS below the target LUFS measurement of the program then it does not contribute to the LUFS calculation. An LUFS calculation is therefore suitable for both steady amplitude noises such as pink noise or white noise and for varying amplitude noises such as regular programming or music.

3 DYNAMIC RANGE COMPRESSION

3.1 INTRODUCTION

Dynamic range compression forms the foundation of modern-day recording, mixing and mastering according to Katz [2007]. Dynamic range compression refers to the resultant effect of a compressor or limiter in this context and is not to be confused with the natural dynamic range compression of the ear also known as the input-output function or to be confused with data compression such as the writing of mp3 files to different file sizes.

3.2 COMPRESSORS

Rumsey & McCormick [2008] summarise a compressor as a device that can change the output level at an altered rate to the input level. They go on to list that compressors have three main variables that are threshold, attack and release. Katz [2007] however suggests that there are three more variables which are ratio, knee and make-up gain.

3.2.1 Threshold

The threshold is the designated level at which the compressor begins to work [Katz, 2007]. If a digital compressor within a DAW has a threshold of -14 dBFS then when the signal level rises above -14 dBFS, the signal becomes compressed.

3.2.2 Attack

Attack is the time for a compressor to begin compressing once over the threshold [Rumsey & McCormick, 2008]. Often measured in milliseconds, the varying attack time can change the manipulation of audio and the attack time is often varied depending on the amplitude characteristic of the audio and desired outcome.

3.2.3 Release

Release is the time taken for the compressor to stop working once the signal level drops back below the threshold [Katz. 2007]. Release and attack times, when considered together, can change the function of the compressor. If a fast attack and medium release is set then only short duration peaks will be compressed, depending on the threshold [Rumsey & McCormick, 2008]. Other uses can be a slow release during final stage compression to compress the dynamics of a track into a narrower window.

3.2.4 Ratio

Ratio is the compression rate once above the threshold. If a ratio of 3:1 is set then if the signal level moves 6 dBFS above the -14 dBFS threshold, the audio would be compressed and output at a signal level only 2 dBFS over the threshold. The ratio means that for every 3 dBFS increase on the input, a 1 dBFS output is yielded [Katz, 2007].

3.2.5 Knee

The knee is the harshness of the compressor when approaching the threshold [Katz, 2007]. If a hard knee is set then when the signal level reaches the threshold level, the compressor will act only as soon as the threshold is met. A soft knee will allow a gentle transition into the compression so the compressor starts just before the threshold with a low ratio and then the ratio increases to the desired ratio and is met just after the threshold.

3.2.6 Make-Up Gain

Make-up gain is the level of which the signal is amplified after the compression [Katz, 2007]. Without make-up gain compressors would only make louder segments quieter.

A DRC can also function as a multi-band compressor which is a compressor that frequency filters the audio into different bandwidths before applying separate levels of DRC. A multi-band compressor according to Vickers [2010b] is more powerful and versatile due to compressing peaks within one frequency band whilst leaving another band untouched. The use of multi-band compressors were also noted as a contributor of the loudness war because of misusing this ability to heavily compress each frequency band at varying levels.

3.3 Limiters

A limiter is a dynamic range compressor that compresses audio to a very compressed state above the threshold via high ratio and fast attack times [Rumsey & McCormick, 2008]. The function of the limiter became useful at preventing signal clipping however when DAWs and digital production became the norm, limiters became more important. Due to playback systems using peak normalisation, digital limiters were able to set the threshold as -0.1 dBFS so to ensure no digital clipping occurred. This capability was as a result of digital limiters able to use extremely fast attack times and even a look ahead function where the audio is buffered slightly in the DAW and uses this buffer time to activate the limiter. This type of limiter is a brickwall limiter. Once the realisation that digital distortion could be avoided it was not too long before compression levels within audio became excessive.

3.4 Hypercompression

As shown within the first chapter in **figure 1.1**, hypercompression loses the natural relationship between peaks and decay found within audio. This visual representation of hypercompression suggest that the audio has lost a lot of the dynamics of the track and Owsinski [2008] suggests that this makes the audio less exciting. Using the brickwall limiter and applying a lot of make up gain achieves hypercompression.

3.5 Dynamic Range Compression and Perceived Loudness

Due to DRC changing the input to output ratio of audio, the inherent loudness of the audio also changes. Katz [2007] divided the application of DRC into two categories; microdynamics and macrodynamics. The microdynamics refer to the relationship between peaks and decay within audio and macrodynamics refers to the different loudness within sections of audio such as a loud chorus and a quiet verse. When considering DRC and the effect on loudness perception it is important to clarify which type of dynamic alteration is being examined. Hypercompression can affect both. It loses the depth of individual passages by ruining the peaks and decay but also ruins the differences between the loudness passages.

By changing the variables of a DRC, the resultant sound can change. Many of the classic analogue DRC hardware units gave characteristics to the audio and many different engineers and producers have favourite units for certain genres because of the tonality. Even though the tonality can differ, the variable change across different DRC can result in similar produced audio. Changing the attack and the release time of the DRC can change the loudness of the audio. Cassidy [2004] suggest that the times should represent the temporal aspects of the human ear as to best perform. This would suggest an attack time over 100ms and a release time of around 200ms depending on the frequency. These attack times would however miss most transient peaks especially for a limiter. Katz [2007] suggested that distortion from signal clipping below 6ms cannot be detected so if a limiter had an attack time of 100ms, this would allow for audible distortion. As previously stated the function of the limiter is to act extremely fast so this would not happen but the theory holds. Depending on the amplitude characteristic of the audio, differing attack and release times would give a different result.

The overall compression aspect of DRC naturally affects the loudness of the audio as Croghan *et al* [2012] found when testing compression levels and the relationship to quality that will be covered further in this thesis. A by-product of the study was that the loudness changed as the DRC level changed. Wendl & Lee [2014] when conducting a similar study further noted this change in loudness. Where Croghan *et al* [2012] did not have a subject change the loudness, Wendl & Lee [2014] did. Croghan *et al* [2012] used RMS matching to match the loudness of the different subjects whereas Wendl & Lee [2014] presented the subject with a reference uncompressed track and then five randomised compressed tracks of which all had different levels of DRC applied. The rate of perceived loudness change was comparable to RMS change of the audio. Though this may seem that more DRC application results in louder audio, this is based upon the peak normalisation process of comparison between audio with inherently different levels of DRC application. However a study by Moore *et al* [2003] found that even when a rough loudness normalisation process took place by matching the RMS of stimulus of speech with different DRC, the most compressed audio signal was still determined as loudest. This may be a result of changing the inherent frequency balance as previously stated where the more linear frequency relationship has the appearance that the audio is louder due to critical band theory.

3.6 Dynamic Range Compression and Perceived Quality

Quality is a much contested area of audio. The judgement of quality varies from person to person and from genre to genre. What may be considered great production and quality for hardcore punk may not be considered great production and quality for jazz. Gabrielsson *et al* [1990] acknowledged that quality is multi-dimensional when trying to determine sound quality of loudspeakers. During the study many different description factors were used and applied to describe the quality such as sharpness, clearness and whether the sound was full. Though these descriptive factors may seem logical in quality description there is little work to suggest what differing changes in audio would constitute a change in the description.

As highlighted earlier, Croghan *et al* [2012] conducted a study to determine whether DRC effects quality perception. The study used audio from two different genres of classical and rock. Each genre had DRC applied to varying levels determined by differing the threshold settings. This gave a set of stimulus which the subject then cross compared against each other and rated via first selecting a preference when cross compared and then ranked into a magnitude of preference. This resulted in clear indication of which levels of compression were preferred in relation to each other. The study found that high levels of DRC were not preferable when the subjects had been presented with RMS matched stimulus as previously mentioned. A low level of DRC gave no significant preference and so suggests that low levels of DRC may be acceptable. However high levels of DRC are detrimental to quality and would affect the listener.

Hjortkjaer& Walther-Hansen [2014] when comparing original masters of popular music against the remasters of the same pieces of popular music, all in the rock or radio friendly popular music genre, contested this. Each stimuli had the peak to average ratio recorded in order to compare effect of DRC and it was found that all the remastered versions had a lower peak to average ratio and so suggests that more DRC had been applied. All the stimulus was matched via RMS normalisation as similar to the Croghan et al study. The subjects were presented with a pairwise comparison test and were asked to choose a preference out of the original and remastered. The results found that there was no significant difference in the choice of either and so suggests that the audio was not of lower quality even though the peak to average ratio was lower. As part of this study the LUFS measurements of each the original and the remastered versions were given in an attempt to prove that RMS calculations for loudness are not too far removed from current objective loudness calculation models. However the LUFS measurements for some of the stimulus showed that the remasters were quieter even though peak to RMS levels had shown it was louder which suggests that more than just compression had taken place during the remastering process. The authors acknowledge that the remastering process may contain more than just DRC and so to conclude that DRC does not change quality judgements, when the controlled audio involved more than just DRC application, raises questions of the validity of the claim.

When Wendl & Lee [2014] conducted a study to determine perceived quality changes in relation to DRC, certain genres lost perceived quality as the DRC increased. Using genres of rock, electronic and jazz subjects assigned scores of different stimulus with different DRC when presented at equal loudness. The equal loudness was determined by using the average loudness level from the same study using the method mentioned previously. The method used to measure the DRC was not based upon variable settings, like Croghan et al, it was based upon the resultant peak to average ratio of the audio listed as the crest factor, a measurement discussed later in this thesis. It was found that the rock and jazz genres had significantly lower perceived quality scores as a result of increased DRC. By comparing the results of how the peak to average ratio changed the perceived quality with no further processing applied, the findings of Hjortkjaer & Walther-Hansen [2014] can be questioned.

An overall study of DRC generally has too many variables to consider. Wagenaars *et al* [1986] ran a study where classical music had a varying set of attack and release times to determine if the perceived quality changed as a result. The stimulus was compared to an uncompressed track and the subjects were asked to grade the stimulus. It was found that longer attack times reduced the perceived quality when heavy DRC is applied. Considering the industry is focusing on loudness normalisation these findings are not too influential however they can be considered when using DRC.

As already mentioned, a type of DRC is a multi-band compressor. As stated earlier in this thesis, multi-band DRC can allow for individual frequency bands to have different DRC applied and therefore creating hypercompressed productions. Fenton *et al* [2011] used a method of measuring different DRC levels within different frequency bands called Inter Band Relationship (IBR). The frequency bands consisted of three bands; a low pass, a band filter and a high pass. The corner frequency of the low pass was 947 with a 6.5 Q and the high pass had a corner frequency of 3186 with a 6.5 Q.

Using five popular music productions, the IBR was recorded and then subjects graded the five productions from 1 (low) to 10 (high). The results found that the subjects judged the popular music production with a low IBR score, that is a score that represents low variation in dynamic range across the frequency bands, as the production with the lowest quality. The limitations were noted in using popular music productions where the subjects had a preference over genre. Further work into the DRC levels of different frequency bands is suggested.

3.7 Measurements of the Effects of Dynamic Range Compression

When considering the effects of DRC, the relationship between the peaks of the audio and the main body of the audio or the RMS is changed. It is important to understand how these two measurements are calculated.

3.7.1 Peak

The peak refers to the peak amplitude of the audio when measured upon a scale. When a peak is measured upon a dBFS scale the measurement will be in relation to Full Scale which means the maximum digital point.

3.7.2 Root Mean Square

The function of the RMS figure is to work out an average of the waveform without having the negative phase cancel out some of the calculations. If a sine wave was to have an average level taken without using the RMS method, the amplitude would equal zero. The RMS value allows for an accurate realistic representation of the average of the waveform.

3.7.3 Crest Factor

The crest factor is the difference between the peak of the audio and the RMS of the audio [Katz, 2007]. If a sample of audio had a peak of -4 dBFS and an RMS of 16 dBFS then the crest factor would equal 12 dBFS. As stated, a DRC will naturally affect the relationship between the peak and the RMS. A lower crest factor can signify a higher level of DRC application. The equation demonstrates this

$$\text{Crest Factor} = \text{Peak Level (dB)} - RMS\ (dB)$$

Deruty & Tardieu [2014] criticise the use of the crest factor as a method of establishing dynamics within audio. The application for macrodynamics is dismissed for measuring a loud segment against a soft segment within the same calculation will keep a high peak level but lower the RMS and therefore not giving a true average calculation of the dynamics. The loud section may have heavy DRC processing which would have a low crest factor but would be ignored in the overall calculation. Therefore the crest factor may be useful to microdynamics, measuring the peaks and average body of the audio. However using the crest factor solely without knowledge of the DRC application can give results that can be misunderstood. If a woodwind instrument gave a steady note then the crest factor calculation would be low for the peak is not far from the RMS however no DRC was applied. Therefore crest factor can be a useful measure of resultant DRC processing such as the Wendl & Lee [2014] methodology but should be used with caution when documenting the level within audio for uses such as suggestive DRC application.

An alternative measurement for recording levels of DRC within audio, especially hypercompression, is the Deruty & Tardieu [2014] created Peak To RMS Regression Coefficient (PRRC). In theory the peak and RMS should follow not too dissimilar patterns along the time scale is there is no hypercompression for peaks are not squashed. During hypercompression the peaks remain constant, around -0.1 dBFS [Owsinski, 2008] however the RMS will still fluctuate. When the audio is not hypercompression the coefficient shows a value nearer to one however when hypercompression is in effect the coefficient is nearer to zero. When the PRRC method is used to measure over 4500 songs ranging across different time scales back to 1967, it accurately tracks the evolution of the loudness wars with different PRRC averages. Within the same study the LUFS measurements were recorded which also accurately traced the evolution of the loudness wars suggesting that the LUFS measurement holds validity in estimations of loudness and DRC application.

4 LISTENING TEST DESIGN

4.1 Introduction

This chapter explains the process of preparation for the experiments that was used to determine if compressed audio is perceived equally across the frequency range and across differentiating amplitude characteristics. Before the comparison between compressed signals was conducted, a procedure to determine the perceived loudness of differing octaves was conducted. This was done partly to examine whether frequency weighting used within objective loudness measures were accurate and also to give data to remove the different perceived loudness of the reference audio within the compressed audio experiment.

Using similar methods of determining equal loudness contours as Robinson & Dadson [1956] whilst adapting the methodology to incorporate bands of noises such as Stevens [1956] and also Bauer & Torick [1966]. The equal loudness contours found for complex noises so far have all differed in methodology. The use of changing the comparison noise and not the reference noise is a model that is prefered due to the measurements of the SPL level using the actual noise and not an different noise. In previous investigations subjects used incrementing steps alongside a "louder" or "softer" judgement system that never gives an accurate perceived loudness level of the comparison tone. Also previous studies used different frequency cut offs for the frequency bands instead of using octaves that are calculated from a 1kHz anchor centre frequency, as detailed further in this chapter. By using these octave bands with a methodology that incorporates the ability of assigning the accurate perceived loudness level of the modified comparison tone, a more valid equal loudness contour was determined for complex noise.

In regards to determining the perceived loudness of octaves bands with different levels of DRC application, no previous test had been conducted. Earlier work had been conducted by Fenton *et al* [2011] to determine if high levels of DRC on broader frequency bands affected the perceived quality of popular music production however the experiment as documented within this thesis investigates if the different DRC levels are perceived equally. If DRC was perceived differently across different frequency bands then the approach to multi-band compression and limiting may change with suggestive conclusions from this thesis.

4.2 Hypothesis

As discussed in section 3.6 the perceived quality of the audio degrades as DRC application increases [Wendl & Lee, 2014]. The previous studies however are for DRC application across the entire frequency range. The concentration on the effects of DRC across certain frequency bands would determine if certain frequency bands could impact the perceived quality more than other frequency bands. However as discussed in section 3.6 the perceived loudness and perceived quality often have a relationship so it is important first to understand the perceived loudness changes of certain frequency bands. Because the limitations often discussed in section 3.6 of conclusions based upon the genre of the audio, a study of fundamental noise would allow for a guideline before a further analysis of the different genre impact could proceed. Therefore this study as previously mentioned aims to discover the differences that DRC application has on different frequency bands.

4.3 Experimental Design

The methodology for the listening tests to produce equal loudness contours for octave bandwidths were inspired by the pairwise tests used to develop the equal loudness contours for both pure tones and noise. The pairwise had a slight variation where multiple comparison octave bands were presented to be compared against the reference. This method allowed for a cross comparison to both the reference and other comparison octaves to give a stronger equal loudness judgement. The methodology to determine if DRC effects perceived loudness for different octaves was inspired by the Multi Stimulus test with Hidden Reference and Anchor (MUSHRA) method which was developed to determine subjective assessment of intermediate audio quality [International Telecommunication Union, 2014a]. The MUSHRA would determine both if DRC could be detected honestly and that subjects did not make level changes because of the situation of a listening test.

4.3.1 Pink Noise

The noise used for the listening tests was pink noise. Pink noise was used due to the inherent properties of equal energy per octave in contrast to white noise [Bauer & Torick, 1966]. This shift in power across the spectrum replicates a closer distribution of equal loudness per octave that gave the advantage as use of a starting point. A 30 second sample of pink noise was used to allow for many variations of the random flicker nature that is inherent within pink noise.

4.3.2 Stimulus Creation

4.3.2.1 Creating Different Bandwidths

The pink noise was filtered into octaves using Adobe Audition CS6 and the “FFT Filter” plugin. The Upper and lower cut off for the filters are demonstrated in **table 1** as well as the centre frequency of the octave. The upper and lower figures were entered into the “FFT Filter” and then the resulting waveform was bounced down.

Table 1 A table demonstrating the lower cut off, centre frequency and upper cut off of the octave bandwidths used for the listening tests

Lower Cut Off Frequency (Hz)	Centre Frequency (Hz)	Upper Cut Off Frequency (Hz)
44	63	87
88	125	176
177	250	353
354	500	706
707	1000	1413
1414	2000	2827
2828	4000	5656
5657	8000	11313
11314	16000	22627

4.3.2.2 Amplitude Characteristics

Using the ten created filtered samples of pink noise, wave shaping was then applied. Due to the intention of finding how compression affected different amplitude characteristics, transient shaped signals were created.

A snare drum hit was first considered as the model for a transient signal and so the on set attack time and decay time was analysed. As can be seen in **figure 4.1**, the snare hit has an almost instant attack time and then a logarithmic decay.

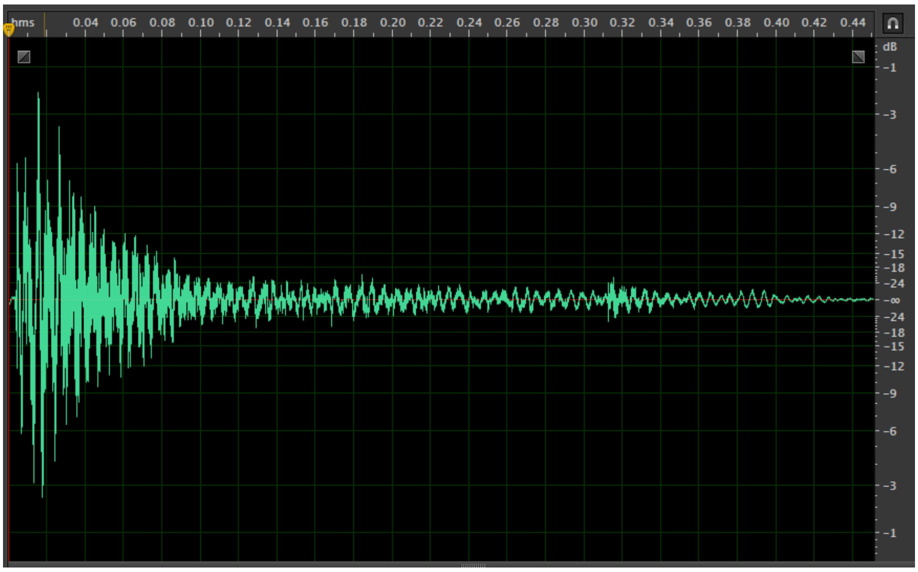


Figure 4.1 The waveform of a snare drum hit demonstrated the fast attack and logarithmic decay

With this considered the different octaves were visually inspected across the waveform using Adobe Audition CS6 for where a peak occurred and then the sample was taken just before the peak. The reason for this was so that when the wave was shaped, the signal had a transient tale. The sample length did not matter at this point as wave shaping would affect the length of the outcome depending on the decided decay time as long as the sample was of a length that allowed for this. Continuing to use Adobe Audition CS6, the sample had a fade in time of 5 ms just before the peak applied so that the sample ramped into the peak. The sample was then left unprocessed for 10 ms to allow for the natural decay after the peak to occur before a 200 ms logarithmic fade out was applied using a 70 dB decay as can be seen in figure

figure 4.2. The difference in perceived loudness between the transient sample and the continuous sample would be present as found by the work of Zwicker [1977].

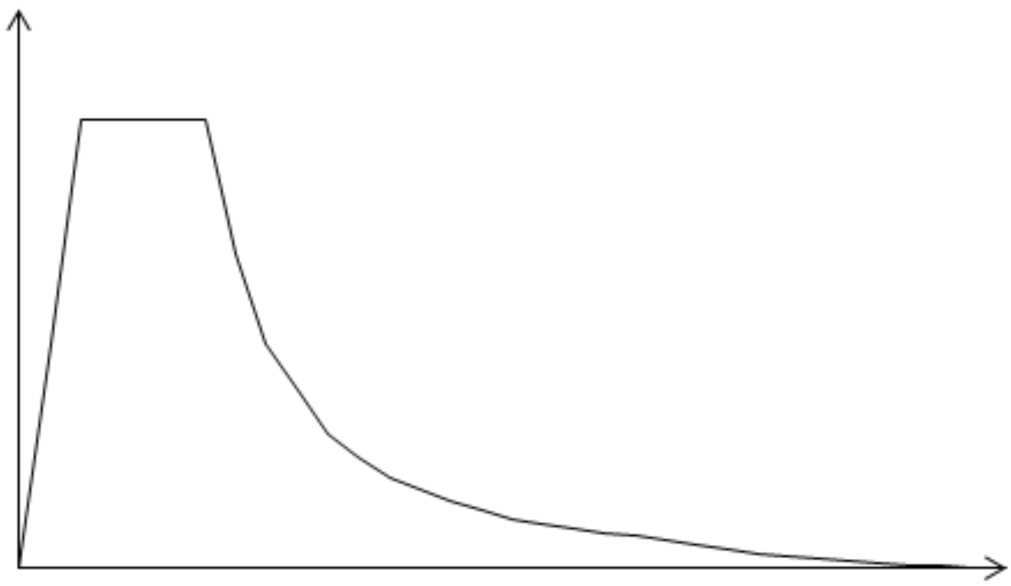


Figure 4.2 Demonstration of the wave shape to create the transient amplitude characteristic

4.3.2.3 Varying Crest Factor Production

As discussed in section 3.7.3 the crest factor can be a measure of microdynamics. Though the use of crest factor measurements can be misleading, when using it as a controlled measurement of DRC application it can be justified. This is especially the case when the focus does not include variable settings of DRC because an experiment like that of Croghan *et al* [2012] could be interpreted as an experiment that is limited to the DRC used.

For each octave band of both amplitude characteristics, a crest factor calculation was taken. This measurement was done by using Adobe Audition CS6 and the amplitude statistics function within the software. The entire sample was analysed and from this the peak level and RMS level were recorded. Using the simple crest factor equation, the crest factor was calculated for each stimulus as can be seen in **Figure A1** to **Figure A4** (see Appendix A). With a starting crest factor calculated for each octave band for both the stimulus of continuous and transient amplitude characteristics, the DRC application could be measured. Instead of creating all equal crest factor values across the octave bands as was done by Wendl & Lee [2014], the crest factor values changed by just 1 dBFS starting at the original crest factor value for the octave band. The decision for this was because the octave bands were already loudness matched using the data from the equal loudness experiment, as will be documented later in this thesis, and so changing the crest factor levels to all be the same value would further amplify octave bands that started with lower crest factor values. This extra amplification would have then altered the equal loudness across the octaves and therefore no comparisons could be made.

The varying crest factor stimulus was created using the Waves L3 Ultramaximizer Peak Limiter plugin within Adobe Audtion CS6. The process follows the flow chart as demonstrated in **figure 4.3**.

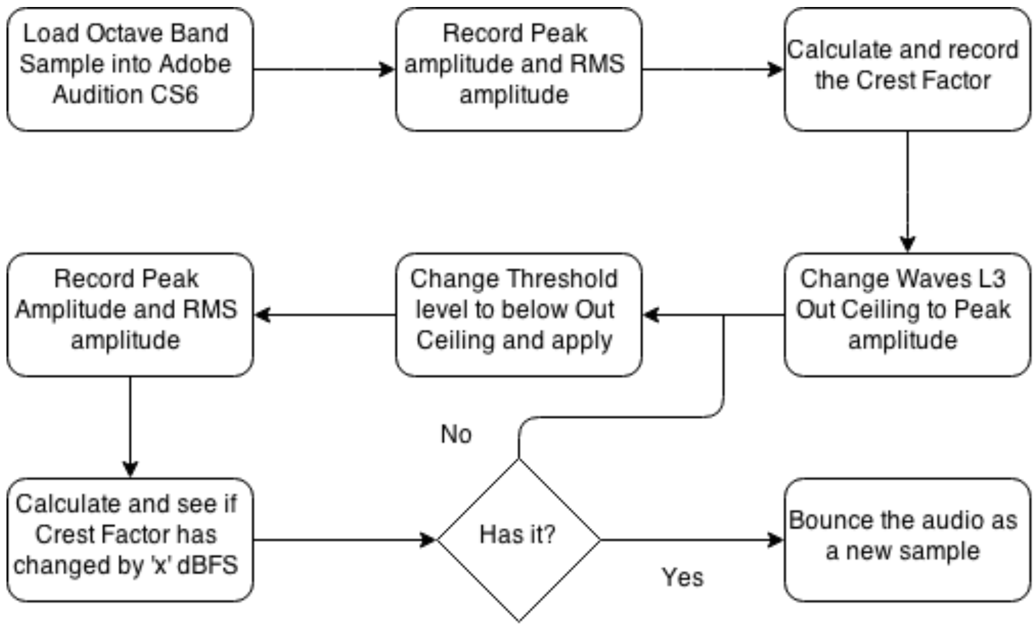


Figure 4.3 Work flow chart on the process of creating stimulus with differing crest factor

As can be seen within **figure 4.3** the peak amplitude was measured for the octave band. Then the Waves L3 Ultramaximizer Peak Limiter had the "Out Ceiling" changed to the nearest figure to one decimal place. This "Out Ceiling" is the dBFS value at which the limiter will not allow for any audio level to pass. This is the brickwall element of the limiter. The release time was set to 0.1 within the plug in across all the stimulus to be consistent for both continuous and transient. The threshold was modified and by cross checking the threshold setting to the resultant crest factor value, each octave band had five crest factor varied stimuli. This change can be denoted as 'n' for the uncompressed stimuli, 'n-1' for the stimuli that has a 1 dBFS change in crest factor, 'n-2' for the stimuli that has a change of 2 dBFS crest factor and so on until 'n-4'. Within **figure 4.3** the dBFS value that the crest factor has changed by is denoted as 'x'. As stated previously, a lower crest factor signifies more DRC.

4.3.3 Reproduction System

4.3.3.1 Room

A semi-anechoic chamber was used to reduce the influence of a room on the listener. Due to the nature of the signals being used during the listening tests, it was suitable to use a monophonic reproduction system as there was no panoramic information. The Monophonic Reproduction system as configured in the ITU-R BS.1116-2 [International Telecommunication Union, 2014b] was used where the loudspeaker was at an angle of 0 ° as perceived from the listener perspective.

Whilst measurements into the recommended distance of the monophonic signal reproduction with a single loudspeaker were considered, the semi-anechoic chamber had limited room and so the distance from the walls to the loudspeaker and listener were prioritised. The resulting room configuration is demonstrated in **figure 4.4** where the distance from the subject to the loudspeaker was measured to be 1.5 metres. The distance from the walls to the loudspeaker and to the listener was measured to be over 1 metre also.

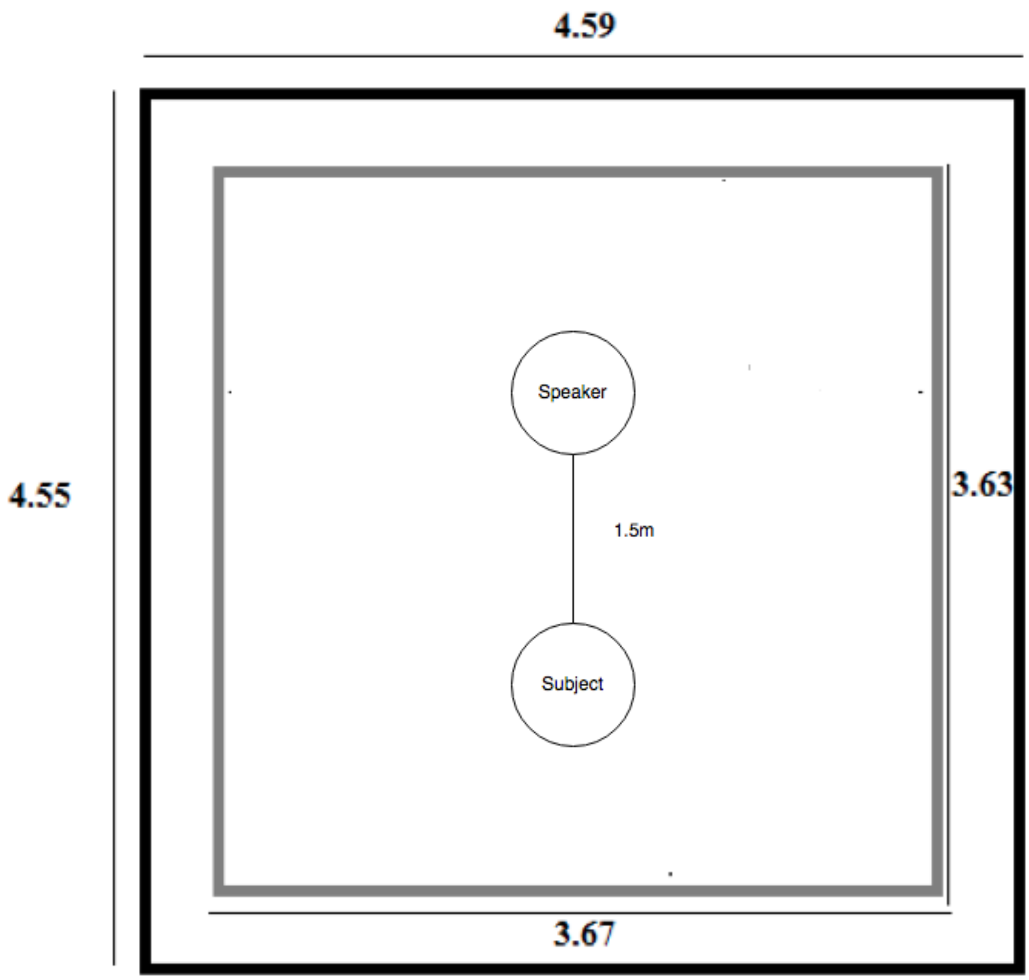


Figure 4.4 Room dimensions and layout for listening tests

The height of the loudspeaker and head of the listener were both measured to be over 1.2 metres as also stated in the ITU-R BS.1116-2 [International Telecommunication Union, 2014b] to be the optimum listening height for a seated listener.

To interface with the listening test a Macbook was used by the listener using the tracking pad. The audio played back via a Motu 4pre sound card connected by Firewire 800. The speaker used was a Genelec 8040A. Whilst this speaker is considered a professional standard speaker, the colouration of the audio was a factor that was considered as had been considered by Robinson and Dadson [1956]. To remove this colouration, a procedure to change the frequency response was performed.

4.3.3.2 Frequency Response Alterations

Using the same distance that the head position of the listener held, an Earthworks QTC40 was placed angled directly at the genelec at the height of 1.2 metres. Using the same Macbook and Motu 4pre configuration as used for the listening tests, a pink noise generator plugin was used through Logic Pro 9 and played back through the Genelec 8040A. The choice of the Earthworks QTC40 was based on the flat frequency response as seen in **figure 4.5** and so added no colouration to signal flow.

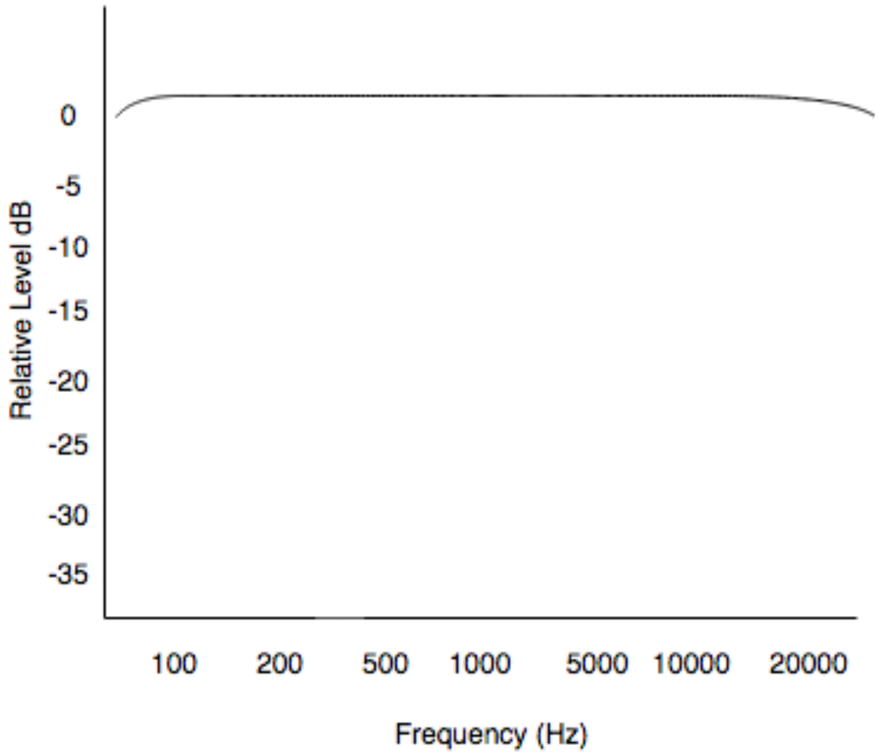


Figure 4.5 Frequency Response of the Earthworks QTC40

This was important because when changing the frequency response of the Genelec 8040A, if the microphone choice added colouration then the procedure of changing the frequency response of the loudspeaker may actually be correcting the frequency response of the microphone.

The frequency response alterations were made with a DBX 1231 Dual 31 band graphic EQ placed in the signal flow between the Motu 4pre soundcard and the Genelec 8040A loudspeaker. The EQ unit started with a flat setting so the frequency response was unchanged. By analysing the pink noise file through the frequency analyser within Adobe Audition CS6 a target was established. The pink noise sample was then played back through the reproduction system and recorded via the microphone and Motu 4Pre into the Macbook. The recorded audio was then analysed with the frequency analyser and compared to the target pink noise audio. The results can be seen in **figure B.1** to **figure B.6** (see Appendix B) and this demonstrates that the Genelec 8040A adds colouration to the audio especially in the 100 Hz and 300 Hz region.

Using the analysed audio as a guide, changes to the DBX 1231 Dual 31 band graphic EQ were made in a controlled process. When a change was made that was deemed to have corrected the discrepancies in the accurate reproduction of pink noise, another recording was made and then processed through the frequency analyser. The stages of change can be seen in **figure B.2** to **figure B.5**.

The final result allowed for the reproduced audio to be the closest representation to pink noise with no colouration possible as seen in **figure B.6.**

This was confirmed by an experienced listener comparing the audio with the frequency response alteration stating that the frequency response altered audio had a better low end and low-mid response which was in accordance with the theory of pink noise.

4.4 Preliminary Listening Tests

A preliminary test was designed to test the created stimulus, the reproduction system and highlight any problems in the process of performing the listening test. The listening test was designed on Max/MSP 6 and followed the pairwise comparison model as previously stated. Subjects listened to a reference track of the band pass filtered pink noise that had the centre frequency of 1kHz and then compare the loudness of a differing band pass filtered pink noise as can be seen in **figure 4.6**.

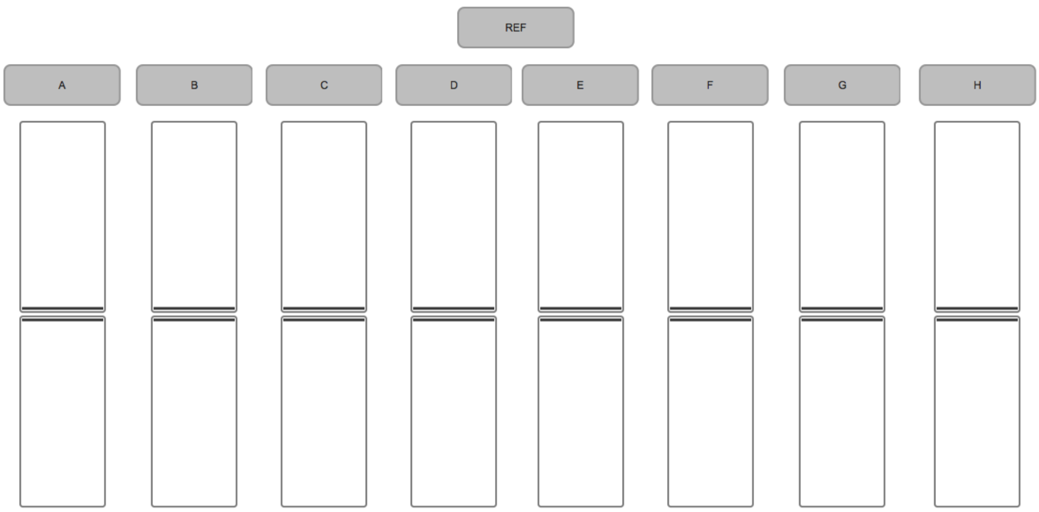


Figure 4.6 Max MSP 6 interface for the listening test

For the preliminary tests only the continuous amplitude characteristic waveforms were used to highlight any playback problems in the reproduction system. The continuous audio stressed the system more due to a constant playback. The preliminary test also tested different SPL to determine if any SPL levels caused problems. Target playback levels of 50 dB SPL, 60 dB SPL, 70 dB SPL, 80 dB SPL and 90 dB SPL were decided. These intervals follow the same dB interval as found in the BS ISO 226:2003 [2003] and gave a reasonable collection of playback levels to determine the optimum levels or highlight any problems in the playback level. Using an SPL meter the playback level was first calibrated to 90 dB SPL through adjusting the gain rotary encoder on the Motu 4pre. To measure the playback level of continuous noise the SPL meter was set to use a Z-weighted filter to ensure no biasing in the equipment. The average level was taken using the function within the SPL meter that calculates the loudness over a calibrated time, in this situation it was 30 seconds. Once the playback level was calibrated to 90 dB SPL, the Max/MSP 6 "patches" had different levels of attenuation programmed into them to create the differing levels of playback as can be seen in **figure 4.6**. This created a series of patches that were of differing octaves and of differing playback levels. The patches were randomised in order.

Nine subjects sat the tests who were deemed expert listeners ranging in age from 22 to 50. The subjects were deemed expert listeners as they were staff at the University of Huddersfield where a requirement was to have regular auditory health checks. All subjects provided proof that they were of normal hearing. Feedback from the tests in the form of consulting the subject allowed to isolate several factors that would need to be considered when constructing the listening tests to establish perceived loudness changes in relation to dynamic range compression.

4.4.1 Analysis

4.4.1.1 Limitations

The feedback from the tests highlighted a problem in the procedure and instruction to the subject. The Max/MSP 6 listening test allowed for the reference stimulus and the tested stimulus to both be played back at the same time. By playing both stimulus at the same time there was not an accurate portrayal of equal loudness across all of the octaves. The octaves either side of the reference stimulus when played back simultaneously were partly masked due to masking effects as detailed in section 2.3.1. Another problem highlighted was the absence of a fade in or a fade out of the stimulus. The ITU-R BS.1116-2 [2014b] guidelines support that the waveforms created artefacts when the subject swapped between the stimuli. These artefacts can impact the judgement of the loudness and so it was paramount to make the appropriate changes.

Finally the playback SPL levels stressed the reproduction system at 80 dB SPL and 90 dB SPL. Both the playback SPL levels could not accurately replicate the octave band with centre frequency of 16 kHz due to severe distortion of the loudspeaker at this level. The 90 dB SPL playback level stressed the system also when playing the octave band with centre frequency of 63 Hz.

4.4.1.2 Subject Feedback

During the subject feedback it was also highlighted that the same octave band stressed the system across all playback levels however not severely. When considering the Genelec 8040A's frequency response there is a drop off in accurate reproduction at 50 Hz which could not be corrected using the DBX 1231 Dual 31 band graphic EQ.

4.4.2 Considerations

The resulting considerations from this was to exclude the 80 dB SPL and 90 dB SPL playback level from the further tests and to change the 2nd octave's lower filter frequency to a frequency that would allow a better reproduction. The lower filter for the octave band with centre frequency of 63 Hz was changed to 50 Hz, therefore the octave band with centre frequency of 63 Hz was not a whole octave band but still had sufficient frequencies to indicate the behaviour of low frequencies.

The fade in and fade outs were added to the test also. The fade in time was set only for the continuous tests because the transient tests already have an inherent 5ms fade in. The fade in time for continuous was set to 250ms. The fade out time was set to 250ms also.

5 Equal Loudness Contours for Octave Bandwidths

5.1 Introduction

To obtain equal loudness levels for octave bands, a listening test was designed as mentioned in the previous chapter instead of using prior results from existing studies. The existing studies had given different accounts of equal loudness for complex tones (Stevens [1956], Jahn [1974] and Bauert & Torick [1966]) and so it was necessary to determine the perceived loudness of each octave band within this thesis. The levels obtained were used to directly contribute to the foundation loudness level of the octave bands when testing varying crest factors.

A Second motivation for obtaining equal loudness contours for octave bands was to investigate if the frequency weighting as used for objective loudness measures were accurate. If the equal loudness contour for 70 dB SPL was different to the K-weighting frequency response when inverted then the objective loudness level could be questioned.

5.2 Test

The Max/MSP 6 file was modified to accommodate the considerations discovered during the preliminary test. The playback level for the tests were limited to 50 dB SPL and 70 dB SPL to re-adjust the focus of the varying crest factor tests to consider a low listening level and a moderate listening level. These two levels simulate the more common listening levels. The Max/MSP 6 patches also were modified to include the fade ins and fade outs of the stimulus. The stimulus included the adjusted octave with centre frequency of 63 Hz.

The SPL level recordings for the transient amplitude characteristic were based upon the peak SPL reading for the signal was too short to take an accurate reading of the average SPL reading. Because all of the octave bands had received the same filtering however it meant that the peak readings would not be too dissimilar from the average reading. This method was used for both the calibration and for the method of acquiring the resultant SPL level for the equal loudness contours.

For the tests to establish equal loudness levels for the stimulus the procedure was similar to the preliminary. Eight subjects were used of which all of them were classed as expert listeners due to their positions as staff at the University of Huddersfield's with association to the Music Technology course. The subjects were all male and ranged from an age of 22 to 50. The listening tests were staged in two sittings where each test contained an equal amount of randomly arranged stimulus of continuous amplitude characteristic played back at the two varying playback levels in a random order.

The subject was taught to manually turn on and off the playback of the reference and the tested stimulus. The reason for this was to allow time between the two playbacks. The subject was asked to change the playback level using either of the two sliders presented. The sliders were placed one above the other to represent the amplification of the playback in moving the top slider up and the attenuation of the playback by moving the bottom slider down. This can be seen in **figure 4.6**. The subject were left alone in the semi-anechoic chamber to complete the test and was asked to retrieve the supervisor of the test upon completion with any feedback on the test.

5.3 Analysis of Results

5.3.1 Spoken Feedback

The verbal feedback from subjects centred around two octave bands. The first was the lowest octave band, with a centre frequency of 63 Hz, where subjects expressed that when comparing it to the 50 dB SPL reference noise they potentially could have amplified the octave band further. This however was more of a preference in the testing to see what a further amplified signal would be perceived as and not because subjects felt it fell short of the equal loudness task.

The second octave band was the octave band with the centre frequency of 16 kHz where subjects felt that the transient amplitude characteristic was less abrasive than the continuous counterpart and that it even resembled the sound of a hi-hat cymbal.

5.3.2 Test for Normality

To determine which average calculation to use on the results, a test to discover the is the results were normally distributed was implemented. Using IBM SPSS Statistics 22, the data was analysed using the Shapiro-Wilk method. The majority of the octave bands found that the results were not normally distributed ($p<0.05$) and so the median average calculation was suitable for the average calculation.

5.3.3 SPL Measurements

The median average results were based upon the gain level that subjects applied to the comparison octave band to match to the reference octave band of 1 kHz. The median gain level therefore was re-applied to the corresponding octave band and a new measurement of SPL was recorded. The SPL was recorded from the distance of the subjects head position using the same SPL meter as so to determine the equal loudness contours. For the continuous amplitude characteristic a an average SPL reading was measured using the Z weighting (no weighting applied), however, for the transient amplitude characteristic a peak SPL reading was measured. This was because of the natural behaviour of the transient signal where an average reading was impossible due to the short length of the transient signal. The methodology was kept uniform respective to the amplitude characteristic so the readings we relative within the amplitude characteristic.

5.3.4 Equal Loudness Contours for Octave Bandwidths

Figure 5.1 shows the equal loudness contours for octave bands using continuous noise.

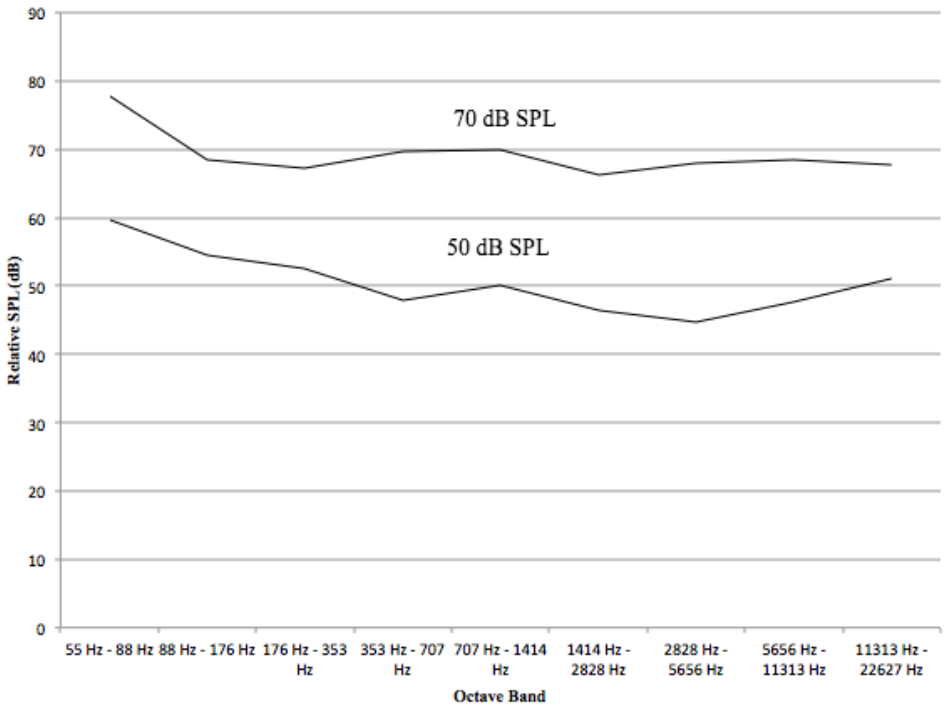


Figure 5.1 Equal Loudness Contour for Continuous Amplitude Characteristic

The equal loudness contours for octave bands using continuous noise show that octave bands with a centre frequency of 63 Hz are not perceived as loud as the rest of the octave bands across the frequency spectrum. The octave bands with centre frequencies of 125 Hz and 250 Hz however show that when compared to a 50 dB SPL reference octave band, they are perceived quieter but when compared to a 70 dB SPL reference octave band they are perceived as louder. This is demonstrated by the equal loudness contour dropping below the 70 dB SPL point on the y axis but remain above the 50 dB SPL point on the y axis for the corresponding contour.

The perceived loudest octave band for continuous noise differed when compared to two different SPL levels of the reference octave band. The octave band with centre frequency of 4 kHz was perceived loudest when compared to a 50 dB SPL noise however an octave band with centre frequency of 2 kHz was perceived as loudest when compared to a 70 dB SPL noise. The response to the octave band with centre frequency of 16 kHz also differed between the two SPL's. When the octave band was compared to a 50 dB SPL reference noise it was perceived as quieter and so required amplification however for the comparison to a 70 dB SPL reference noise the octave band was perceived as louder and so required attenuation.

Figure 5.2 shows the equal loudness contours for octave bands using transient noise.

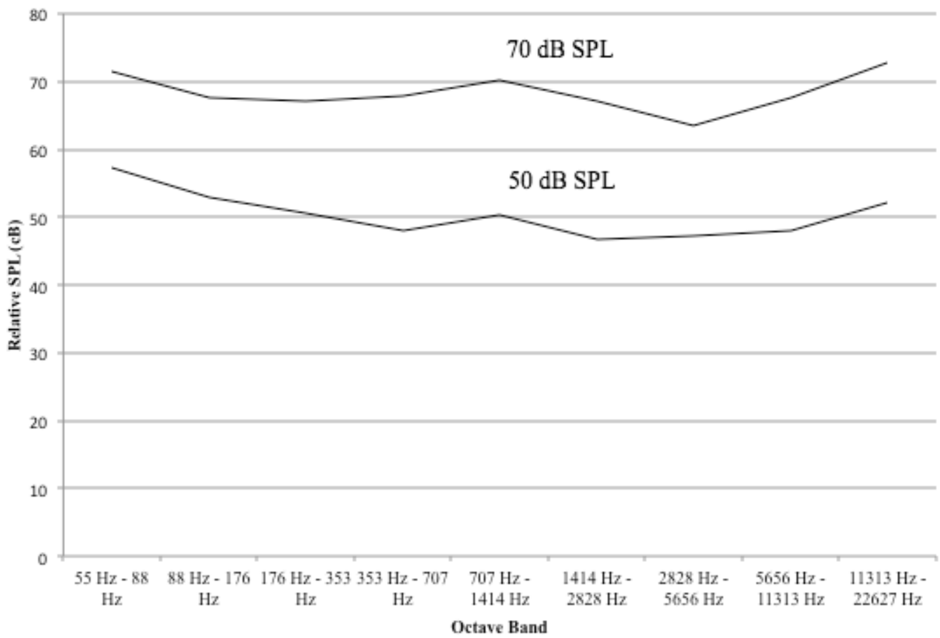


Figure 5.2 Equal Loudness Contour for Transient Amplitude Characteristic

The equal loudness contours for octave bands using transient noise show that octave bands with a centre frequency of 63 Hz are perceived as quieter than the reference noise. The octave band with centre frequency of 125 Hz shows that it is perceived quieter when compared to the 50 dB SPL reference noise but when the playback level is changed to be based upon a 70 dB SPL reference noise the octave band is perceived as louder. The octave band with centre frequency 250 Hz is perceived as equally loud when the reference noise is 50 dB SPL however it is perceived as louder when the reference noise is 70 dB SPL. The loudest octave band for the 70 dB SPL contour is the octave bandwidth with a 4 kHz centre frequency whereas the loudest octave band for the 50 dB SPL contour is the octave band with a centre frequency of 2 kHz. Both contours show that the octave band with the centre frequency of 16 kHz is perceived as quieter due to the amplification applied.

Figure 5.3 shows the combined equal loudness contours for octave bands incorporating both amplitude characteristics.

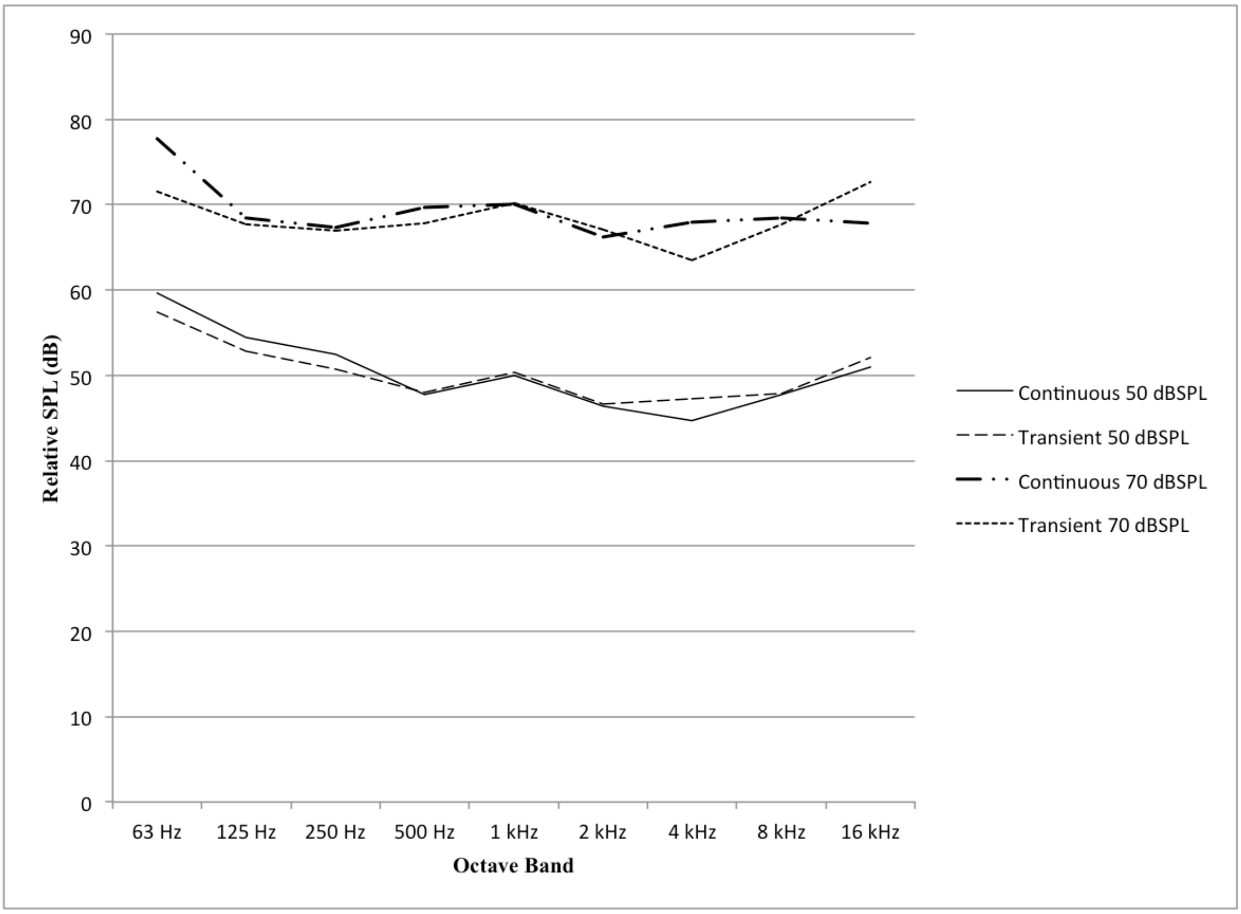


Figure 5.3 Combined equal loudness contours for both continuous amplitude characteristic and transient amplitude characteristic

The differences between the continuous and the transient amplitude characteristics are not too dissimilar for the 50 dB SPL contours. The trends are similar apart from the octave band with centre frequency of 4 kHz where the continuous amplitude characteristic is perceived as louder. The lowest three octave bands there were included in the listening test, those of centre frequencies of 63 Hz, 125 Hz and 250Hz, display how the continuous amplitude characteristic was perceived as quieter.

The 70 dB SPL contour shows three octave bands where the amplitude characteristics differ in regards to loudness perception. The low frequency octave band with a centre frequency of 63 Hz suggests that the continuous amplitude characteristic was perceived as quieter due to more amplification. The octave band with centre frequency of 4 kHz also appears to have a difference between the two amplitude characteristics. The transient amplitude characteristic was perceived as louder compared to the continuous amplitude characteristic. Finally the last large discrepancy between the two amplitude characteristics was in the octave band with a centre frequency of 16 kHz. The continuous amplitude characteristic was perceived as louder than the reference noise whereas the transient amplitude characteristic was perceived as quieter than the reference noise.

5.3.5 Discussion of Results

The continuous amplitude characteristic equal loudness contour for 50 dB SPL displayed similarities to the equal loudness contours for pure tones as shown in **figure 2.2**. However the contour for 70 dB SPL has noticeable differences in what can be classed as the low mid range are of frequencies, the octave bands with centre frequencies of 125 Hz and 250 Hz, and the high frequencies, the octave band with 16 kHz centre frequency.

5.3.5.1 Low Mid Frequency Octave Bands

The low mid range gave an indication that it was perceived as louder, which contradicts previous equal loudness contours for complex noise (Stevens [1956], Jahn [1974] and Bauer & Torick [1966]). The corrective process of the reproduction system may explain this low mid range of frequencies. The low mid frequency range can be regarded as “mud” when mixing and so as can be seen in **figure B.1** (see

Appendix B) the Genelec 8040A attenuates a lot of energy around these frequencies. This may the case for a lot of playback systems and so previous models may not have taken into account the corrective characteristics of the speaker or headphones used. Stevens [1956] constructed equal loudness contours for complex noise however highlighted the colouration from the playback system and so voided the results. This frequency range therefore should be considered as potentially louder than the 1 kHz reference noise.

5.3.5.2 High Frequency Octave Band

The octave band with a centre frequency of 16 kHz gave interesting results because they differed for each amplitude characteristic. The continuous amplitude characteristic was judged as louder than the reference noise and this may be a contribution of the annoyance factor. Stevens [1956] noticed that high frequency content can be judged as annoying and this may have had an influence on the loudness perception because annoyance could cause a reaction with the subject to instinctively lower the volume. This also adheres to the equal loudness contours for complex noise as previously found (Stevens [1956], Jahn [1974] and Bauer & Torick [1966]) and so it would suggest that high frequency content is considered louder. This however contradicts the equal loudness contours for pure tones. So why would complex tones be perceived different to pure tones? ERB theory [Moore *et al,* 1997] may explain this because the octave band with centre frequency of 16 kHz has a lot of content and therefore the auditory nerve will respond more due to more of the inherent filters having overlap with this broad frequency range. For a pure tone there may not be enough of the fibres responding but for complex noise there would.

The transient amplitude characteristic may suggest that this was not the case as the transient amplitude characteristic was judged as quieter however, this does adhere to the equal loudness contours for pure tones. The difference between the amplitude characteristics may raise the question of whether the ERB theory does apply. A solution could be offered with the temporal integration studies for the transient amplitude characteristic has the majority of the signal shorter than the 100ms threshold that contributes to loudness perception. This potentially could also correlate to the annoyance factor. The verbal feedback from subjects had highlighted that the transient amplitude characteristic was less abrasive than the continuous amplitude characteristic. The annoyance factor may be related to the 100ms threshold because the sound does not have long enough to register the complete loudness and so it may be perceived as less annoying. This would explain the difference shown in **figure 5.3** and suggest that ERB theory fits.

5.3.5.3 Differences Between Playback Levels

The lower playback level helps to reinforce the annoyance factor theory for a quieter level would not cross over into as many inherent filters within the auditory system. This may suggest however that there is an annoyance threshold that may be a useful research target. The low mid range for the lower playback level may not stress the playback system as much and so the corrections made to the frequency response may not have been as apparent. The changes between playback level can raise questions as to why the pattern changes more than previous equal loudness contours for both pure tones and complex noise.

The differences could provide evidence for mixing engineers in regards to how the audio could potentially differ when played back at different levels. An example would be for drums with popular music or rock music where if the cymbals are left ringing out then this may not be a problem at lower playback levels however when listened back at a louder playback level, the high frequency content becomes loud and possibly annoying. However the hi-hat of a drum kit would appear to be unaffected due to the short temporal nature of the sound. Therefore whilst continuous high frequency noises should be perhaps lowered in the mix, the hi-hat can avoid any special considerations.

5.3.5.4 K Weighting

The filter for the LUFS calculation is based upon the K weighting as previously discussed. The K weighting filter was an adaptation of the B weighting filter that was based upon the equal loudness contour for pure tones at 70 dB SPL. By comparing the 70 dB SPL equal loudness contour of octave bands a review into the accuracy of the LUFS calculation can be made. The equal loudness contour as found as part of this hypothesis suggests that the K weighting filter is not wildly inaccurate. The only differences are within the octave bands with centre frequency of 125 Hz and 250 Hz. These octave bands however may make large contributions for objective loudness measures so a proposed further study is possible from this review.

Though the K weighting has accuracy for the moderate playback level, the differences found when the playback level is lower suggests that the LUFS calculation does not accurately represent loudness when the playback level is lower. This is a weakness of the objective loudness measure because the playback level changes the equal loudness contours of octave bands which could suggest that complex audio changes the perceived loudness of different frequency bands in a way that the LUFS calculation does not comprehend. Though LUFS gives a good indication of relative loudness between broadcasters and programs, it fails to determine the differences when the playback level changes.

5.3.6 Limitations

As discussed the limitations of this experiment are the subject number and the uncertainty around the potential impacts of the alterations made to the frequency response of the playback system.

The subject number could be increased as suggested by the ITU-R BS.1116-2 [International Telecommunications Union, 2014b] when making judgements of audio via listening tests. Each listener will be different and therefore more numbers would create a better average between the subtle differences between individuals.
The alterations made to the playback system allowed for an unbiased study however further testing into creating a flat frequency response playback system could be executed. Differing playback systems however add colouration to the playback system so the equal loudness contours can never be applied without an appreciation to the effects of this colouration.

6 LOUDNESS PERCEPTION CHANGES IN RELATION TO CREST FACTOR VARIATION

6.1 Introduction

This chapter describes the experiment to determine if DRC affects the perceived loudness of frequency bands uniformly or whether certain frequency bands have a greater response in terms of perceived loudness than others. As discussed in section 3.4 DRC affects loudness perception and this has a suggested link to quality perception. By determining the frequency bands response a possible DRC practise for music production could be established.

6.2 Application of Equal Loudness Measurements

Part of the purpose of the test to find equal loudness of the octave bands was to apply the loudness levels to create a uniform loudness as discussed in section 4.3.2.3. The process consisted of using the median average value of the gain changes made for both amplitude characteristics at each octave band at the different SPL levels. The average levels were applied within Adobe Audition CS6 creating a new set of stimuli that were used for the basis of the crest factor varying stimulus. The new stimulus contained all the different octave bands at both amplitude characteristics where each octave band had an uncompressed sample and then four samples with different crest factors where the difference incremented by 1 dBFS. This gave 5 samples for each octave band for two different SPL levels and two different amplitude characteristics resulting in a total of 180 samples within the stimulus group.

6.3 Objective Loudness Measurements

Once the stimulus had been processed to include the equal loudness considerations, an objective loudness measurement was taken. Each sample was analysed using the amplitude statistics within Adobe Audition CS6 as similar to the crest factor calculation however the measurement taken was the LUFS measurement in dBFS. These measurements are represented in **figure C.1** to **figure C.4** (see Appendix C).

6.4 Test

For this listening test the MUSHRA testing process was used. The implementation of this test used Max/MSP 6 as cam be seen in **Figure 6.1**.

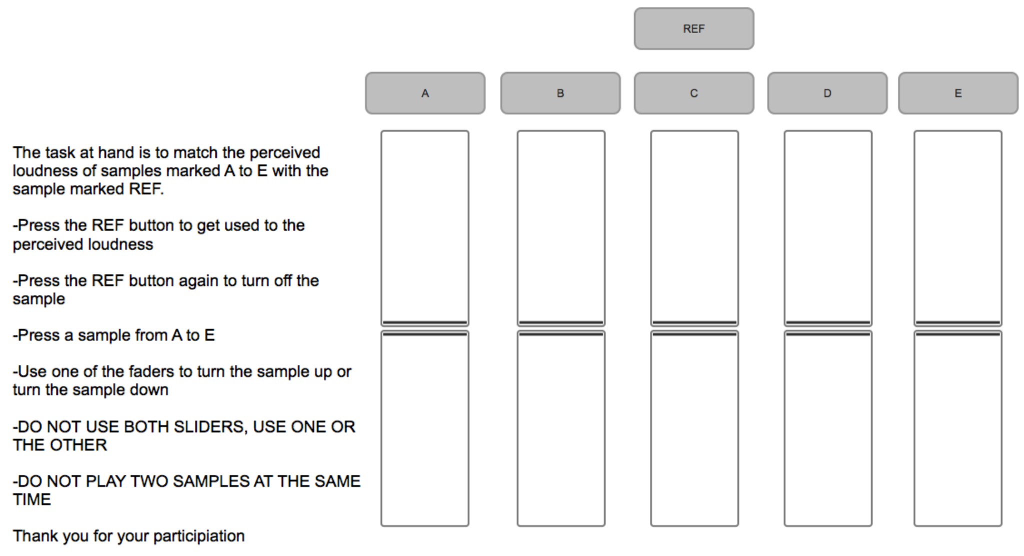


Figure 6.1 Max MSP 6 patch for the varying crest factor listening test

Four Max/MSP 6 files were configured to allow for the continuous vs transient factor and the 50 dBSPL vs 70 dBSPL factor. The tests used the modified set of octave bands and fade ins and outs as discussed in section 4.2.2. Each Max/MSP 6 file contained level adjustments programmed into the file to account for the different SPL conditions that the subject was partaking in which allowed for a constant level between tests. The order of the stimulus was randomised so the ‘A’ to ‘E’ order was different for each octave band and therefore complied with the hidden reference aspect of the MUSHRA testing process.

The instructions for the listening test were on screen for the subjects as well as the instructions delivered vocally with the allowance to raise any questions. Eight subjects were used of which all of them were classed as expert listeners due to their positions as staff at the University of Huddersfield’s with association to the Music Technology course. The subjects were all male and ranged from an age of 22 to 50. The subject was taught to manually turn on and off the playback of the reference and the tested stimulus. The reason for this was to allow time between the two playbacks. The subject was asked to change the playback level using either of the two sliders presented. The sliders were placed one above the other to represent the amplification of the playback in moving the top slider up and the attenuation of the playback by moving the bottom slider down. The subject was left alone in the semi-anechoic chamber to complete the test and was asked to retrieve the supervisor of the test upon completion with any feedback on the test.

6.5 Analysis of Results

6.5.1 Test for Normality

A Shapiro-Wilk tst was conducted using IBM SPSS Statistics 22 to determine if the data was normally distributed. The tests were conducted on each octave band for each amplitude characteristic and each playback level. The results found that there was not a normal distribution for the results ($p<0.05$). This indicates that none parametric tests should be considered for any statistical analysis and that the median average should be used for average calculations.

6.5.2 Hidden Reference

Across all of the tested octave bands with changing variables of playback level and amplitude characteristics, the hidden reference gave an anchored median value of 0 dbFS gain level change.

6.5.3 Inter Crest Factor Perceptual Changes

The median averages for the change in gain as a result of the crest factor changing gave an indication into the perceived loudness of each crest factor varied stimulus. The results were plotted graphically and analysed statistically.

The graphs can be explained by examining **figure D.1** to **figure D.36** (see Appendix D) which show the median gain level difference in dBFS on the y-axis as applied by subjects across the crest factor varying stimulus for each octave band on the x-axis. The y-axis has mostly negative figures because the gain applied by the subjects was mostly to attenuate the signal and therefore required a negative gain application. The x-axis has the different crest factor levels as denoted by a reference to 'n'. The 'n' value of the x-axis denotes the hidden reference which is the uncompressed octave band which had no DRC processing applied. The 'n-1' value is the octave band with DRC processing that has changed the crest factor by 1 dBFS therefore making the crest factor 1 dBFS less than the reference and hence the 'n-1' label. The 'n-2' value is the octave band with heavier DRC processing that has changed the crest factor by 2 dBFS therefore making the crest factor 2 dBFS less than the reference and hence the 'n-2' label. The 'n-3' value is the octave band with further DRC processing that has changed the crest factor by 3 dBFS therefore making the crest factor 3 dBFS less than the reference and thus the 'n-3' label. The 'n-4' value is the octave band that had the most DRC processing applied and so the crest factor has changed by 4 dBFS therefore making the crest factor 4 dBFS less than the reference and so has the 'n-4' label.

For the statistical analysis considerations into the averages and test type were considered. Due to the nature of the results being calculated with the median average as a result of non-normal distribution, a non-parametric alternative to the repeated measures analysis of variance (ANOVA) test was applied. The test used was the Friedman test within IBM SPSS Statistics 22 which allowed for the individual crest factor values to be analysed in a pairwise format. This gave a clear indication of how the crest factor affected the loudness perception.

6.5.3.1 Continuous Amplitude Characteristic at 50 dB SPL

The median results are demonstrated in table 2 and relate to the graphs in **figures D.1** to **figure D.9** (see Appendix D).

Table 2 A table displaying the median average of gain application for continuous amplitude characteristic at 50 dB SPL.

Centre Frequency of Octave Band (Hz)	**Median Crest Factor Level (dBFS)**				
	n	**n-1**	**n-2**	**n-3**	**n-4**
63	0	-0.25	-1.09	-2.53	-3.20
125	0	-0.92	-1.55	-2.48	-3.70
250	0	-0.73	-1.98	-2.76	-3.64
500	0	-0.97	-1.89	-2.59	-3.63
1000	0	-0.42	-1.65	-2.29	-3.02
2000	0	-0.69	-1.35	-1.75	-3.14
4000	0	-0.60	-1.28	-2.61	-3.60
8000	0	-0.56	-1.57	-2.01	-3.38
16000	0	-0.32	-1.32	-2.17	-2.75

As can be seen from **table 2** and the **figures D.1** to **figure D.9** the stimulus with the most DRC application was routinely identified as the loudest across all of the octaves. This is shown by the 'n-4' value have the most gain reduction. It can also be observed by tracing the median values that the application levels of DRC changed the perceived loudness almost linearly. There are fluctuations in the results that make it non-linear but despite not always being linear it is clear that loudness increases for every 1 dBFS that the crest factor lowered. Interestingly the gain application did not match the gain change within the stimulus as seen that 4 dBFS change in the crest factor did not ever produce 4 dBFS. This is noticeable across all of the different crest factor values across every octave band.

Though the gain application never matches the inherent gain change there is an anomaly where the gain application is noticeably further from the expected level. The gain application level for 'n-3' of the octave band with centre frequency of 2 kHz displays an unusual value as seen in table 2 for the attenuation was not as high as expected when considering the 'n-3' values for the octave bands. This could suggest that this octave band may have little or no response to perceived loudness changes as a result of DRC application. Statistical analysis could however prove otherwise.
A Friedman test was conducted with the further use of the pairwise comparison being the key focus. By running this pairwise comparison an analysis of whether there is significant difference between the differing crest factor values in regards to loudness perception could be established. **Figures E.1** to **figure E.9** (see Appendix E) show the pairwise results and the significant difference with a Bonferroni correction as highlighted on the right hand side of the figures. The Bonferroni correction is applied for this calculation of significant difference because when multiple pairwise

comparisons are made, the chance of significant difference increased proportionally to the number of comparisons [Napierala, 2012]. To avoid giving incorrect statistical evaluations, this practise was adhered to. The statistical significant difference was generally found to be between the ‘n-3’ value and the ‘n’ value. The ‘n-4’ value also had significant difference from the ‘n’ and ‘n-1’ values for every octave. This suggests that the difference in loudness perception and potentially the perception of DRC can be noticed when the DRC application reaches a high enough level to produce a reduction of 3 dBFS in crest factor when the playback is at a low level and when the amplitude characteristic is of a continuous nature. Analysing the octave band with a centre frequency of 2 kHz, the significant difference actually suggests that the threshold for a change in loudness perception is when the crest factor has a difference of 2 dBFS. Considering this octave band had given a lower attenuation level for the ‘n-3’ value, this counteracts the proposed poor perceived loudness response to DRC application by not only disagreeing but by also suggesting the opposite. The statistical analysis suggests that the octave band with a 2 kHz centre frequency is more susceptible to perceived changes as a result of DRC application.

6.5.3.2 Continuous Amplitude Characteristic at 70 dB SPL

The median results are demonstrated in table 3 and relate to the graphs in **figures D.10** to **figure D.18** (see Appendix D).

Table 3 A table displaying the median average of gain application for continuous amplitude characteristic at 70 dB SPL.

Centre Frequency of Octave Band (Hz)	Median Crest Factor Level (dBFS)				
	n	n-1	n-2	n-3	n-4
63	0	-0.25	-2.25	-3.29	-5.88
125	0	-1.73	-4.26	-5.47	-7.86
250	0	-0.53	-1.50	-2.67	-4.02
500	0	-0.60	-1.55	-2.56	-3.83
1000	0	-0.62	-1.62	-2.68	-3.74
2000	0	-0.85	-1.25	-2.64	-3.89
4000	0	-1.04	-1.28	-2.87	-3.60
8000	0	-1.35	-1.83	-2.93	-4.83
16000	0	-0.49	-1.25	-2.45	-3.29

For this playback level the loudest perceived crest factor level was again the ‘n-4’ crest factor as was found for the lower playback level. Similar to the lower playback level the results suggest that there is an almost linear relationship between DRC application and perceived loudness. The results for this playback level also show that the gain application in dBFS generally did not match the inherent crest factor dBFS change apart from two values. The 'n-4' value for the octave band with a centre frequency of 250 Hz and the 'n-1' value for the octave band with a centre frequency of 4 kHz. Though gain application does not match there were a few noticeable patterns relating to relationship between gain application and DRC. For the octave bands with centre frequencies of 250 Hz, 500 Hz and 1 kHz the difference in dBFS between the 'n-1', 'n-2' and 'n-3' ranges were proportional to the difference in dBFS of the crest factor level. As can be seen the levels for the 'n-1' crest factor level for all of the octave bands are generally around the -0.6 dBFS region and then as the crest factor changes by 1 dBFS, the gain application also changes by 1 dBFS to change the levels to be generally around -1.6 dBFS. However the 'n-4' crest factor value for all of these octave bands loses the trend.

The results for this amplitude characteristic at this playback level seem to show a lack of trends however though there are a lack of trends, a few results can still be identified as anomalies because the results are largely different from what would be expected or seen in other octave bands. The two lowest octave bands, those with centre frequencies of 63 Hz and 125 Hz, showed that subjects on average gave gain application levels much greater attenuation levels than expected. The octave band with a centre frequency of 63 Hz only displayed this for the 'n-4' crest factor value where the gain application was -5.88 dBFS, a figure nearly 2 dBFS lower than inherent dBFS change, but the octave band with a centre frequency of 125 Hz displayed this from 'n-1' onwards. The 'n-1' figure had -1.73 dBFS gain application which when compared to the inherent dBFS change was 0.73 dBFS different however this is small compared to the ‘n-2’ gain application. The gain application level for this was -4.26 dBFS which is over 2 dBFS from the inherent diference. The ‘n-3’ gain application follows the same difference but then for ‘n-4’ the gain application was -7.86 dBFS which was almost 4 dBFS different from inherent dBFS change. However on inspection of the spread of data within **figures D.10** to **figure D.18** (see Appendix D) the octave band with centre frequency of 125 Hz shows that the crest factor values that gave the unexpected results also had a large spread of data with the spread of data even suggesting that at least one subject used a more expected gain application of around -4 dBFS for the ‘n-4’ crest factor. This however can be seen as only a rarity for the quartiles still show a range of gain application that at either end can be classed as outside the expected range. It could be suggested that for moderate listening levels, sound with continuous amplitude characteristics create a difficult situation of identifying DRC application in lower octave bands.

The statistical analysis however could contradict this if there is no significant difference found from the ‘n-4’ and the ‘n-3’ crest factor levels from the uncompressed (‘n’) crest factor level. Upon inspecting the pairwise comparisons as part of the Friedman test it was found that there was significant difference found between ‘n-4’ and ‘n’ and also from ‘n-3’ and ‘n’ as can be seen in **figures E.10** to **figures E.18** (see Appendix E). This suggests that the statistical evidence does not contradict the previous statement and so it could be suggested that for moderate listening levels, sound with continuous amplitude characteristics create a difficult situation of identifying DRC application in lower octave bands. As for the other octave bands, the same significant difference trend was found as was found for the low playback level of continuous amplitude characteristic. This suggests that the difference in loudness perception and potentially the perception of DRC can be noticed when the DRC application reaches a high enough level to produce a reduction of 3 dBFS in crest factor when the playback is at a moderate level also.

6.5.3.3 Transient Amplitude Characteristic at 50 dB SPL

The median results are demonstrated in table 4 and relate to the graphs in **figures D.19** to **figure D.27** (see Appendix D).

Table 4 A table displaying the median average of gain application for transient amplitude characteristic at 50 dB SPL.

Centre Frequency of Octave Band (Hz)	**Median Crest Factor Level (dBFS)**				
	n	**n-1**	**n-2**	**n-3**	**n-4**
63	0	-1.35	-1.63	-3.38	-4.47
125	0	-0.78	-1.80	-3.02	-3.79
250	0	-0.85	-1.98	-2.99	-3.57
500	0	-0.99	-1.47	-2.59	-3.54
1000	0	-0.85	-1.62	-2.50	-3.48
2000	0	-1.04	-2.01	-3.73	-3.83
4000	0	-1.08	-2.28	-3.35	-4.02
8000	0	-1.11	-2.22	-3.20	-3.93
16000	0	-1.04	-1.93	-3.26	-3.99

As can be seen in **table 4** the findings made within the continuous amplitude characteristic in regards to the octave bands with the most DRC application perceived as the loudest also hold true when the amplitude characteristic is of a transient nature at low playback level. All of ‘n-4’ gain application levels are the most attenuated in their respective octave bands. The relationship between the perceived loudness and DRC application is almost linear where the loudness perception increases when the crest factor is reduced. For this amplitude characteristic at this playback level, the gain application on dBFS followed the inherent dBFS change in crest factor closer than the continuous amplitude counterparts. Though there is still a lot of variation, which did lead to difficulties in making an absolute conclusion, it could suggest that transient amplitude characteristics at low playback levels are better to judge DRC application. The only anomaly from this can be seen for the ‘n-2’ value for the octave band with centre frequency of 500 Hz where the gain application is closer to -1 dBFS despite the inherent crest factor change being 2 dBFS. Comparing this to the **figures D.19** to **figure D.27** (see Appendix D) the plot suggests that there is not a large spread on the data and so it would appear that this judgement was fairly consistent. The spread of the data across most of the plots appear to fairly small suggesting that transient amplitude characteristics cause less differences between subjects.

When considering the significant difference results from the post hoc comparisons within the Friedman tests as seen in **figures E.19** to **figure E.27** (see Appendix E), the same trend as discovered in both continuous amplitude characteristic data sets is apparent. The noticeable difference that DRC application makes is prevalent once the DRC application creates a difference of 3 dBFS within the crest factor.

6.5.3.4 Transient Amplitude Characteristic at 70 dB SPL

The median results are demonstrated in table 5 and relate to the graphs in **figures D.27** to **figure D.36** (see Appendix D).

Table 5 A table displaying the median average of gain application for transient amplitude characteristic at 70 dB SPL

Centre Frequency of Octave Band (Hz)	Median Crest Factor Level (dBFS)				
	n	n-1	n-2	n-3	n-4
63	0	-1.18	-2.28	-3.99	-5.92
125	0	-1.57	-2.61	-5.38	-6.74
250	0	-1.13	-2.12	-3.35	-4.61
500	0	-1.52	-2.25	-3.47	-5.48
1000	0	-0.67	-1.75	-2.99	-4.75
2000	0	-1.04	-1.78	-2.42	-4.16
4000	0	-0.90	-2.09	-3.02	-3.96
8000	0	-1.11	-1.78	-2.96	-4.57
16000	0	-1.45	-2.33	-2.87	-4.16

As can be seen within **table 5** the findings made in the three other controlled listening tests relating to the DRC application level and perceived loudness also holds for this playback level. The octave bands with the most DRC application were perceived as the loudest. The results also show the near linear relationship as was found in the continuous amplitude characteristics for both playback levels and for the transient amplitude characteristic with a low playback level. Again the fluctuations prevent a total linear relationship and there are more anomalies within these sets of results. The octave band with centre frequency of 63 Hz had a greater amount of attenuation applied for the 'n-'4 crest factor level as shown by the -5.92 dBFS gain level applied which is 1.92 dBFS out from the expected level. This was found to the octave band with centre frequency of 250 Hz also where the 'n-4' crest factor was attenuated closer to a gain level application of -5 dBFS than -4 dBFS. There were two anomalies for the octave band with centre frequency of 500 Hz as can be seen where the 'n-1' and the 'n-4' variables have more attenuation than expected however because of the 'n-2' and 'n-3' holding to expectation this does not suggest anything. THe 'n-4' crest factor level had more attenuation for the octave bands with 1 kHz centre frequency and for the octave band with 8 kHz centre frequency.The repetition of the anomaly may suggest that when transient amplitude characteristics are played back at a moderate level, the application of DRC changes the loudness perception more severely when past the threshold of a change in crest factor level of 3 dBFS. Studying the plots of the data in **figures D.28** to **figures D.36** (see Appendix D) shows that the spread of data for certain octave bands at this crest factor level may be too large to make this suggestion. Notably the octave bands with centre frequencies of 63 Hz, 500 Hz and 8 kHz. The statistical analysis as discussed later in this section will support or undermine this suggestion.

Another octave band which had shown this anomaly in the ‘n-4’ crest factor was the octave band with centre frequency of 125 Hz however the results for this entire octave band were outside the expected results. None of the dBFS gain application matched the inherent changes in dBFS of the crest factor. The ‘n-1’ crest factor had a gain application level of -1.57 dBFS which is closer to a change of 2 dBFS than 1 dBFS. The ‘n-2’ crest factor had a similar result where the gain level application was closer to 3 dBFS than 2 dBFS however the ‘n-3’ and ‘n-4’ crest factor levels gave much greater differences between the gain level application and the dBFS level change of the crest factor. For the ‘n-3’ crest factor level the signal had -5.38 dBFS of gain applied which meant that this was 2.38 dBFS past the expected level. The ‘n-4’ crest factor level displayed a further difference from the expected level by 2.74 dBFS. A study of the corresponding plot within **figures D.28** to **figure D.36** (see Appendix D) show that the spread of the data is quite vast, especially for the crest factor levels of ‘n-2’, ‘n-3’ and ‘n-4’. The spread of this data suggest that there is a degree of uncertainty in judging the perceived loudness. This, like the continuous amplitude characteristic, suggests that at moderate listening levels sounds with transient amplitude characteristics create a difficult situation of identifying DRC application in lower octave bands.

The statistical analysis, like performed for the continuous amplitude characteristic at moderate playback level, could contradict this suggestion if there is no significant difference found from the ‘n-4’ and the ‘n-3’ crest factor levels from the uncompressed (‘n’) crest factor level. The pairwise comparison from the Friedman test showed that there was significant difference found between ‘n-4’ and ‘n’ and also from ‘n-3’ and ‘n’ as shown in figures (**INSERT APPENDIXB**). The pairwise comparison also suggests that unlike the continuous amplitude characteristic there was a significant difference between ‘n-3’ and ‘n-1’ which hints that the threshold for perceiving change as result of DRC happens around the ‘n-2’ level. As well it also suggests that the statistical evidence does not contradict the suggestion made prior and that for moderate listening levels, sound transient amplitude characteristics create a difficult situation of identifying DRC application in lower octave bands. The statistical analysis also shows that for the other octave bands there is significant difference between the ‘n-4’ and the ‘n-3’ level to the ‘n’ level which suggests that the difference in loudness perception and potentially the perception of DRC can be noticed when the DRC application reaches a high enough level to produce a reduction of 3 dBFS in crest factor when the playback is at a moderate level.

6.5.4 COMPARISON TO OBJECTIVE LOUDNESS CALCULATION

As can be seen in **figure C.1** to **figure C.4** (see Appendix C) the corresponding LUFS calculations for the stimulus is presented. Considering that the samples had been altered in inherent level to give an equal perceived loudness across the ‘n’ crest factor value, the LUFS measurements suggest that the octave bands are inherently different in regards to loudness. This contradicts the results from the equal loudness test and suggests that the LUFS calculation may be inaccurate when played back through a flat response playback system. The LUFS measurements track the crest factor change accurately by changing by 1 LUFS for one 1 dBFS crest factor change.

6.6 Discussion of Results

6.6.1 Gain Level Application as a Response to Dynamic Range Compression at Low Level Playback

The general relationship between crest factor measured DRC application and loudness perception for different octave bands at different playback levels suggests that as the DRC application increases, the perceived loudness increases. The testing method used peak normalisation so the increase in loudness is essentially a relationship with the RMS of audio. Soulodre [2004] had commented on the perceived loudness changes relating to changes of RMS levels. This relationship supports the findings made by Croghan *et al* [2012] and Wendl & Lee [2014] though similar to the findings of Wendl & Lee [2014], the gain application did not strictly follow the inherent RMS change. From the controlled independent variables it was found that a transient amplitude characteristic with a low level playback (50 dB SPL) had the closest gain application level to the RMS change. For every 1 dBFS change of the RMS as a result of the crest factor change, the gain application level was close to a 1 dBFS change. This is noticeable when comparing the results from the continuous amplitude characteristic at low playback level it was found that the continuous amplitude characteristic did not have this change as often or as close to the 1 dBFS figure. This would suggest that the transient amplitude characteristic is a better indicator of crest factor changes and this could be due to temporal aspects as discussed in section 2.3.3. Zwicker [1977] suggested that sounds with a steady amplitude characteristic allow for better judgements of the perceived loudness. This would suggest that continuous amplitude characteristics should give results that show a relationship between the perceived loudness and crest factor changes. Although this may be apparent, it can be interpreted another way. Because the DRC changed the waveform of the transient amplitude characteristic it gave an extra marker for the judgement of loudness changes. When DRC is applied, the compressed peak creates a longer continuous RMS time. As Zwislocki [1969] displayed in **figure 2.9** the longer the duration of the sound, or in this case the longer the RMS time, the louder a sound is perceived. This extra aspect of perceived loudness partnered with the inherent ability to perceive RMS changes gave the changes of crest factor levels across transient amplitude characteristics greater perceived differences and therefore subjects gave more accurate gain level changes.

6.6.2 Gain Level Application as a Response to Dynamic Range Compression at Moderate Level Playback

The general relationship between perceived loudness changes and crest factor level changes as a result of DRC at moderate playback levels (70 dB SPL) was fairly similar to the trend found in the low level playback where as the DRC application increased, so did perceived loudness. The difference observed however was that the gain level application from subjects had a greater spread in more octave bands across both the continuous and the transient amplitude characteristics. Not only was the spread greater but the median gain level application values had significantly different levels as demonstrated by running the Wilcoxon signed rank tests between the different playback level results as seen in **table F.1** and **table F.2** (see Appendix F). This can be interpreted as suggesting that DRC application is more noticeable as the playback level increases. Though the crest factors were the same, the results from the subjects suggest that at a higher playback level, more attenuation was needed because the audible difference was greater. When considering that the auditory system becomes more linear as suggested by equal loudness contours for pure tones as discussed in section 2.2.2.3 and by equal loudness contours for complex noise as discussed in section 2.3.2.2, this could suggest that sensitivity increases as the playback level increases. Also for an increased playback level, the auditory system has an increased activity due to inherent auditory filters having greater cross over as discussed in section 2.2.1.1. This increase would detect smaller changes in crest factor levels due to the increased overlapping filters. **Figure 2.1** demonstrated this. Around the 50 dB SPL region an octave band with centre frequency would fall within three of the filters however if the octave band increases to the 70 dB SPL region then the octave band would fall within four or five filters. **Figure 2.1** only shows a simplified version of the theory for the number of filters is upwards of 20,000 and so for increased playback levels, the number of filters that an octave band would fall within is considerably greater. This therefore suggests that at louder playback levels the perception of DRC application is more noticeable as a result of increased loudness perception sensitivity.

6.6.3 Comparison of Differing Playback Level

To determine how the effect of different playback level and different amplitude characteristics had on the perception of DRC application as result of loudness perception changes, pairwise comparisons were made using IBM SPSS Statistics 22. The pairwise comparisons were done by using a Wilcoxon signed rank test to determine if there was significant difference between the same crest factor levels within all octave bands across the different independent variables.

Table F.1 (see Appendix F) shows the specific octave bands that gave a significant difference between the 50 dB SPL playback level and the 70 dB SPL playback level for the continuous amplitude characteristic. As can be see there are seven overall crest factor levels in four different octave bands that displayed significant difference which were the octave bands with centre frequencies of 63 Hz, 125 Hz, 2 kHz and 8 kHz. The most interesting octave band is that with the centre frequency of 125 Hz. As discussed previously the moderate playback level (70 dB SPL) had given a greater spread of results and also a more attenuated median gain level. As can be seen in **figure 6.2** the crest factor contours also show the perceived difference between playback levels.

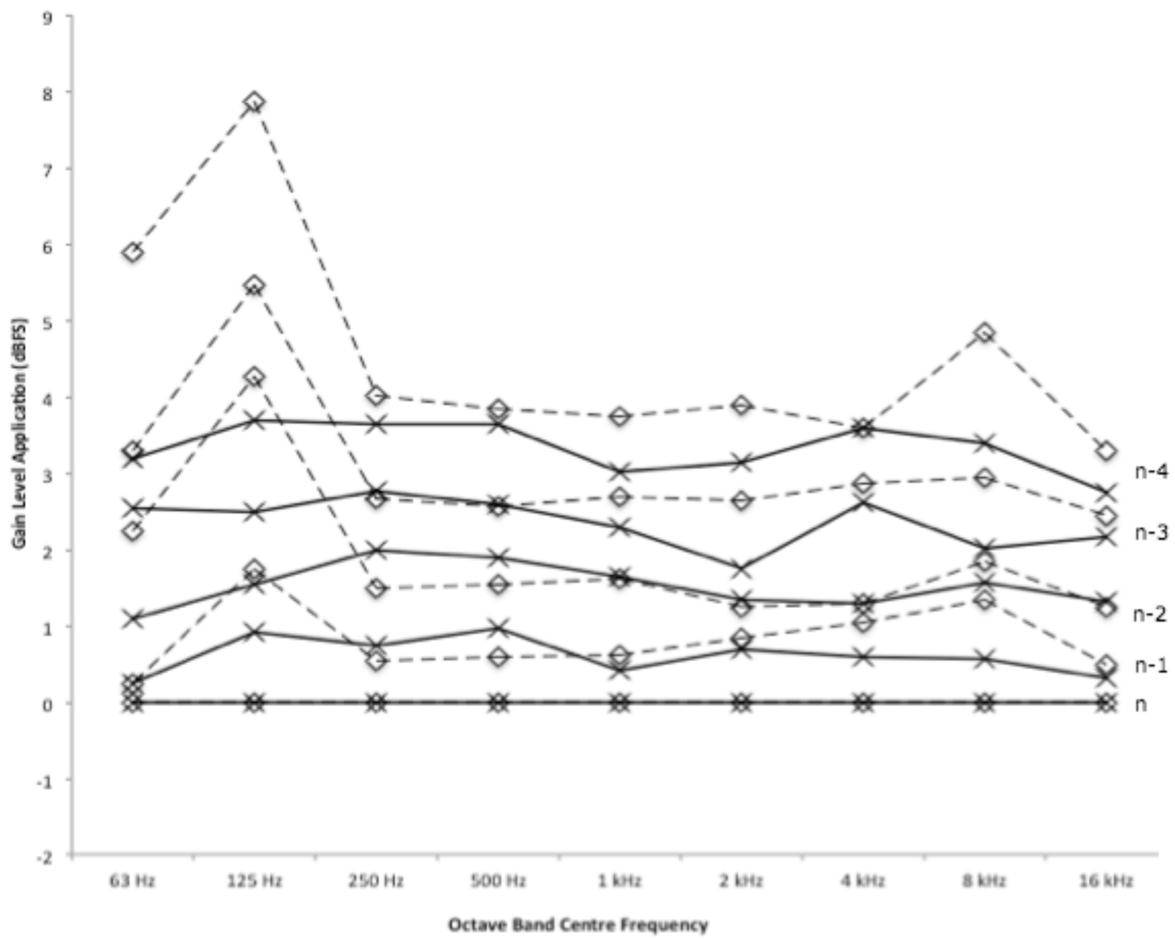


Figure 6.2 shows the contours of the different crest factor levels across the different octave bands for continuous amplitude characteristics. The **−** represents the playback level of 50 dB SPL and the --- represents the playback level of 70 dB SPL. The × shows the median values for the contours for the 50 dB SPL playback level and the ◇ shows the median values for the contours for the 70 dB SPL playback level. The graph is flipped to demonstrate when an octave band is determined as louder.

Figure 6.2 shows the contours where the ‘n’ crest factor contour remains flat due to the accurate gain level application of 0 dBFS. Each contour above the ‘n’ crest factor contour is the next crest factor level so ‘n-1’ and then ‘n-2’ and so on. The y axis is flipped to show clearly when an octave band is louder because the subject will have attenuated the noise more to match the reference noise. This flipping of the y axis demonstrates how for a louder playback level the loudness perception for different crest factor values across different octave bands do not follow the inherent RMS changes.

Table F.2 (see Appendix F) demonstrates the specific octave bands that gave a significant difference between the 50 dB SPL playback level and the 70 dB SPL playback level for the transient amplitude characteristic. There are nine overall crest factor levels that had significant different across five different octave bands. These octave bands are the ones with centre frequencies of 125 Hz, 250 Hz, 500 Hz, 2 kHz and 4 kHz. Again the most interesting octave band is the centre of 125 Hz octave band. Similar to the continuous amplitude characteristic, at moderate playback levels (70 dB SPL) there was a greater spread in the results and a more attenuated median gain level across the different crest factor levels. **Figure 6.3** shows the crest factor contours for transient amplitude characteristics and the perceived difference as a result of different playback levels.

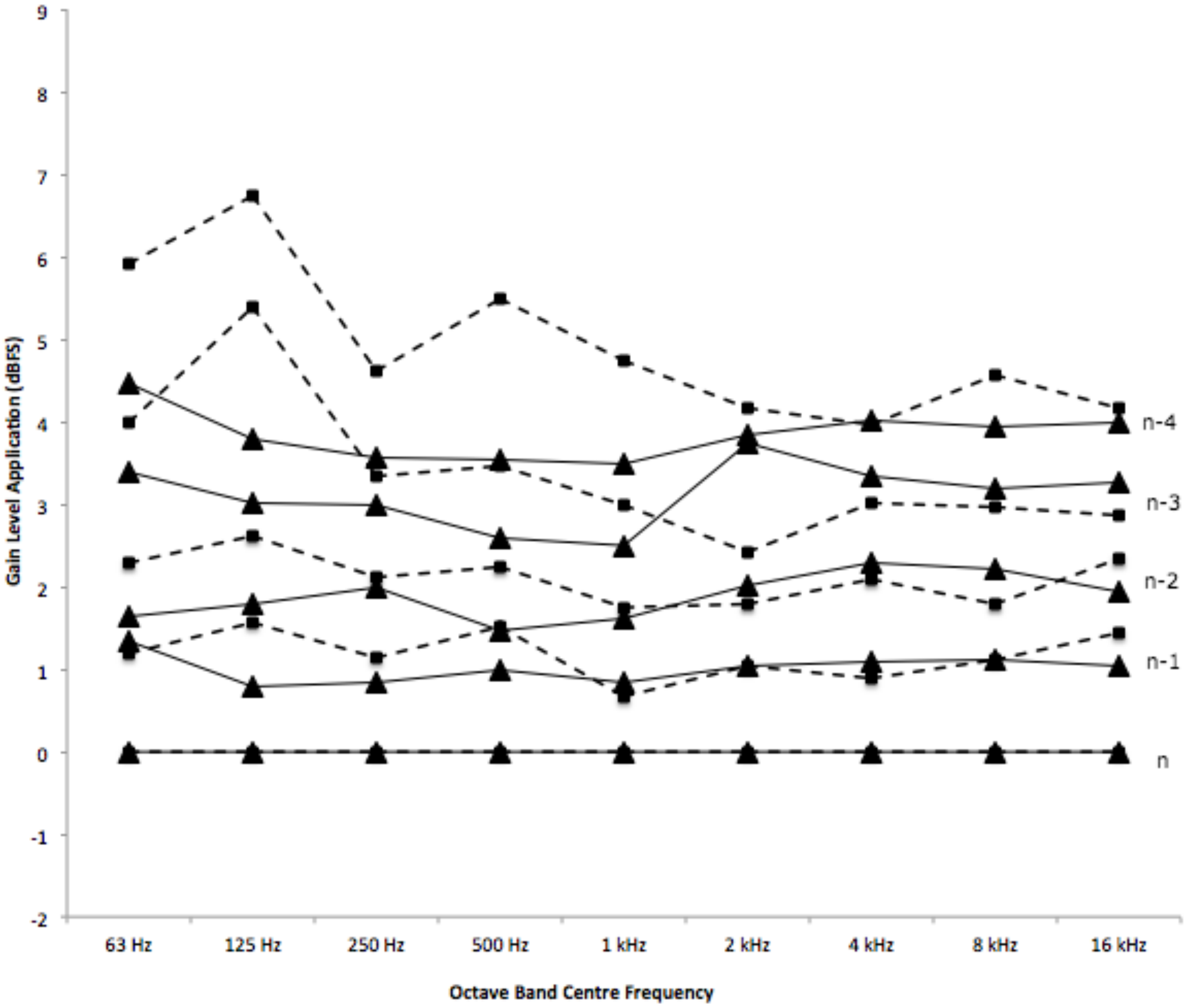


Figure 6.3 shows the contours of different crest factor levels across the different octave bands for transient amplitude characteristics. The **–** represents the playback level of 50 dB SPL and the --- represents the playback level of 70 dB SPL. The ▲ shows the median values for the contours for the 50 dB SPL playback level. The ■ shows the median values for the contours for the 70 dB SPL playback level. The graph is flipped is demonstrate when an octave band is determined as louder.

Within **figure 6.3** it can be noted that when the playback level is louder, the gain level application did not match the inherent RMS change as a result of the crest factor change. This follows the same trend as found for continuous amplitude characteristics.

Interestingly when comparing different amplitude characteristics at the same playback level there were few significant differences. This suggests that at static playback level, the resultant DRC application in terms of crest factor level is equally noticeable independent of whether the noise has a continuous amplitude characteristic or a transient amplitude characteristic

6.6.4 Irregularity of Loudness Perception in Lower Frequency Octave Bands

The lower frequency octave bands with centre frequencies of 63 Hz and 125 Hz gave median gain application levels past both what was expected and what was found in other octave bands but only within the moderate playback level listening tests. The octave band with a centre frequency of 63 Hz displayed irregular results at the ‘n-4’ crest factor only however the octave band with a centre frequency of 125 Hz displayed irregular results for the ‘n-2’, ‘n-3’ and ‘n-4’ crest factor levels. This could suggest that this octave band does not have a linear relationship between loudness perception changes and the crest factor level changes. This could suggest that for final stage DRC the frequency band around 125 Hz should be considered more carefully and may require further work to explore why this frequency band behaved differently. The irregularity is potentially important when considering the objective loudness measurements. The LUFS values found in **figure C.1** to **figure C.4** suggest that the loudness only changes by 1 dBFS whereas the subjective values have variations between the different crest factor levels and can be closer to changes of 3 dBFS. This suggests that the LUFS measurements do not accurately represent the perceived change in loudness of audio as a result of DRC.

6.6.5 Potential Implications on Perceived Quality

Croghan *et al* [2012] demonstrated how the quality of audio degraded as the DRC application increased and this was supported in 2014 by Wendl & Lee [2014. Though both studies have similar conclusions the method used differs and the methodology used for this thesis has comparisons with the Wendl & Lee [2014] methodology. Croghan *et al* [2012] focused the changes in DRC by using the threshold parameter instead of a calculated output of the DRC. The limitation this created was that the study may be limited to the characteristics of the certain DRC used for the test because whilst the threshold level changed it could be the case that the DRC is more aggressive than other DRC or less aggressive. By measuring the resultant change that the DRC had as done by Wendl & Lee [2014] and as done within this thesis, the DRC characteristic is eliminated. This is the why the crest factor level is used for this thesis. The crest factor level can be achieved by any DRC but the relationship to loudness perception and quality perception is therefore more conclusive that using just the threshold.

Even with the differences, the studies suggested that DRC can be detrimental to audio quality. If this is assumed then it could be suggested that certain octave bands after DRC application either have a greater impact on audio quality or that certain octave bands are more prone to quality degradation individually. The lower octave bands as discussed in section 6.6.4 could potentially degrade the quality. Therefore a mastering engineer when using DRC either across the entire audio or when isolating using a multi-band compressor should be cautious about the differences in DRC perception when the playback level changes. For the moderate level, which could be assumed as more common, the lower octave bands could potentially degrade the audio greater than other octave bands.

6.7 LIMITATIONS

A limitation of the study is that the frequency response of the playback system may need to be compared to multiple playback systems to account for the colourisation that audio via a playback system. The alteration of the frequency response for this study aimed to remove any colourisation of the equipment used and to therefore determine a fundamental understanding of loudness perception changes across octave bands and amplitude characteristics as a result of DRC. The alterations however may have impacted the results as suggested by irregular results at lower octave bands. The lower octave bands had received more EQ application to correct the playback system. Though the removal of the playback systems colourisation was executed as accurately and methodically as possible it may be that playback systems carefully adjust the response at these frequencies to not give these irregularities for audio. This could impact the claim made in regards to LUFS not accurately predicting the loudness of the octave band with centre frequency of 125 Hz.

Another limitation was that subject number. Only eight expert listeners were used whereas a greater number would give a more accurate equal loudness contour and a more accurate median average for gain level application. The expert listeners do give a strong set of results however due to their experience with analysing audio and having no hearing damage but this low number could impact the results.

Finally the SPL measurements for the transient levels gave difficulties in the recording of accurate measurements due to short sample time causing problems when measuring the average SPL. The chosen method was to use the peak SPL level which was different from the average SPL level as used for the continuous amplitude characteristic. The SPL levels were kept consistent across the transient amplitude characteristic however so the comparisons can still be made.

7 SUMMARY AND CONCLUSIONS

7.1 Summary

The overall aim of the research presented within this thesis was to find the how DRC fundamentally affects different frequency bandwidths with changing amplitude characteristics and playback levels. The thesis involved:

- A background of the DRC application within the audio industry including the loudness war and progressions of the use
- A brief outline of the functionality of the auditory system including inherent filters and the auditory nerve fire rate
- A review of equal loudness contours for pure tones including methodologies, implications of the equal loudness contours, phon and sone calculations and resultant frequency weightings
- A review of calculating the loudness of complex noise with focus on critical bands, equal loudness contours for complex noise including methodologies, temporal aspects and objective measurements for loudness
- A study of dynamic range compression including types of DRC, parameters to control DRC, the impact it has on loudness, the impact it has on quality and methods to measure the effects of DRC
- The development of listening tests to acquire equal loudness contours for octave bands and to acquire results to determine the loudness perception changes as a result of DRC. This included:

    - Experimental design
    - Stimulus creation
    - Reproduction system design and alterations required
    - Preliminary Test
- The equal loudness contour for octave bands which included the methodology to acquire results, analysis of results, discussion of results and limitations
- The study of loudness perception changes in relation to crest factor variation which included objective loudness measurements, a methodology to acquire the results, an analysis of the results, a comparison to objective loudness measurements a discussion of results and limitations

The compression levels within audio has seen popular music productions being labelled as hypercompressed. A hypercompressed track has been labelled poorer quality and lead to loudness normalisation being implemented across many playback forums via the use of objective loudness calculations.

The review of the functionality of the auditory system highlighted how the inherent auditory filters are related to the auditory nerve fibres based along the basilar membrane. There are many auditory filters which overlap however for a quieter noise less fibres respond and for louder noises more auditory fibres respond due to the noise falling within the cutoff frequencies of more filters. This was related to the auditory fire rate and especially to the ERB method of calculating loudness as a result of the excitation across the basilar membrane.

This helped to understand how different frequencies are perceived and how different loudness of different frequencies are received though the actual equal loudness contours demonstrate the non-linearity of the ear. By reviewing different methods including the use of a 1 kHz reference tone and the change in amplitude of comparison frequencies, an understanding into how to evaluate the non-linearities of the ear was accomplished. From these studies dB SPL levels were better understood especially via the analysis of the phon and sone calculations. The non-linearities of the auditory system based upon equal loudness contours of pure tones helped to develop different frequency weightings in regards to calculating the loudness of frequencies. Different frequency weighting were based upon the contours for different playback levels and by inverting the contour of a weighting, considerations in the loudness calculations were established.

For the loudness of complex noise the work on critical bands as reviewed in this thesis, it was found that there was a threshold of the bandwidth of the noise where when breached would contribute to loudness judgements. For different frequencies the bandwidth varied and can be calculated with ERB theory. The listening tests considered to determine equal loudness contours for complex noise were quite different in the process. The reference noise was a factor that changed often though to judge the relationship between octave bands there had been no study that used an octave band with 1 kHz octave band as the reference noise. This could partially undermine these listening tests.

The temporal aspects of calculating loudness have shown that if the noise is shorter than a 100 ms time frame, the noise is perceived quieter than if the noise was longer than 100 ms for the same playback level. The relationship appears to be a linear relationship between the length of the audio and the loudness level up to the 100 ms. The objective measurements for loudness mainly focused on the recent LUFS measurement and the method to calculate in LUFS. It was discovered that the first process is the use of a custom frequency weighting called the K-weighting. K-weighting is a modified inverted equal loudness contour of pure tones at 70 dB SPL. The modification was the increased weighting towards higher frequencies which actually appeared similar to an inverted equal loudness contour for complex noise. This suggests that the weighting may be accurate.

Different DRC parameters include threshold, attack, release, ratio, knee and make-up gain. A type of DRC called a limiter was identified as one of the main tools used to hypercompress audio and the use of DRC via limiters was the basis of many studies. These studies found that DRC does change the perceived loudness of audio when peak normalised but more importantly studies found that the quality degraded as DRC application increased.

From reviewing previous test designs it was decided to use an octave band with centre frequency of 1 kHz as a reference noise across the tests. The equal loudness tests used octave bands as the comparison tones with the octave bands being in the form of a continuous amplitude characteristic and a transient characteristic. The reproduction system had a Genelec 8040A as the speaker but the audio was processed through an external graphic EQ to correct the colouration of the speaker. After a preliminary test it was decided that two playback levels should be used with these being 50 dB SPL and 70 dB SPL recorded with the reference noise.

The subjects used an interface with sliders to amplify or attenuate the comparison signals and the levels recorded. This gave equal loudness contours for octave bands which were found to behave similar to equal loudness contours for pure tones at the 50 dB SPL level but for the 70 dB SPL there behaviour of higher frequencies. For the continuous amplitude characteristic the equal loudness contour supported the K-weighted filter used for the LUFS measurement but for the transient amplitude characteristic this was not so. The difference was suggested that the annoyance factor of the higher frequencies is found for continuous amplitude characteristic but not for the transient amplitude characteristic due to temporal aspects.

Using the results for the equal loudness contours test, the median average level was applied to the stimulus to create an equally loud set of stimulus. This stimulus then had DRC applied at varying levels to produce further stimulus with varying levels of crest factor 1 dBFS apart. Subjects then used a similar slider system to amplify or attenuate the stimulus when compared to the uncompressed stimulus. The results suggest that that across all octave bands the perceived loudness increases with DRC application however the transient amplitude characteristic gave a better indication of DRC application. It was also suggested that the octave band with a centre frequency of 125 Hz gave the best indication to DRC application. When comparing this octave band with the objective loudness measurements, the subjective loudness measurements suggested that the LUFS method may not be totally accurate at predicting the loudness.

7.2 CONCLUSIONS

7.2.1 Main Points

The conclusions from this hypothesis are:

- Equal loudness contours for octaves bands at low playback levels such as 50 dB SPL are similar to equal loudness contours found for pure tones
- Equal loudness contours for octave bands at moderate playback levels such as 70 dB SPL are different from equal loudness contours for pure tones. The higher frequencies (octave band with a centre frequency of 16 kHz) were perceived as louder instead of quitter.
- The high frequency content of continuous amplitude characteristics is perceived as louder than the high frequency content of transient amplitude characteristics at a moderate playback level
- The equal loudness contour for octave bands at 70 dB SPL for a continuous amplitude characteristic supports the K-Weighted filter used for LUFS
- The equal loudness contours for octave bands at a lower playback suggest that LUFS measurements are not accurate for low level playback
- The perceived loudness of all octave bands increases as DRC increases
- Transient amplitude characteristics at low level playback has the most accurate indicator of crest factor changes due to accurate perceptions of loudness changes
- The loudness perception of crest factor changes as a result of DRC at moderate playback levels do not match the inherent RMS change and

therefore indicate that at moderate playback levels there is a non-linear relationship between RMS change and loudness perception change

- The octave band with a centre frequency of 125 Hz has a non-linear relationship between loudness perception changes and crest factor level changes. As DRC increased ad therefore the crest factor increased, the median gain level did not correspond to the inherent RMS change of the stimulus (RMS changes due to the crest factor calculation). The median gain levels were near 4 dBFS over what the RMS change corresponded to.
- The octave band with a centre frequency of 125 Hz was not correctly portrayed with objective loudness measurements

7.2.2 Implications of Conclusions

The equal loudness contours for octave bands differ from previous equal loudness contours for complex noise. Importantly differences between the transient amplitude characteristic and the continuous amplitude characteristic for the higher frequencies show that the temporal aspects effect loudness perception for this frequency range. The transient amplitude characteristic was found to be quieter whereas the continuous amplitude characteristic was found to be louder. The feedback from subjects suggested this may be due to an annoyance factor. For a practical implication this would suggest that continuous high frequency content should be attenuated for when the content is played back at a greater level, the annoyance factor could ruin the quality of audio. An example would be the mixing of drums in popular music or rock music where the cymbals may ring out for a long time and cause annoyance. However the hi-hat of a drum kit would not express this annoyance and so could sit higher in the mix without problems if desired.

The differences in the equal loudness contours for octave bands when played back at different levels also highlighted how LUFS calculations may be inaccurate. K-Weighting is based upon a contour at a playback at 70 dB SPL which the 70 dB SPL contour found as part of this thesis supported however the changes of the playback level shows how the contours change. LUFS therefore uses inaccurate calculations to determine the loudness of content. A solution needs to be within the metadata of the format of the audio that communicates with the playback system. This would allow for the changing of the content for the playback level or situation and therefore give a better objective loudness level that could in turn potentially give a better quality across differing platforms.

The changes in loudness perception to crest factor changes as a result of DRC suggests that at moderate playback levels the loudness of lower frequency octave bands do not correspond to inherent changes to the crest factor. This could therefore mean that if audio is compressed then when listened back to at a moderate level the lower frequency ranges could be a lot louder than anticipated. Mastering engineers therefore should be cautious with DRC application and should monitor the changes made to the lower frequencies. With previous studies suggesting that as perceived loudness increases as a result of DRC, the perceived quality degrades, this could potentially suggest that quality perception may be related to DRC levels of lower frequency ranges.

7.3 FURTHER WORK

Conducting the same tests across a larger subject number would enable a further study into whether the conclusions made are accurate.

The scope for further work includes applying DRC to produce crest factors past a 4 dBFS difference. By examining the DRC effect for levels past a 4 dBFS change a study may find how octave start to behave differently in regards to loudness perception. If this were found then certain frequency bands when heavily compressed would affect the loudness perception but potentially leave objective measurements with a lower reading.

Another study would be to apply the findings from this thesis to popular music productions of different genres to see if the fundamental translates well to actual audio. If there is a change it may suggest that there is more than just inter frequency bandwidths that contribute towards loudness perception. Also it would be useful to study the change in quality perception as the different frequency bandwidths receive different levels of DRC to determine if DRC degrades quality on a localised basis like how it does for when DRC is applied across the entire frequency spectrum.

A study into how playback systems respond to different DRC levels could be followed up from this thesis from the questions raised by the frequency response alterations and whether that affected the final results. This information would potentially help audio engineers avoid mistakes within final stage DRC because of how playback systems translate the DRC application.

Appendix A Crest Factor Calculations for Stimulus with Playback Level Considerations and Amplitude Characteristics

Figure A.1 to figure A.4 tables to show crest factor calculations for the different amplitude characteristics and the different playback levels.

Continuous at 70 dB SPL	Centre Frequency								
	65	**125**	**250**	**500**	**1000**	**2000**	**4000**	**8000**	**16**
Peak (dBFS)	-2.1	-4.06	-11.93	-8.29	-10.95	-13.05	-13.24	-10.97	-1
RMS (dBFS)	-13.99	-18.1	-24.96	-21.95	-24.36	-26.53	-26.8	-24.29	-2
Crest Factor (dBFS)	11.89	14.04	13.03	13.66	13.41	13.48	13.56	13.32	1

Figure A.1

Transient at 70 dB SPL	Centre Frequency								
	65	**125**	**250**	**500**	**1000**	**2000**	**4000**	**8000**	**160**
Peak (dBFS)	-13.46	-12.09	-20.58	-14.7	-13.45	-15.49	-17.66	-14.89	-9.
RMS (dBFS)	-28.79	-27.67	-35.62	-33.34	-32.84	-35.17	-37.79	-34.17	-28.
Crest Factor (dBFS)	15.33	15.58	15.04	18.64	19.39	19.68	20.13	19.28	1

Figure A.2

Continuous at 50 dB SPL	Centre Frequency								
	65	125	250	500	1000	2000	4000	8000	16000
Peak (dBFS)	-0.17	-0.1	-6.98	-10.94	-10.95	-12.99	-15.81	-12.64	-8.34
RMS (dBFS)	-12.06	-12.31	-20.01	-24.61	-24.36	-26.47	-29.38	-25.96	-20.99
Crest Factor (dBFS)	11.89	12.21	13.03	13.67	13.41	13.48	13.57	13.32	12.65

Figure A.3

Transient at 50 dB SPL	Centre Frequency								
	65	125	250	500	1000	2000	4000	8000	16000
Peak (dBFS)	-6.86	-6.24	-16.63	-15.19	-13.45	-16.74	-18.08	-15.07	-10.12
RMS (dBFS)	-22.18	-21.82	-31.68	-33.83	-32.84	-36.42	-38.21	-34.35	-29.02
Crest Factor (dBFS)	15.32	15.58	15.05	18.64	19.39	19.68	20.13	19.28	18.9

Figure A.4

Appendix B Frequency Analysis of the Process to Correct the Frequency Response of the Playback System

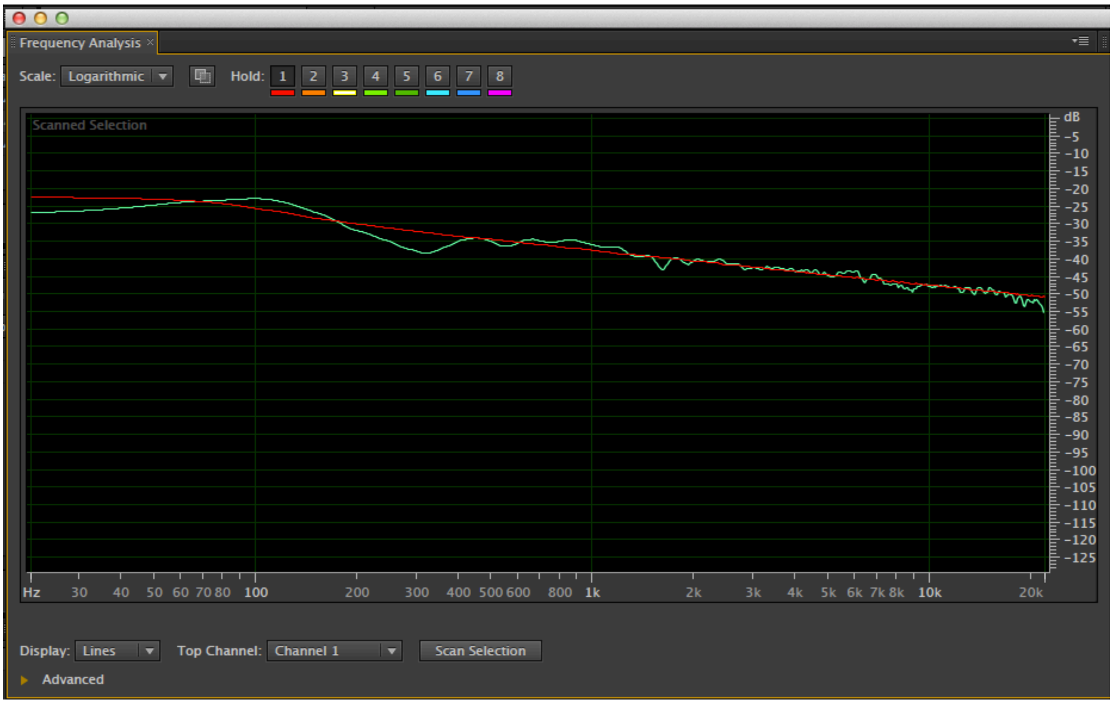


Figure B.1 Frequency Analyser screenshot of the starting Genelec 8040A response. Note the large attenuation around 300Hz and the boost around 125 Hz.

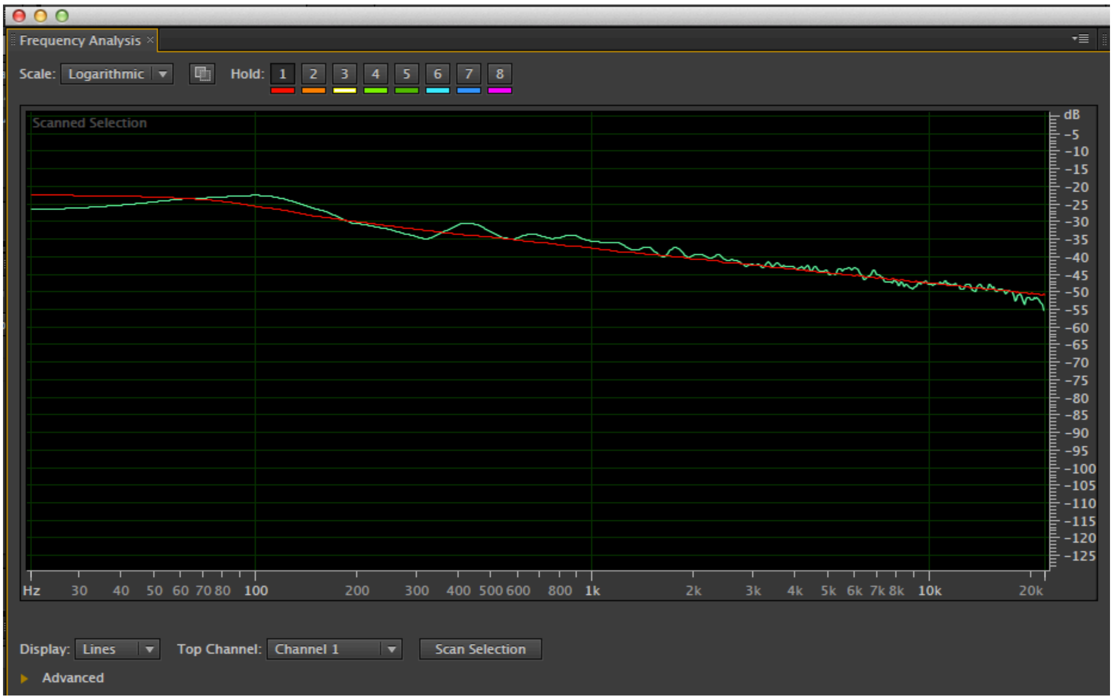


Figure B.2 First EQ corrections

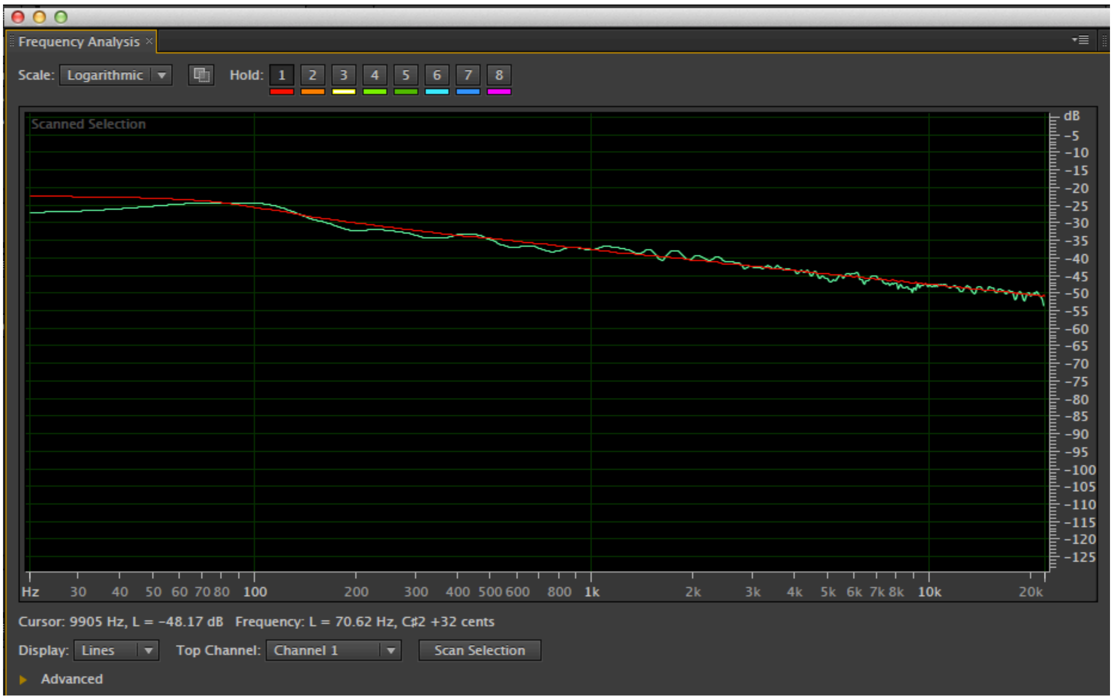


Figure B.3 Second EQ corrections. Note the 300 Hz attenuation has been changed.

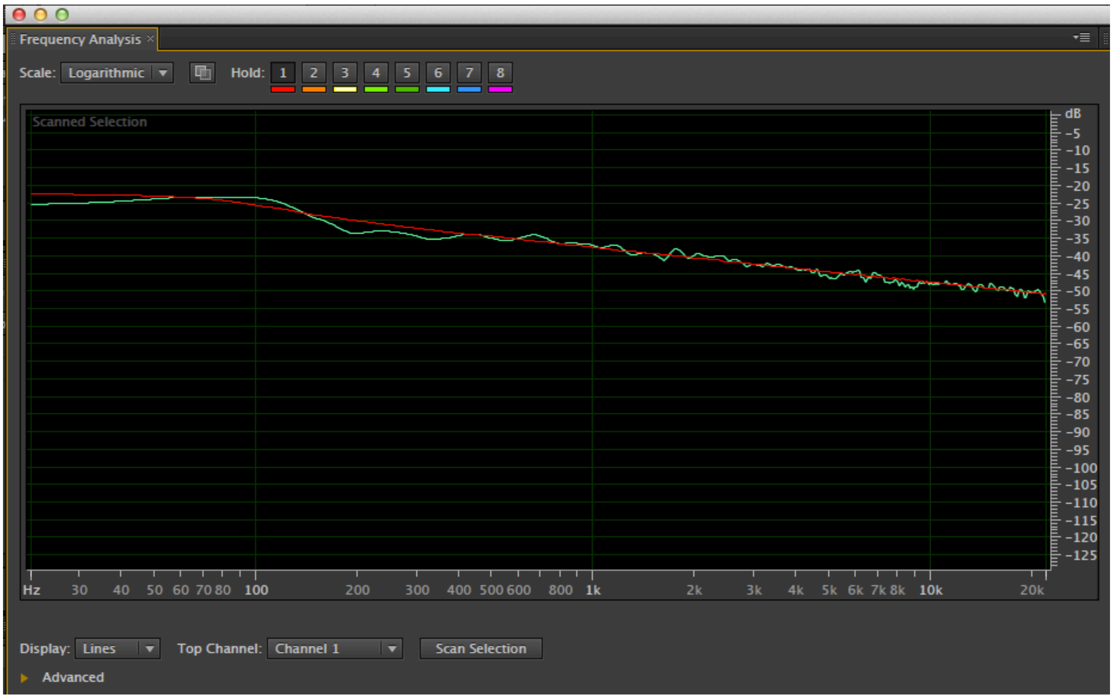


Figure B.4 Further EQ corrections however the sensitivity to slight changes becomes apparent

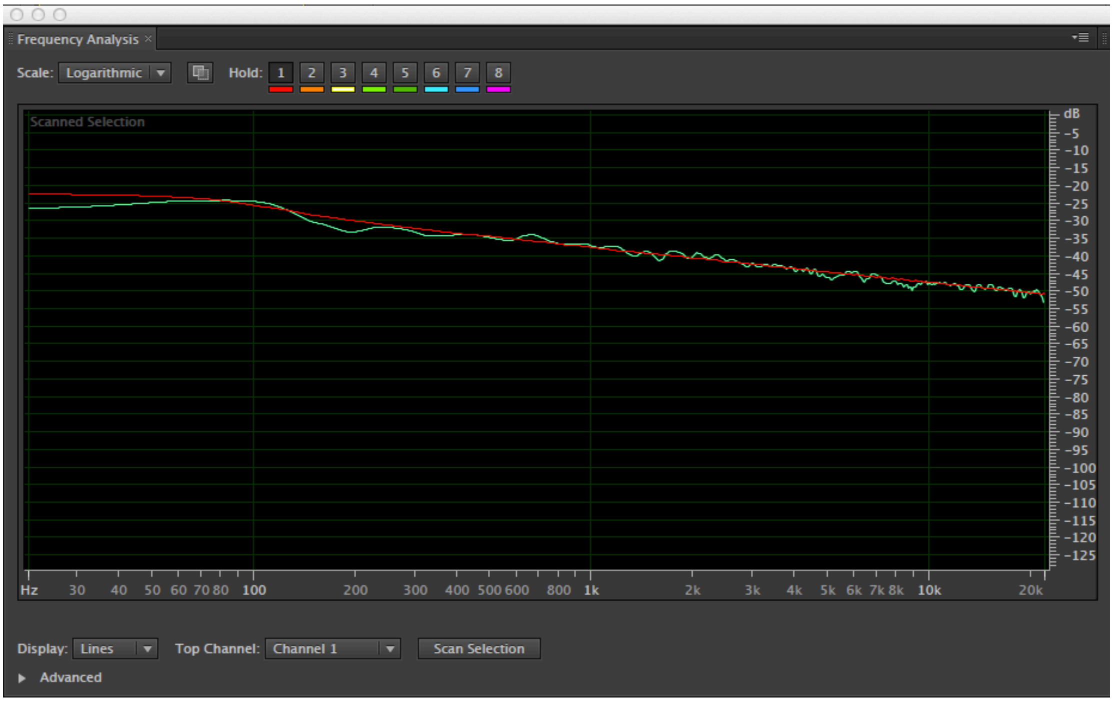


Figure B.5 FInalising EQ adjustments though the 300 Hz attenuation is returning

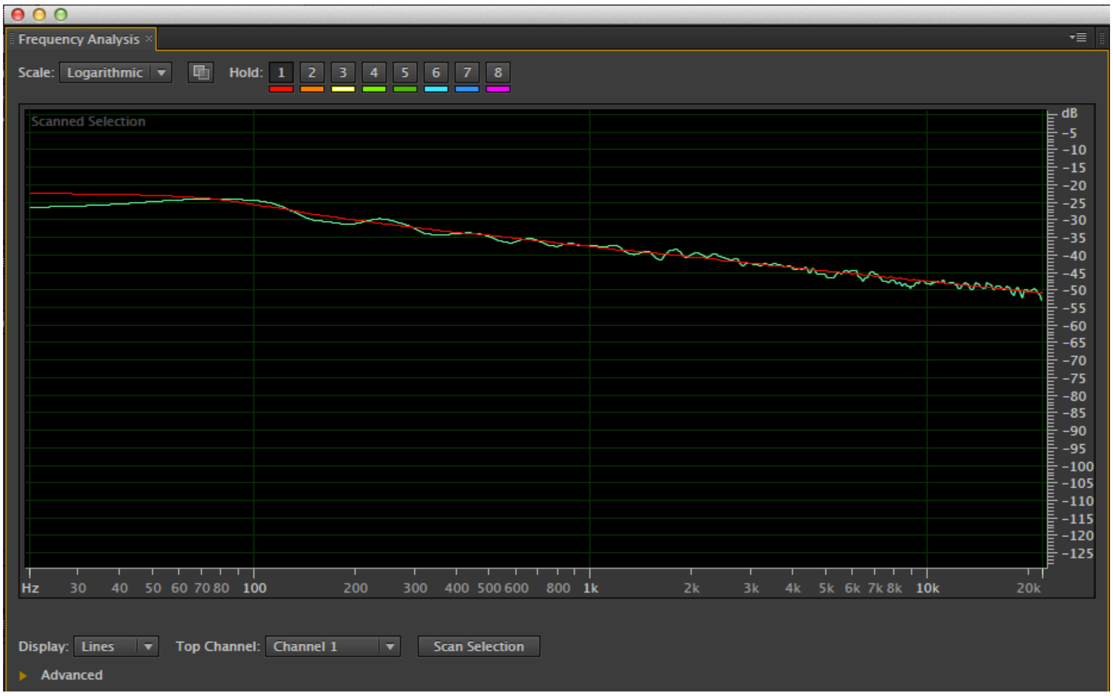


Figure B.6 Resultant EQ corrections

Appendix C LUFS Measurements for the Stimulus

Figure C.1 LUFS measurements for the continuous amplitude characteristics with a 50 dB SPL playback

	50 dBSPL				
Centre Frequency of Octave Band	**Continuous Crest Factors**				
	n	**n-1**	**n-2**	**n-3**	**n-4**
65	-15.11	-14.12	-13.23	-12.09	-11.11
125	-13.71	-12.71	-11.73	-10.68	-9.69
250	-20.89	-19.9	-18.91	-17.93	-16.89
500	-25.28	-24.29	-23.29	-22.3	-21.24
1000	-24.27	-23.37	-22.28	-21.3	-20.36
2000	-24.2	-23.2	-22.21	-21.23	-20.22
4000	-26.17	-25.17	-24.17	-23.2	-22.19
8000	-22.65	-21.65	-20.66	-19.59	-18.62
16000	-17.68	-16.69	-15.69	-14.74	-13.84

50 dB SPL

Centre Frequency of Octave Band	Transient Crest Factors n	n-1	n-2	n-3	n-4
65	-28.42	-28.39	-28.34	-28.29	-24.46
125	-26.14	-25.1	-24.12	-23.15	-22.26
250	-35.72	-34.68	-33.69	-32.72	-31.76
500	-37.5	-36.5	-35.51	-34.52	-33.55
1000	-35.83	-34.85	-33.87	-32.85	-31.91
2000	-37.2	-36.17	-35.23	-34.27	-33.28
4000	-38.11	-37.13	-36.14	-35.17	-34.15
8000	-34.11	-33.11	-32.14	-31.14	-30.21
16000	-28.82	-27.79	-26.83	-25.84	-24.91

Figure C.2 LUFS measurements for the transient amplitude characteristics with a 50 dB SPL playback

70 dB SPL					
Centre Frequency of Octave Band	**Continuous Crest Factors**				
	n	**n-1**	**n-2**	**n-3**	**n-4**
65	-15.11	-14.12	-13.23	-12.09	-11.11
125	-19.5	-18.5	-17.5	-16.5	-15.5
250	-25.84	-24.84	-23.85	-22.78	-21.84
500	-22.63	-21.63	-20.63	-19.65	-18.68
1000	-24.74	-23.27	-22.28	-21.3	-20.36
2000	-24.26	-23.26	-22.26	-21.29	-20.28
4000	-23.59	-22.59	-21.6	-20.53	-19.53
8000	-20.98	-19.98	-18.99	-18.02	-17.04
16000	-21.88	-20.89	-19.89	-18.94	-18.04

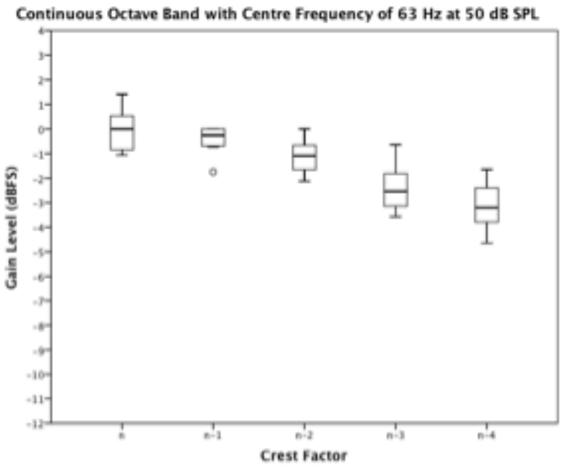
Figure C.3 LUFS measurements for the continuous amplitude characteristics with a 70 dB SPL playback

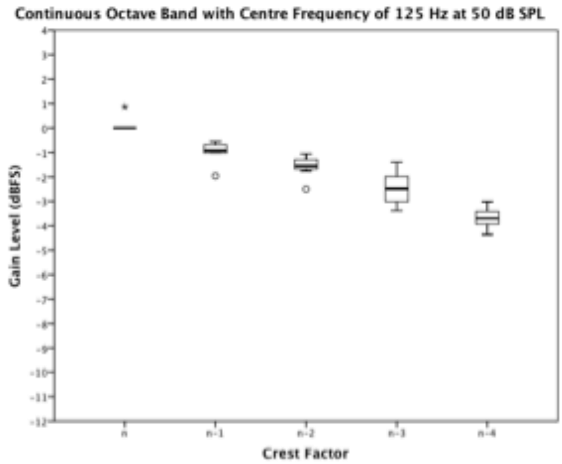
70 dB SPL

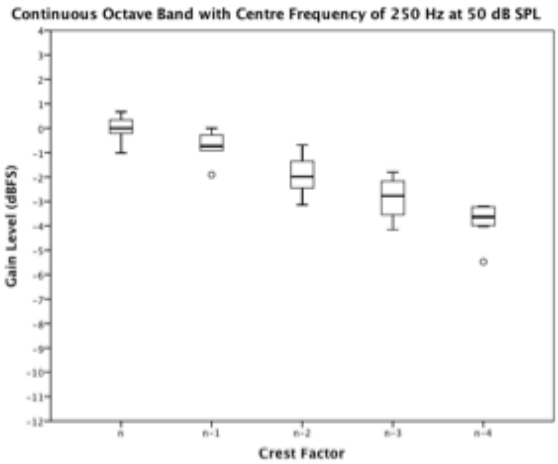
Centre Frequency of Octave Band	Transient Crest Factors				
	n	**n-1**	**n-2**	**n-3**	**n-4**
65	-35.05	-34.07	-33.05	-32.05	-31.1
125	-31.94	-31.02	-30.04	-29.04	-28.15
250	-39.67	-38.72	-37.65	-36.69	-35.72
500	-37.02	-36.01	-35.02	-34.03	-33.01
1000	-35.83	-34.85	-33.87	-32.9	-31.91
2000	-35.95	-35.04	-34.06	-33.1	-32.1
4000	-37.69	-36.72	-35.72	-34.76	-33.73
8000	-33.93	-32.93	-31.95	-30.95	-30.02
16000	-28.45	-27.52	-26.49	-25.53	-24.62

Figure C.4 LUFS measurements for the transient amplitude characteristics with a 70 dB SPL playback

Appendix D Graphs of Median Gain Levels with Inter Quartile Ranges







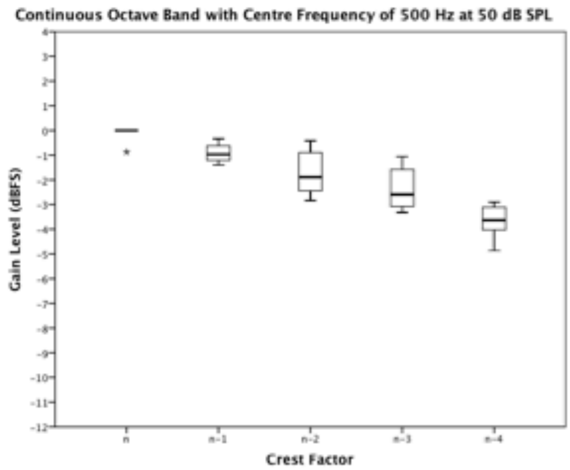


Figure D.1 (top left) to **figure D.4** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

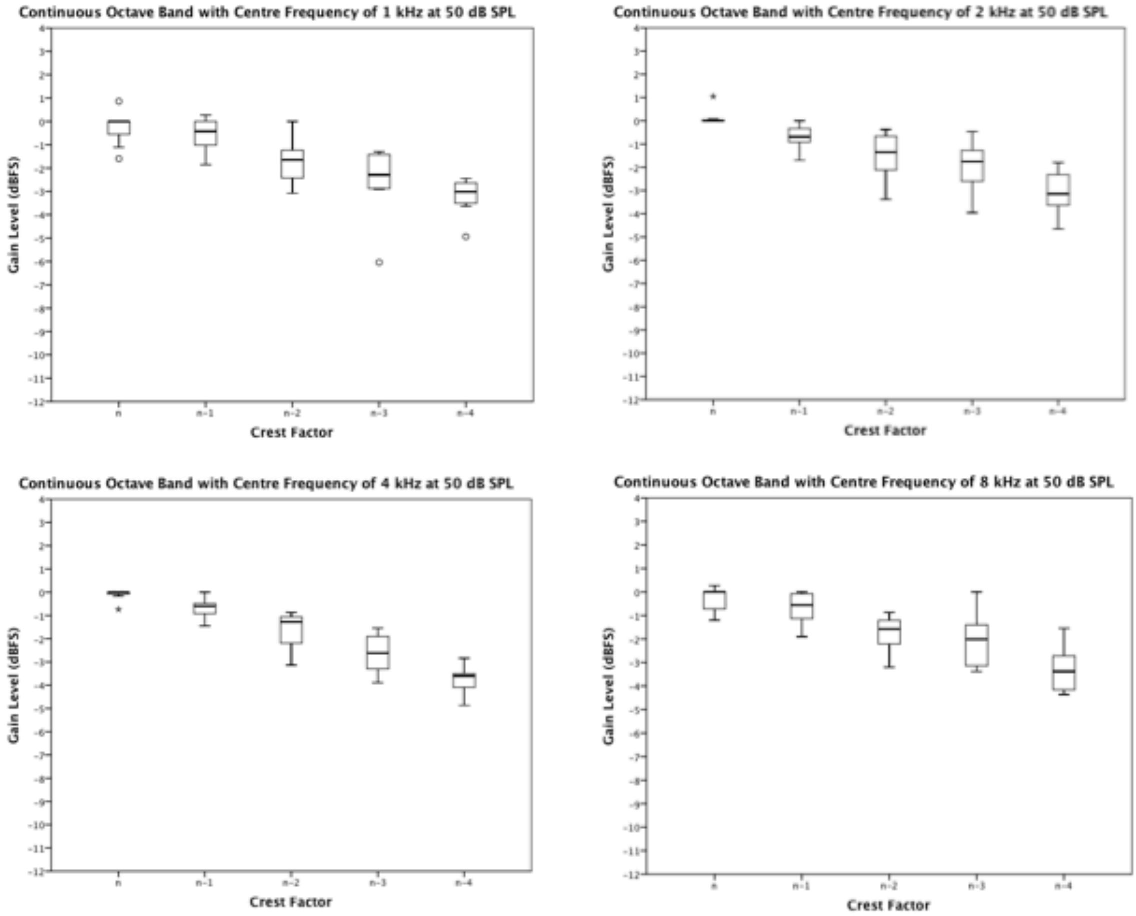


Figure D.5 (top left) to **figure D.8** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

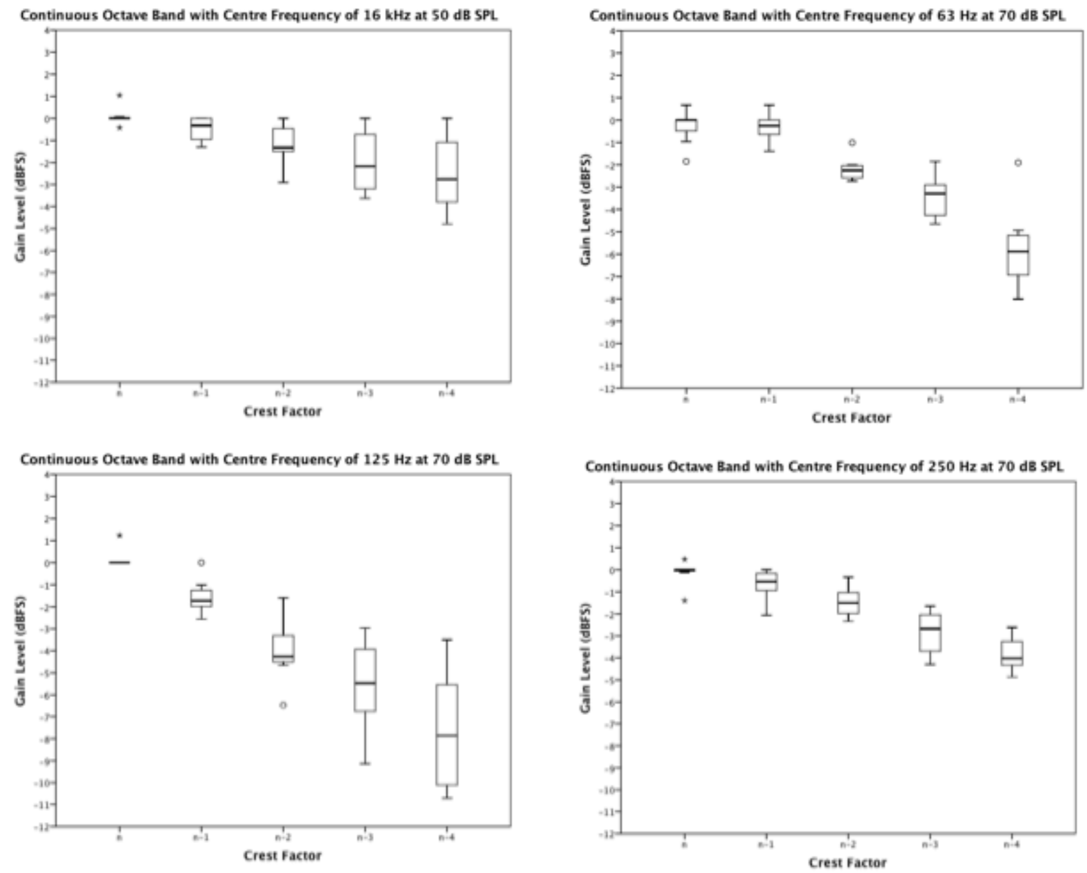


Figure D.9 (top left) to **figure D.12** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

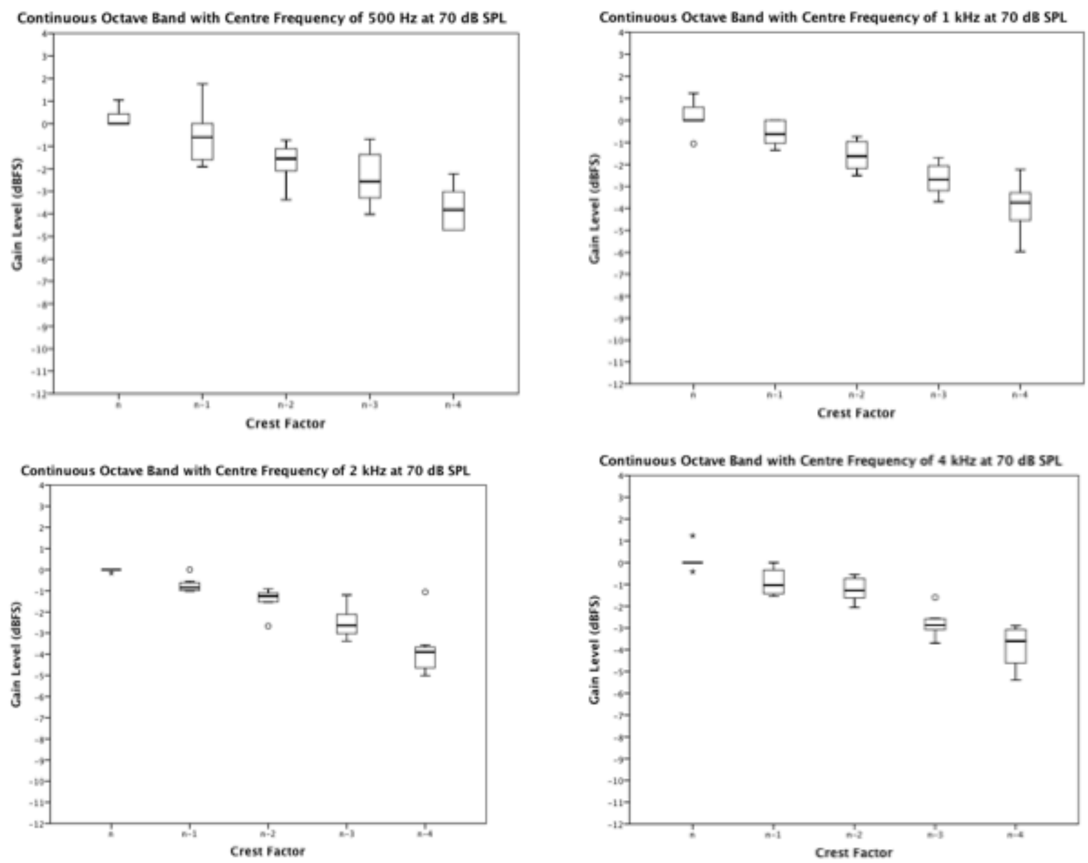


Figure D.13 (top left) to **figure D.16** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

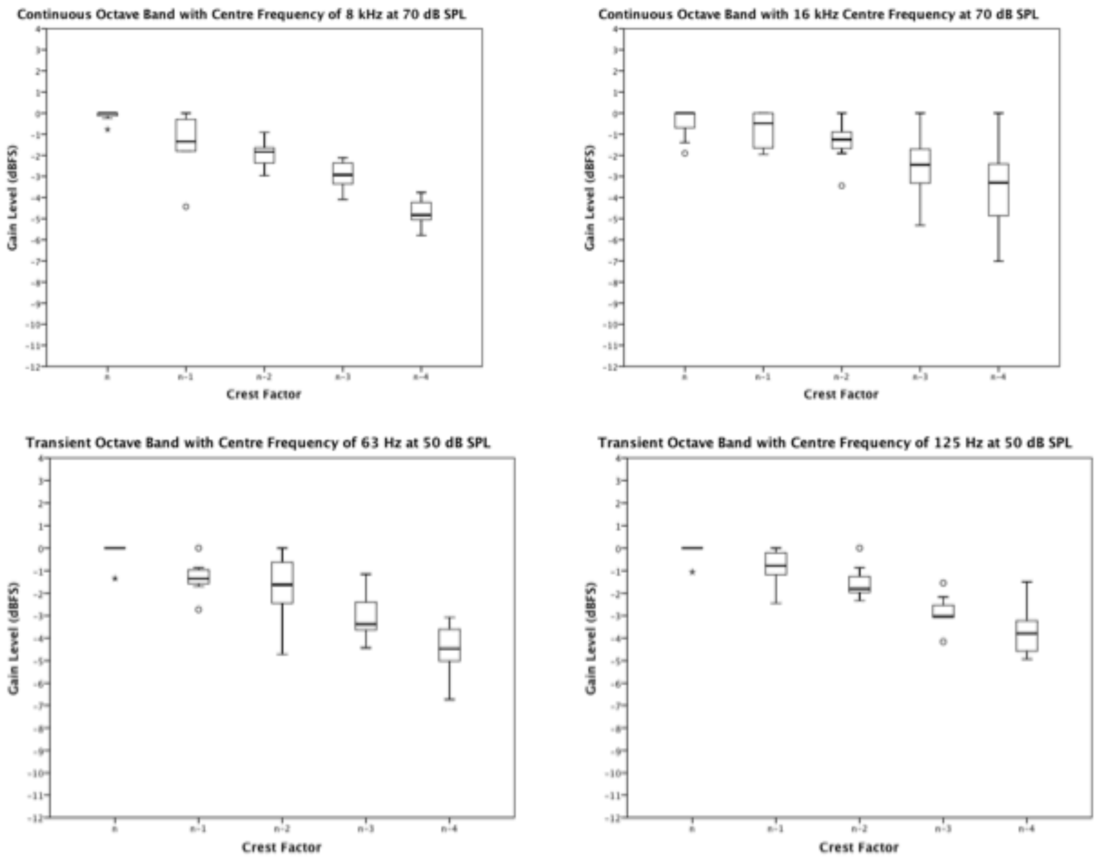


Figure D.17 (top left) to **figure D.20** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

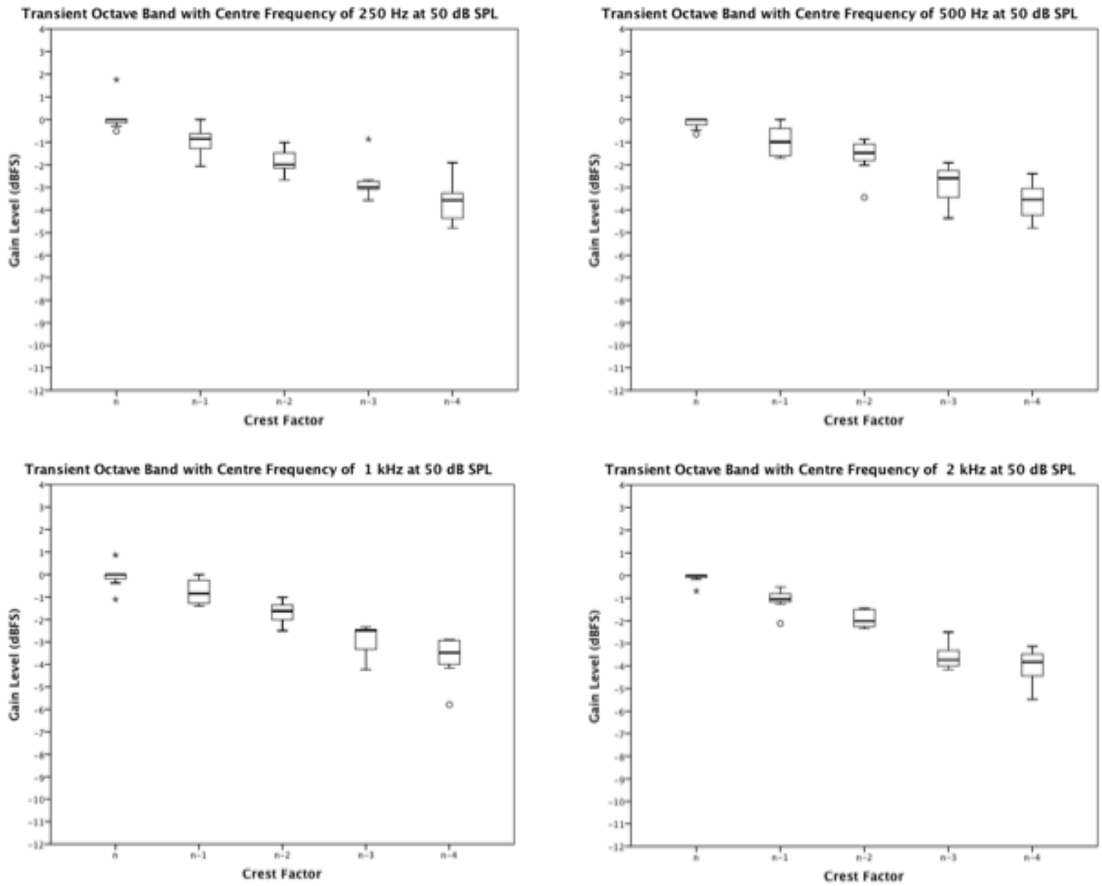


Figure D.21 (top left) to **figure D.24** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

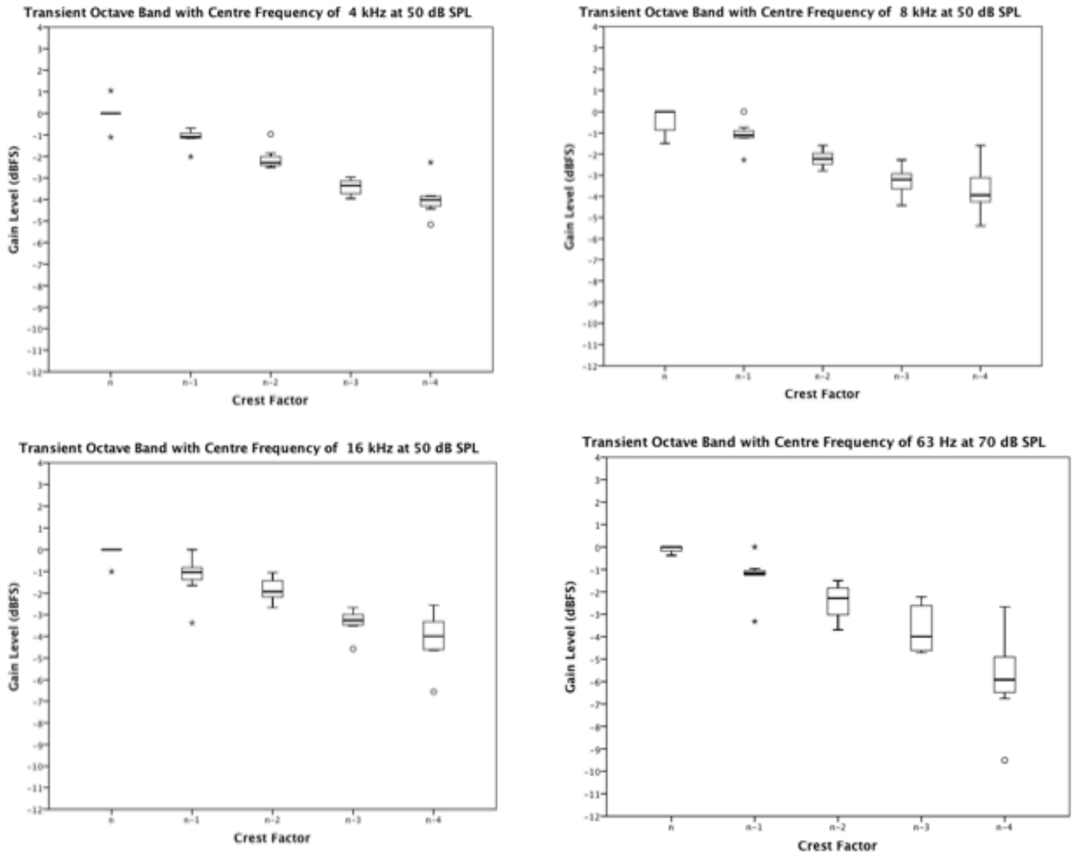


Figure D.25 (top left) to **figure D.28** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

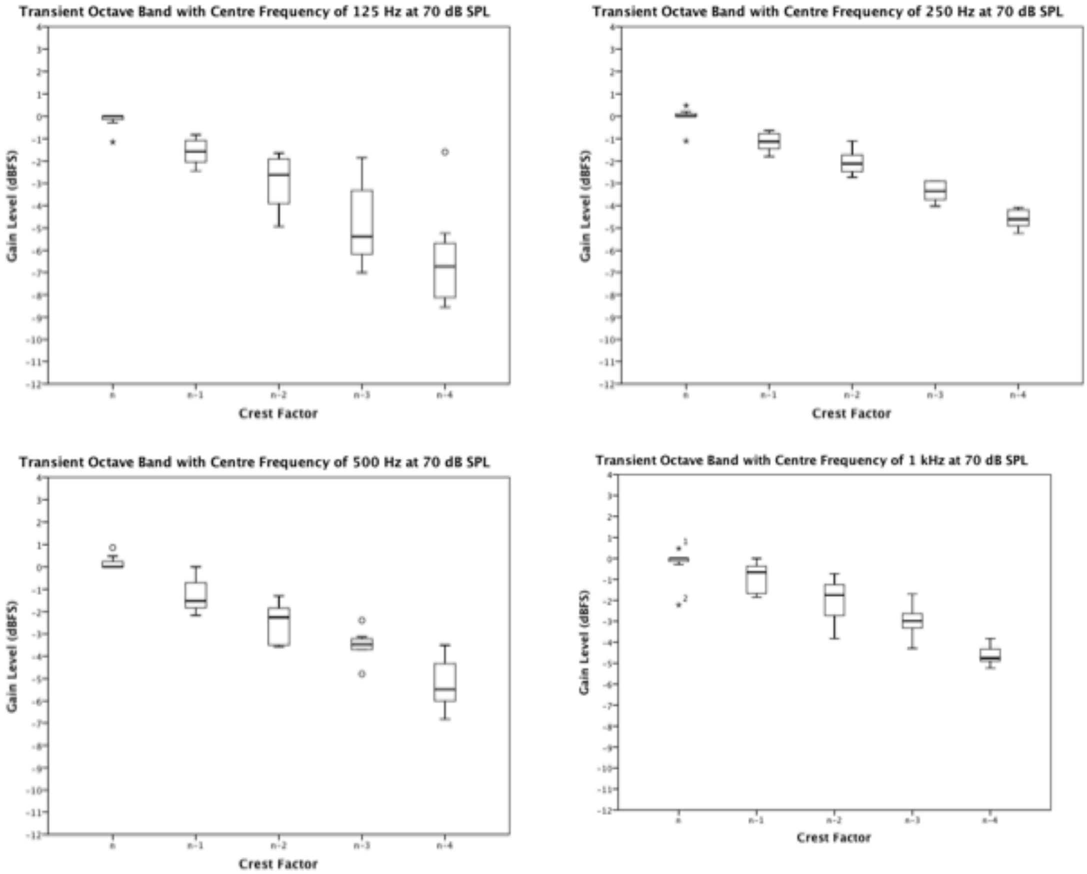


Figure D.29 (top left) to **figure D.32** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

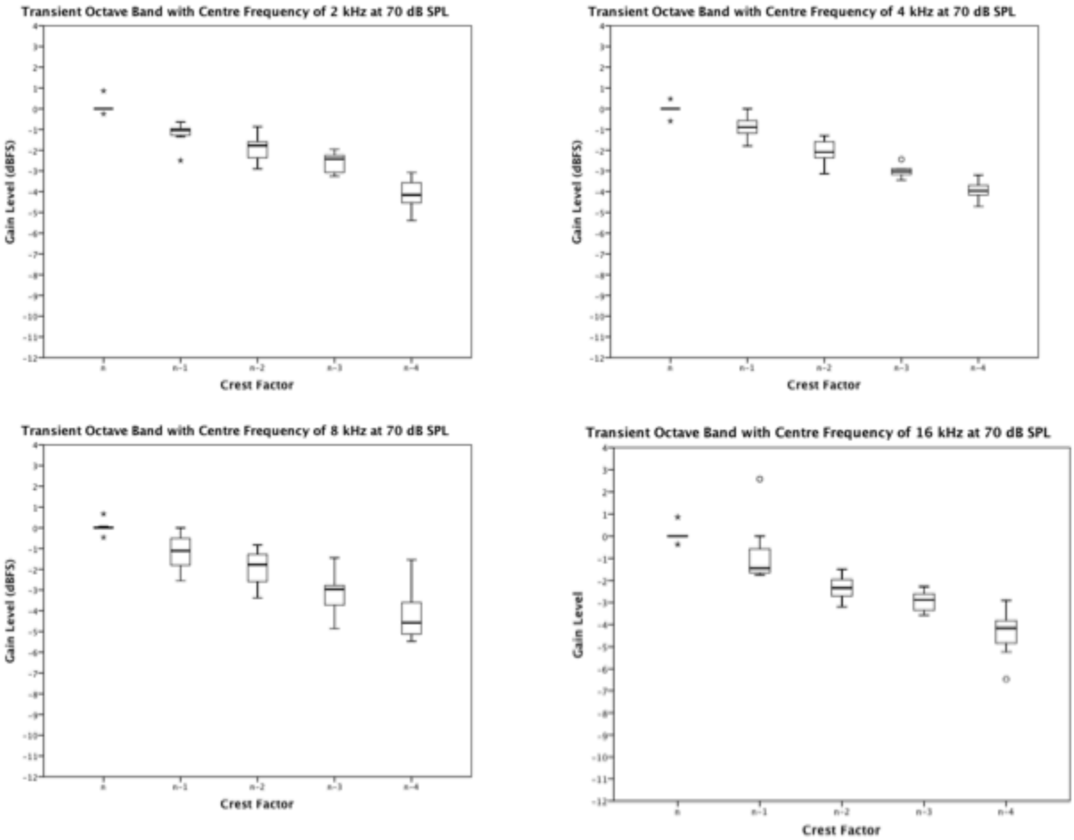


Figure D.33 (top left) to **figure D.36** (bottom right). Plots showing median gain levels and inter quartiles ranges for different amplitude characteristics, octave bands and playback levels.

Appendix E Friedman Pairwise Tests for Median Gain Level Changes for Different Crest Factors

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_2nd_n_4-C_50_2nd_n_3	1.188	.791	1.502	.133	1.000
C_50_2nd_n_4-C_50_2nd_n_2	2.500	.791	3.162	.002	.016
C_50_2nd_n_4-C_50_2nd_n_1	3.125	.791	3.953	.000	.001
C_50_2nd_n_4-C_50_2nd_n	3.188	.791	4.032	.000	.001
C_50_2nd_n_3-C_50_2nd_n_2	1.312	.791	1.660	.097	.969
C_50_2nd_n_3-C_50_2nd_n_1	1.938	.791	2.451	.014	.143
C_50_2nd_n_3-C_50_2nd_n	2.000	.791	2.530	.011	.114
C_50_2nd_n_2-C_50_2nd_n_1	.625	.791	.791	.429	1.000
C_50_2nd_n_2-C_50_2nd_n	.688	.791	.870	.385	1.000
C_50_2nd_n_1-C_50_2nd_n	.062	.791	.079	.937	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.1 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 2nd means octave band with centre frequency of 63 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_4th_n_4-C_50_4th_n_3	.625	.791	.791	.429	1.000
C_50_4th_n_4-C_50_4th_n_2	1.625	.791	2.055	.040	.398
C_50_4th_n_4-C_50_4th_n_1	2.875	.791	3.637	.000	.003
C_50_4th_n_4-C_50_4th_n	3.625	.791	4.585	.000	.000
C_50_4th_n_3-C_50_4th_n_2	1.000	.791	1.265	.206	1.000
C_50_4th_n_3-C_50_4th_n_1	2.250	.791	2.846	.004	.044
C_50_4th_n_3-C_50_4th_n	3.000	.791	3.795	.000	.001
C_50_4th_n_2-C_50_4th_n_1	1.250	.791	1.581	.114	1.000
C_50_4th_n_2-C_50_4th_n	2.000	.791	2.530	.011	.114
C_50_4th_n_1-C_50_4th_n	.750	.791	.949	.343	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.2 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 3rd means octave band with centre frequency of 125 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_3rd_n_4-C_50_3rd_n_3	1.188	.791	1.502	.133	1.000
C_50_3rd_n_4-C_50_3rd_n_2	2.000	.791	2.530	.011	.114
C_50_3rd_n_4-C_50_3rd_n_1	2.812	.791	3.558	.000	.004
C_50_3rd_n_4-C_50_3rd_n	4.000	.791	5.060	.000	.000
C_50_3rd_n_3-C_50_3rd_n_2	.812	.791	1.028	.304	1.000
C_50_3rd_n_3-C_50_3rd_n_1	1.625	.791	2.055	.040	.398
C_50_3rd_n_3-C_50_3rd_n	2.812	.791	3.558	.000	.004
C_50_3rd_n_2-C_50_3rd_n_1	.812	.791	1.028	.304	1.000
C_50_3rd_n_2-C_50_3rd_n	2.000	.791	2.530	.011	.114
C_50_3rd_n_1-C_50_3rd_n	1.188	.791	1.502	.133	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.3 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 4th means octave band with centre frequency of 250 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_5th_n_4-C_50_5th_n_3	1.125	.791	1.423	.155	1.000
C_50_5th_n_4-C_50_5th_n_2	1.750	.791	2.214	.027	.269
C_50_5th_n_4-C_50_5th_n_1	2.750	.791	3.479	.001	.005
C_50_5th_n_4-C_50_5th_n	3.750	.791	4.743	.000	.000
C_50_5th_n_3-C_50_5th_n_2	.625	.791	.791	.429	1.000
C_50_5th_n_3-C_50_5th_n_1	1.625	.791	2.055	.040	.398
C_50_5th_n_3-C_50_5th_n	2.625	.791	3.320	.001	.009
C_50_5th_n_2-C_50_5th_n_1	1.000	.791	1.265	.206	1.000
C_50_5th_n_2-C_50_5th_n	2.000	.791	2.530	.011	.114
C_50_5th_n_1-C_50_5th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.4 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 5th means octave band with centre frequency of 500 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_6th_n_4-C_50_6th_n_3	.625	.791	.791	.429	1.000
C_50_6th_n_4-C_50_6th_n_2	1.438	.791	1.818	.069	.690
C_50_6th_n_4-C_50_6th_n_1	2.812	.791	3.558	.000	.004
C_50_6th_n_4-C_50_6th_n	3.250	.791	4.111	.000	.000
C_50_6th_n_3-C_50_6th_n_2	.812	.791	1.028	.304	1.000
C_50_6th_n_3-C_50_6th_n_1	2.188	.791	2.767	.006	.057
C_50_6th_n_3-C_50_6th_n	2.625	.791	3.320	.001	.009
C_50_6th_n_2-C_50_6th_n_1	1.375	.791	1.739	.082	.820
C_50_6th_n_2-C_50_6th_n	1.812	.791	2.293	.022	.219
C_50_6th_n_1-C_50_6th_n	.438	.791	.553	.580	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.5 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 6^{th} means octave band with centre frequency of 1 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_7th_n_4-C_50_7th_n_3	1.125	.791	1.423	.155	1.000
C_50_7th_n_4-C_50_7th_n_2	1.625	.791	2.055	.040	.398
C_50_7th_n_4-C_50_7th_n_1	2.750	.791	3.479	.001	.005
C_50_7th_n_4-C_50_7th_n	3.875	.791	4.902	.000	.000
C_50_7th_n_3-C_50_7th_n_2	.500	.791	.632	.527	1.000
C_50_7th_n_3-C_50_7th_n_1	1.625	.791	2.055	.040	.398
C_50_7th_n_3-C_50_7th_n	2.750	.791	3.479	.001	.005
C_50_7th_n_2-C_50_7th_n_1	1.125	.791	1.423	.155	1.000
C_50_7th_n_2-C_50_7th_n	2.250	.791	2.846	.004	.044
C_50_7th_n_1-C_50_7th_n	1.125	.791	1.423	.155	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.6 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 7^{th} means octave band with centre frequency of 2 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_8th_n_4-C_50_8th_n_3	1.000	.791	1.265	.206	1.000
C_50_8th_n_4-C_50_8th_n_2	2.125	.791	2.688	.007	.072
C_50_8th_n_4-C_50_8th_n_1	3.062	.791	3.874	.000	.001
C_50_8th_n_4-C_50_8th_n	3.812	.791	4.822	.000	.000
C_50_8th_n_3-C_50_8th_n_2	1.125	.791	1.423	.155	1.000
C_50_8th_n_3-C_50_8th_n_1	2.062	.791	2.609	.009	.091
C_50_8th_n_3-C_50_8th_n	2.812	.791	3.558	.000	.004
C_50_8th_n_2-C_50_8th_n_1	.938	.791	1.186	.236	1.000
C_50_8th_n_2-C_50_8th_n	1.688	.791	2.135	.033	.328
C_50_8th_n_1-C_50_8th_n	.750	.791	.949	.343	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.7 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 8^{th} means octave band with centre frequency of 4 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_9th_n_4-C_50_9th_n_3	1.250	.791	1.581	.114	1.000
C_50_9th_n_4-C_50_9th_n_2	1.625	.791	2.055	.040	.398
C_50_9th_n_4-C_50_9th_n_1	3.062	.791	3.874	.000	.001
C_50_9th_n_4-C_50_9th_n	3.438	.791	4.348	.000	.000
C_50_9th_n_3-C_50_9th_n_2	.375	.791	.474	.635	1.000
C_50_9th_n_3-C_50_9th_n_1	1.812	.791	2.293	.022	.219
C_50_9th_n_3-C_50_9th_n	2.188	.791	2.767	.006	.057
C_50_9th_n_2-C_50_9th_n_1	1.438	.791	1.818	.069	.690
C_50_9th_n_2-C_50_9th_n	1.812	.791	2.293	.022	.219
C_50_9th_n_1-C_50_9th_n	.375	.791	.474	.635	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.8 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 9th means octave band with centre frequency of 8 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_50_10th_n_4-C_50_10th_n_3	.500	.791	.632	.527	1.000
C_50_10th_n_4-C_50_10th_n_2	1.375	.791	1.739	.082	.820
C_50_10th_n_4-C_50_10th_n_1	2.250	.791	2.846	.004	.044
C_50_10th_n_4-C_50_10th_n	2.750	.791	3.479	.001	.005
C_50_10th_n_3-C_50_10th_n_2	.875	.791	1.107	.268	1.000
C_50_10th_n_3-C_50_10th_n_1	1.750	.791	2.214	.027	.269
C_50_10th_n_3-C_50_10th_n	2.250	.791	2.846	.004	.044
C_50_10th_n_2-C_50_10th_n_1	.875	.791	1.107	.268	1.000
C_50_10th_n_2-C_50_10th_n	1.375	.791	1.739	.082	.820
C_50_10th_n_1-C_50_10th_n	.500	.791	.632	.527	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.9 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 10^{th} means octave band with centre frequency of 16 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_2nd_n_4-C_70_2nd_n_3	.875	.791	1.107	.268	1.000
C_70_2nd_n_4-C_70_2nd_n_2	1.875	.791	2.372	.018	.177
C_70_2nd_n_4-C_70_2nd_n_1	3.188	.791	4.032	.000	.001
C_70_2nd_n_4-C_70_2nd_n	3.438	.791	4.348	.000	.000
C_70_2nd_n_3-C_70_2nd_n_2	1.000	.791	1.265	.206	1.000
C_70_2nd_n_3-C_70_2nd_n_1	2.312	.791	2.925	.003	.034
C_70_2nd_n_3-C_70_2nd_n	2.562	.791	3.241	.001	.012
C_70_2nd_n_2-C_70_2nd_n_1	1.312	.791	1.660	.097	.969
C_70_2nd_n_2-C_70_2nd_n	1.562	.791	1.976	.048	.481
C_70_2nd_n_1-C_70_2nd_n	.250	.791	.316	.752	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.10 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 2nd means octave band with centre frequency of 63 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_3rd_n_4-C_70_3rd_n_3	1.125	.791	1.423	.155	1.000
C_70_3rd_n_4-C_70_3rd_n_2	1.875	.791	2.372	.018	.177
C_70_3rd_n_4-C_70_3rd_n_1	3.062	.791	3.874	.000	.001
C_70_3rd_n_4-C_70_3rd_n	3.938	.791	4.981	.000	.000
C_70_3rd_n_3-C_70_3rd_n_2	.750	.791	.949	.343	1.000
C_70_3rd_n_3-C_70_3rd_n_1	1.938	.791	2.451	.014	.143
C_70_3rd_n_3-C_70_3rd_n	2.812	.791	3.558	.000	.004
C_70_3rd_n_2-C_70_3rd_n_1	1.188	.791	1.502	.133	1.000
C_70_3rd_n_2-C_70_3rd_n	2.062	.791	2.609	.009	.091
C_70_3rd_n_1-C_70_3rd_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.11 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 3rd means octave band with centre frequency of 125 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_4th_n_4-C_70_4th_n_3	.750	.791	.949	.343	1.000
C_70_4th_n_4-C_70_4th_n_2	2.062	.791	2.609	.009	.091
C_70_4th_n_4-C_70_4th_n_1	2.875	.791	3.637	.000	.003
C_70_4th_n_4-C_70_4th_n	3.688	.791	4.664	.000	.000
C_70_4th_n_3-C_70_4th_n_2	1.312	.791	1.660	.097	.969
C_70_4th_n_3-C_70_4th_n_1	2.125	.791	2.688	.007	.072
C_70_4th_n_3-C_70_4th_n	2.938	.791	3.716	.000	.002
C_70_4th_n_2-C_70_4th_n_1	.812	.791	1.028	.304	1.000
C_70_4th_n_2-C_70_4th_n	1.625	.791	2.055	.040	.398
C_70_4th_n_1-C_70_4th_n	.812	.791	1.028	.304	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.12 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 4th means octave band with centre frequency of 250 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_5th_n_4-C_70_5th_n_3	1.375	.791	1.739	.082	.820
C_70_5th_n_4-C_70_5th_n_2	1.875	.791	2.372	.018	.177
C_70_5th_n_4-C_70_5th_n_1	2.938	.791	3.716	.000	.002
C_70_5th_n_4-C_70_5th_n	3.812	.791	4.822	.000	.000
C_70_5th_n_3-C_70_5th_n_2	.500	.791	.632	.527	1.000
C_70_5th_n_3-C_70_5th_n_1	1.562	.791	1.976	.048	.481
C_70_5th_n_3-C_70_5th_n	2.438	.791	3.083	.002	.020
C_70_5th_n_2-C_70_5th_n_1	1.062	.791	1.344	.179	1.000
C_70_5th_n_2-C_70_5th_n	1.938	.791	2.451	.014	.143
C_70_5th_n_1-C_70_5th_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.13 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 5th means octave band with centre frequency of 500 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_6th_n_4-C_70_6th_n_3	1.125	.791	1.423	.155	1.000
C_70_6th_n_4-C_70_6th_n_2	2.000	.791	2.530	.011	.114
C_70_6th_n_4-C_70_6th_n_1	3.125	.791	3.953	.000	.001
C_70_6th_n_4-C_70_6th_n	3.750	.791	4.743	.000	.000
C_70_6th_n_3-C_70_6th_n_2	.875	.791	1.107	.268	1.000
C_70_6th_n_3-C_70_6th_n_1	2.000	.791	2.530	.011	.114
C_70_6th_n_3-C_70_6th_n	2.625	.791	3.320	.001	.009
C_70_6th_n_2-C_70_6th_n_1	1.125	.791	1.423	.155	1.000
C_70_6th_n_2-C_70_6th_n	1.750	.791	2.214	.027	.269
C_70_6th_n_1-C_70_6th_n	.625	.791	.791	.429	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.14 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 6th means octave band with centre frequency of 1 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_7th_n_4-C_70_7th_n_3	.875	.791	1.107	.268	1.000
C_70_7th_n_4-C_70_7th_n_2	1.500	.791	1.897	.058	.578
C_70_7th_n_4-C_70_7th_n_1	2.688	.791	3.399	.001	.007
C_70_7th_n_4-C_70_7th_n	3.688	.791	4.664	.000	.000
C_70_7th_n_3-C_70_7th_n_2	.625	.791	.791	.429	1.000
C_70_7th_n_3-C_70_7th_n_1	1.812	.791	2.293	.022	.219
C_70_7th_n_3-C_70_7th_n	2.812	.791	3.558	.000	.004
C_70_7th_n_2-C_70_7th_n_1	1.188	.791	1.502	.133	1.000
C_70_7th_n_2-C_70_7th_n	2.188	.791	2.767	.006	.057
C_70_7th_n_1-C_70_7th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.15 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 7th means octave band with centre frequency of 2 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_8th_n_4-C_70_8th_n_3	.500	.791	.632	.527	1.000
C_70_8th_n_4-C_70_8th_n_2	2.000	.791	2.530	.011	.114
C_70_8th_n_4-C_70_8th_n_1	2.688	.791	3.399	.001	.007
C_70_8th_n_4-C_70_8th_n	3.562	.791	4.506	.000	.000
C_70_8th_n_3-C_70_8th_n_2	1.500	.791	1.897	.058	.578
C_70_8th_n_3-C_70_8th_n_1	2.188	.791	2.767	.006	.057
C_70_8th_n_3-C_70_8th_n	3.062	.791	3.874	.000	.001
C_70_8th_n_2-C_70_8th_n_1	.688	.791	.870	.385	1.000
C_70_8th_n_2-C_70_8th_n	1.562	.791	1.976	.048	.481
C_70_8th_n_1-C_70_8th_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.16 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 8th means octave band with centre frequency of 4 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_9th_n_4-C_70_9th_n_3	1.188	.791	1.502	.133	1.000
C_70_9th_n_4-C_70_9th_n_2	2.062	.791	2.609	.009	.091
C_70_9th_n_4-C_70_9th_n_1	2.938	.791	3.716	.000	.002
C_70_9th_n_4-C_70_9th_n	3.812	.791	4.822	.000	.000
C_70_9th_n_3-C_70_9th_n_2	.875	.791	1.107	.268	1.000
C_70_9th_n_3-C_70_9th_n_1	1.750	.791	2.214	.027	.269
C_70_9th_n_3-C_70_9th_n	2.625	.791	3.320	.001	.009
C_70_9th_n_2-C_70_9th_n_1	.875	.791	1.107	.268	1.000
C_70_9th_n_2-C_70_9th_n	1.750	.791	2.214	.027	.269
C_70_9th_n_1-C_70_9th_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.17 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 9th means octave band with centre frequency of 8 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C_70_10th_n_4-C_70_10th_n_3	.750	.791	.949	.343	1.000
C_70_10th_n_4-C_70_10th_n_2	1.750	.791	2.214	.027	.269
C_70_10th_n_4-C_70_10th_n_1	2.438	.791	3.083	.002	.020
C_70_10th_n_4-C_70_10th_n	3.188	.791	4.032	.000	.001
C_70_10th_n_3-C_70_10th_n_2	1.000	.791	1.265	.206	1.000
C_70_10th_n_3-C_70_10th_n_1	1.688	.791	2.135	.033	.328
C_70_10th_n_3-C_70_10th_n	2.438	.791	3.083	.002	.020
C_70_10th_n_2-C_70_10th_n_1	.688	.791	.870	.385	1.000
C_70_10th_n_2-C_70_10th_n	1.438	.791	1.818	.069	.690
C_70_10th_n_1-C_70_10th_n	.750	.791	.949	.343	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.18 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the C means continuous, 10th means octave band with centre frequency of 16 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_2nd_n_4-T_50_2nd_n_3	.875	.791	1.107	.268	1.000
T_50_2nd_n_4-T_50_2nd_n_2	2.188	.791	2.767	.006	.057
T_50_2nd_n_4-T_50_2nd_n_1	2.625	.791	3.320	.001	.009
T_50_2nd_n_4-T_50_2nd_n	3.688	.791	4.664	.000	.000
T_50_2nd_n_3-T_50_2nd_n_2	1.312	.791	1.660	.097	.969
T_50_2nd_n_3-T_50_2nd_n_1	1.750	.791	2.214	.027	.269
T_50_2nd_n_3-T_50_2nd_n	2.812	.791	3.558	.000	.004
T_50_2nd_n_2-T_50_2nd_n_1	.438	.791	.553	.580	1.000
T_50_2nd_n_2-T_50_2nd_n	1.500	.791	1.897	.058	.578
T_50_2nd_n_1-T_50_2nd_n	1.062	.791	1.344	.179	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.19 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 2nd means octave band with centre frequency of 63 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_3rd_n_4-T_50_3rd_n_3	.500	.791	.632	.527	1.000
T_50_3rd_n_4-T_50_3rd_n_2	2.000	.791	2.530	.011	.114
T_50_3rd_n_4-T_50_3rd_n_1	2.688	.791	3.399	.001	.007
T_50_3rd_n_4-T_50_3rd_n	3.562	.791	4.506	.000	.000
T_50_3rd_n_3-T_50_3rd_n_2	1.500	.791	1.897	.058	.578
T_50_3rd_n_3-T_50_3rd_n_1	2.188	.791	2.767	.006	.057
T_50_3rd_n_3-T_50_3rd_n	3.062	.791	3.874	.000	.001
T_50_3rd_n_2-T_50_3rd_n_1	.688	.791	.870	.385	1.000
T_50_3rd_n_2-T_50_3rd_n	1.562	.791	1.976	.048	.481
T_50_3rd_n_1-T_50_3rd_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.20 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 3rd means octave band with centre frequency of 125 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_4th_n_4-T_50_4th_n_3	1.000	.791	1.265	.206	1.000
T_50_4th_n_4-T_50_4th_n_2	1.875	.791	2.372	.018	.177
T_50_4th_n_4-T_50_4th_n_1	2.750	.791	3.479	.001	.005
T_50_4th_n_4-T_50_4th_n	3.750	.791	4.743	.000	.000
T_50_4th_n_3-T_50_4th_n_2	.875	.791	1.107	.268	1.000
T_50_4th_n_3-T_50_4th_n_1	1.750	.791	2.214	.027	.269
T_50_4th_n_3-T_50_4th_n	2.750	.791	3.479	.001	.005
T_50_4th_n_2-T_50_4th_n_1	.875	.791	1.107	.268	1.000
T_50_4th_n_2-T_50_4th_n	1.875	.791	2.372	.018	.177
T_50_4th_n_1-T_50_4th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.21 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 4th means octave band with centre frequency of 250 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_5th_n_4-T_50_5th_n_3	.625	.791	.791	.429	1.000
T_50_5th_n_4-T_50_5th_n_2	1.688	.791	2.135	.033	.328
T_50_5th_n_4-T_50_5th_n_1	2.562	.791	3.241	.001	.012
T_50_5th_n_4-T_50_5th_n	3.562	.791	4.506	.000	.000
T_50_5th_n_3-T_50_5th_n_2	1.062	.791	1.344	.179	1.000
T_50_5th_n_3-T_50_5th_n_1	1.938	.791	2.451	.014	.143
T_50_5th_n_3-T_50_5th_n	2.938	.791	3.716	.000	.002
T_50_5th_n_2-T_50_5th_n_1	.875	.791	1.107	.268	1.000
T_50_5th_n_2-T_50_5th_n	1.875	.791	2.372	.018	.177
T_50_5th_n_1-T_50_5th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.22 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 5th means octave band with centre frequency of 500 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_6th_n_4-T_50_6th_n_3	.750	.791	.949	.343	1.000
T_50_6th_n_4-T_50_6th_n_2	1.875	.791	2.372	.018	.177
T_50_6th_n_4-T_50_6th_n_1	3.000	.791	3.795	.000	.001
T_50_6th_n_4-T_50_6th_n	3.750	.791	4.743	.000	.000
T_50_6th_n_3-T_50_6th_n_2	1.125	.791	1.423	.155	1.000
T_50_6th_n_3-T_50_6th_n_1	2.250	.791	2.846	.004	.044
T_50_6th_n_3-T_50_6th_n	3.000	.791	3.795	.000	.001
T_50_6th_n_2-T_50_6th_n_1	1.125	.791	1.423	.155	1.000
T_50_6th_n_2-T_50_6th_n	1.875	.791	2.372	.018	.177
T_50_6th_n_1-T_50_6th_n	.750	.791	.949	.343	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.23 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 6th means octave band with centre frequency of 1 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_7th_n_4-T_50_7th_n_3	.500	.791	.632	.527	1.000
T_50_7th_n_4-T_50_7th_n_2	1.875	.791	2.372	.018	.177
T_50_7th_n_4-T_50_7th_n_1	2.688	.791	3.399	.001	.007
T_50_7th_n_4-T_50_7th_n	3.688	.791	4.664	.000	.000
T_50_7th_n_3-T_50_7th_n_2	1.375	.791	1.739	.082	.820
T_50_7th_n_3-T_50_7th_n_1	2.188	.791	2.767	.006	.057
T_50_7th_n_3-T_50_7th_n	3.188	.791	4.032	.000	.001
T_50_7th_n_2-T_50_7th_n_1	.812	.791	1.028	.304	1.000
T_50_7th_n_2-T_50_7th_n	1.812	.791	2.293	.022	.219
T_50_7th_n_1-T_50_7th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.24 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 7th means octave band with centre frequency of 2 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_8th_n_4-T_50_8th_n_3	.625	.791	.791	.429	1.000
T_50_8th_n_4-T_50_8th_n_2	1.812	.791	2.293	.022	.219
T_50_8th_n_4-T_50_8th_n_1	2.812	.791	3.558	.000	.004
T_50_8th_n_4-T_50_8th_n	3.812	.791	4.822	.000	.000
T_50_8th_n_3-T_50_8th_n_2	1.188	.791	1.502	.133	1.000
T_50_8th_n_3-T_50_8th_n_1	2.188	.791	2.767	.006	.057
T_50_8th_n_3-T_50_8th_n	3.188	.791	4.032	.000	.001
T_50_8th_n_2-T_50_8th_n_1	1.000	.791	1.265	.206	1.000
T_50_8th_n_2-T_50_8th_n	2.000	.791	2.530	.011	.114
T_50_8th_n_1-T_50_8th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.25 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 8th means octave band with centre frequency of 4 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_9th_n_4-T_50_9th_n_3	.375	.791	.474	.635	1.000
T_50_9th_n_4-T_50_9th_n_2	1.500	.791	1.897	.058	.578
T_50_9th_n_4-T_50_9th_n_1	2.750	.791	3.479	.001	.005
T_50_9th_n_4-T_50_9th_n	3.500	.791	4.427	.000	.000
T_50_9th_n_3-T_50_9th_n_2	1.125	.791	1.423	.155	1.000
T_50_9th_n_3-T_50_9th_n_1	2.375	.791	3.004	.003	.027
T_50_9th_n_3-T_50_9th_n	3.125	.791	3.953	.000	.001
T_50_9th_n_2-T_50_9th_n_1	1.250	.791	1.581	.114	1.000
T_50_9th_n_2-T_50_9th_n	2.000	.791	2.530	.011	.114
T_50_9th_n_1-T_50_9th_n	.750	.791	.949	.343	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.26 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 2nd means octave band with centre frequency of 63 Hz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_50_10th_n_4-T_50_10th_n_3	.500	.791	.632	.527	1.000
T_50_10th_n_4-T_50_10th_n_2	1.938	.791	2.451	.014	.143
T_50_10th_n_4-T_50_10th_n_1	2.750	.791	3.479	.001	.005
T_50_10th_n_4-T_50_10th_n	3.562	.791	4.506	.000	.000
T_50_10th_n_3-T_50_10th_n_2	1.438	.791	1.818	.069	.690
T_50_10th_n_3-T_50_10th_n_1	2.250	.791	2.846	.004	.044
T_50_10th_n_3-T_50_10th_n	3.062	.791	3.874	.000	.001
T_50_10th_n_2-T_50_10th_n_1	.812	.791	1.028	.304	1.000
T_50_10th_n_2-T_50_10th_n	1.625	.791	2.055	.040	.398
T_50_10th_n_1-T_50_10th_n	.812	.791	1.028	.304	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.27 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 10th means octave band with centre frequency of 16 kHz and 50 means 50 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_2nd_n_4-T_70_2nd_n_3	1.000	.791	1.265	.206	1.000
T_70_2nd_n_4-T_70_2nd_n_2	2.000	.791	2.530	.011	.114
T_70_2nd_n_4-T_70_2nd_n_1	3.062	.791	3.874	.000	.001
T_70_2nd_n_4-T_70_2nd_n	3.938	.791	4.981	.000	.000
T_70_2nd_n_3-T_70_2nd_n_2	1.000	.791	1.265	.206	1.000
T_70_2nd_n_3-T_70_2nd_n_1	2.062	.791	2.609	.009	.091
T_70_2nd_n_3-T_70_2nd_n	2.938	.791	3.716	.000	.002
T_70_2nd_n_2-T_70_2nd_n_1	1.062	.791	1.344	.179	1.000
T_70_2nd_n_2-T_70_2nd_n	1.938	.791	2.451	.014	.143
T_70_2nd_n_1-T_70_2nd_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.28 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 2nd means octave band with centre frequency of 63 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_3rd_n_4-T_70_3rd_n_3	.375	.791	.474	.635	1.000
T_70_3rd_n_4-T_70_3rd_n_2	1.625	.791	2.055	.040	.398
T_70_3rd_n_4-T_70_3rd_n_1	2.625	.791	3.320	.001	.009
T_70_3rd_n_4-T_70_3rd_n	3.500	.791	4.427	.000	.000
T_70_3rd_n_3-T_70_3rd_n_2	1.250	.791	1.581	.114	1.000
T_70_3rd_n_3-T_70_3rd_n_1	2.250	.791	2.846	.004	.044
T_70_3rd_n_3-T_70_3rd_n	3.125	.791	3.953	.000	.001
T_70_3rd_n_2-T_70_3rd_n_1	1.000	.791	1.265	.206	1.000
T_70_3rd_n_2-T_70_3rd_n	1.875	.791	2.372	.018	.177
T_70_3rd_n_1-T_70_3rd_n	.875	.791	1.107	.268	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.29 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 3rd means octave band with centre frequency of 125 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_4th_n_4-T_70_4th_n_3	1.000	.791	1.265	.206	1.000
T_70_4th_n_4-T_70_4th_n_2	2.000	.791	2.530	.011	.114
T_70_4th_n_4-T_70_4th_n_1	3.000	.791	3.795	.000	.001
T_70_4th_n_4-T_70_4th_n	4.000	.791	5.060	.000	.000
T_70_4th_n_3-T_70_4th_n_2	1.000	.791	1.265	.206	1.000
T_70_4th_n_3-T_70_4th_n_1	2.000	.791	2.530	.011	.114
T_70_4th_n_3-T_70_4th_n	3.000	.791	3.795	.000	.001
T_70_4th_n_2-T_70_4th_n_1	1.000	.791	1.265	.206	1.000
T_70_4th_n_2-T_70_4th_n	2.000	.791	2.530	.011	.114
T_70_4th_n_1-T_70_4th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.30 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 4th means octave band with centre frequency of 250 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_5th_n_4-T_70_5th_n_3	.875	.791	1.107	.268	1.000
T_70_5th_n_4-T_70_5th_n_2	1.875	.791	2.372	.018	.177
T_70_5th_n_4-T_70_5th_n_1	2.812	.791	3.558	.000	.004
T_70_5th_n_4-T_70_5th_n	3.812	.791	4.822	.000	.000
T_70_5th_n_3-T_70_5th_n_2	1.000	.791	1.265	.206	1.000
T_70_5th_n_3-T_70_5th_n_1	1.938	.791	2.451	.014	.143
T_70_5th_n_3-T_70_5th_n	2.938	.791	3.716	.000	.002
T_70_5th_n_2-T_70_5th_n_1	.938	.791	1.186	.236	1.000
T_70_5th_n_2-T_70_5th_n	1.938	.791	2.451	.014	.143
T_70_5th_n_1-T_70_5th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.31 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 5th means octave band with centre frequency of 500 Hz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_6th_n_4-T_70_6th_n_3	1.250	.791	1.581	.114	1.000
T_70_6th_n_4-T_70_6th_n_2	2.000	.791	2.530	.011	.114
T_70_6th_n_4-T_70_6th_n_1	3.125	.791	3.953	.000	.001
T_70_6th_n_4-T_70_6th_n	3.625	.791	4.585	.000	.000
T_70_6th_n_3-T_70_6th_n_2	.750	.791	.949	.343	1.000
T_70_6th_n_3-T_70_6th_n_1	1.875	.791	2.372	.018	.177
T_70_6th_n_3-T_70_6th_n	2.375	.791	3.004	.003	.027
T_70_6th_n_2-T_70_6th_n_1	1.125	.791	1.423	.155	1.000
T_70_6th_n_2-T_70_6th_n	1.625	.791	2.055	.040	.398
T_70_6th_n_1-T_70_6th_n	.500	.791	.632	.527	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.32 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 6th means octave band with centre frequency of 1 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_7th_n_4-T_70_7th_n_3	1.375	.791	1.739	.082	.820
T_70_7th_n_4-T_70_7th_n_2	1.875	.791	2.372	.018	.177
T_70_7th_n_4-T_70_7th_n_1	2.750	.791	3.479	.001	.005
T_70_7th_n_4-T_70_7th_n	4.000	.791	5.060	.000	.000
T_70_7th_n_3-T_70_7th_n_2	.500	.791	.632	.527	1.000
T_70_7th_n_3-T_70_7th_n_1	1.375	.791	1.739	.082	.820
T_70_7th_n_3-T_70_7th_n	2.625	.791	3.320	.001	.009
T_70_7th_n_2-T_70_7th_n_1	.875	.791	1.107	.268	1.000
T_70_7th_n_2-T_70_7th_n	2.125	.791	2.688	.007	.072
T_70_7th_n_1-T_70_7th_n	1.250	.791	1.581	.114	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.33 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 7th means octave band with centre frequency of 2 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_8th_n_4-T_70_8th_n_3	1.125	.791	1.423	.155	1.000
T_70_8th_n_4-T_70_8th_n_2	1.875	.791	2.372	.018	.177
T_70_8th_n_4-T_70_8th_n_1	3.188	.791	4.032	.000	.001
T_70_8th_n_4-T_70_8th_n	3.812	.791	4.822	.000	.000
T_70_8th_n_3-T_70_8th_n_2	.750	.791	.949	.343	1.000
T_70_8th_n_3-T_70_8th_n_1	2.062	.791	2.609	.009	.091
T_70_8th_n_3-T_70_8th_n	2.688	.791	3.399	.001	.007
T_70_8th_n_2-T_70_8th_n_1	1.312	.791	1.660	.097	.969
T_70_8th_n_2-T_70_8th_n	1.938	.791	2.451	.014	.143
T_70_8th_n_1-T_70_8th_n	.625	.791	.791	.429	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.34 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 8^{th} means octave band with centre frequency of 4 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_9th_n_4-T_70_9th_n_3	.875	.791	1.107	.268	1.000
T_70_9th_n_4-T_70_9th_n_2	2.000	.791	2.530	.011	.114
T_70_9th_n_4-T_70_9th_n_1	2.750	.791	3.479	.001	.005
T_70_9th_n_4-T_70_9th_n	3.750	.791	4.743	.000	.000
T_70_9th_n_3-T_70_9th_n_2	1.125	.791	1.423	.155	1.000
T_70_9th_n_3-T_70_9th_n_1	1.875	.791	2.372	.018	.177
T_70_9th_n_3-T_70_9th_n	2.875	.791	3.637	.000	.003
T_70_9th_n_2-T_70_9th_n_1	.750	.791	.949	.343	1.000
T_70_9th_n_2-T_70_9th_n	1.750	.791	2.214	.027	.269
T_70_9th_n_1-T_70_9th_n	1.000	.791	1.265	.206	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.35 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 9^{th} means octave band with centre frequency of 8 kHz and 70 means 70 dB SPL

Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
T_70_10th_n_4-T_70_10th_n_3	1.000	.791	1.265	.206	1.000
T_70_10th_n_4-T_70_10th_n_2	1.625	.791	2.055	.040	.398
T_70_10th_n_4-T_70_10th_n_1	3.062	.791	3.874	.000	.001
T_70_10th_n_4-T_70_10th_n	3.688	.791	4.664	.000	.000
T_70_10th_n_3-T_70_10th_n_2	.625	.791	.791	.429	1.000
T_70_10th_n_3-T_70_10th_n_1	2.062	.791	2.609	.009	.091
T_70_10th_n_3-T_70_10th_n	2.688	.791	3.399	.001	.007
T_70_10th_n_2-T_70_10th_n_1	1.438	.791	1.818	.069	.690
T_70_10th_n_2-T_70_10th_n	2.062	.791	2.609	.009	.091
T_70_10th_n_1-T_70_10th_n	.625	.791	.791	.429	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Figure E.36 Friedman Pairwise test showing significant difference between median gain level for different crest factors. Note that the T means transient, 10th means octave band with centre frequency of 16 kHz and 70 means 70 dB SPL

Appendix F Wilcoxon Post Hoc Pairwise Comparison Across Independent Variables

Table F.1 A table demonstrating the octave bands of continuous amplitude characteristic that gave a significant difference between different playback levels when a Wilcoxan Post Hoc Pairwise comparison was conducted.

Centre Frequency of Octave Bands	Crest Factor Value of Octave Bands	P-Value
63 Hz	n-2	0.018
63 Hz	n-4	0.017
125 Hz	n-2	0.018
125 Hz	n-3	0.018
125 Hz	n-4	0.025
2 kHz	n-3	0.05
8 kHz	n-4	0.025

Table F.2 A table demonstrating the octave bands of transient amplitude characteristic that gave a significant difference between different playback levels when a Wilcoxan Post Hoc Pairwise comparison was conducted.

Centre Frequency of Octave Bands	**Crest Factor Value of Octave Bands**	**P-Value**
125 Hz	n-2	0.025
125 Hz	n-3	0.05
125 Hz	n-4	0.017
250 Hz	n-3	0.036
250 Hz	n-4	0.036
500 Hz	n-3	0.012
500 Hz	n-4	0.036
2 kHz	n-3	0.036
4 kHz	n-3	0.036

References

Bauer, B. & Torick, E. (1966). Researches in loudness measurement. *IEEE Transactions on Audio and Electroacoustics*, 14(3), 141-151. doi:10.1109/TAU.1966.1161864

Bauer, B., Torick, E., Rosenheck, A. & Allen, R. (1967). A loudness-level monitor for broadcasting. *IEEE Transactions on Audio and Electroacoustics*, 15(4), 177-182. doi:10.1109/TAU.1967.1161920

Baumgarte, F. (2001, May). A computationally efficient cochlear filter bank for perceptual audio coding. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on (Vol. 5, 3265-3268) Salt Lake City, UT, USA*. IEEE. doi:10.1109/ICASSP.2001.940355

BS ISO 226:2003: Acoustics. Normal equal-loudness-level contours (2003). *British Standards Institute.* Retrieved from https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000030094804

Cassidy, R. J. (2004). Dynamic range compression of audio signals consistent with recent time-varying loudness models. In *Acoustics, Speech, and Signal Processing,* *2004. Proceedings.(ICASSP'04). 2004 IEEE International Conference on (Vol. 4, 213-216)* doi:10.1109/ICASSP.2004.1326801

Croghan, N. B. H., Arehart, K. H. & Kates, J. M. (2012) Quality and Loudness Judgements for Music Subjected to Compression Limiting. *The Journal of Acoustical Society of America,* (132), 1177-1188. doi: 10.1121/1.4730881

Deruty, E., & Tardieu, D. (2014). About dynamic processing in mainstream music. *Journal of the Audio Engineering Society*, 62(1-2), 42-56. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17085

Doucet, J. R., & Relkin, E. M. (1997). Neural contributions to the perstimulus compound action potential: Implications for measuring the growth of the auditory nerve spike count as a function of stimulus intensity. *Journal of the Acoustical Society of America,* 101(5 I), 2720-2734. doi:10.1121/1.418560

European Broadcasting Union (2011) *EBU Tech 3343 Practical Guidelines for Production and Implimentation in accordance with EBU R 128.* Retrieved from https://tech.ebu.ch/docs/tech/tech3343.pdf

Fenton, S., Fazenda, B. & Wakefield, J. (2011, May). Objective measurement of music quality using inter-band relationship analysis. In *Audio Engineering Society*

130th Convention, London, UK. Retreived from http://www.aes.org/e-lib/browse.cfm?elib=15861

Fletcher, H. & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5(2), 82-108. doi:10.1121/1.1915637

Florentine, M. Buus, S. & Poulsen, T. (1996). Temporal integration of loudness as a function of level. *The Journal of the Acoustical Society of America*, 99(3), 1633-1644. doi:10.1121/1.415236

Gabrielsson, A., Hagerman, B., Bech-Kristensen, T. & Lundberg, G. (1990). Perceived sound quality of reproductions with different frequency responses and sound levels. *Journal of the Acoustical Society of America*, 88(3), 1359-1366. doi:10.1121/1.399713

Hammershøi, D. & Møller, H. (1996). Sound transmission to and within the human ear canal. *Journal of the Acoustical Society of America*, 100(1), 408-427. doi:10.1121/1.415856

Hjortkjær, J. & Walther-Hansen, M. (2014). Perceptual effects of dynamic range compression in popular music recordings. *Journal of the Audio Engineering Society*, *62*(1/2), 37-41. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17084

Howard, D.M. & Angus, J.A.S. (2009). *Acoustics and Psychoacoustics* (4th ed.). Oxford: Elsevier.

International Telecommunication Union (2012). *Recommendation ITU-R BS.1770-3.* Retrieved from http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-3-201208-I!!PDF-E.pdf

International Telecommunication Union (2014a) *Recommendation ITU-R BS.1534-2* Retrieved from http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-2-201406-I!!PDF-E.pdf

International Telecommunication Union (2014b) *Recommendation ITU-R BS.1116-2* Retrieved from http://www.itu.int/rec/R-REC-BS.1116-2-201406-I/en

Jahn, G. (1974). Curves of Equal Loudness For Octave Bands. *NASA Technical Translation.* Retrieved from https://archive.org/details/nasa_techdoc_19740024056

Katz, B. (2007). *Mastering Audio the art and the science* (2nd ed.). Oxford: Elsevier Science.

Lachs, G., Al-Shaikh, R., Bi, Q., Saia, R. A. & Teich, M. C. (1984). A neural-counting model based on physiological characteristics of the peripheral auditory system. V. application to loudness estimation and intensity discrimination. *IEEE Transactions on Systems, Man, and Cybernetics.* SMC-13(6), 819-836. doi:10.1109/TSMC.1984.6313310

Lund, T. (2007, October) Audio Delivery Specification. In *Audio Engineer Society 123rd Convention, New York, NY, USA* Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=14240

Meadows, G. (2008). What Exactly Is Mastering?. In Owsinski, B. (Ed.) *The Audio Mastering Handbook* (2nd ed.) (pp. 3-8). Boston: Course Technology Cengage Learning.

Moore, B. C. J. & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74(3), 750-753. doi:10.1121/1.389861

Moore, B. C. J., Glasberg, B. R. & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4), 224-239. Retreived from http://www.aes.org/e-lib/browse.cfm?elib=10272

Moore, B.C.J., Glasberg, B.R., Stone, M.A. (2003). Why Are Commercials so Loud? - Perception and Modeling of the Loudness of Amplitude-Compressed Speech. *Journal of the Audio Engineering Society, 51* (12), 1123-1132. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=12190.

Napierala, M.A. (2012). *What Is the Bonferroni Correction?*. Retrieved from http://www.aaos.org/news/aaosnow/apr12/research7.asp.

Nielsen, S. H. & Lund, T. (1999, September). Level Control In Digital Mastering. In *Audio Engineering Society 107th Convention, New York, NY, USA*. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=8161

Norcross, S. G. & Soulodre, G. A. (2003, October). Objective Measures of Loudness. In *Audio Engineering Society 115th Convention, New York, NY, USA*. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=12382

Owsinski, B. (2008). *The Audio Mastering Handbook* (2nd ed.). Boston: Course Technology Cengage Learning.

Pohlmann, K.C. (2005). *Principles of Digital Audio* (5th ed.). New York: McGraw-Hill.

Relkin, E. M. & Doucet, J. R. (1997). Is loudness simply proportional to the auditory nerve spike count? *Journal of the Acoustical Society of America*, 101(5 I), 2735-2740. doi:10.1121/1.418561

Robinson, D. W. & Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics,* 7(5), 166-181. doi:10.1088/0508-3443/7/5/302

Rumsey, F. & McCormick, T. (2008). *Sound and Recording An Introduction* (5th ed.). Oxford: Elsevier.

Simpson, A. J. R., Terrell, M. J. & Reiss, J. D. (2013, May). A Practical Step-by-Step Guide to the Time- Varying Loudness Model of Moore, Glasberg and Baer. In *Audio Engineering Society 134th Convention, Rome, Italy.* Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=16774

Soulodre, G. A. (2004, May). Evaluation of Objective Loudness Meters. In *Audio Engineering Society 116th Convention, Berlin, Germany.* Retrived from http://www.aes.org/e-lib/browse.cfm?elib=12790

Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review*, 8, 405. Retrieved from http://search.proquest.com.libaccess.hud.ac.uk/docview/615061433/8A80FB523F964A24PQ/2?accountid=11526

Stevens, S. S. (1956). The direct estimation of sensory magnitudes: Loudness. *The American Journal of Psychology*, 69(1), 1-25. Retrieved from http://www.jstor.org/discover/10.2307/1418112?sid=21105759439943&uid=2&uid=4&uid=3738032

Stevens, S. S. (1956). Calculation of the loudness of complex noise. *The Journal of the Acoustical Society of America*, 28(5), 807. doi:10.1121/1.1908487

Stevens, S. S. (1961). Procedure for calculating loudness: Mark VI. *The Journal of the Acoustical Society of America*, 33(11), 1577. doi:10.1121/1.1908505

Vickers, E (2010a, November). The Loudness War: Background, Speculation and Recommendations. In *Audio Engineering Society 129th Convention, San Francisco, CA, USA.* Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=15598

Vickers, E. (2010b, November). The Non-flat and Continually Changing Frequency Response of Multiband Compressors. In *Audio Engineering Society 129th Convention, San Francisco, CA, USA*. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=15633

Wagenaars, W. M., Houtsma, A. J. M. & Van Lieshout, R. A. J. M. (1986). SUBJECTIVE EVALUATION OF DYNAMIC COMPRESSION IN MUSIC. *Journal of the Audio Engineering Society*, 34(1-2), 10-18. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=5286

Wendl, M. & Lee, H. (2014, April) The Effect of Dynamic Range Compression on Loudness and Quality Perception in Relation to Crest Factor. In *Audio Engineering Society 136th Convention, Berlin, Germany.* Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17168

Yates, G.K. (1995). Cochlear Structure and Function. In Moore, B.C.J. (Ed.)*Hearing* (2nd ed.) (pp. 41-74). London: Academic Press Limited.

Zwicker, E., Flottorp, G. & Stevens, S. S. (1957). Critical band width in loudness summation. *The Journal of the Acoustical Society of America*, 29(5), 548. doi:10.1121/1.1908963

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2), 248. doi:10.1121/1.1908630

Zwicker, E. & Fastl, H. (1972). On the development of the critical band. *The Journal of the Acoustical Society of America*, 52(2B), 699. doi:10.1121/1.1913161

Zwicker, E. (1977). Procedure for calculating loudness of temporally variable sounds. *The Journal of the Acoustical Society of America*, 62(3), 675. doi:10.1121/1.381580

Zwislocki, J. J. (1969). Temporal summation of loudness: An analysis. *The Journal of the Acoustical Society of America*, 46(2), 431-441. doi:10.1121/1.1911708