Robotham, Thomas

The effects of a vertical reflection on the relationship between listener preference and timbral and spatial attributes

THE EFFECTS OF A VERTICAL REFLECTION ON THE RELATIONSHIP BETWEEN LISTENER PREFERENCE AND TIMBRAL AND SPATIAL ATTRIBUTES

THOMAS ROBOTHAM

A Thesis submitted to the University of Huddersfield in partial
fulfilment of the requirements for the degree of Master of Science by Research

Applied Psychoacoustic Lab

April 2016

ABSTRACT

Early reflections play a large role in our perception of sound and as such, have been subject to various treatments over the years due to changing tastes and room requirements. Whilst there is research into these early reflections, arriving both vertically and horizontally in small rooms regarding critical listening, little research has been conducted regarding the beneficial or detrimental impact of early vertical reflections on listener preference, in the context of listening for entertainment.

Two experiments were conducted through subjective testing in a semi-anechoic chamber and listening room in order to assess subjects' preference of playback of a direct sound against playback with the addition of the first geometrical vertical reflection. Program material remained constant in both experiments, employing five musical and one speech stimuli.

Experiment one used a paired comparison method assessing a subjects' preference, and perceived magnitude of timbral and spatial difference provided by a frequency *independent* ceiling reflection. Each comparison was followed by a free verbalisation task for subjects to describe the perceived change(s). Experiment two investigated this further by focusing specifically on subjects' preference with a frequency *dependent* reflection. A more controlled verbalisation task provided a list of descriptive terms which the subject's used to describe which attribute(s) influenced their preference.

The results show that preference for playback with the inclusion of a vertical reflection was highly varied across both subjects and samples. However both experiments suggest that the main perceptual attribute with which subject's based their preference was timbre, common spatial attributes (image shift/spread) cannot be used to predict preference. Experiment two suggests that the alteration of the frequency content of a vertical reflection, may also provide a more consistent level of preference for certain stimuli. It is also shown that while certain attributes occur frequently (brilliance/fullness) for describing preference, others less frequently used (nasal/boxy), may influence preference to a greater extent.

CONTENTS

FIGURES

TABLES

EQUATIONS

NOTATION

TDS	Time Delay Spectrometry
ms	Milliseconds
cm	Centimetres
m	Meters
LEDE	Live End Dead End
RFZ	Reflection Free Zone
RT	Reverberation Time
α	Average absorption coefficient
EDT	Early Decay Time
ITDG	Initial Time Delay Gap
J_{LF}	Lateral Energy Fraction
IACC	Interaural Cross Correlation
JND	Just Noticeable Difference
SI	Speech Intelligibility
$C_{50/80}$	Clarity Index - with respective time division in *ms*
D_{50}	Definition
TS	Centre Time
ft	Foot
Hz	Hertz
kHz	Kilohertz
ILD	Inter-aural Level Difference
ITD	Inter-aural Time Difference
ASW	Apparent Source Width
LEV	Listener Envelopment
ESI	Early Spatial Impression

ACKNOWLEDGEMENTS

Many thanks to my supervisor Matthew Stephenson, to whom I am extremely grateful for his generous guidance, patience and understanding, both prior and throughout this research.

Thanks to all University of Huddersfield academic staff and post-graduate students who participated in either experiments.

Finally, thanks to my family and friends for your support and encouragement.

CHAPTER 1
INTRODUCTION

1.0 INTRODUCTION

Reflections within concert halls have always been an integral part of the performance. Extensive research has studied the use of absorbing or promoting early reflections within a large diffuse sound field. These changes result in different spatial attributes which, depending on the desired sound, are made to ultimately enhance enjoyment. When focus shifted to understanding our perception of reflections within small rooms, this was done mainly through our knowledge of concert hall acoustics and measurements. However, little research has been done to understand early reflections within small rooms from a subjective perspective, to discover if they provide a positive or negative spatial *and/or* timbral impact, on our listening pleasure. This research presents two experiments investigating the effect(s) of the first geometrical vertical reflection from above on subject preference in the context of *listening for entertainment or pleasure* within a small room. With the lack of research in terms of subject preference in this field, this work opens further questions within small room acoustics for investigation. The author's intent throughout this research was to make any conclusions as relevant as possible to practical application and as such, listening tests conducted in both experiments utilised five musical and one speech stimuli.

1.1 SCOPE OF THIS THESIS

Small rooms referred to in this thesis are those typically found in a domestic environment utilised for home theatre systems. Reflections arriving vertically may also originate from a low hanging surface (reflector/absorber) as opposed to a ceiling. Therefore, stronger reflections are considered here for the purposes of investigating their perceptual effect(s). In order to remove perceived unintentional lateral effects, a single mono sound source was used to play the *direct sound* on axis to all subjects at head height. Although it is acknowledged that this (mono source) may not be applicable to everyday scenarios, an understanding of a simple vertical reflection without inter-aural differences from stereophonic reproduction was necessary in order to propose further research.

The main aim of this research is to achieve a greater understanding of a vertical reflection's impact on a listeners' preference of reproduced sound, through perceived timbral and/or spatial attributes. Two experiments were conducted in order to meet this aim:

- Experiment 1 – Frequency independent semi-anechoic room test
  a. Assessment of subject's level of preference for playback with a frequency *independent* reflection present, against playback with just the direct sound.
  b. Subjective experiment assessing the level of timbral and/or spatial difference perceived by subjects in the presence of a frequency *independent* vertical reflection in comparison to just the direct sound. This was followed by a brief free verbalisation task after each comparison discussing any perceived difference.
- Experiment 2 – Frequency dependant listening room test
  a) Subjective experiment conducted in an ITU-R BS.1116 compliant listening room assessing subject's level of preference for playback with a frequency *dependant* reflection present, against playback with just the direct sound. This was followed by a constrained verbalisation task identifying reasons for subjects' preference via provided descriptive attributes (Appendix D).

1.2 Structure of This Thesis

The literature review presented in Chapter 2 aims to provide the reader with a brief overview of the development of small room acoustics and designs for critical listening. Although in a different context, understanding critical listening environments provides useful information regarding the control of early reflections within small rooms. In addition, there is a discussion of objective measurements and perceptual attributes and the sensory relationship between the two from the literature within this field. While experiments conducted in this thesis do not specifically analyse objective measurements (RT60, C_{80}, D_{50}) against subjective preference, understanding of measurement values is important in the discussion of perceptual descriptions and resulting preference (Kaplanis, Bech, Jensen, & van Waterschoot, 2014). Measurements taken within experiment two (section 5.1), showing certain physical parameters are included to provide understanding of the reproduced sound field in the context of the literature discussed.

Chapter 3 discusses the simulation process of a single vertical reflection. Testing was conducted and checked against the response of a real reflection with minimal absorption, taking into consideration loudspeaker frequency dispersion of the direct sound. Chapter 4 introduces the first experiment with a breakdown of the methodology and test administration, followed by a discussion of these results concluded with a summary. Some observed limitations of the experiment are discussed. Chapter 5 follows the same discussion process for experiment two. Chapters 6 and 7 further evaluates the results of both experiments and discusses the novel contributions to this field of study regarding vertical reflections and preference.

CHAPTER 2

LITERATURE REVIEW

1.0 INTRODUCTION

The following chapter presents literature and theory surrounding reflections in both small rooms and concert halls, the applicability of measurements used in concert hall acoustics for small rooms and how these relate to perceptual attributes and descriptors. Auditory spatial and timbral sensations of lateral reflections are then covered, followed by a discussion of elevated sound sources and reflections. Little research has been done solely investigating *vertical* reflections, therefore literature reviewed in the context of elevated sound reproduction will be highlighted, along with any limitations in the application to this thesis.

In an outdoor environment, sound is allowed to propagate freely with minimal reflections reaching our ears. In a closed environment sound is constrained by boundaries and objects which, when summed at a singular point alter its original characteristics. Depending on a room's design and its intended purpose, these alterations could be perceived to be beneficial or detrimental. Large auditoria and concert halls employ the use of natural reverb to carry the music across to each listener position. The reflections created by an environment, add tonal and spatial qualities to the sound that we (up to a point) find pleasant (Olive & Toole, 1988). A small rooms' acoustics are different and theory cannot be directly applied from concert hall acoustics in the same way. Small listening spaces with too much reverb would be disadvantageous to the room's purpose of enjoying audio. Vice versa, listening in an acoustically 'dead' room could be strenuous and un-enjoyable. It is important to note in comparison to a large concert hall, a cinema could be viewed as small room. However in the context of this thesis, small rooms are those typically found in a domestic home, office spaces, small classrooms and recording studios.

The amount of reverberation generated in concert halls is not *naturally* possible in small rooms (Toole, 2006), however the level of reverberation is still a key aspect, much to the point where design requirements limit reverb to certain amplitudes at specific frequencies (ITU-R BS.1116-1:1997)(ITU-R BS.6840-13:1998). At lower frequencies, the internal dimensions cause standing waves (room modes) around the room and are often said to be harmful to the perception of an audio signal through frequency response errors (Reiley & Grimani, 2003).

Therefore, rather than considering overall reverberation time, focus of this investigation is to examine the influence of *early reflections* on the way we perceive audio through timbre and/or spatial attributes. More specifically vertical reflections, since first reflections typically originate from the floor and ceiling rather than laterally (Kaplanis et al., 2014).

2.1 Control of Reflections in Studios

Although people's perception of room modes can differ, the general consensus in studios is to control them as much as possible, to create as smoother frequency response at the listening position. Research of early reflections and reverberation at higher frequencies in these small rooms has resulted in many different designs over the years. With changes in trends and opinions, it is impossible that one single design will meet the requirements of everybody's tastes. The need for an *effective* design became apparent when music was beginning to travel between locations and with the introduction of stereo playback. Prior to this, mono loudspeakers and the listening position could be moved to achieve a reasonably desirable sound (Newell, 2007).

Tom Hidley's designs became one of the first commercially viable studios where interestingly, he chose to *promote* early ceiling reflections with a 'compression ceiling' (Cox & D'Antonio, 2009). A design with a low hanging ceiling at the listener position, then elevated towards the front above the speakers and at the rear behind the listening position. Voetmann (2007) also states that Hidley's early designs incorporated a ceiling canopy however, this was later removed.

In the late 1960's Time Delay Spectrometry (TDS) was introduced by Richard Heyser which in turn led to the objective measurements of Hidley's designs by Don and Carolyn Davis (Davis, 1979). TDS was an improvement of the current sine sweep method of a rooms frequency response (sometimes known as house curve) whereby a microphone records from a specified offset of time from the initial signal and processed through a narrow-band filter, thus allowing the selective spatial data of a sound field (Heyser, 1967). These measurements revealed significant regular inconsistencies along the frequency response that we now know as comb filtering[1]. This use of TDS led to the development of the 'Live-End, Dead-End' (LEDE) in the late 1970's (Davis & Davis, 1980). The control of reflections is highly contrasting to those of Hidley's, proposing the front should be highly absorptive (dead) and the rear reverberant (live).

[1] Discussed further in section 2.4.4

The reason behind this change is the perception of the early reflections, recognising that they provided the listener with important information about the environment. Direction, density and frequency of these first reflections affect our perception of t*imbre* and tonal *imbalance* (Newell, 2007) that helps create an impression of our surroundings and so the LEDE aspires to place us in a neutral environment by *removing* these properties.

This design was well received and later built upon by D'Antonio utilising Schroder diffusers on the rear wall and broadband absorption of early reflections at the front. A continuation of this design was the Reflection Free Zone (RFZ), which utilised the room's geometry in such a way that no reflections entered the listening position without abundant need for absorbers (Everest & Pohlmann, 2009). This was effectively able to create a stereo image across a frequency range of 500 to 5000 Hz (Voetmann, 2007). In addition to this, Angus (1997) demonstrated that the use of diffusers can also be implemented to achieve the desired RFZ. He states that due to the diffusion of the early reflections, the listener is granted more freedom in terms of '*listener position*' and a reduction in comb filtering at higher frequencies and offers the potential for a better control room in combination with other techniques.

It is apparent that the consensus in modern designs is to remove early reflections from the listener's position as they hinder the judgement of playback material during critical assessment. These early reflections add properties of an environment that would otherwise be absent in another room. However, a number of questions remain: *How* do these early reflections impact our perception of audio? How can we objectively observe and subjectively describe these auditory sensations? Are these auditory sensations categorically present or absent, or we able to perceive varying magnitudes of change?

2.2 Objective Metrics

The perceived differences of the same sound played in different environments may be related to objective measurements of the sound field. The change sound undergoes from propagation at source, to interpretation by listener, may be observed as a filter model (Figure 1). It is therefore useful to compare what measures of the sound field *h(t)* correspond to what perceived aspect of *y(t)*. ISO 3382-1:2009 outlines five specific subjective listener aspects [2] in

[2] 1 – Subjective *level* of sound, 2 – Perceived *reverberance*, 3 – Perceived '*clarity*' of sound, 4 – Apparent source width (ASW), 5 – Listener envelopment.

performance spaces which, within each, hold one or more acoustic quantity. However it seems difficult to categorically state that any one variable affects one perceived perceptual effect, and more so the application of objective measurements used in concert halls to small rooms. The following section will briefly introduce objective measures used in concert hall acoustics and their applicability to small rooms.

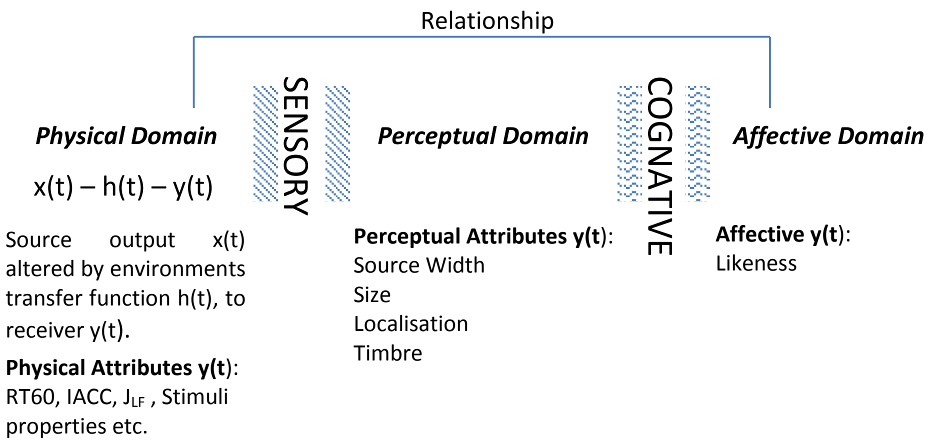


FIGURE 1: FILTER MODEL ADAPTED FROM KAPLANIS ET AL. (2014)

2.2.1 REVERBERATION TIME

One of the most common objective measurements is reverberation time (RT). Documents such as ANSI S12.60-2010, ITU BS.1116-1:1997 and ISO 3382-2:2009, all discuss limitations and tolerances of reverberation times for different uses and criteria. In RT60, the use of 60dB is used as a ratio for how much the energy must dissipate. However, in small rooms ratios of 20dB (-5 to -25dB) and 30dB (-5 to -35dB) are often used and then scaled up relative to a 60dB ratio, for a uniform measurement no matter room size. Room impulse responses (Figure 2) are measurements taken to visually investigate decay rate, density and amplitude of the reverberant sound field rather than just a value. Griesinger, (1996) and Seetharaman & Tarzia, (2012) both discuss techniques for capturing RT, along with visualised 3-dimensional method by Dunn & Protheroe, (2014). This may be beneficial in identifying areas of a sound field that propagate certain attributes associated with a standard impulse response, especially in smaller rooms such as from a ceiling or floor.

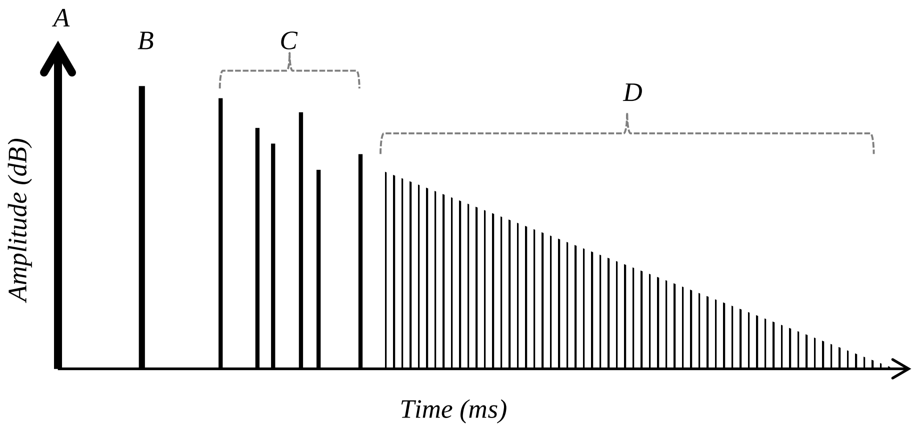


FIGURE 2: SIMPLISTIC REPRESENTATION OF AN IMPULSE RESPONSE. A = INITIAL SOUND. B = DIRECT SOUND. C = EARLY REFLECTION FROM CEILING, FLOORS AND WALLS. D = LATE REFLECTIONS

2.2.2 First and Early Reflections

Between the direct sound and the first reflection is a notable gap for the time taken for sound to travel the second shortest distance to the receiver: the 'Initial Time Delay' , or 'Initial Time Delay Gap' (ITDG – Figure 3) (Beranek, 2008). The understanding of ITDG resulted in the success of the LEDE studio design, whereby this control of early reflections in the control room did not mask the ITDG of the studio (Davis, 1979). ITDG however, is still criticised as a higher value *could* imply a greater distance laterally however it is often reflections from the median plane that arrive first (Ouis, 2003). This is especially true in small rooms where (from the centre of the room) distances to the floor/ceiling are generally shorter than to the sides. Therefore, the use of this measurement within small rooms could be misinterpreted.

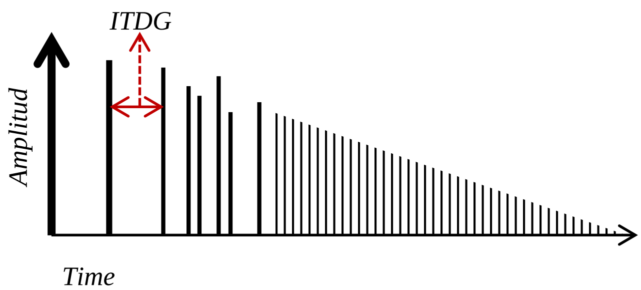


FIGURE 3: INITIAL TIME DELAY GAP

Progressing through the room impulse response we see discrete early reflections. A rooms' early lateral reflections can be objectively measured through two methods: lateral energy fraction and inter-aural cross correlation. Lateral energy fraction J_{LF} (sometimes denoted as LF, LF_E, J_{LFC}, ELEF) (Equation 1) is a ratio of energy between measurements taken with figure-of-eight (between 0.005*s* and 0.080*s*) and omnidirectional microphones (between 0*s* and 0.080*s*), where *pL(t)* is the sound pressure of the figure-of-eight microphone and *p(t)* of the omni-directional microphone (Figure 4) (ISO 3382-1:2009).

IACC (sometimes seen as $IACC_E$ for energy arriving *early*, before 80ms) is measure of difference in sound arriving at the two ears (Beranek, 2008) and also associated with lateral arriving energy. Cross correlation is a measurement rated from values 1.0 - exactly the same, to 0.0 - being totally different and via a binaural dummy head. The uses of these metrics provide information about how we may perceive a rooms spatial attributes (Section 2.4).

[1] – *Lateral Energy Fraction*

$$J_{LF} = \frac{\int_{0.005}^{0.080} pL^2(t)dt}{\int_{0}^{0.080} p^2(t)dt} \qquad [1]$$

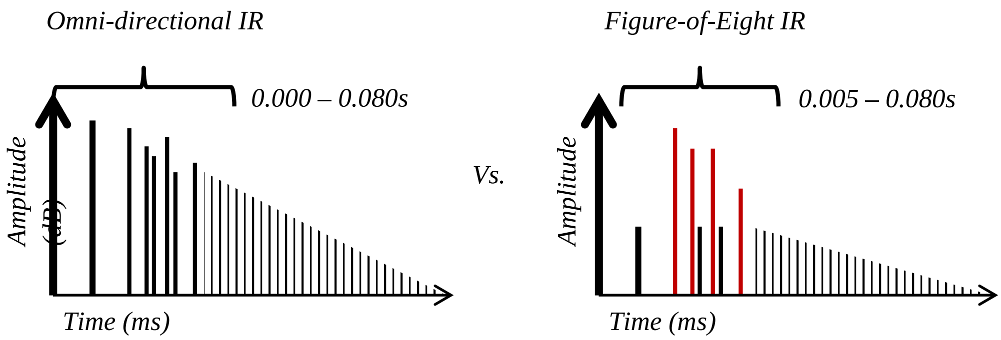


FIGURE 4: COMPARISON OF ENERGY BETWEEN OMNI-DIRECTIONAL AND FIGURE OF EIGHT IMPULSE RESPONSES

2.2.3 Early to Late Reflections

Another objective measurement used to analyse a room's acoustic is 'Clarity'. However, this can be ambiguous in terms of how we define what clarity actually is. Depending on the stimuli used, this may be the ability to perceive musical detail, the onset of notes or the intelligibility of speech (Speech intelligibility S.I) etc. These perceptual attributes are discussed in further Sections. As a standard objective measure, clarity is defined as a ratio of early to *late arriving* energy (Equation 2), and noted as C_{50} or C_{80}, expressed in dB (John S. Bradley & Soulodre, 1995a). Depending on the stimuli and room, 50 or 80ms is chosen as the dividing point. The perceived aspect of 'clarity' (in ISO 3382-1:2009) is also associated as a function of *definition* (Equation 3) and *centre time* (Equation 4).

Definition (D_{50}) is a ratio of the early to *total* sound energy expressed as a percentage using 50ms as a dividing point. Centre Time (*TS*) is the *"time of the centre gravity of the squared impulse response"* (ISO 3382-1:2009) and expressed in milliseconds. The greater level of early energy (or reduction in later energy) increases the levels of C_{80}/C_{50} and D_{50} and therefore generally improves clarity, suggesting that early reflections are beneficial for this attribute.

With more/stronger early reflections *TS* would decrease as the centre of gravity shifts more towards the direct sound and away from the reverberance. In Bradley and Soulodre's (1995a) paper, the use of C_{80} in conjunction with an adaptation of sound *strength* (denoted as *G*(A) - with an A weighting frequency response more typical to that of human hearing sensitivity), was found to be more accurate in the prediction of clarity than singularly C_{80} or *TS*. This however was in the context of concert hall and was not considered in small room acoustics.

[2] – *Clarity Index (80ms division)*

$$C_{80} = 10 \cdot Log\left(\frac{\int_0^{0.080} p^2(t)dt}{\int_{0.080}^{\infty} p^2(t)dt}\right) \qquad [2]$$

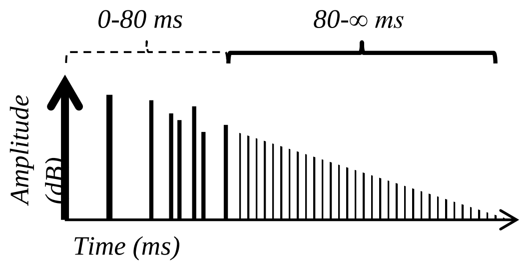


FIGURE 5: ENERGY DIVIDE IN CLARITY

[3] – *Definition Index*

$$D_{50} = \frac{\int_0^{0.050} p^2(t)dt}{\int_0^{\infty} p^2(t)dt} \qquad [3]$$

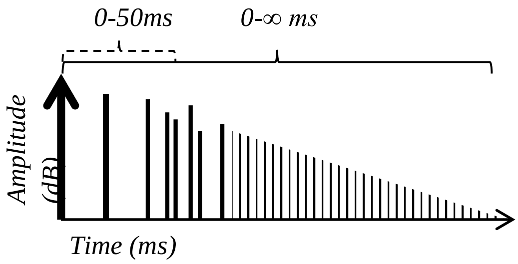


FIGURE 6: ENERGY DIVIDE IN DEFINITION

[4] – *Centre Time Index*

$$TS = \frac{\int_0^\infty t \cdot p^2(t)dt}{\int_0^\infty p^2(t)dt} \qquad [4]$$

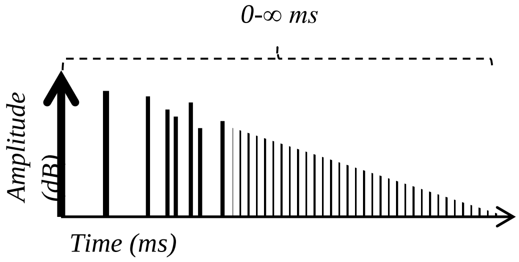


FIGURE 7: ENERGY USED IN TS

Figures 5, 6 & 7 – View of energy used along the impulse response in respective equations. For all equations, p(t) is the instantaneous sound pressure of the impulse response measured at that point.

It is important to note the applicability of these parameters in the context of small room acoustics. Most of the definite integrals above have divisions corresponding at times that separate early and late arriving reflections. This is for concert halls and large auditoria, between 50-80ms. Given that a small rooms RT will seldom achieve this naturally, some metrics such as L_{JF}, C_{50}/D_{50} and C_{80} may not be representative of the real subjective impression, being based on the time division of early and late energy (Kaplanis et al., 2014). Metrics such as TS may be a more useful tool in seeing the distribution of energy and not relying upon a finite point in time to yield an effective result.

2.3 The Effects of Early Reflections

Changes of first reflections have a great impact on our perception of the incident signal, producing differences, whilst still having a negligible effect on the total reverberation time (Niaounakis & Davies, 2002). The following section discusses research that has investigated the effect of early reflections from varying directions.

Kishinaga, Shimizu, Ando, and Yamaguchi, (1979) conducted subjective testing to provide supporting evidence that early reflections strongly impact the sound field. Their experiment included the set-up of multiple configurations of absorptive and reflective material on all walls. The floor, was constructed of parquet and the ceiling, of rock wool absorbing board (floor and ceiling materials remained constants throughout). Subjects were instructed to record their impression on certain attributes and compared against IACC measurements. Results showed that for *critical listening*, absorptive side walls are desirable yielding an IACC of 0.44. To *enjoy* the music, reflective walls are preferred with the IACC value measured at 0.26. This would imply that early reflections have a detrimental impact for the critical evaluation of audio and beneficial for enjoyment.

Investigations into the effect of singular reflections on the timbre of reproduced audio within a sound field have proven reflections from above impact our perception. Bech's (1994a) research investigated which early reflections are strong enough to impact perceived timbre *individually*, and the required level of these reflections needed to produce this change. Results showed that only floor and ceiling reflections could possibly be heard as individual reflections affecting timbre within the sound field. Unlike previous experiments considering individual reflections (Olive & Toole, 1988), Bech simulated a reverberant sound field using six additional loudspeakers. Bech (1995a) later reinforced his findings confirming that floor reflections contribute to the overall timbre of the sound field given *noise* stimuli. Bech's work and methodology is discussed further in detail in Section 2.7 regarding the effects of vertical reflections.

Griesinger (2009) states that *lateral* reflections from 10 to 50ms contribute to the feeling of "*distance*" to the sound image creating space between the listener and source, mentioning that it is not specific reflections by themselves that generate this effect, but collectively. Griesinger also highlights an 'Insitution of Acoustics' (IoA) conference, Krokstad[3] explains the sense of "*involvement*" is actually what acousticians are seeking to accomplish in concert halls and that the direct sound to reverberation ratio (D/R), has a strong impact on this perception of *distance* rather than "*envelopment*" (see Section 2.4.3). Although related to concert halls, Griesinger does go on to mention that the addition of early reflections in smaller halls often *hinders* clarity rather than improving it and the direction of these early reflections is *not* audible where there is dominant reverberation energy. His experiment on treating a small concert hall (350 capacity, RT occupied 1.0s at 100Hz) showed that the addition of 700ft^2 of absorption on stage and absorbing *first order* lateral reflections above 1kHz dramatically improved the sound, implying that even in rooms larger than critical listening spaces, the attenuation of lateral reflections has a *positive* impact on the sound field. However, in small rooms the energy of late reflections will seldom dominate those of early reflections, thus the directionality of these early reflections *could* potentially alter a person's preference.

[3] No reference given in Griesinger (2009)

Imamura, Marui, Kamekawa, and Nakahara, (2013) recently carried out an experiment to see how this directionality of early reflections within a sound field is perceived by a listener. Using a dummy head, nine impulse responses were taken with various patterns of acoustic treatments using absorptive panels. These were then convolved with three music stimuli. The author notes that the difference of RT was almost equal throughout the patterns and therefore, specifically focused on early reflections. Through headphones, subjects were asked to rate stimuli against each other using evaluative terms on individual listening impressions. Terms used were: timbre brightness, width of sound image, envelopment, clarity, timbre naturalness, reverb suitability and listeners' preference. Out of these, width, envelopment and clarity displayed significant differences throughout the nine arrangements. It was concluded that as more first *lateral* reflection points are covered with absorption, width of sound image will narrow and also *envelopment* will lower. Vertical panels were not added individually, but simultaneously with front wall absorption thus the influence of early ceiling reflections cannot be solely assessed. However, the author points out that absorption at these points not only decreased sound image width but increase *clarity*. Imamura, Marui, Kamekawa, and Nakahara, (2014) further investigated perceived *clarity* based on previous results and simulated reflections via a lateral loudspeaker array varying directionality and delay times of reflections. The paper concludes by stating lateral reflections have a strong effect on perceived spatial clarity, apparent source width and listener envelopment. However this investigation did not incorporate vertical reflections, when previous results stated that increased clarity was also in conjunction with ceiling absorption.

Subjective studies assessing the effect of lateral energy on the *adaptability* of an engineer carried out by King, Leonard, & Sikora, (2011) asked users to create a simple mix between orchestral backing and solo soprano stereo stems within *critical listening* spaces. Subjects were instructed to change the level of a stem by 0.5dB increments until satisfied with the mix. Two control rooms were utilized to increase the subject pool with RT60's of 200ms and 175ms. The surfaces of first geometrical reflections points were altered three times using: absorptive, diffusive and reflective treatment. It is important to note acoustical alterations were made behind acoustically transparent fabric, and the treatment unemployed remained in the room to maintain RT. After three training attempts, each acoustic treatment contained three trials of three music excerpts totalling 27 individual mixes per engineer. The results showed that the treatment had no significant impact on the variance of level but interestingly when asked which treatment was easiest to mix with, subject's *preferred* reflective treatment. A continuation of this work shows that the room treatment has a significant effect on the levels of reverberation set within a mix (Leonard, King, & Sikora, 2012). Reflective panels added laterally at ear height (increasing the high frequency reverberation time by 20ms) resulted in a lower mean mix of added reverberation. Therefore, although the human brain may be able to adapt to varying sound fields, lateral energy can still affect judgement in reverb related mixing tasks.

From the papers discussed it is clear that early reflections can alter the way in which we perceive the direct sound as well as altering the way in which we perform certain tasks. However, it is necessary to try quantify the effects of these reflections within the perceptual domain into specific attributes. As pointed out by Kaplanis et al. (2014), although many of the aspects are from concert hall acoustics, assessment of these attributes could identify common characteristics which may also affect perception in smaller rooms.

2.4 Perception of Lateral Sounds

In the previous section, papers were discussed showing that early reflections impact the way in which we perceive sound through subjective testing. Aspects such as: apparent source width, listener envelopment, spatial impression, timbre, depth, distance, colouration and clarity are some of the many terms which comprise auditory sensations of an environment. In relation to Figure 1, this is the perceptual domain of the psychoacoustic relationship. The following section outlines numerous perceptual attributes we use to *describe* sound fields and the relevant theory behind the auditory processes.

The ability to localise where a sound is coming from is dependent upon a number of factors. As seen in Figure 8, a direct sound in an enclosed space will bounce off surfaces and create multiple reflections with varying characteristics which will ultimately arrive at the receiver. The summation of these reflections, allows us to localise signals. Consequently, these reflections will provide spatial differences which will alter our spatial impression of a sound. This may also be an attribute on which subjects may base their preference.

The precedence effect (sometimes known as 'law of first wavefront' or 'haas effect') is the binaural phenomenon that enables us to determine the origin of a sound in an acoustically complex environment (Olive & Toole, 1988). When the direct sound is followed by a delayed version of that sound within less than 1*ms*, then '*summing localisation'* occurs and the incident and lagged sound are fused together. From 1*ms* – 5*ms*, sounds are still fused together (depending on the stimuli used) and the image is still perceived as one, this stage is called '*fusion*'. Above a certain point, roughly 5ms, referred to as the '*echo threshold'*, the image begins to split into two spatially separate sources (Litovsky, Colburn, Yost, & Guzman, 1999). The binaural cues used ascertain the information needed from these reflections to process the perceived direction of the auditory event using two receivers. Inter-aural level difference (ILD) is the amplitude difference of a sound arriving at two ears. Inter-aural time difference (ITD) is the difference in time is takes for a sound to reach one ear from the other. As our ears are positioned along the horizontal plane, these cues are paramount in comprising what we conclude to be the '*spatial impression'*. These ITDs and ILDs can also be provided by a direct sound and a delayed and attenuated version of that sound from a reflection which we have control over, and as such may be able to alter our preference.

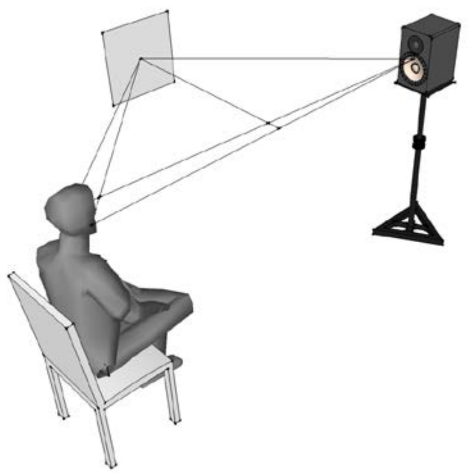


FIGURE 8: ITD'S AND ILD'S PROVED BY A LATERAL REFLECTION

2.4.1 SPATIAL IMPRESSION

Upon entering a sound field, we build an impression of our surroundings based upon the properties of reflections given by geometrical boundaries. The arriving angles of certain reflections contribute to different auditory sensations, it is understood that the energy received laterally is beneficial to our understanding of '*spatial impression*' (Toole, 2008). Spatial impression has been subject to more than one definition over the years and thus presents some confusion. Describing a *physical aspect*, and to perceive a *phenomenon* based on the relationship between said aspect and phenomenon, are different things (Lehnert, 1993). For instance, one may describe a sound field as *spacious* attempting to define a rooms' physical scale. However, this may be interpreted as a feeling of *distanc*e and the space between listener and source as a *result* of the rooms' scale. In concert hall acoustics, spatial impression is often broken up into two main components: apparent source width (ASW) and listener envelopment (LEV). These are standard components referred to in criteria of performance spaces such as ISO 3382-1:2009.

2.4.2 Apparent Source Width

ASW is influenced by the relative strength of early reflections, it is the combination of the direct sound and the *lateral* energy that gives the perception of how horizontally wide a sound source is. This attribute can be objectively interpreted though lateral energy fraction (J_{LF}), or an inter-aural cross correlation (IACC) measurement at the listener position. Bradley & Soulodre (1995a) state that increasing early lateral reflection energy only leads to a broadening of the source image and those late arriving reflections (typically above 80m) contribute to LEV, though once again this is in the context of concert hall acoustics. Griesinger (1998) however, proposed new metrics: '*Diffuse Field Transfer'* function (DFT) related to envelopment, and '*Average Inter-aural Time Delay*' (AITD) relating to early lateral reflections within small rooms. Griesinger also describes spatial impression as '*Early* Spatial Impression*'* (ESI), given that most reflected energy in small spaces occur within 50ms and is perceived to be predominantly frontal.

Data from Ando (1977), Hidaka (1997) and Barron & Marshall (1981) is collated by Toole (2008) regarding the directionality of reflections, ASW and preference. Results from Ando and Barron and Marshall show that a decreasing IACC values are much *preferred*, and (given that IACC is a measure of ASW) subsequently relates to greater spatial impression. Frequency dependant studies on ASW have also shown some frequencies may be more effective at creating a wider image than others. Hidaka also concludes that ASW strongly corresponds to frequencies from around 350 to 2800Hz. He created $IACC_{E3}$ to represent this as a function of average IACC over three octave bands (500Hz – 2kHz). He also concluded that lateral reflections arriving at 60° provided the greatest contribution to sense of ASW. Morimoto & Iida's (2005) show that objective measures of IACC (referred to as ICC) are *not* affected by frequency content above 1kHz, *but* subjective assessments of ASW *did* increase as lateral reflections moved closer to 90° and not 60°. This was found by increasing low pass filter cut-off points, on broadband noise from 200Hz - 1, 2, 4 and 8kHz.

ASW has been noted by many studies within concert halls to be a major factor in subject's *preference*. Kuusinen, Pätynen, Tervo, & Lokki (2014) mention that early energy arriving laterally comprising of higher frequency content, correlated well with listener's preference for classical music excerpts in nine varying concert halls. It is also stated that the increase of 'width' is also in combination with 'depth' when broadband lateral reflections are provided. In multi-channel reproduction Choisel and Wickelmaier (2007) demonstrate the change in subject preference through four musical stimuli of varying loudspeaker configurations and rating of attributes comprising of spatial and timbral listings. Although difficult to pinpoint preference with a singular auditory sensation, is it noted that for a centred listening position stereo configuration was among the most preferred playback formats than that of mono, supporting evidence that a greater ASW is more preferable.

2.4.3 Listener Envelopment

The other component to spatial impression is listener envelopment (LEV). And while the ASW is influenced by angle, intensity and frequency of early lateral reflections, LEV is impacted by the reverberance of a room's impulse response – typically above 80ms for concert halls (J.S Bradley & Soulodre, 1995b). It is the sense of being surrounded by, or being in the centre, of a reverberant sound field succinctly expressed by Blesser & Salter (2006) as "*analogous of swimming underwater, [rather] than being sprayed by a water hose*". The reverberant field is constructed via the increasingly shortening spatial distribution of reflections from all geometrical boundaries. However, reflections arriving from the rear of the listener are more enveloping than frontal energy (Griesinger, 1999). With regards to frequency, reverberation time at both low and high frequencies significantly affect the listener envelopment (Morimoto, Jinya, & Nakagawa, 2007). In small rooms however, it is reasonable to say that true envelopment cannot be achieved due to their small volume incapable of producing longer reverberation times, and can only be achieved via the use of multichannel loudspeaker arrays (Toole, 2008).

2.4.4 Early Lateral Reflections And Timbre

Timbre is often associated with descriptors such as; *rich, dull, bright, harsh, coloured, smooth, mellow* and so on (Howard & Angus, 2013). In terms of sound reproduction in a listening environment, acoustical barriers and use of treatments influence the way in which one sound can be perceived in multiple ways. Although late reflections have been shown to impact our perception of timbre, with changing echo density and intensity resulting in different characteristics (Huang & Abel, 2007) (Huang, Abel, Terasawa, & Berger, 2008), it is the early reflections which provide the 'unique fingerprint' of the room that are of interest here. As briefly mentioned in Section 2.3, research has shown that early reflections contribute to our sense of timbre and clarity (Bech, 1994b) (Imamura et al., 2014).

Comb filtering is one of the most discussed effects of lateral energy in relation to timbre. The effect is a result of the repetition of a signal added in rapid succession to its predecessor which can be objectively seen from frequency response measurements of a room. In the context of this thesis, this would be the summation of individual delayed reflections arriving at one receiver. Regularly spaced peaks and troughs, constructive and destructive interference respectively (Toole, 2008), are the result of early reflections being added to the direct sound. Two phenomenon may occur in the presence of comb filtering: at low frequencies resonances are generated, but will be mostly dominated by room modes, and at higher frequencies provide alteration to the timbre of the sound. Using white noise as stimuli, this periodic repetition constructive and destructive filtering (Figure 9) is known as *harmonic* and *anharmonic cosine noise* respectively (Rubak & Johansen, 2003).

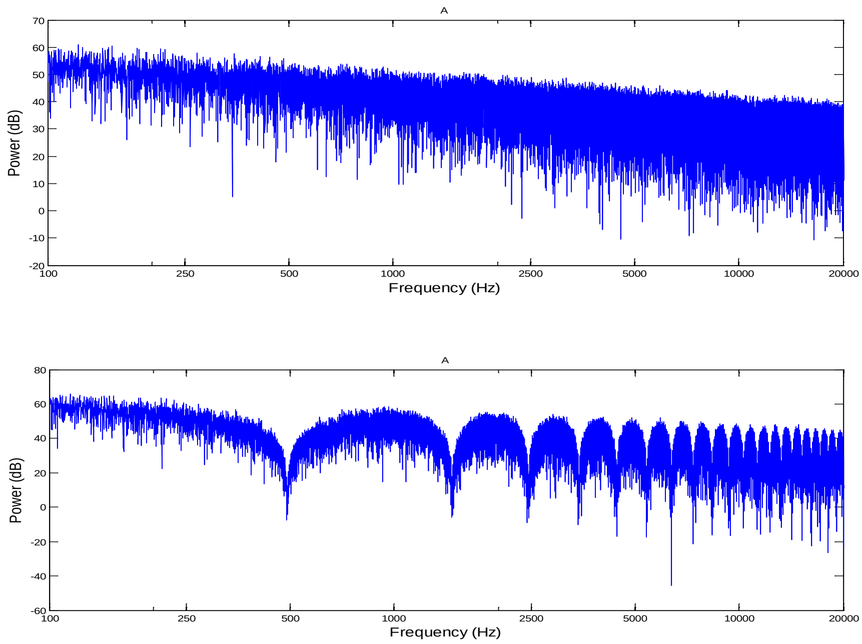


FIGURE 9: COMB FILTER FREQUENCY RESPONSE.

However, even though this comb filter pattern may occur annalistically it cannot be immediately condemned and can, as mentioned by Clark (1983), may actually be a preferable response. In his experiments the assessment of a delayed signal reaching two ears was employed in multiple ways - through stereo playback, mono playback with a single reflection (vertically then horizontally) and finally mono playback with a delay replicated though the same speaker. His conclusions provided evidence that comb filtering through stereo playback was *pleasing* to the listener and can be preferred over a flat response. A single lateral reflection produces minimal audible notches and the delayed sound through the same loudspeaker produced an unpleasant *degrading* effect. Interestingly, notches produced by a delayed reflection arriving *vertically* became more noticeable. Clark suggests this is a result of the reflections paths arriving from the same horizontal angle, but still poses thought for the contribution of vertical reflections to comb filtering, timbre and consequently subject preference. Overall, this paper highlights significant audible differences can occur even when objective measures of comb filtering look very similar, dependant on angle.

In binaural hearing, there is also evidence to suggest the human auditory system may possess the ability to disregard such comb filtering distortions (Blauert, 1997), altering our timbral perception of the event. The use of a "*central spectrum*", summed from two subtly different responses significantly reduce colouration from lateral reflections (Toole, 2008). In monaural listening (listeners with one ear plugged), the colourations in timbre are far more distinct. One consideration involving ILD, is the acoustic shadowing effect produced by a centre body between the two receivers resulting in the head becoming an obstacle. The acoustic shadowing provided by the head means that early lateral reflections reaching the furthest receiver will be attenuated at higher frequencies. Given the average breadth of the human head (measured above and behind the ears) is 14-16 cm, frequencies with smaller wavelengths above 2450Hz would be attenuated thus meaning the effect of comb filtering will also be reduced. The idea that objective observation of comb filtering only gives us a small impression of what we will actually hear is quite apparent.

Research into audibility of comb filtering (Brunner, Maempel, & Weinzierl, 2007) shows that under good listening conditions, noticeable differences still occur when a reflection's level difference is 18dB below the direct sound. Research from Barron & Marshall (1981) regarding *spatial impression* also highlighted lateral reflections' contribution to colouration, but in the context of concert halls. Although the predominant effect with a single reflection was indeed '*spatial impression*', reflections between 10 and 20ms were noticeable for producing '*tone* *colouration'*. For noise signals, reflections arriving from a left wall are suggested to individually alter timbre (Bech, 1995b) with a delay of 9.94*ms* and attenuation of 9.7dB relative to the direct sound. Although this paper does not critically assess the detection of comb filtering, it is pointed out this effect is a result of an added delay reflection, which produced timbral alterations.

This section has identified perceptual attributes provided when a direct is sound summed with a delayed version of itself. In small domestic rooms, unaltered reflections will be louder than those in a concert hall and could therefore, be increasingly destructive (or constructive) to the initial signal. Altered reflections from frequency dependant surfaces will also add varying levels of timbral change due to their spectral content. It has also been shown that binaural listening may have the ability to attenuate comb filtering effects. Therefore it may be seen that comb filtering along the median plane may have a greater impact on our perception timbre, as both ears will be provided with the same signal.

2.5 PERCEPTION OF VERTICAL SOUNDS

Section 2.4 focused on early lateral reflections, this chapter aims to review the literature of sound arriving from a vertical source. There is little literature specifically discussing vertical reflections therefore, the following chapter also incorporates discussions of multichannel elevated audio systems. The applicability if not in the context of reflections will be highlighted. Throughout research into vertical reflections, variations in terminology have been used regarding the axis and the reflection surface. Therefore, the term 'median plane' will henceforth be referred to in this thesis as the vertical plane split symmetrically down a listeners head, front to back (Figure 10).

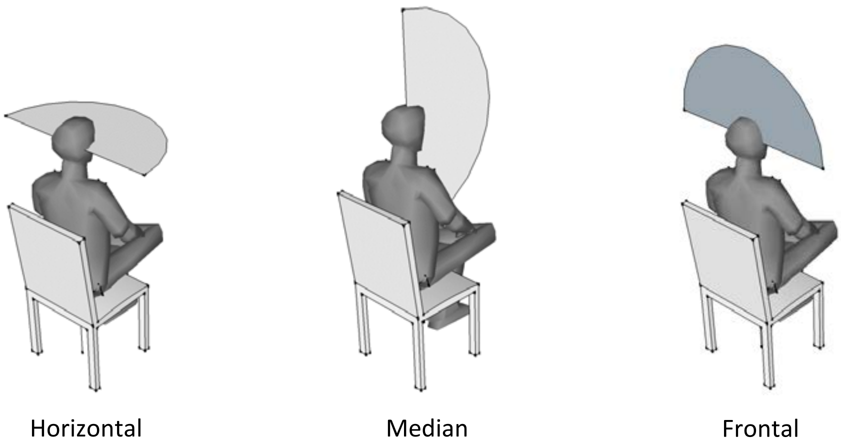


FIGURE 10: PLANES USED THROUGHOUT THESES TO DESCRIBE AS TAKEN FROM HARTMANN (1993)

2.5.1 DETECTION OF VERTICAL REFLECTIONS

The work of Olive & Toole, (1988) provides some fundamental information about vertical reflections related to a "*domestic or control-room size*" environment. By way of sound field synthesis, the aim was to ascertain the threshold at which reflections within a room affected aspects of the sound or, sound field. Users were given control of over a multi-turn potentiometer (so that no positional cues could be used) and instructed to adjust the level of a test reflection somewhere between '*just audible*' and '*just not audible*'. Subjects could also switch the test reflection off, and to its maximum level of 10dB above the direct sound. No time constraints were imposed. The audibility target was to identify *any* change to the sound or the sound field. This would be considered the **absolute** threshold. Pulse noise, pink noise, speech and castanet signals were used as stimuli throughout testing. Results showed the angle of reflection (i.e. side wall or ceiling) produced *no* notable difference in level of *detection*, but the use of stimuli proved to be very sensitive. Absolute thresholds for delays below 10*ms* were considerably lower (subject's required minimal level to identify a change) for continuous sounds – (noise). Above 10*ms*, pulse stimuli retained a lower threshold revealing a crossover of around 10*ms*. Interestingly, the level for the discontinuous stimuli to provide any change, could be as low as -40dB below the direct sound at 20*ms* delay regardless of angle of reflection. The paper reinforces findings for the absolute threshold of a single vertical reflection with *three* different listeners, all showing the same pattern using pink noise. Observing the results Figure 11, it can be seen that from up to 5*ms*, the absolute threshold for the relative stimuli used is roughly at its *lowest*, meaning any change is most noticeable up to this point. It is also noted by the authors that the change in timbre was quite apparent for the *vertical* reflection.

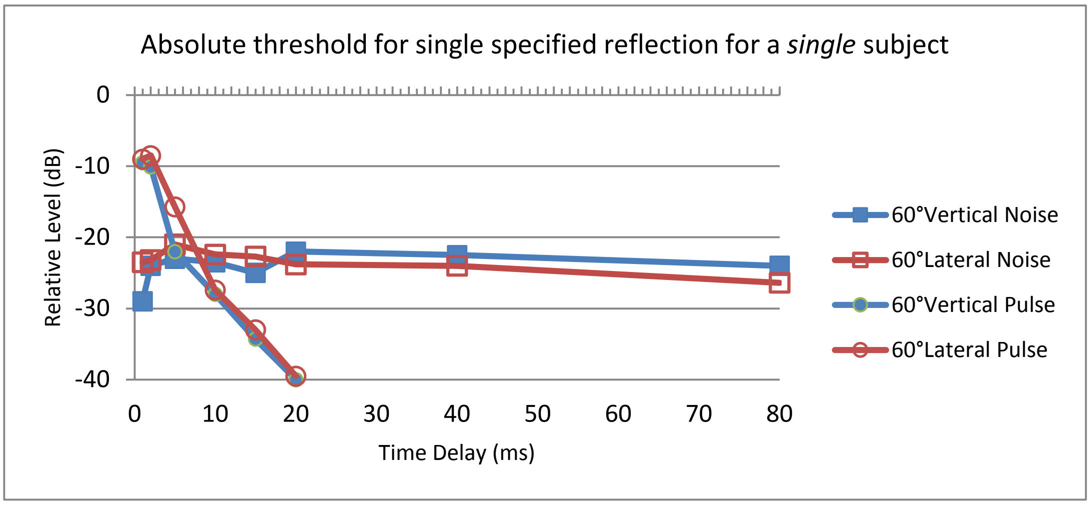


FIGURE 11: ABSOLUTE THRESHOLD RESULTS FOR A SINGULAR SUBJECT. ADAPTED FROM OLIVE AND TOOLE (1988)

The paper also recognised that previous work has often used delayed signals which contain the same spectral content as the incident sound. With surfaces absorbing/reflecting varying energy levels across the frequency spectrum it is necessary to consider reflection as frequency dependant signal.

Bech (1995) produced a number of studies regarding timbre within small rooms with the aim of investigating: which early reflections can individually contribute to changes in *timbre* and what is the required level needed to produce this change? His methodology included electroacoustic simulations in an anechoic chamber similar to that of Olive and Toole. The sound field simulated was typically separated into three components: direct sound, early reflections up to 21*ms*-22*ms* and a reverberant field with reflections greater than 21*ms* -22*ms*. The experiment did *not* simulate an individual reflection solely in the presence of a direct sound, but unlike Olive and Toole, was incorporated into a sound field using multiple loudspeaker sources. Therefore, the amount of speakers needed to individually produce early reflections up to 22*ms* would be too great and were restrained under the following rules. Only reflections above -20dB relative to the direct sound were implemented and multiple early reflections would be produced by the same loudspeaker resulting in: the direct sound, 17 reflections produced by 15 loudspeakers and the reverberant sound field simulated by six, evenly spaced laterally distributed speakers (Figure 12).

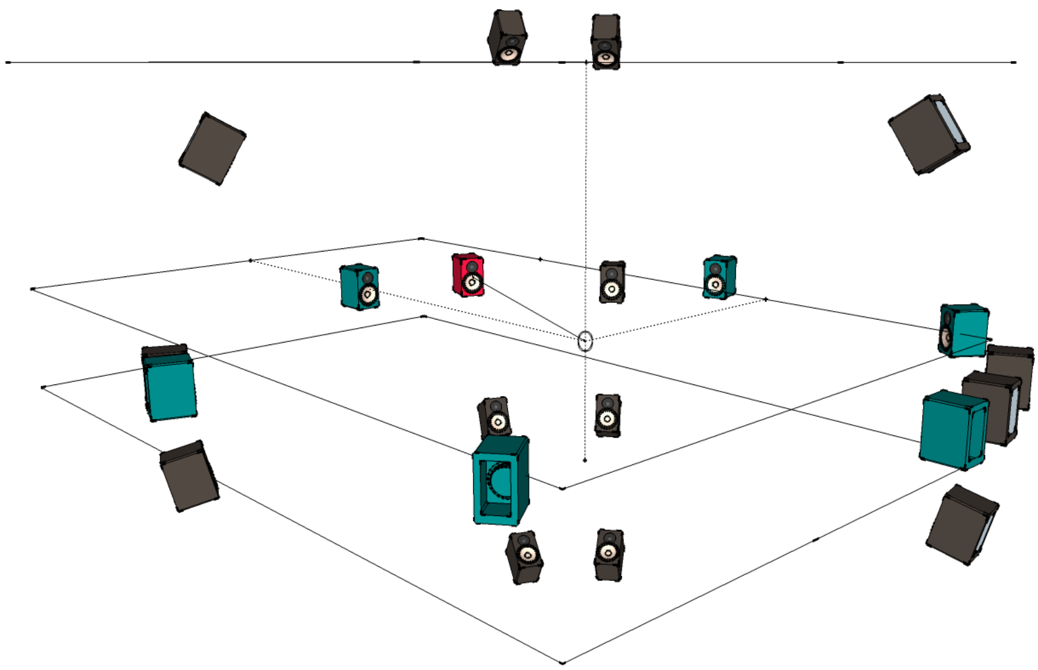


FIGURE 12: BECH EXPERIMENTAL SET-UP DERIVED FROM INFORMATION IN TABLE 1 IN - BECH, 1995

Incident sound is propagated from the **RED** loudspeaker (-22°H, 0°V). **BLACK** loudspeakers represent the reflections. The reverberant sound field is generated from the **CYAN** loudspeakers.

Importantly, the directivity of the direct loudspeaker was modelled in a cardioid pattern *independent* of frequency, thus SPL was radiated evenly at all frequencies. In conjunction with this, absorption coefficients applied to the reflections were also independent of frequency with values: ceiling = 0.05, floor = 0.3 and walls 0.44 resulting in an RT of 0.4s. Loudspeakers were positioned on a 3m hemispherical radius to the listener position and supporting structures treated with acoustic absorbent material to reduce reflection interference. Processing of reflections ranged from 1.64*ms* – 14.98*ms* delay and 3.6dB – 15.5dB level attenuation. The listener was situated on a motorised chair to position the listener's ears and SPL measured at the listener position was 66dB and 50dB for noise and speech stimuli respectively. The stimulus used was 1s pink noise (20Hz – 20kHz) and 3.8*s* sample of male speech. Eight subjects participated (five male / three female) and were free to move their heads as this experiment did not involve localisation.

Each of the 17 reflections were assessed in the presence of the sound field whereby subjects were to ascertain two psychoacoustic properties corresponding to the two aims mentioned previously: The threshold of detection (TD), and just-noticeable difference (JND). The interpretation of timbre was given to the subjects as the American Standards Institution (Section 2.6.2).

Bech's results indicate reflections one, three (median plane), eight and twelve (left wall) resulted in a TD *lower* in dB, or not significantly higher, than the natural levels of reflections in a standard listening room (Table 1). This suggests these reflections are likely to individually contribute to timbre within the context of a sound field. More specifically only reflections one and three will be potentially audible for speech signals *and* noise signals.

Reflection No#	Delay	Attenuation	Lateral Position	Vertical Position
1 – Floor	1.64ms	3.6dB	-25°	-28°
3 – Ceiling	4.16ms	9.2dB	-25°	+48°

TABLE 1: VERTICAL REFLECTIONS CONTRIBUTING TO TIMBRAL CHANGE WITHIN A SOUND FIELD (BECH, 1995)

After identifying the applicability of this work, frequency response characteristics were then taken into consideration (Bech, 1996). The experimental set-up remained the same however, loudspeaker simulating reflections were then altered by added frequency dependant characteristics. *Six* filters were applied as a function of frequency dependant absorption coefficients at octave-bands 125Hz – 8KHz to selected loudspeakers simulating individual reflections. The method and stimuli used was the same as detailed in the previous experiment.

The results from this experiment build on those from the first report, that only the floor reflection had a TD lower than that of a natural room for a *noise signal.* TD was not ascertained for reflection '*3*' (ceiling) but is stated that this is also likely to be lower than that of a natural reflection. A TD test, with and without the transfer function filters for the loudspeaker revealed that for a noise signal, detection values for individual reflections *five* (floor), *seven* (ceiling) and *nine* (left wall) increased significantly in dB with the filter *on*. Bech highlights that this is due to the removal of energy in the frequency region 500Hz – 2kHz. Whilst these reports conclude that the first order floor reflection (and possibly ceiling reflection) is most likely to individually contribute to timbre, of great importance to this study is that Bech notes these are only threshold detection tests and not a prediction in terms of timbral *quality*.

The impact of vertical reflections from above have also been the focus of a study regarding *auditory envelopment* (Furuya, Fujimoto, Takeshima & Nakamua, 1995). In the context of concert hall acoustics, three experiments were conducted to subjectively assess the contributions of:

- A single reflection from above along the median plane on "auditory size of sound image"
- Energy of multiple early reflections from above and "auditory impression of envelopment"
- A repetition of experiment two in the presence of a reverberant sound field.

All experiments utilised musical stimuli and the reflection and sound field were electro acoustically simulated. In relation to this thesis investigating a singular reflection, experiment one demonstrated that as the delay time of a singular reflection increased, the sound *image* grows vertically in size. Regarding auditory envelopment, experiments two and three show that as long as the ratio of lateral and vertical energy remains constant up to 200ms, envelopment becomes stronger as energy arriving from above increases. The author does note however that lateral arriving energy must be the "*predominant factor to perceive envelopment*".

2.5.2 Vertical Localisation

Section 2.4 discussed auditory cues needed for localisation mainly on the lateral plane. The following chapter reviews literature studying localisation along the vertical plane. Outlined in

Section 1.1, this thesis investigates the subjective preference of an early vertical reflection through timbral *and* spatial differences. Therefore, the understanding of spectral and spatial cues used in localisation along the median plane will be useful in discussing results. Although much of the reviewed research is in the context of audio reproduction, the process of localising an elevated source along the median plane remains the same with a reflection.

Early research into localisation with source elevation was conducted by Pratt (1930). Using tonal stimuli he concluded that subjects could *not* locate the incident sound along the vertical plane. However, they *did* observe that when a signal is presented '*diotically*'[4], the auditory event was systematically perceived to be physically low, for low tones, and higher for high tones. This experiment was repeated by Roffler and Butler, (1968a) who observed the same phenomenon. Subjects were asked to localise the sound source and were unaware of loudspeaker quantity or placement. Program material used was varying tonal and filtered noise signals. The results of the experiment confirmed those by Pratt (1930) that the ability to localize tonal stimuli and broadband noise is poor from frequency content below 7kHz. Also the perceived auditory event was located higher along the vertical plane with respect to higher frequency and that for accurate localisation along this axis, the signal must be complex. Roffler and Butler (1968b) investigated this further whilst subjects lay in different orientations and distances from a loudspeaker array and also using blind people and young children. Using frequency bursts ranging from 250Hz – 7200Hz the subjects were asked as before to localise the sound source, but even in different positions localisation was still poor.

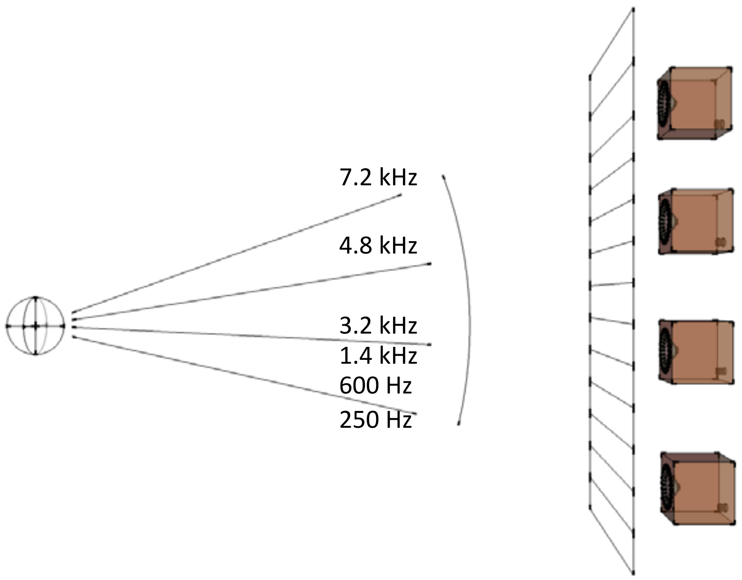


FIGURE 13: ROLFER AND BUTLER EXPERIMENT OF PITCH HEIGHT EFFECT

[4] A Monaural signal – To be presented at both ears from a single source along the median plane

A great amount of work in this area has been conducted by Bleuart (1997), who has carried out a number of investigations with regards to localisation along the medium plane. The test resulted in clues about the angle of incidence with regards to the frequency content of the signal. Broken into three sections *h, o* and *v* (*behind*, *overhead*, *forwards* respectively), 1/3rd octave noise stimuli was presented once from the direct speaker and rear speaker alternately and subjects were asked to localise the auditory event based on the three locations presented, the results of which can be seen in Figure 14 showing the relative probability of subjects answering "*behind*", "*overhead*" or "*forward*", with respect to frequency. The results show that the probability of someone perceiving an auditory event from above with frequencies roughly between 7 – 10kHz is great even if the sound source is from the front or rear. Most importantly, Blauert has demonstrated that the localisation of an auditory event may be influenced *independent* of direction and more through frequency content. Regarding this thesis, this may impact subject's preference when strong frequency content is perceived to be located at a particular angle when emphasised with a reflection.

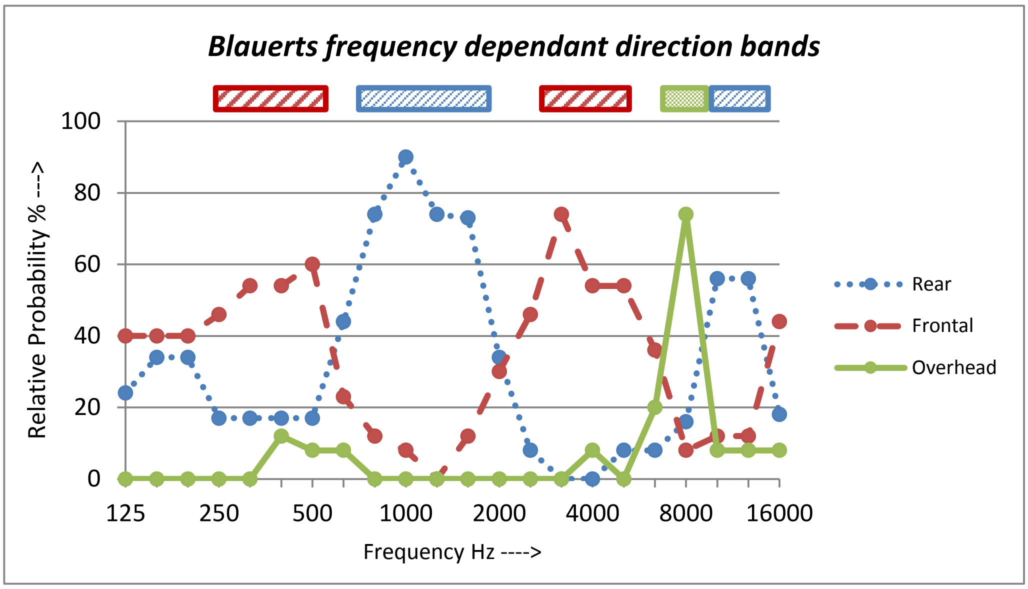


FIGURE 14: RESULTS OF BLAUERTS' DIRECTIONAL BANDS EXPERIMENT, ADAPTED FROM BLAUERT (1997 P.109 FIG 2.46)

In the context of multichannel reproduction, experiments by Lee, (2011) investigated both the masking and localisation thresholds of a sound, using a vertically placed loudspeaker with a delayed signal. The masking threshold being - the level (dB, *not* height) at which the vertical loudspeaker had no audible effects. The localisation threshold - the level the vertical speaker needed to be so the sound source is localised only from the primary speaker. The ICTD's (Inter-channel Time Difference's) used were 0, 0.25, 0.4, 1.0, 2.5, 5.0, 10, 25, 50*ms* and presented at an average 75dB(A) to the subject. Results of the localisation threshold test revealed for time delays greater than 1.1*ms* and below the echo threshold (roughly 5ms), the precedence did *not* operate. All assessments of delay times *needed* to be attenuated for the source to be *fully* located at the direct speaker and could not be localised on just ICTD.

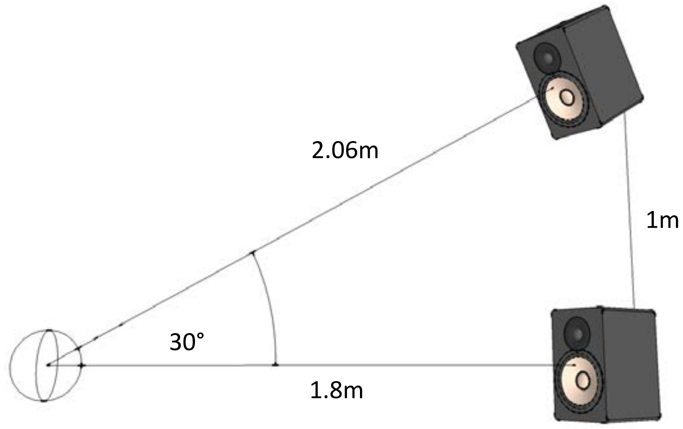


FIGURE 15: LOCALISATION EXPERIMENTAL SETUP IN LEE (2011)

The vertical channel's level attenuation up to 5ms, was consistently around -9/-10dB for localisation threshold, and -6/-7 dB attenuation for the masking threshold. With regards to timbre in this investigation, Lee points out that from informal discussions with the subject post-test, it was clear that the most prominent factors were indeed tonal colourations *and* localizability. This further supports work discussed, regarding the effect of elevated sources and reflection on the median plane altering our perception of timbre. However still no research has identified is this is a positive or negative effect on the listener.

In contrast to Lee's results (2011), Hartmann (1993) previously suggested that the precedence effect *does* operate within all planes. A simple experiment using click stimuli with delays of 0, 0.1, 0.2, 0.5 1.0, 2.0, 5.0 and 10*ms* along the same plane, with three loudspeakers to represent each plane (*frontal*, *lateral* and *median*) was used to investigate if the precedence effect was a higher order cognitive process, not seldom based upon frequency. His results concluded that the ability to localise along the sagittal plane was still achieved through the precedence effect *without* the need for Interaural differences. In the context of this paper, this would indicate reflections arriving from a vertical source may still be localised through the precedence effect and independently, frequency content may be manipulated to achieve beneficial or detrimental timbral effects.

2.6 PERCEPTUAL ASSESSMENT

Previous chapters have discussed some of the literature relating to the both the physical domain, and then the perceptual domain regarding both lateral and vertical early reflections. Whilst these papers discuss the timbral and spatial effects of a rooms early reflections, a way of identifying *what* these changes are, and how we describe them in more detail needs to be explored through the use of specific descriptions. It is also important to consider the aspects of the methodology that enables subjects to record the magnitude of a sensation to a physical value such as rating scales.

2.6.1 ATTRIBUTE SCALING

In order to quantify a user response to a given question, a format must be chosen that can be interpreted equally throughout subjects with no bias. Commonly in audio assessment, ratings scales are employed to retrieve a subject's response to a perceptual attribute. The use of '*direct*' scaling procedures lends a simple way to convert the magnitude of a sensation to a corresponding scale whilst '*indirect*' judges the degree in which a sensation is different in one stimulus, compared to another (Bech & Zacharov, 2006). The most common of which are Difference threshold (DL) and Paired comparison methods (PC). Difference threshold tests, as discussed throughout papers in previous chapters (Olive & Toole, 1988) (Lee, 2011), identify values of Just Noticeable Difference by increasing or decreasing program material with a single controllable parameter in specific increments (*e.g 1dB*). Paired comparison testing (Imamura et al., 2013) is the assessment of two stimuli or playback systems against one another, whereby the user rates an individual attribute such as fidelity, or preference. A number of scales have been developed to asses sensory attributes and are discussed below.

A series of papers investigating the problematic trade-off between bandwidth limitation and down-mixing algorithms in delivery systems show insightful methodologies on subjective testing. Psychoacoustic testing of subjects provided information on which attributes may be less/more desirable to retain with limitations transmission conditions (Zeilingski, et al, 2003). In Zeilinski et al's (2005) paper, subjects were asked to grade three attributes: timbre, frontal and surround spatial fidelity. This method of identifying *fidelity* requires users to rate the 'trueness' of which a stimulus is replicated in comparison to the original. This can be likened to that of an *impairment scale* (Table 2) and as mentioned "*quality of processed items used was degraded considerably*". Therefore, the use of a double-blind multi-stimulus test method with a hidden reference and anchors (MUSHRA)(ITU-R BS.1534-3:2015) was used as a response format. This allowed rapid comparison of multiple program material. All stimuli were equalised in order to eliminate any bias due to loudness change and presented in a randomised order to reduce carryover effect. Whilst such a methodology could be applied to the context of this research into vertical reflections, as shown in the literature, reflections may be beneficial thus a degradation scale may not be suitable.

TABLE 2: ITU – R 5-POINT CONTINUOUS IMPAIRMENT SCALE

80-100	Excellent
60-80	Good
40-60	Slightly Better
20-40	About the Same
0-20	Slightly Worse

Zacharov & Lorho's (2004) investigation into home theatre systems and multichannel algorithms is of interest due to the experimental design and response scale. One of two experiments consisted of a loudspeaker test whereby six algorithms, chosen as the dependant variables, are assessed in terms of '*reproduction quality*'. A Comparison Category Rating (CCR – Table 3 & 4) is chosen with a paired comparison (also referred to as an A/B comparison) methodology. ITU-R P.800:1998 has an extensive overview of all *Absolute*, *Degradation* and *Category* Comparison Ratings. However, the CCR method employed allows the user to compare unprocessed stimuli against processed stimuli whereby the order of processed and unprocessed for each pair is randomised.

The CCR method unlike *degradation* comparison rating (DCR), also allows ratings of *improvement*. This may be used in conjunction with modified MUSRHA style testing (Fenton, Bruno & Wakefield, 2009) to consider the possibility that the assessment stimuli may exceed the reference in criteria such as audio quality. In this investigation (Zacharov & Lorho, 2004), 14 subjects for the loudspeaker experiment were instructed to grade their preference in terms of '*overall quality'* considering both spatial and timbral characteristics. The program material selected used excerpts of "*Music*, *Movie Sound* and *Gaming sound*". These were selected for specific timbre and spatial cues and averaged at 78 dB(A) SPL across each program material for loudspeaker reproduction. Before subjective testing, administrative familiarisation took place using the '*GuineaPig 2 listening system*'.

TABLE 3 ITU P.800 CCR GRADING SCALE

3	Much Better
2	Better
1	Slightly Better
0	About the Same
-1	Slightly Worse
-2	Worse
-3	Much Worse

TABLE 4: MODIFIED 9-POINT HEDONIC SCALE (ZACHAROV & LORHO, 2004)

4	Prefer B extremely
3	Prefer B very much
2	Prefer B moderately
1	Prefer B slightly
0	Neither Prefer A or B
-1	Prefer A slightly
-2	Prefer A moderately
-3	Prefer A very much
-4	Prefer A extremely

As mentioned by Bech & Zacharov (2006), subject familiarisation is key to ensuring understanding of variables under examination and that consistency of instructions both verbal and written should be maintained across all subjects. Bias was also eliminated by adding extra 0.5 “run-offs” at opposing ends of the rating scale (extending the sable range to -4.5 to +4.5) to eliminate subject reservation in using extreme end points. This methodology allows comparison of each individual stimulus directly against each other, rather than a group of stimuli against a single reference as with a MUSHRA test.

A way of rapid comparison of a large number of samples is that of the rank order method or, *round robin*. Discussed in Zacharov & Huopaniemi's (1999) paper, the aim was to be able to quickly compare numerous VHT (Virtual Home Theatre) systems using a large number of samples assessing sound quality split into timbral and spatial attributes. This rank order method is advantageous due to its simplicity in acquiring data and involves little preliminary subject training, requiring only the rating of program material from 1 to N (N being number of samples) based on a criteria and direction of ranking specified. The major disadvantage of this method that it provides no scaling information between comparisons (Otto, 1997) and therefore, are only used when an *indication* of how sounds compare is needed.

Adaptations of this rank order method have also been investigated and compared to paired comparison with respect to speed and accuracy of results by way of ‘ranking by elimination’ (Wickelmaier, Umbach, Sering, & Choisel, 2009). Although this may yield similar results to that of the rank order method, the scaling *between* ranks is still an issue that only paired comparison reveals. Whilst all scales have been established and used in research for some time, recent work possibly suggests that some of these scales could still be misrepresentative of how our psychoacoustic assessment (both sensory and cognitive filters – Figure 1) is processed. Even though that scales may be presented in a linear fashion by anchoring labels at an equidistant point along an axis, this does not mean that the interpretation of these labels are not perceptually linear (S. Zielinski, Brooks, & Rumsey, 2007).

2.6.2 Attribute Descriptions

When an attribute change such as quality, degradation or fidelity is the result of a perceived characteristic and we want to know *why*, the attribute response becomes far more complex as the number of individual adjectives needed just to describe timbre is extensive. Most of the literature discussed throughout has assessed audio quality through general timbral change. However, rather than a global quality, it can be useful to identify certain attributes within timbre that may separately impact preference.

For the assessment of sound quality Gabrielsson & Sjögren (1979) used over 50 adjectives to describe playback systems. These were a result of questionnaires given to 170 people, consisting of roughly 200 descriptors and were used in scales varying from 0-9 to indicate that particular attributes '*quality*'. The assessment of loudspeakers conducted by Staffeldt, (1974) utilized 35 descriptors and were recorded in a binary format from a paired comparison test, '1' – indicating system ***'i'*** possess this characteristic, and '0' indicating system ***'j'*** possess this characteristic. The process of attribute selection can be done through a combination of personal experience, interviews, literature research, elicitation etc. However, an elicitation is not within the scope of this research therefore, the following sections discuss literature assessing timbral definitions to use during experiments.

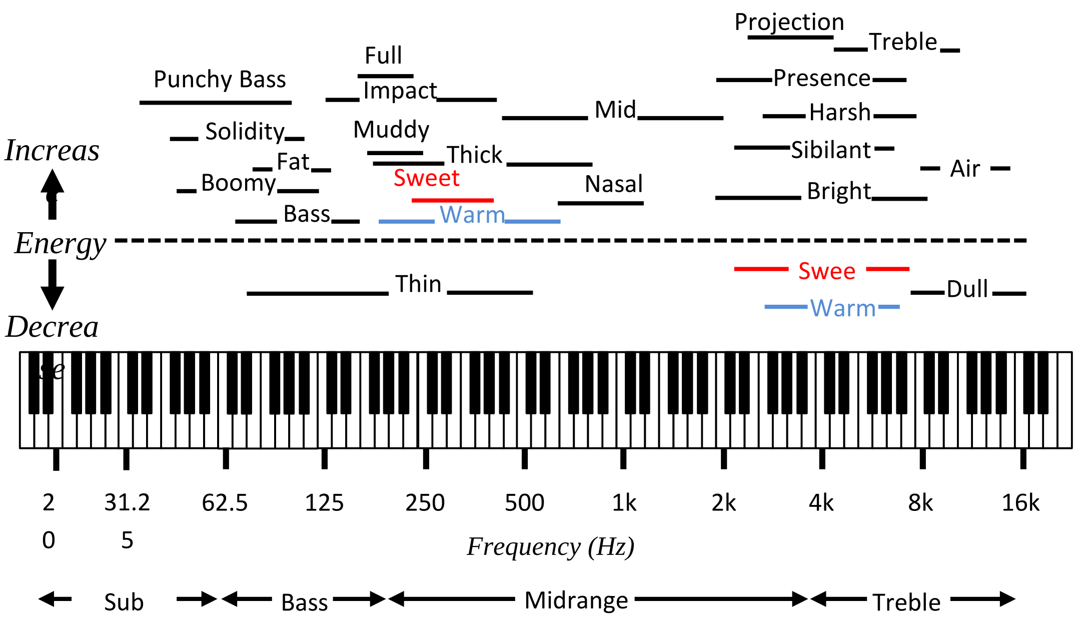


FIGURE 16: FREQUENCY RESPONSE SHOWN AGAINST A KEYBOARD 20HZ – 20KHZ AND RESPECTIVE TIMBRAL DESCRIPTORS - NOTE THAT ONE ATTRIBUTE MAY BE ACHIEVED ("WARM" OR "SWEET") BY REDUCTION OR AMPLIFICATION AT DIFFERENT FREQUENCIES. ADAPTED FROM HOWARD & ANGUS (2013)

Words such as *bass*, *mid* and *treble* are often seen amongst audio engineers and musicians but are somewhat broad-stroke in terms of frequency range. Figure 16, adapted by Howard & Angus (2013) from Katz (2007), demonstrates the relationship between frequency content and timbral descriptors. It is interesting to see that the perception of a physical change may be induced in multiple ways (‘*warm*’, ‘*sweet*’) by exciting different frequencies. Reflecting on this scale of pitch and descriptors how is it, that two instruments demonstrating the same pitch and loudness exert completely a different timbre? The classic, much-quoted definition of timbre by the American National Standards Institute (ANSI S1.1-1960)[5] is as follows:

“…*that attribute of sensation in terms of which a listener can judge that two sounds having the* ***same loudness*** *and* ***pitch*** *are* ***dissimilar.***”

This however, implies that sounds must possess a pitch for the definition to apply (Bregman, 1994), and that sounds which do not contain *pitch* such as “*scraping a shovel in a pile of gravel*” cannot contain timbre. Bregman describes timbre as an “*ill-defined wastebasket category*” and that the only reason loudness and pitch are accounted for is that they are easy to manipulate on a musical instrument. Regardless of its definition, it is clear that timbre in some way incorporates the spectral content of a signal. As it cannot be scaled on a singular axis such as low-high or quiet-loud, it should therefore be recognised as a multi-dimensional attribute.

Erickson (1975) has elicited a list of some subjective parameters of timbre and their counterpoints within the physical domain (Table 5) with regards to “*music-orientated*” sounds. These are based upon five dimensions from Schouten (1968) which he describes as an excellent classification of perceptual analysis. These objective features take a step further than just frequency content into identifying timbre and include dynamics and ‘*musicality*’ of the signal. One technique employed to analyse the timbre of a sound is a *spectrogram,* whereby the whole envelope, frequency content, duration and steady-state[6] changes are captured. This allows identification of not only the harmonic content and amplitude (as with an FFT) but the time at which these harmonics occur and their duration. The onset phase of a musical note is particularly important at perceiving timbre, as colouration of the direct sound from early

[5] It is recognised that this document is now superseded by revision ANSI S1.1-2013, however subjective testing assessing timbre (Bech - Section 2.5) are all consistent with the definition in ANSI S1.1-1960.
[6] Steady-state does not mean that no changes are present but the sustain period to which a note is held.

reflections may only impact the perceived timbre *after* the note-on has occurred (Howard & Angus, 2013).

SUBJECTIVE	*OBJECTIVE*
Tonal Character, usually pitched	Periodic Sound
Noisy, with or without some tonal character	Noise, including random pulses
Colouration	Spectral Envelope
Beginning/Ending	Physical rise and decay time
Colouration glide or format glide	Change of spectral envelope
Micro-intonation	Small change (one up and down) in frequency
Vibrato	Frequency modulation
Tremolo	Amplitude modulation
Attack	Prefix
Final sound	Suffix

TABLE 3: LIST OF SUBJECTIVE ASPECT AND THEIR PHYSICAL COUNTERPOINTS REGARDING MUSICAL INSTRUMENTATION TIMBRE

Assessments of audio quality are often broken up into two main sub-categories comprising of spatial attributes (Section 2.4), and timbral attributes. However with so many descriptors available for timbral characteristics, subjective testing can prove to be a difficult task. Previous studies have assessed a subject's perception of timbral quality as a whole (S. K. Zielinski, Rumsey, Kassier, & Bech, 2005), whilst other research delves further identifying *which* descriptive terms can best describe this perceived change (Torben Holm Pedersen, 2008). Experiments conducted with the use of individual vocabulary profiling (IVP) have also been conducted (Kuusinen et al., 2014). This gives subjects the freedom to develop their own set of descriptions in the assessment of stimuli.

With perceptual evaluation and sound quality assessment becoming more popular in recent years, there has been much need to try to consolidate verbal descriptors from across the literature in an attempt to provide uniformity throughout research and as such, is the focus of Pedersen & Zacharov's (2015) paper. In the context of *reproduced sound*, this study takes a step closer in delivering a universal list of descriptors for general usage. Taken across English, German and Nordic material including scientific literature, papers, product descriptions and hi-fi magazines were 200 words. These were selected after the removal of repeated words and words relating to preference and subjective liking, the end result was a "*sound wheel*" (Figure 17) This also included assessing loudspeaker systems (recorded with a Bruel and Kjaer head and torso simulator) through headphones and 4 mono loudspeakers across a broad price range, describing timbral/spatial attributes and differences that were *most* prominent.

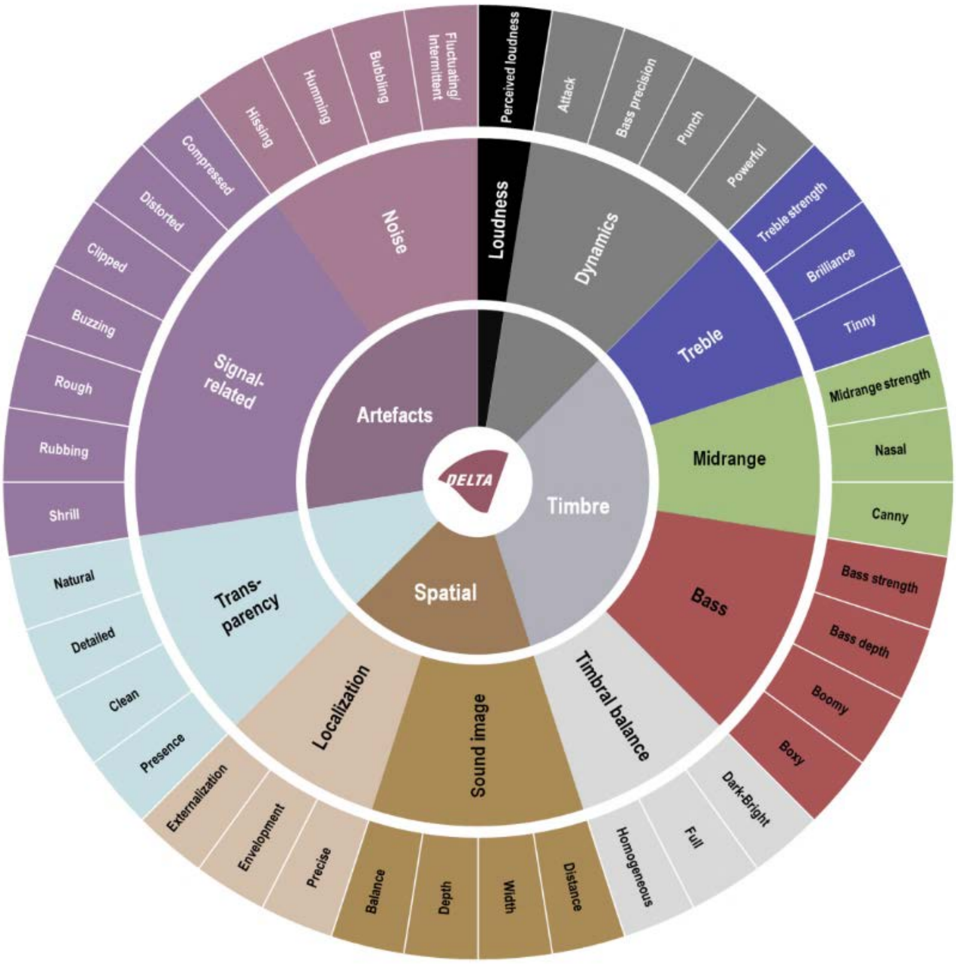


FIGURE 17: SOUND WHEEL AFTER PEDERSEN & ZACHAROV (2015). INNER RING IS MAIN PERCEPTUAL ATTRIBUTES. MIDDLE RING CATEGORISES AND OUTER RING PROVIDES ADJECTIVES.

Whilst there are many descriptors and even debates regarding the definition of timbre, most of the evaluative terms seem to be in assessment of a sound source and not in describing the timbral difference a reflection provides. Where first reflections *have* been assessed on the influence on timbre (Søren Bech, 1994b) (Bech, 1995b) (Bech, 1996), definitions of timbre have been that of the American Standards Institute. Therefore, the description of timbre given to subjects will also follow the majority of research in using ANSI definition.

2.7 SUMMARY

Through the review of this literature, a number of research gaps have been identified. A large proportion of studies discussed have demonstrated that early reflections affect our timbral and spatial perception of a sound. However, little has been investigated regarding their effect on listener's enjoyment (Kishinaga et al. 1979). Particularly for early vertical reflections (and elevated sound sources along the median plane), research considering their impact on listener preference is even less apparent, even though a number of studies have discussed their ability to alter the perceived timbre of the direct/incident sound (Clark 1983)(Bech 1995)(Bech 1996)(Lee 2011). Much of the literature discussed has also employed minimal musical stimuli and where musical stimuli is assessed, it is not in the context of listeners' preference. Therefore, the need to investigate the perceptual effect of these timbral (and spatial) changes on preference provided by a vertical reflection using musical stimuli, will provide useful information in our understanding of sound within small rooms.

CHAPTER 3

REFLECTION MODELLING

3.0 INTRODUCTION

Literature discussed throughout Chapter 2 has shown that room reflections have a great impact upon the way we perceive sound. However, there is still a clear gap within research needing investigation to determine if these early reflections have a positive or negative timbral or spatial effect when listening to audio for entertainment. In order to assess the effect of a singular reflection on our perception in comparison to playback with no reflections, material with the desired properties could be placed at the point of reflection. However, in order for quick assessment of different properties of this single reflection, changing material would be unfeasible potentially introducing bias. The reliability of a subject's acoustic memory in any time gap also reduces any accuracy of any comparison (Pike, Mason, & Brookes, 2014). Therefore, a reflection was electro-acoustically simulated by a loudspeaker. This chapter demonstrates the setup of a secondary loudspeaker for correct simulation of a reflection along with validation of the processing used.

3.1 INITIAL MEASUREMENTS

For the purpose of electroacoustic simulation of vertical reflections a number of preliminary tests were needed to collect data and simulate the required set-up. The testing took place in the University of Huddersfield's semi-anechoic chamber. A brief experiment was conducted to clarify the audible timbral effect discussed throughout Chapter 2, whereby a 16mm plywood panel with reflective veneered surface was installed at the first calculated geometrical reflection point acting as a low hanging reflector. Sitting at the listening position, the author could clearly hear an audible difference in timbre along with certain spatial attributes for musical stimuli.

For comparison, impulse responses were taken with the above measurements with accumulating acoustic treatment along the floor and ceiling panel trussing. Once floor absorption had been incorporated, no interfering reflections were observed above -18dB to that of the direct sound. (An impulse response of the set-up dimensions below can be seen in Figure 21).

- *Listening height – 1.15m,*

- *Listening distance – 2m,*
- *Ceiling height – 1.95m*
- *Loudspeaker – Genelec 8040a*
- *RoomEQ Wizard Measurement Software*
- *Interface – Focusrite Safire Pro 14*
- *Dbx measurement Microphone*

3.2 LOUDSPEAKER DISPERSION

To prevent confusion, the following terms will be given to loudspeakers used. The loudspeaker used to play the direct sound will be referred to as the '*direct loudspeaker'*. The loudspeaker used to replicate the reflection from above, will be referred to as the '*reflection loudspeaker*'. These can be seen in Figure 18.

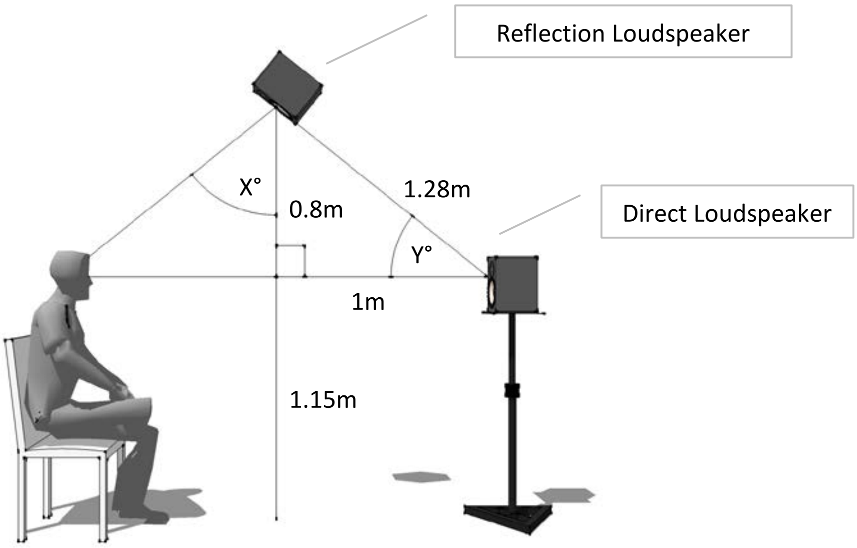


FIGURE 18: REFLECTION SIMULATION SET-UP

To make the ceiling reflection realistic to replicate, it must also be representative of what is being projected by the direct loudspeaker (Bech, 1990). If a reflection is simply replaced by a loudspeaker pointing on axis to the receiver, the spectral content will not be accurate due to the direct loudspeakers' frequency dispersion. Published with the Genelec 8040a, are the frequency responses at horizontal angles 15°, 30°, 45° and 60° but no vertical information. Therefore, measurements at a 1m radius at 15°, 30°, 45° and 60° were taken of the frequency response vertically (Appendix B). Finally, the angle of projection for the dimensions above was calculated and measured (Figure 19).

$$Y = cos^{-1}\left(\frac{1}{1.28}\right) = 38.62° \quad [5]$$

$$X = 90° - 38.62° = 51.38° \quad [6]$$

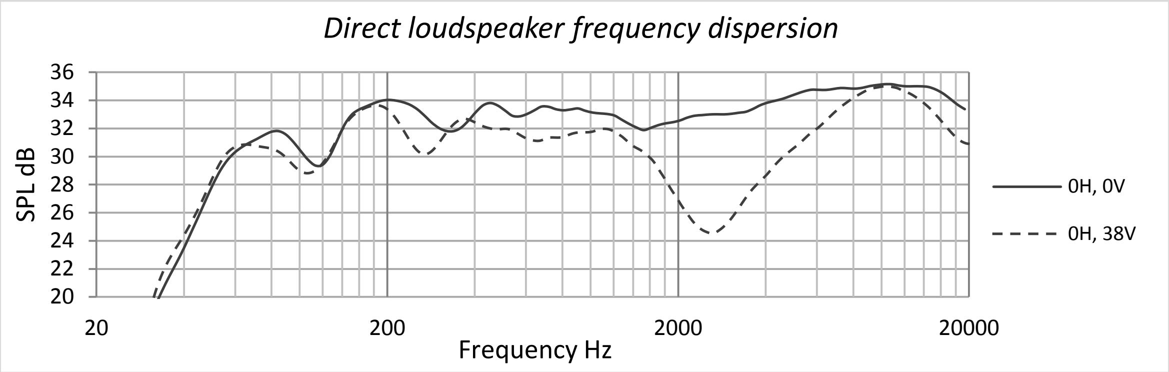


FIGURE 19: FREQUENCY DISPERSION OF GENELEC (0°H 0°V - SOLID, 0°H 38°V – DASHED)

119 values of equal logarithmic distance were exported representing the difference between the loudspeakers on-axis response and calculated angle across 20Hz – 20kHz. This difference was then applied to cascade filters within MAX MSP to ensure frequencies delivered at the correct amplitude by the direct loudspeaker, are replicated by the reflection loudspeaker. This can be seen as a Delta Spectrum in Figure 20. The dip seen between 2-3kHz is most likely due to the Genelec 8040a crossover point around 3kHz.

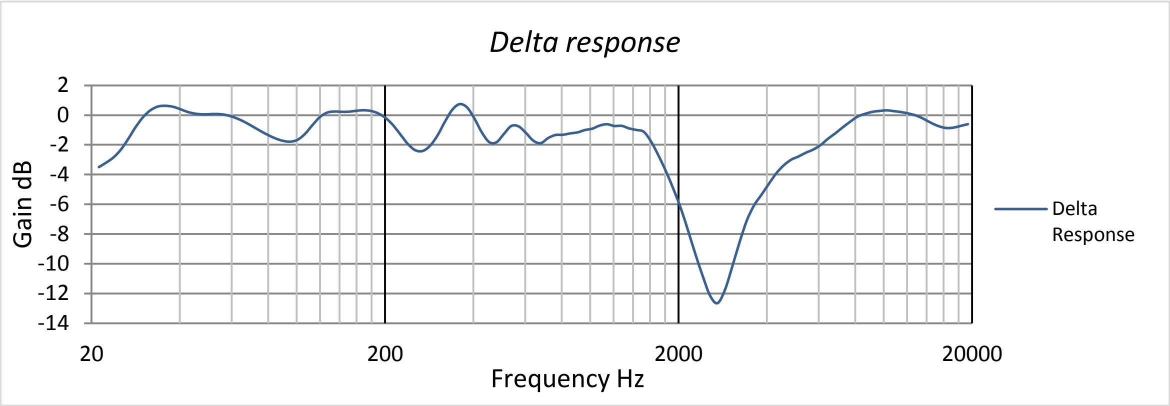


FIGURE 20: DELTA SPECTRUM OF FREQUENCY DIFFERENCE APPLIED TO SIMULATING REFLECTION CHANNEL

3.3 CALCULATION ACCURACY

Finally calculations were used to ascertain the required delay and global attenuation[7]. Using dimensions specified in Section 3.1, the calculations result in a time interval of **1.63*ms*** (Equation 9) and level difference of **2.2dB** (Equation 12) between the direct and reflected sound as calculated below.

Direct sound delay (loudspeaker to listener)

$$\frac{Direct_{Distance}}{344} = 5.81ms \qquad [7]$$

Where '$Direct_{Distance}$' = 2m

Ceiling reflection delay (loudspeaker – reflection point – listener)

$$\frac{Ceil_{Distance}}{344} = 7.44ms \qquad [8]$$

Where '$Ceil_{Distance}$' = 1.28m × 2 (see Figure 18)

Therefore the resulting time interval between the direct sound delay [7] and ceiling reflection delay [8] is:

$$7.44ms - 5.81ms = \mathbf{1.63ms} \qquad [9]$$

Direct sound attenuation

$$(20 * \log(Direct_{Distance})) = 6.0dB \qquad [10]$$

Where '$Direct_{Distance}$' = 2m

Ceiling reflection attenuation

$$(20 * \log(Ceil_{Distance})) + (10 * \log(1 - \alpha)) = 8.2dB \qquad [11]$$

Where; '$Ceil_{Distance}$' = 1.28m × 2 and a boundary absorption level 'α' = 0.01. This absorption level was chosen to represent a highly reflective surface with the least amount of absorption possible.

Therefore the resulting level difference of attenuation between the direct sound [10] and the reflected sound with minimal absorption at point of reflection [11] is:

[7] Associated with the amount of attenuation produced by inverse square law, without frequency dependant alterations made by reflecting surface properties or loudspeaker dispersion.

$$8.2dB - 6.0dB = \mathbf{2.2dB} \qquad [12]$$

These calculated results were then compared against the genuine measured reflection (Figure 21), and show a small difference of **0.21***ms* delay and **0.37**dB attenuation difference. It is recognised that whilst these differences *are* present between the calculated results and the genuine reflection, changes in room temperature and absorption may possibly account for the discrepancy. As these factors may always fluctuate, the author deems it acceptable to simulate the vertical reflection using the calculations above.

As the reflection loudspeaker is placed halfway along the reflection path, values of **3.72***ms* delay (7.44ms from Equation [8] ÷ 2) and **4.1**dB attenuation (8.2dB from Equation [11] ÷ 2) were applied, with the inclusion of the dispersion filter (Figure 20) for the simulated reflection. The results from the impulse response with an electro-acoustically simulated loudspeaker using these values of attenuation and delay, show an accurate simulation with a delay interval = **1.48***ms* and level difference = **2.52**dB. A comparison of the genuine reflection against the simulated reflection using the calculated values can be seen in Figures 21 and 22, along with related frequency response measurements in Figure 23.

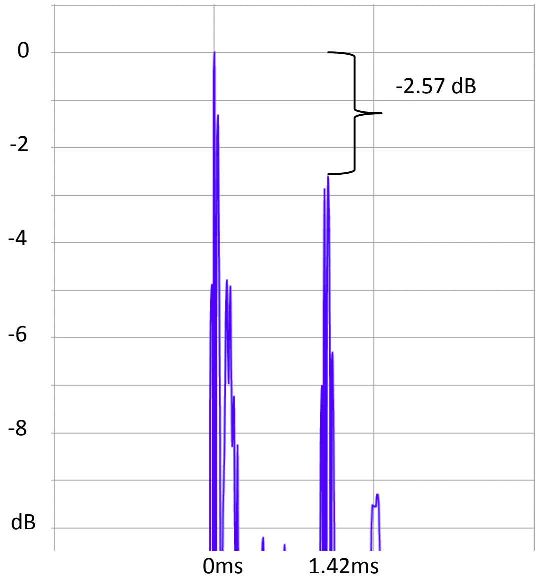


FIGURE 21: IMPULSE RESPONSE TAKEN USING GENUINE REFLECTION

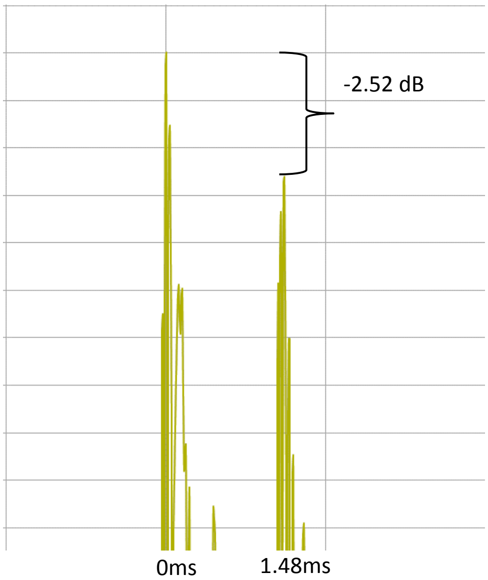


FIGURE 22: IMPULSE RESPONSE TAKEN USING SIMULATED REFLECTION WITH VALUES OF - **4.1**DB ATTENUATION AND **3.72***MS* DELAY

It is also noted that the differences between the genuine reflection (Figure 21) and simulated reflection (Figure 22), are *smaller* than those between genuine reflection and what was calculated in Equations 9 and 12. Therefore it is reasonable to assume that subtle changes within a rooms' environment may be the cause of small differences however, the calculations can still be used to replicate an accurate reflection.

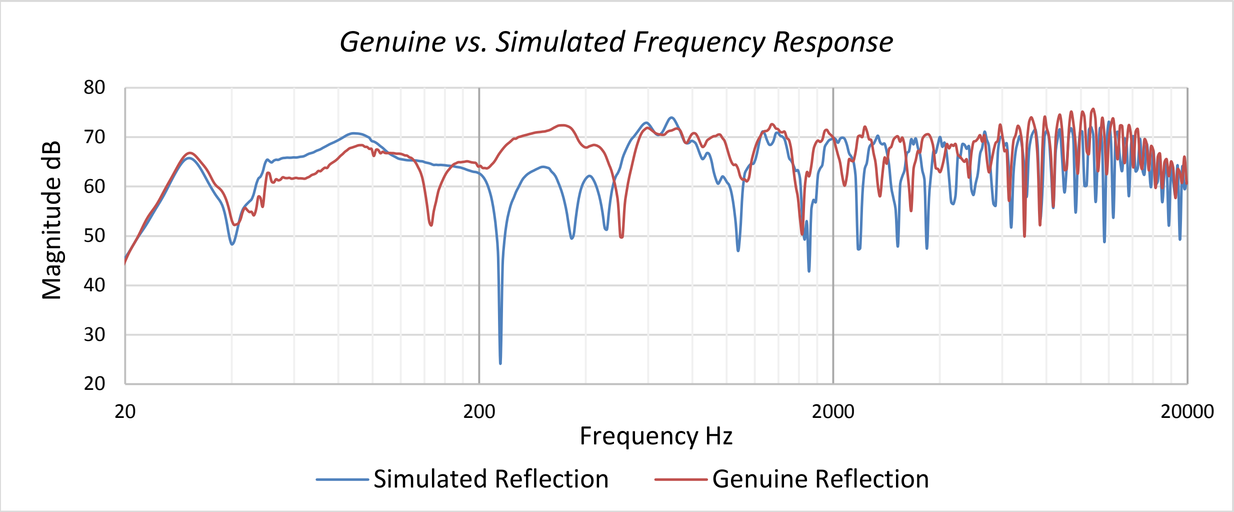


FIGURE 23: FREQUENCY RESPONSE OF DIRECT LOUDSPEAKER WITH A) GENUINE REFLECTION VS B) SIMULATED REFLECTION

This chapter has shown that following these equations and calibration process, an effective ceiling reflection can be modelled using a loudspeaker. This will allow quick comparison of playback with and without a reflection, along with further manipulation of the ceiling reflection's frequency characteristics.

CHAPTER 4

EXPERIMENT ONE

4.0 INTRODUCTION

With little literature in the field concerning the preference of a vertical reflection, the aim of this first experiment was to provide new data regarding the magnitude of perceived timbral and spatial differences and a listeners' preference of a singular ceiling reflection. As reviewed in the literature throughout Chapter 2, the importance of first geometrical reflection points have been of particular interest in studio/control rooms, concert halls and more recently small rooms. It has also been demonstrated that these floor and ceiling reflections add timbral and also spatial alterations to our sensory process of the direct sound. Therefore, this study will focus on the first geometrical ceiling reflection. Floor reflections will not be covered in this thesis but will be a topic of discussion later (Chapter 7).

4.1 METHODOLOGY

Subjective testing was employed to study the level of preference of a listener between the reproduction of sound with or without a reflection as well as investigating any correlation between this preference and the magnitude of perceived timbral and spatial differences. The motivation behind this experiment was to ascertain if this increased or decreased magnitude of perceived timbral or spatial change could indicate a beneficial or detrimental contribution to our *enjoyment* of audio and consequently, if the reflection it actually *needs* to be removed or not. A vertical reflection was electro-acoustically simulated following the procedure shown in Chapter 3 within the University of Huddersfield's semi-anechoic chamber (Figure 24). This was to ensure that subjects would be assessing a single vertical reflection without the presence of any other early reflections or reverberation. The reflection replicated was based on the following dimensions:

Listener distance – 2*m*, Listener height – 1.15*m*, Reflection Point – 1.95*m*

A listener distance of 2m was chosen as minimum outlined in ITU-R BS.1116:1997, however with certain height limitations of the semi-anechoic chamber a slightly shorter value of 1.15*m* for listener height was chosen rather than the recommended 1.2*m*. The value of a 1.95*m* vertical boundary was chosen as an extreme case. The author recognises that ceilings are seldom this low, or that reflectors/absorbers are very rarely hung from the ceiling at such a low height. Due to available resources of an acoustically dead environment, this was a physical limitation of the semi-anechoic chamber's size. However, for this first experiment an exaggerated reflection would be useful to determine initial results and provide a comparison for further studies. This reflection path of 2.56*m* (source to receiver) would result in a delay of **1.63***ms* and attenuation of **2.2**dB below the direct sound which is similar to that of the frequency independent floor reflection of that reviewed in the literature (Bech, 1996)(Section 2.7.1), but with differing angle of incidence and frequency content.

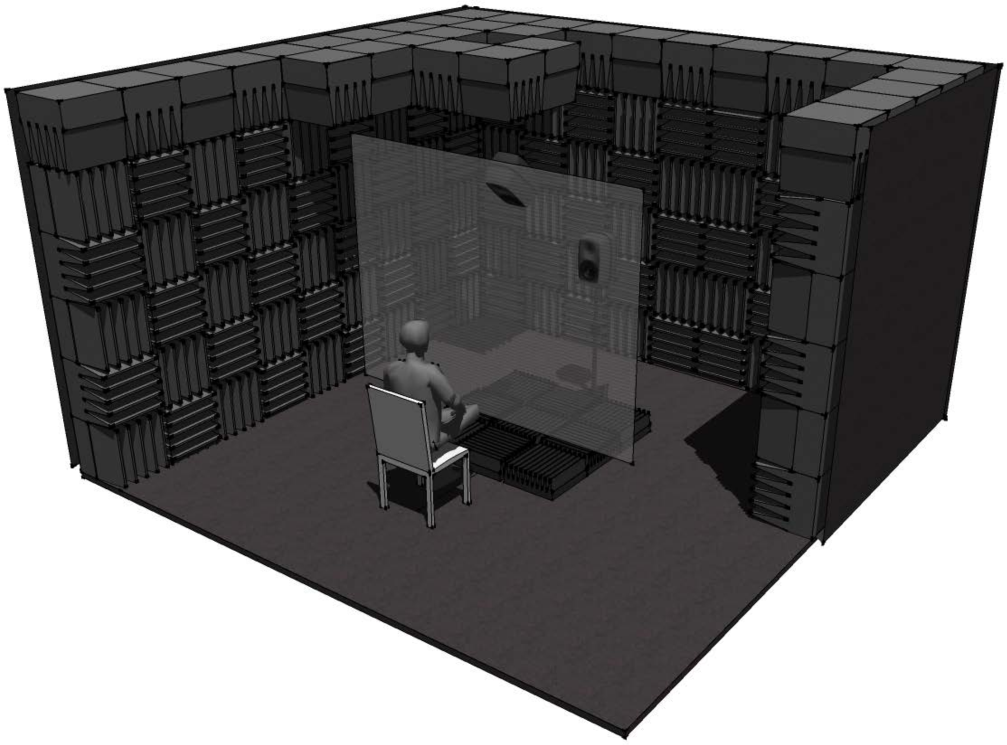


FIGURE 24: UNIVERSITY OF HUDDERSFIELD SEMI ANECHOIC CHAMBER – SET-UP OF EXPERIMENT ONE

A paired comparison method of assessment was chosen (Section 2.6) whereby subjects had the ability to switch between two playback methods of a stimuli, as many times they wished. Broken into two sections, the first assessment was to simply ascertain subject's preference between stimuli. The second section focused on the magnitude of perceived timbral and spatial difference between stimuli. The labelling of stimuli was presented as '*Reference*' and '*Processed*'. Reference triggered playback of program material through just the direct loudspeaker, and processed through both the direct loudspeaker and its respective simulated reflection through the reflection loudspeaker. The author understands the use of such labelling may possibly bias subjects' response (as a processed sample might imply an improvement).

However, participants were unaware as to the nature of the experiment and all equipment was obscured via an acoustically transparent curtain to eliminate any visual bias (Toole & Olive, 1994). In addition to this, subjects were told that playback of the 'processed' sample *could* result in better, worse or even *no* change in reproduction. A tablet interface was designed and used throughout all testing to minimise any possible reflections that would otherwise be present when using laptop on a table. The ergonomic design of the tablet software[8] (Figure 25) was such that subjects required minimal head movement whilst assessing audio, with stimuli triggers located at thumb locations. All data was sent/received wirelessly via a MAX/MSP patch where all processing and stimuli control took place.

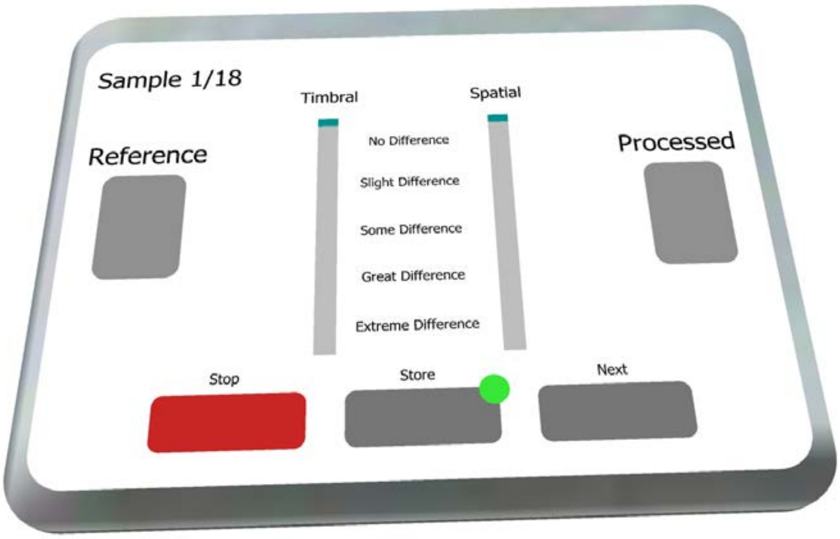


FIGURE 25: TABLET INTERFACE DESIGN

Main sample triggers are placed in thumb locations for ease of use without subjects having to move their head.

4.1.1 Subjects and Administration

Subjects participating in this experiment were mixed ability from selected assessors to expert assessors (ISO 8586-2:1994 recommended application to the field of audio (Bech & Zacharov, 2006)) within the University of Huddersfield. Thirteen subjects in total took part in this experiment and all reported normal hearing acuity. All subjects undertook a familiarisation exercise at the beginning of the experiment. This consisted of three paired comparisons of different genres presented in a similar style to that in Figure 25 with the absence of any scales. Samples used in this exercise were *not* used for the following tests but are representative of the

[8] Touch OSC software developed by Hexler Ltd.

extreme differences of timbral and spatial changes of samples shown in Section 4.1.2. Comparing these stimuli, subjects were instructed to listen to any perceived changes the '*Processed'* sample had in comparison to the '*Reference'* sample. This was to confirm that: subjects could both hear differences between the two and so they may understand what *kind* of differences may be perceived.

4.1.2 STIMULI

Few subjective experiments in the field of vertical reflections have focused on the use of musical signals and have mainly used speech, noise or tonal samples. This investigation hopes to establish results more applicable to the preference of a signal commonly heard in everyday listening of audio. It is highly unlikely for people to be listening to pink noise or sinusoids in the context of listening for entertainment, or to base a preference. Therefore, five music signals and one speech signal (for comparison) were chosen (Table 6), all possessing different spectral and temporal characteristics (see Appendix A for sample FFTs).

	Excerpt	Duration	Characteristics
A	***Artist***: Amy Winehouse ***Track***: You know I'm No Good	4.96s	Transient kick and snare drum hits and snare rolls
B	***Artist***: Newton Faulkner ***Track***: Feels Like Home	11.87s	Sustained guitar notes and transient percussion on guitar body. Guitar and string noise
C	***Artist***: Joe Satriani ***Track***: Satch Boogie	4.45s	Hi-hats sample chosen for isolation of high frequency content
D	***Artist***: Sam Hulick ***Track***: From The Wreckage	13.75s	Full range orchestral sample with sustained notes
E	***Artist***: Caro Emerald ***Track***: That Man	4.95s	Low-fi style piano
F	***Artist***: ***Track***:	13.0s	Foreign speech signal

TABLE 4: TABLE OF STIMULI

4.1.3 PREFERENCE TESTING

Verbal and written instructions were given to subjects to compare stimuli '*Reference*' and '*Processed*' and assess them simply based on their preference. It was highlighted that the context of this assessment should be thought of as, "*listening for entertainment and pleasure*," i.e. within a home theatre. The level of preference given would indicate how preferable the playback *with* the reflection (processed) sample was over playback with just the direct loudspeaker (reference). A bipolar rating scale was employed ranging from -50 to +50. '*Highly not preferred*' corresponded to -50 and conversely, +50 indicated '*Highly preferred*'. '*No preference*' was located at the middle equal to 0. Subjects had full freedom of adjustment with a step size of 1.0, and were told to think of this scale as a linear progression between the two opposing ends. Each comparison was presented three times and randomised throughout both tests of preference and perceived magnitude of change. 18 comparisons were made in total for each test.

4.1.4 Perceived Difference Testing

Following the preference test, subjects were then instructed to assess the same set of stimuli for the perceived timbral and spatial differences. The stimuli were presented in a randomised order from the previous test to minimise carryover effect. Unlike studies by Bech (1995, 1996) where the amplitude of a reflection required to *produce* a (timbral) change is investigated, this takes a static level and focuses on the perceived magnitude of change the reflection provides timbrally *and* spatially. Descriptions of spatial and timbral characteristics (Appendix C) were presented to the subjects and used to aid the assessment, providing attributes the user might feel best describe the timbral or spatial change perceived. Subjects were instructed to identify the nature of this difference through these descriptions provided (or their own description) and state if this was a *positive* or *negative* change. The description of timbre was presented to subjects in accordance with ANSI standard (Section 2.6.2).

In replacement of the preference scale for the second test, was a '*perceived difference'* scale. Although this is not a recognised scale within the literature, scales discussed in Section 2.7 and in further literature all have negative or positive connotations regarding their labelling. For instance, a 'degradation comparison rating' (DCR) scale could have been employed however, this would imply that the subject is assessing the '*Processed*' stimuli in terms of it being worse and not how much change has occurred. The use of negative and positive labelling of scales like this could possibly bias subject's response therefore, the scale shown in Table 7 was employed. It is acknowledged that this scale includes levels of 0-20 as '*No Difference'* however, subjects were clear this was the lowest category on this continuous scale. The order of test one and two was conducted specifically so that no preconceived thoughts about timbral and spatial attributes would influence subject's preference, although a subject may prefer one stimuli over another *based* on these attributes, it was purely through their own thought process.

TABLE 5: ADAPTED 5-POINT GRADING SCALE OF PERCEIVED DIFFERENCE

80-100	Extreme Difference
60-80	Great Difference
40-60	Some Difference
20-40	Slight Difference
0-20	No Difference

4.2 RESULTS

4.2.1 GENERAL OBSERVATIONS

Shaprio-Wilks analysis of distribution show that for **preference**, only 2/6 samples were normally distributed, 5/6 samples for **spatial** difference were normally distributed and 6/6 samples for **timbral** difference were normally distributed (Table 8). With a mix of normal and non-normally distributed data, standard parametric analysis of data would be unsuitable therefore non-parametric statistics were employed (Figures 26 and 27).

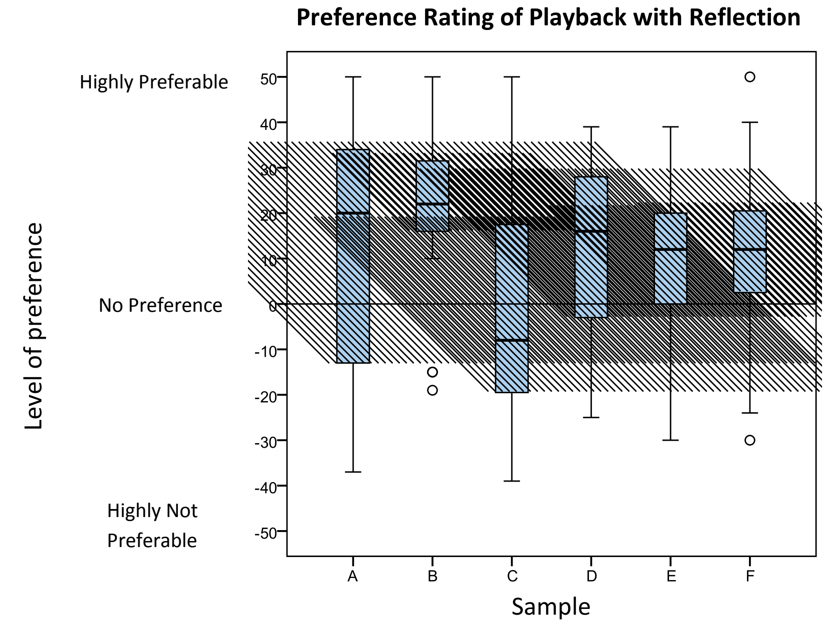


FIGURE 26: PREFERENCE RATINGS ACROSS ALL SUBJECTS FOR ALL STIMULI

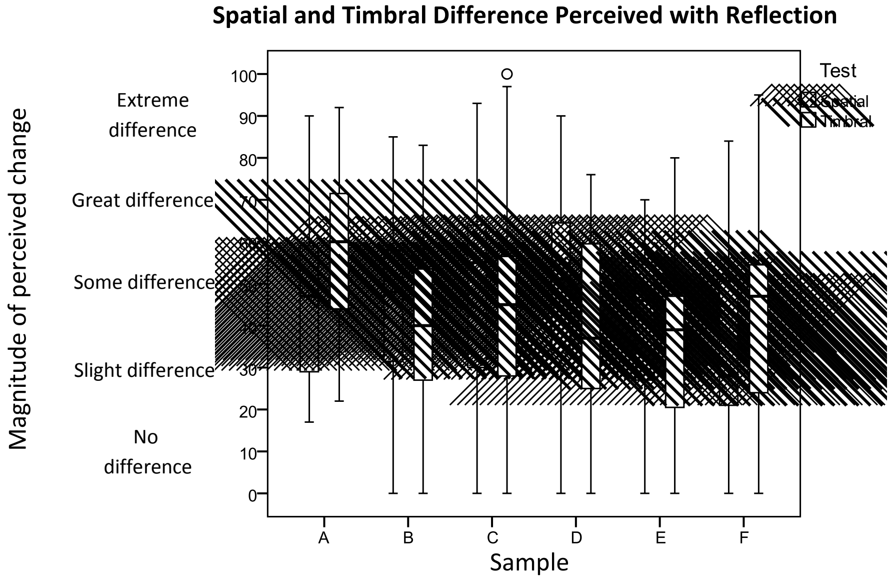


FIGURE 27: PERCEIVED MAGNITUDE OF CHANGE RATINGS ACROSS ALL SUBJECTS FOR ALL STIMULI

		Shaprio-Wilks Statistical Analysis of Normality					
		Preference		Timbral Difference		Spatial Difference	
Sample	DF	*Sat.*	*Sig. (**p**)*	*Sat.*	*Sig. (**p**)*	*Sat.*	*Sig. (**p**)*
Drums	39	0.916	*0.007*	0.975	0.522	0.955	0.125
Guitar		0.92	*0.009*	0.977	0.605	0.95	0.176
Hi-Hats		0.96	0.178	0.968	0.314	0.959	0.167
Orchestra		0.919	*0.008*	0.963	0.231	0.972	0.437
Piano		0.939	*0.035*	0.95	0.08	0.906	*0.003*
Speech		0.917	0.401	0.967	0.299	0.953	0.101

TABLE 6: SHAPRIO-WILKS STATISTICAL TEST FOR NORMALITY WITH 5% SIGNIFICANCE LEVEL

Observing the results of subject's preference of the '*Processed*' sample against the '*Reference*' sample, five of the six samples have a median greater than zero. This would initially imply that in the majority of cases, playback *with* the reflection was favoured by subjects. In turn, this would suggest that the reflection provided beneficial timbral and/or spatial differences. However, this is with the assumption that *only* spatial and timbral differences contributed to the cognitive process for subject's preference (Figure 1). Sample 'C' (Hi-Hats) was the only sample whereby playback was preferred *without* the reflection, with the median lying beneath '*No preference*'. As no procedural error (Bech & Zacharov, 2006) was observed throughout testing, the outliers observed in Figures 26 and 27 cannot be discounted. Regarding the magnitude of perceived change, nearly all subjects *did* perceive a change both timbrally and spatially. However, both tests display a spread of data over a large margin of error and through visual investigation of Figures 26 and 27, it is clear no significant difference can be found. Therefore, the use of significance testing for non-parametric data such as Wilcoxon or, Mann-Whitney U significance testing is not employed.

4.2.2 SUBJECT RATING CONSISTENCY

Some responses for a repeated sample were seen to have a large range therefore, it was within the author's interest to observe each subject's '*intrarater reliability*'[9] (Zacharov & Mattila, 2001). As mentioned in Section 4.1, six different stimuli were used throughout testing and repeated two more times in a randomised order, resulting in 18 ratings per subject. Subjects were unaware if any varying processing had been applied to these repeating samples (which had not). The main reason for the inclusion of these repetitions was to observe the reliability in subject's responses.

To identify inconsistent ratings of a subject, a margin of difference was applied in order to address any large ranging responses for any one sample. The scale employed ranged from -50 to +50 therefore, subject's whose rating for a single sample differed by 25 or more points were initially removed from the data set. Although this may seem a large degree of inaccuracy, a subject's ability to reliably and accurately place an indicator on a scale, within a smaller margin based on a perceptual attribute was not the focus of this investigation. A range 1/4th of the scale would also allow freedom of response. With thirteen subjects taking part in total assessing six samples, 78 consistency ranges were reviewed. Highlighting subjects that exceed this discrepancy of 25, the following points can be drawn:

- 21/78 results for subject's preference exceeded 25.
- 33/78 results for subject's magnitude of perceived timbral change exceeded 25.
- 32/78 results for subject's magnitude of perceived spatial change exceeded 25.
- Split of preference can also be observed between positive and negative preference across all subjects.

Shapiro-Wilks normality test and non-parametric analysis was then re-conducted with the removal of these results (referred to as '*selected data*') and were seen to make little difference to the normality of results shown in Figures 26 and 27. The range of selected data still spans the extremes of the two scales, and median values remaining positive and negative for the same

[9] The consistency of results for repeated comparisons *per* subject.

samples as before. Therefore, the author sees no reason to exclude responses greater than a discrepancy of 25. Although all subjects can be quantified in terms of '*selected assessors*' to '*expert assessors*' (Section 4.1.1), one cannot argue that subject's *did not* hear a different level of timbral and spatial difference or preference throughout the assessment for the same stimuli. These inconsistent ratings may well be intentional and could be due to a number of factors such as: subconscious effects – thinking or presuming each sample must be different, simply perceiving something in a repetition not heard previously and/or auditory adaptation (Pike, Brookes, & Mason, 2013). Preferential change throughout the testing could be an area of further investigation.

4.2.3 PREFERENCE SPLIT

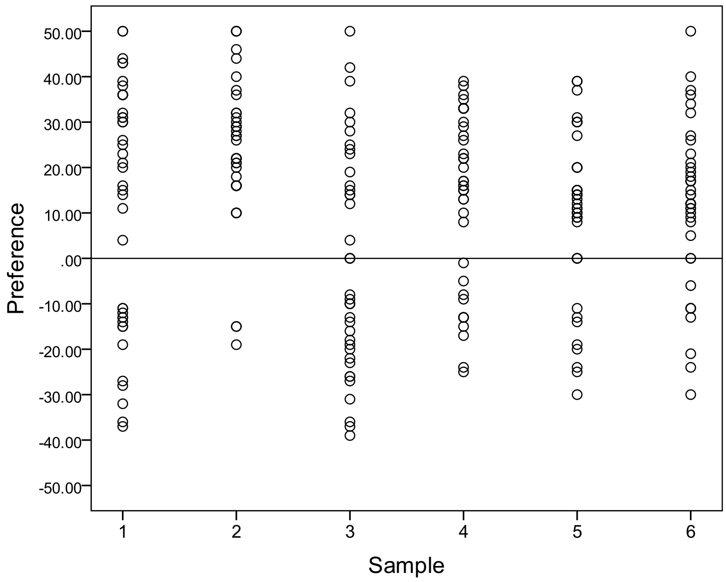


FIGURE 28: PREFERENCE VS. PERCEIVED TIMBRAL AND SPATIAL DIFFERENCE FOR ALL SAMPLES

Investigating the data further, a scatter analysis of subject's **preference** was plotted (Figure 28) whereby a split can be observed. It was therefore hypothesised that by splitting these results of subject's preference, two directions of the '*cognitive processes'* (Figure 1) between the '*perceptual domain'* and '*affective domain'* may be seen. Results of preference were segregated into two graphs representing subject's preference level, with their corresponding values of perceived timbral and spatial difference. All values of positive preference were averaged[10], along with timbral and spatial differences and plotted in Figure 29 and vice versa for values of

[10] *Mean* averaged values

negative preference in Figure 30. Pearson correlation coefficient analysis reveals the following relationships between all three response ratings:

- Correlation of +**0.81** for musical stimuli (A-E) between positive preference level and perceived **timbral** change
- Correlation of +**0.85** for musical stimuli (A-E) between positive preference level and perceived **spatial** change

With the inclusion of sample 'F' (speech), correlation coefficients significantly reduced. The only correlation noteworthy being between preference and timbre yielding **0.72**. These values could suggest that when subjects vote increasingly ***positive*** for the inclusion of a reflection, they do so based on the increasing magnitude of perceived timbral and spatial change.

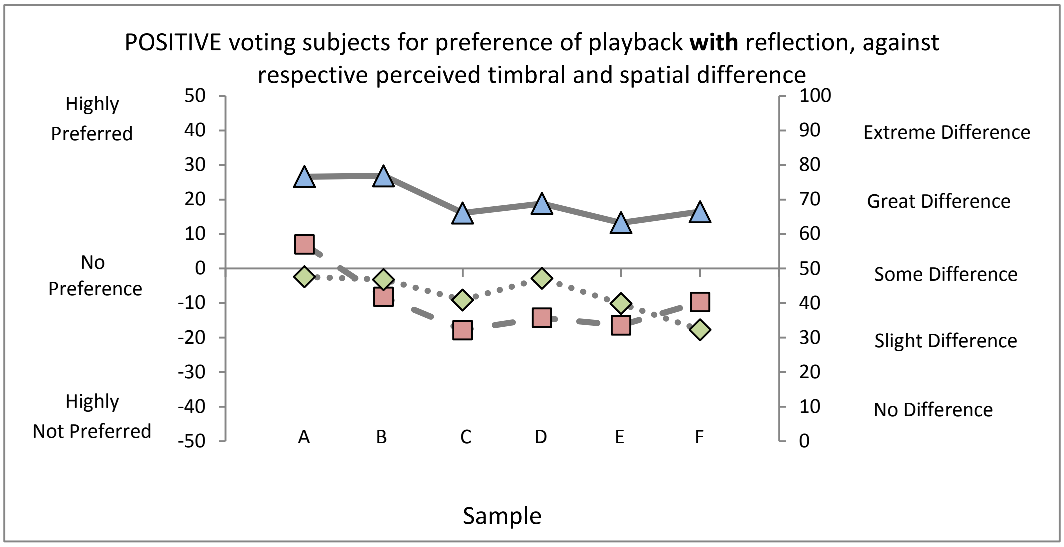


FIGURE 29: MEAN POSITIVE PREFERENCE LEVELS AND RESPECTIVE MEAN LEVELS OF PERCEIVED TIMBRAL AND SPATIAL CHANGE

When a subject's preference was ***negative*** with the inclusion of a reflection (Figure 30), the perceived level of timbral and spatial differences were very similar throughout all musical stimuli (A- E) with high correlation of **+0.89**. However, this final observation is not in relation to *preference* level, and can therefore, only imply that when subjects do not prefer a reflection they may judge the magnitude of timbral and spatial difference it provided at a similar level.

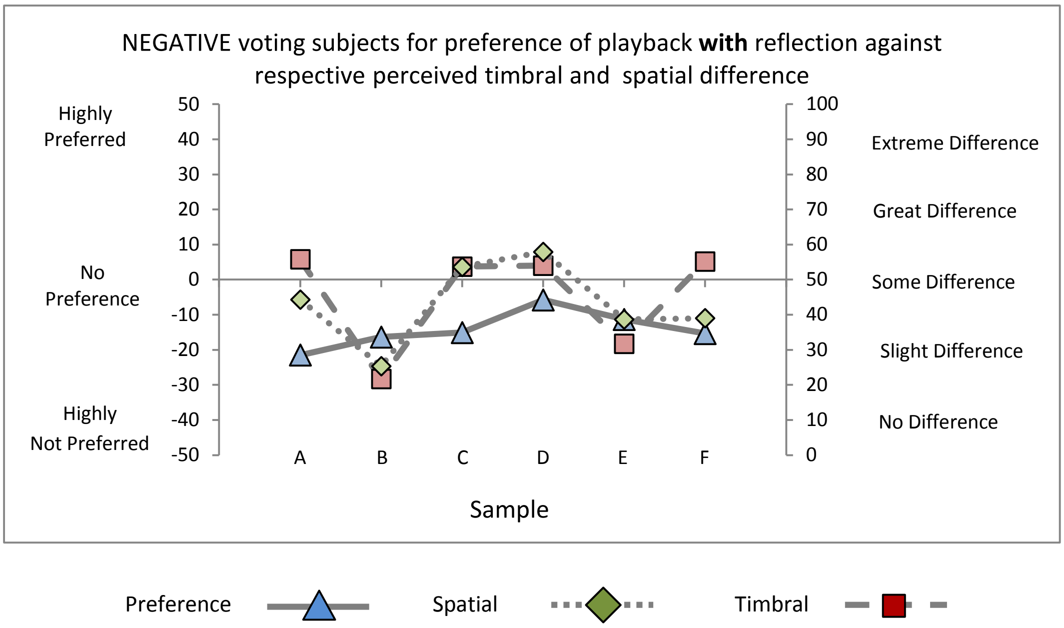


FIGURE 30: MEAN NEGATIVE PREFERENCE LEVELS AND RESPECTIVE MEAN LEVELS OF PERCEIVED TIMBRAL AND SPATIAL CHANGE

4.2.4 Descriptive Terms

The experiment thus far has shown that both spatial and timbral differences are heard with the presence of a vertical reflection, supporting work discussed in the literature review. Also, subject's did not perceive a static amount of change, but that this change may increase and decrease dependant on stimuli. Following the results of the correlation analysis, which suggested the preference of musical stimuli could be based upon a cognitive assessment of timbral and spatial attributes, this section aims to identify *which* perceived timbral or spatial attributes impacted subject preference. A free verbalisation task took place after each comparison to describe any timbral and/or spatial differences heard and state if this was a *positive* or *negative* effect. Semantic analysis of descriptions was conducted to lemmatize descriptions and auditory sensations to their base level (e.g. the response of "*fuller*", "*fullness*" and "*fullest*" all possess different suffix, but can all be categorised by their stem adjective "*full*"). By consolidating all subjects' preference levels and these timbral and/or spatial descriptions for each stimulus, a connection was observed regarding subject's descriptive terms and the preference level for that sample.

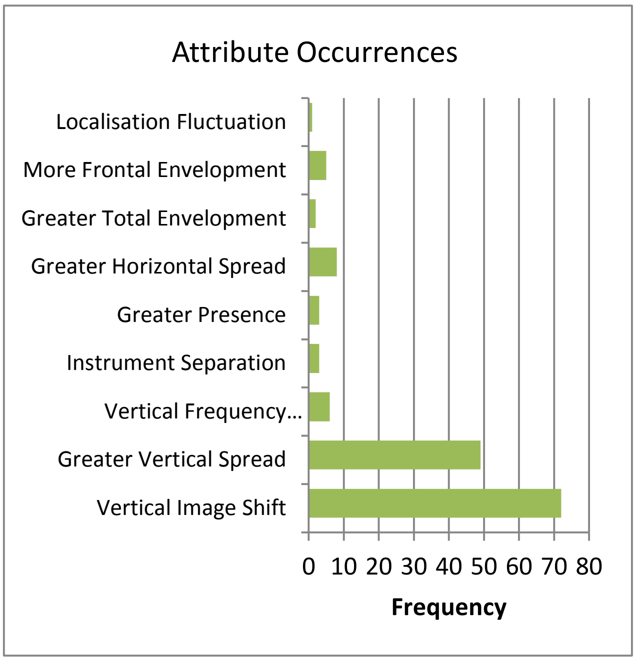


FIGURE 31: FREQUENCY OF SPATIAL ATTRIBUTES – (1)

Throughout all descriptions of spatial attributes, the most commonly used to describe auditory sensations were "*vertical image shift*" and "*greater vertical spread*' (Figure 31). As the experiment took place within a semi-anechoic chamber, it is reasonable to assume that the addition of a vertical reflection provided these spatial cues. However, in cases where either of these two attributes were consistently mentioned from a single subject for all samples, the subject's preference level varies from positive to negative. This could therefore imply that another factor is influencing subject's preference to a *greater extent* than spatial attributes causing it to vary from negative to positive. Or that because these two attributes were consistently observed, they were *consistently* either a positive or negative perceived effect and therefore, provided an '*offset*' to subject's preference.

Descriptive terms used by subjects to describe timbre varied (Figure 32). The most commonly used adjectives to describe a negative attribute were: "*thin*", "*nasal*" and "*boxy*". In contrast, those used to describe positive characteristics were: "*full*", "*rich*" and "*clear*". Elicitation and individual vocabulary profiling are beyond the scope of this paper; the main goal here was to see if these negative and positive words corresponded with subject's preference. Noticeably, the use of "*bright*" was the most frequent but interestingly, was used to describe both negative and positive perceived changes.

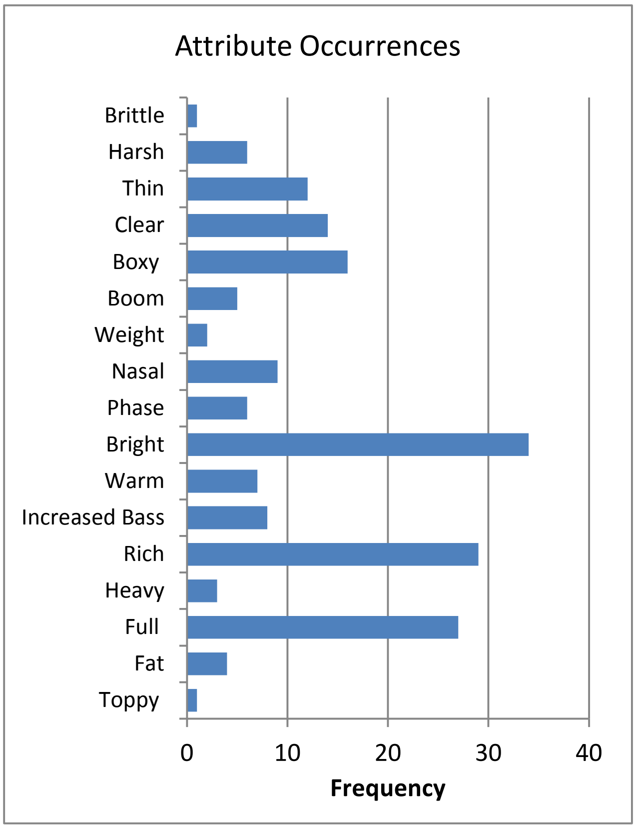


FIGURE 32: FREQUENCY OF TIMBRAL ATTRIBUTES - (1)

When a subject's level of preference of playback with reflection was positive, the majority of timbral descriptions were also positive characteristics (as mentioned by subjects), and vice versa when preference was negative. Overall, 10/13 subject's timbral descriptions followed subjects' negative and positive preference for 80% of stimuli, unlike spatial attributes that remained largely consistent throughout. Therefore, it is not unreasonable to assume that regardless of a positive (or negative) spatial impression, most subjects' preference of a reflection is largely based on their perception of a timbral characteristic.

4.3 Experiment One Summary

This experiment comprises the first of two studies focusing on preference and magnitude of perceived timbral and spatial differences. The results have provided new data regarding these attributes when playback includes a first geometrical ceiling in the context of listening for entertainment. Further discussion of these results is included following the analysis of the second experiment in Chapter 6.

Initial conclusions from this experiment are:

- No *single* correlation can be found between magnitude of perceived timbral and spatial difference and subject's preference to estimate the enjoyment of playback with a reflection.
- No significant difference is observed of the perceived timbral or spatial difference when a reflection is present between all stimuli.
- No significant difference can be seen through variation of program material with regards to subjects' preference of a reflection being present.

However, the experiment has highlighted some interesting results, with the following points possibly leading to further research to provide clarification:

- Positive and negative levels of preference could possibly be based upon two different *cognitive processes* in a semi-anechoic listening environment. Mean positive voting preference highly correlated with mean perceived levels of timbre and spatial attributes for all musical stimuli. This Implies that *positive* preference may be based upon the magnitude of perceived changes.
- When subjects did *not* prefer playback with the reflection, perceived mean timbral and spatial differences were rated similar throughout musical stimuli.
- Subject's description of negative and positive *timbral* attributes generally corresponded to subject's preference of playback with the reflection. This implies that timbral attributes contributed highly to a subject's preference.
- Where consistent use of spatial attributes was observed, preference of playback with reflection would still vary from negative to positive. This suggests either spatial change did not have a great enough impact to sufficiently alter a subject's preference. Or, subject's *consistently* perceived these attributes as a beneficial or detrimental effect regardless of stimuli.

4.4 LIMITATIONS OF THIS EXPERIMENT

This experiment suggests some possible results regarding preference and the influence of timbral changes. As previously stated, experiment one was chosen to be conducted within a semi-anechoic chamber to remove all other reflections therefore, assessing criteria solely based on the contribution of a vertical reflection. However this is a scenario seldom seen in domestic environments and is addressed in experiment two in Chapter 5.

The loudspeaker used to simulate a reflection had processing of SPL attenuation, dispersion characteristics and delay, representative of a true reflection. Nevertheless, playback of the program material with the reflection increased the SPL levels at the listener position by roughly 1.5dB above playback with just the direct sound. Arguments can be made both for and against this loudness difference. On one hand, it is often considered that when people are asked which sounds ‘better’ out of the same piece of music played at different volumes, they’re likely to choose the louder one (Vickers, 2010)(Milner, 2010). This may suggest the majority of people would prefer playback with the reflection as the volume at listener position was increased. On the other hand, this level difference is representative of what would happen with and without a reflection present. Therefore, calibrating playback *with* the reflection loudspeaker to equal SPL levels of just the direct sound would technically not be assessing the addition of a true reflection, but of an elevated sound source. This is addressed in the second experiment of this research.

CHAPTER 5

EXPERIMENT TWO

5.0 INTRODUCTION

The aim of this experiment is to investigate the effects of a frequency dependant reflection on a listeners' preference in a more realistic listening environment. From the free verbalisation task in experiment one, the descriptors given by subjects throughout each paired comparison suggested that negative and positive adjectives used to describe timbre, in most cases corresponded with subject's level of preference. No limitations were imposed on descriptive words subjects could use to describe any differences heard. As long as subjects used the correct terms to describe the auditory sensation, this seemed suitable[11]. The following experiment continues this, focusing mainly on the use of adjectives and subjects' preference more in-depth. As no *single* correlation could not be found between the magnitude of perceived timbral/spatial difference and preference, magnitude of global spatial and timbral differences is not investigated here. For this experiment, frequency content of the reflected sound is manipulated to investigate if this affects subject's preference and responding attributes.

5.1 METHODOLOGY

Subjective testing was used to investigate levels of preference of sound reproduction with and without the presence of a vertical ceiling reflection as before. Subjects were instructed to verbally feedback to the assessor descriptive terms, directly describing *why* they preferred their chosen playback option. The previous experiment used adjectives to describe the magnitude of perceived timbral and/or spatial change, and then cross-referenced with subjects preference. Therefore, an indirect observation could be made as to whether this preference was based on the negative or positive perceived changes. In this experiment, directly describing why subject's preferred their choice of playback eliminates any error with interpreting results between the two.

A vertical ceiling reflection was electro-acoustically simulated following the process shown in Chapter 3. This time, the experiment was conducted in the University of Huddersfield's

[11] Discussion during the free verbalisation task with each subject ensured that the responded attribute represented the subject's perceived effect. This was purely *clarification* and the assessor was cautious not to bias attribute response.

'Applied Psychoacoustics Lab' listening room (Figure 33) compliant with ITU-R BS.1116 regulations with additional floor absorption. Experiment one was an exaggerated scenario conducted in a semi-anechoic chamber with a low ceiling reflection which, when assessing a single reflection in a room with *no* reflections, clearly affected the subjects' spatial response. Therefore, the addition of this sound field will provide better applicability of results to a real world listening scenario. The ceiling reflection was replicated based on the following dimensions:

Listener Distance - 2m, Listener Height – 1.15m, Reflection Point – 2.2m

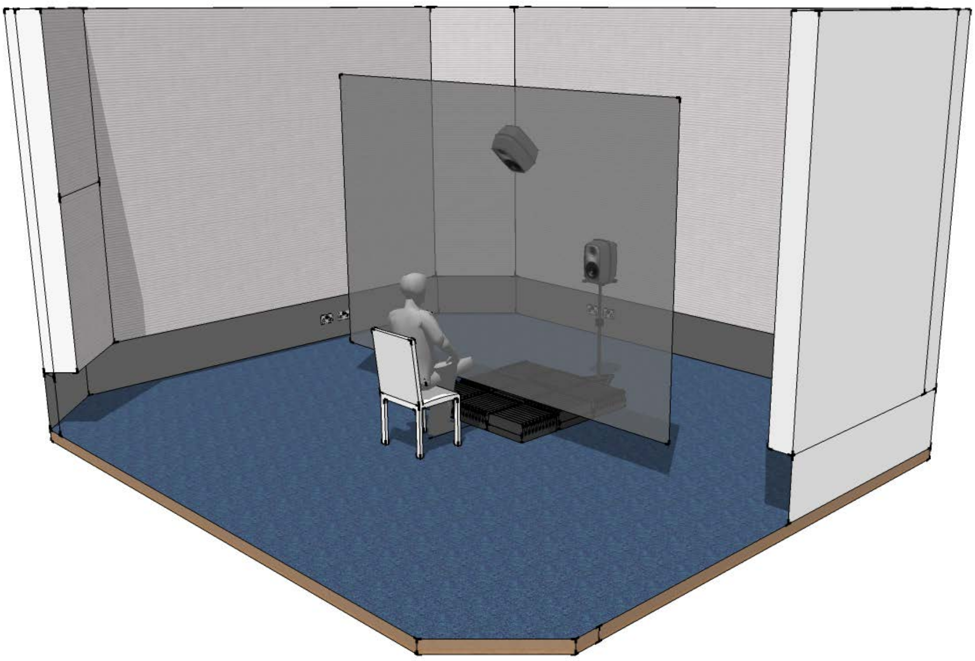


FIGURE 33: UNIVERSITY OF HUDDERSFIELD'S APPLIED PSYCHOACOUSTICS LAB ITU-R BS.1116 LISTENING ROOM

Following the calibration process in Chapter 3 whereby '$Direct_{Distance}$' = 2m and '$Ceil_{Distance}$' = 1.45m × 2 (defined from the dimensions above), Equation [8] results in a ceiling reflection delay of 8.4*ms* and Equation [11] in 9.2dB ceiling reflection attenuation. As before, these values are halved to simulate the reflection from halfway along the reflection path, resulting in processing of **4.2***ms* delay and **4.6**dB attenuation. Using these values, an impulse was taken and a discrepancy of **-0.57**dB level and **0.296***ms* delay was observed between calculated and simulated results. Whilst not exact, the author feels as though due to subtle differences in room temperature and absorption, calculated values may always have a small margin of error when compared to a genuine reflection as seen in Section 3.3. Head movements were also not restricted by the use of a head clamp for either experiment. Therefore, while subjects are instructed to remain as still as possible facing forward, minor head movements will always incur small changes to delay and attenuation at the listener position. For these reasons, the discrepancies observed for both experiments were deemed small enough to be acceptable for this Thesis.

Regarding the simulated height, although still not as high as domestic ceilings, the author did not want to make too greater change to height in conjunction with the rooms natural characteristics. The addition of too many changes to the experiment may prove difficult when discussing reasons in any noticeable differences between results.

A paired comparison test was employed as before, to assess playback with just a direct loudspeaker, against a direct loudspeaker with its respective ceiling frequency *dependant* reflection. However unlike the previous experiment, where stimuli were presented as '*Reference*' and '*Processed*', stimuli were presented as '*A*' and '*B*' to subjects, eliminating any potential bias when choosing preference. Playback with and without the presence of the ceiling reflection was also randomised between '*A*' and '*B*' ensuring subjects could not become accustomed to a certain playback, with a particular stimuli selection and can therefore be considered a blind AB comparison test. As before, all equipment was obscured from view with an acoustic curtain to retain no visual bias. Subject's response of preference was performed on a tablet (Figure 25) for reasons discussed in Section 4.1. The stimuli chosen for experiment two were consistent with those of experiment one (see Section 4.1.2).

5.2.1 Subjects and Administration

Subjects participating in experiment two consisted of academic staff, post-graduate students and lecturers at the University of Huddersfield spanning assessment abilities outlined in Section 4.1.1. Eleven subjects in total took part, six of whom participated in the previous experiment. All subjects took part in a familiarisation exercise before beginning the test. The six samples to be assessed, were presented as a paired comparison with and without the reflection - no labelling was necessary ('*A'*/'*B'*) as no subject response was needed. Subjects were instructed to go through all pairs and listen for differences between the two playback options per sample understanding: frequency content, temporal characteristics, spatial and timbral changes. The reflections being simulated for this training exercise were frequency *independent*, with only the direct loudspeaker dispersion altering spectral content of the reflection loudspeaker. To gain a controlled response of descriptors, subjects were also handed an adjective response sheet (see

Appendix D). This list of auditory sensations and corresponding descriptions meant that interpretation of each sensation would remain consistent across all subjects, rather than being dependent upon each subject's own interpretation. The main content of this is taken from Torben H. Pedersen & Zacharov's paper (2015) discussed earlier in Section 2.6. However, the inclusion of vertical spatial sensations had to be included for the nature of this work, along with omitting artefact descriptions such as signal related issues and noise which are not assessed. The test began once any remaining questions were answered and subjects agreed that they understood all descriptive terms to be used.

5.2.2 Preference Testing

Subjects were given verbal instructions to assess stimuli '*A*' against stimuli '*B*' and use the presented scale to indicate their level of preference. As in experiment one, the context of this assessment should be thought of in terms of '*listening for entertainment or pleasure'* and not critical assessment of mixing. The scale employed was an adapted version of that in ITU-R P.800:1996 used in Zacharov & Lorho (2004) seen in Section 2.6.1. The scale employs 9-points and full freedom of adjustment, a step size of 0.1, with the addition of 0.5 tails to eliminate end bias (subject reservations in going to extreme values). The anchors of the words were presented in Figure 34.

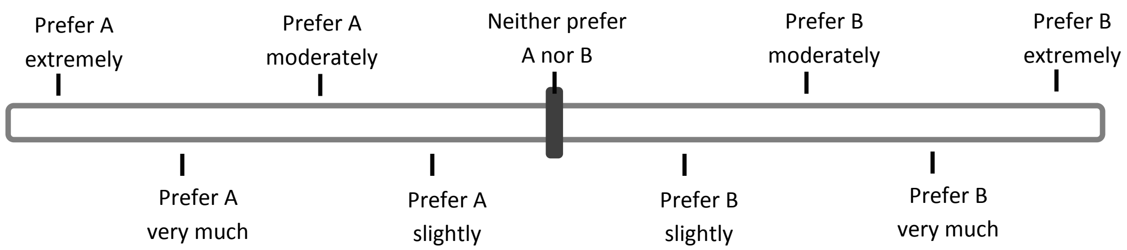


FIGURE 34: REPRESENTATION OF PREFERENCE RATING SCALE EMPLOYED

After each paired comparison, subjects were asked to describe why they preferred A/B using the descriptive sheet provided and instructed to adhere to these as much as possible. If any auditory sensation was not categorised by given words subjects felt needed to be used, a note was taken by the assessor.

Throughout the experiment, subjects would be assessing playback with the direct sound vs. playback with direct sound and an associated *frequency dependant* reflection. The varying frequency dependant signal would be the removal of one of eight single octave-bands ranging

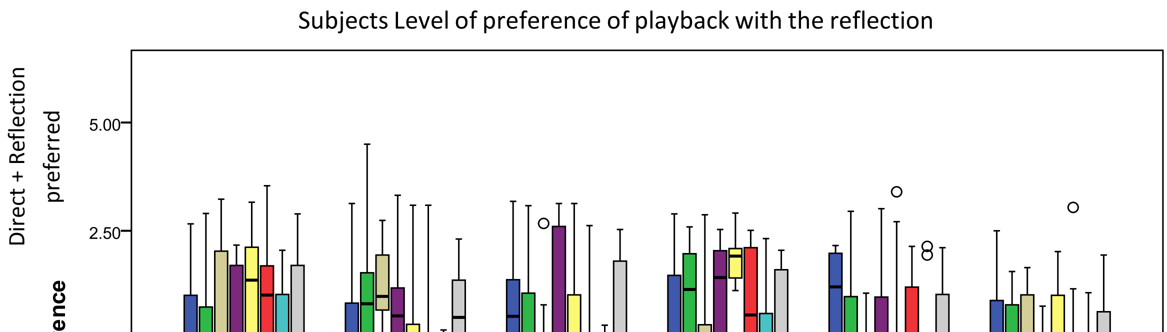
125Hz – 16kHz from the reflection. The applicability of this frequency dependant reflection can be thought of as an absorber with an octave-band target frequency, with an absorption coefficient of 0.99. This was reasoned to be a more realistic scenario than its counterpoint of a reflector, reflecting just a specific frequency. Octave-band removal was achieved using a 12th order stop band Butterworth filter applied in MAX MSP, allowing variable input of upper and lower bounds, at the -3dB point of the rise and fall at each side. With six samples and eight octave filters being applied, 48 comparisons were made per subject. Unlike experiment one, each comparison *was not* repeated three times as this would have resulted in 144 comparisons resulting in subject fatigue. Both samples and all octave-band filters were randomised throughout the experiment.

5.2 Results

This section will provide an overview of the results from experiment two followed by a discussion of both experiments in Chapter 6. Initially, the raw results must be sorted in terms of how much each subject preferred playback with/without the reflection. Throughout this A/B comparison, playback with/without the reflection altered between ‘A’ and ‘B’. Therefore, the results need to be sorted so that one particular playback option is anchored to ‘*A*’ and the other to ‘*B*’. After sorting, values above zero correspond to playback *with* the reflection (‘*B*’), negative values *without* (‘*A*’). Whilst sorting, when this was not observed and positive values were associated with playback *without* the reflection, these values were reversed (e.g. Table 9). This technique (similar to that of Lorho and Zacharoz (2004)), assumes that subject’s treated the response scale symmetrically.

Example	Preference level	B =		Preference level	B =
1	2.17	With Reflection	->	2.17	With Reflection
2	-2.53	Without reflection	->	2.53	With Reflection

TABLE 7: EXAMPLE OF ANCHORING RESULTS BETWEEN PLAYBACK OPTIONS TO POSITIVE AND NEGATIVE VALUES



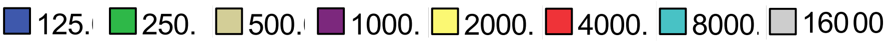


FIGURE 35: PREFERENCE LEVEL OF PLAYBACK WITH THE REFLECTION – GROUPED BY **SAMPLE**. POSITIVE VALUES INDICATE REFLECTION WAS PREFERRED

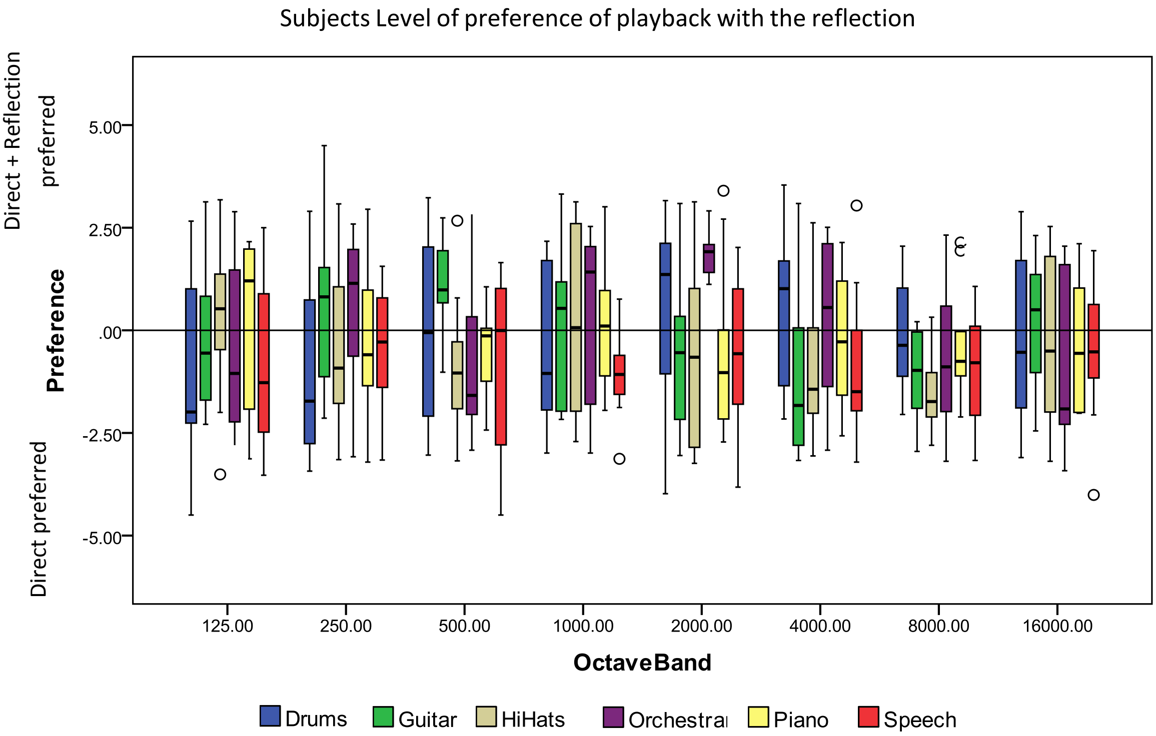


FIGURE 36: PREFERENCE LEVEL OF PLAYBACK WITH THE REFLECTION – GROUPED BY **OCTAVE-BAND**. POSITIVE VALUES INDICATE REFLECTION WAS PREFERRED

Shapiro-Wilks analysis of distribution for preference values show a split of normal and non-normally distributed data therefore, following analysis is non-parametric. Observing Figure 35 and 36, three observations can be made:

- A large degree of spread can be viewed across most comparisons, with the exception of some frequency dependant reflections for particular samples providing: a more concise quartile range (e.g. Guitar - 500Hz, Orchestra - 2kHz and Hi-Hats - 8kHz) and the most extreme median of preference (discussed further in Section 5.2.2)
- The removal of one single octave band in a reflection is neither consistently preferred nor not preferred across all samples. Although *removal* of octave-band 8kHz does provide a *median* below 'no preference' for all samples.
- Only 14/48 median values are *positive* for playback with the reflection.

5.2.1 DESCRIPTIVE TERMS

For the verbalisation task, a list of descriptive terms and their associated auditory sensation was given so that no subjects would misinterpret meanings. However, these descriptive terms were not labelled positive or negative – this was down to subject interpretation. As before with preference values, the descriptions were organised so that the response was always associated with playback with the reflection. For instance in a scenario where '*A*' = playback with reflection, and '*B*' = direct loudspeaker only, one may moderately prefer 'B' because of *increased* brilliance and clarity. To correspond with preference values being always in relation to *the addition* of the reflection, this would then be reversed to - playback *with* the reflection *decreased* clarity and brilliance. The author's decision to do this was for the organisation of data and to provide a consistent way of interpreting the results. Similarly to preference values, this also assumes that subjects would treat the increase/presence and decrease/absence of a sensation symmetrically. This was done for all 48 paired comparisons across 11 subjects resulting in 528 paired comparisons spanning all octave bands and stimuli.

In combination with these descriptive terms it may be possible to identify the reasons as to why a particular preference score is given for a sample. As previously mentioned in Section 4.2.4, subjects almost always perceived a level of spatial *and* timbral change. This experiment demonstrates which ones influenced subject's preference directly. One subject in particular had no preference for 26 out of the 48 paired comparisons and thus gave no descriptive terms for over half the data. Therefore, with very little preference data and verbal descriptions contributing to overall results, this subjects' data was excluded from further analysis leaving a total of 480 paired comparisons.

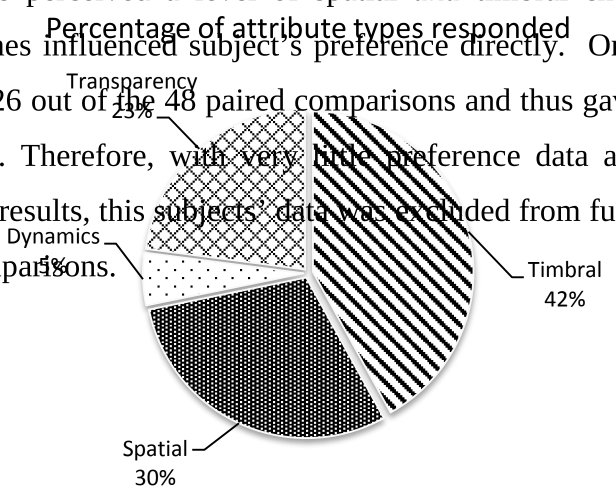


FIGURE 37: RESPONDED ATTRIBUTE TYPES AS A PERCENTAGE

As multiple attributes could be given per comparison, 741 attribute responses were given in total.

Across all comparisons made: 311/480 preference ratings were based on 'timbral' descriptors, 221/480 were based on 'spatial', 170/480 were 'transparency' descriptors and 39/480 were 'dynamic' (Figure 37). This data could initially suggest that overall, timbre had the greatest impact for most subjects' preference. One could *also* argue that transparency characteristics such as '*naturalness*' and '*detail*' given their descriptions (Appendix D), contain characteristics that could class them as a timbral quality.

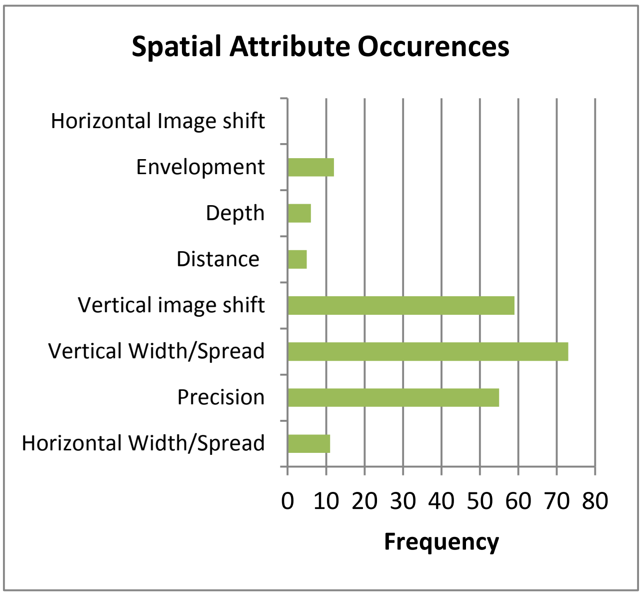


FIGURE 38: FREQUENCY OF SPATIAL ATTRIBUTES - (2)

Out of 221 responses of spatial attributes, only 74 influenced subject's preference *positively*. When preference *was* based on spatial attributes (regardless of whether the rating was positive or negative), the most commonly described auditory sensations were, '*vertical spread*' and '*vertical image shift*' (Figure 38). Conversely, the third most common spatial attribute -

increased and decreased '*precision'*, *did* correspond with *positive* and *negative* preference respectively suggesting greater 'precision' will lead to a higher preference. This implies that while the increase and decrease of *some* spatial attributes are always perceived as detrimental or beneficial across subjects, the most commonly used attributes, vertical spread and image shift are not.

The use of timbral descriptions varied greatly. Responded attributes that were in the descriptive column (third column) of the sheet provided (Appendix D), were stemmed back to the attribute group (second column) to remain consistent with responded spatial attributes. The most common sensations included increased/decreased: '*brilliance*', '*fullness'* and '*treble strength*' (Figure 39). Unlike spatial descriptions, the increase and decrease of timbral attributes always correlated with positive or negative preference of playback with the reflection. For instance, the use of "*increased treble resonance – tinny*" always corresponded with negative preference, conversely "*increased brilliance – clarity*" was positive. This observation also highlights that whilst descriptions were given for the attributes, subjects were unanimous in interpreting certain attributes as a negative and others as positive.

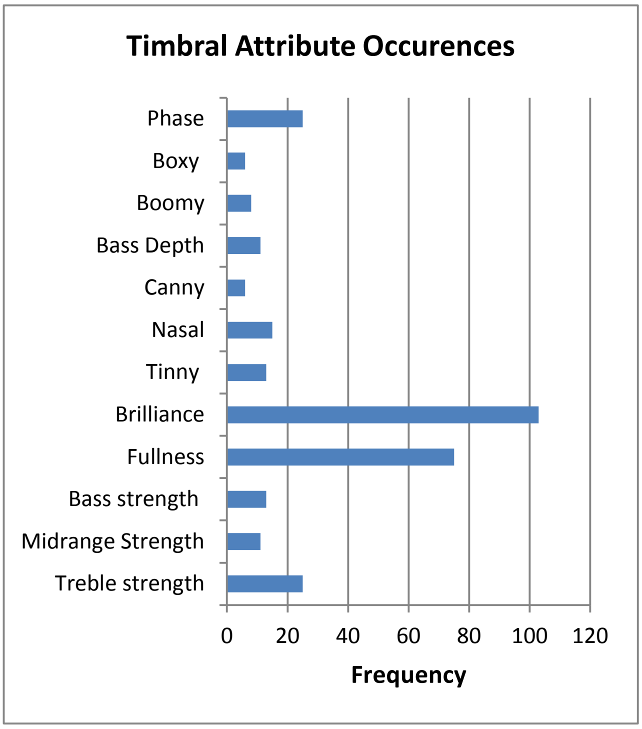


FIGURE 39: FREQUENCY OF TIMBRAL ATTRIBUTES - (2)

Unlike experiment one, where *global* 'spatial' and 'timbral' difference was measured, the perceived magnitude of *specific* sensations was not studied in experiment two therefore, it is not possible to statistically correlate their effects to ascertain *how* preferable individual attributes are (i.e. Is 'brilliance' more preferable that 'fullness'?). This could be the basis of a more focused study employing techniques such as 'direct attribute scaling' (Stephenson, 2012). However, as subjects were undivided regarding the positive/negative connotations of responded timbral attributes associated with preference, it *is* possible to investigate the occurrences of these attributes (Figure 39) against the mean preference value *for* that attribute (Table 10).

The preference values for all 311 timbral responses were split in two groups: when subject's perceived an **increase** in a sensation and a **decrease**. Within these groups, the number of times subject's perceived each attribute was counted, along with the cumulating preference value for the attribute. This sum preference value was then divided by the number of occurrences of that attribute to give a mean **preference** *increase* – **green** or *decrease* - **red**, when a subject perceives a particular sensation.

Attribute	Total Occurrences (*Figure 39*)	***Increased*** Sensation		***Decreased*** Sensation	
		Occurrences	Mean *preference increase* of attribute	Occurrences	Mean *preference decrease* of attribute
Treble strength	25	9	0.49	14	-0.71
Midrange Strength	11	4	0.68	7	-1.56
Bass strength	13	8	1.54	5	-1.37
Fullness	75	33	1.94	42	-1.62
Brilliance	103	62	1.91	41	-1.61
Bass Depth	11	6	2.11	5	-2.34
Tinny	13	9	-1.89	4	0.85
Nasal	15	7	-2.56	8	0.98
Canny	6	6	-2.40	0	0.0
Boomy	8	4	-2.42	4	0.58
Boxy	6	5	-2.43	1	1.02
Phasey	25	21	-1.98	4	1.43
Total	311	174		94	

TABLE 8: MEAN PREFERENCE *INCREASE* OR *DECREASE* FOR INDIVIDUAL ATTRIBUTES FOR PLAYBACK WITH THE PRESENCE OF A REFLECTION

This data builds on Figure 39 suggesting that although '*brilliance*' and '*fullness*' were the most commonly reported timbral attributes, the perception of other attributes may alter preference to a greater extent. Regarding playback with the reflection, subject's that perceived an *increase* in brilliance and fullness, on average, rated their preference +1.9. However, subjects that perceived an *increase* of a 'nasal' sound, rated their preference -2.56. Table 10 also shows that a *direction* of a perceived sensation (increase/decrease) may also have a greater impact on preference than the other. For instance where four subjects said that playback with the reflection **increased** a '*Boomy*' sensation, on average decreased their preference by -2.42, whilst four subjects that perceived a **decrease** of this sensation only improved preference by +0.58.

5.2.2 SAMPLE VS. FREQUENCY

This section will discuss the samples used and related frequency content reflected, associated with subject's level of preference for experiment two. Experiment one used a frequency *independent* reflection to assess subject's preference whereas in experiment two, reflections were frequency *dependent* with the removal of one of eight individual octave-bands from 125Hz – 16kHz. Frequency content of all samples can be seen through FFT analysis. Although the removal of certain octave bands on different samples will have a negligible effect (i.e. the removal of low frequency content on sample C – Hi-Hats which contains predominantly higher frequencies), for consistency across all samples these comparisons were still assessed. The most notable frequency dependant results are shown in Figure 40 (a subset of Figure 35).

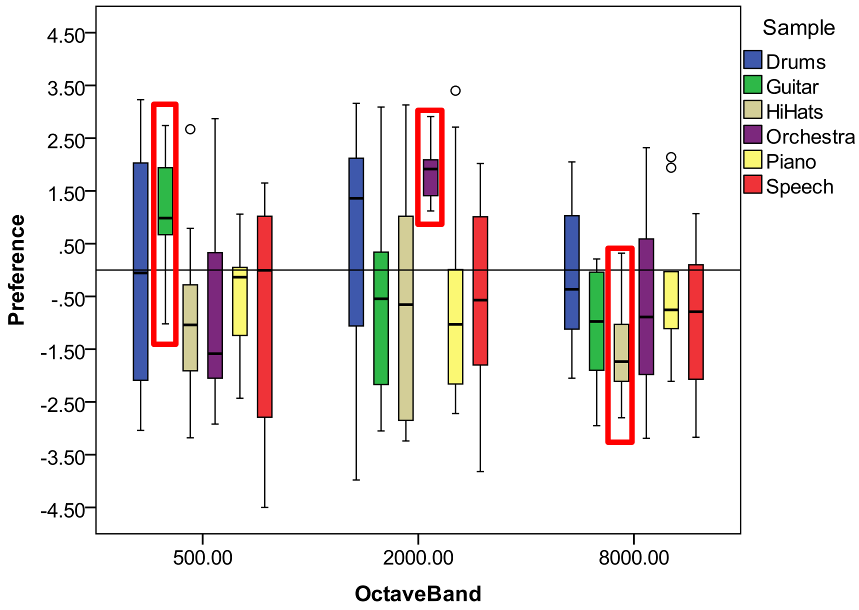


FIGURE 40: NOTABLE FREQUENCY DEPENDANT RESULTS
500Hz - Guitar, **2 kHz - Orchestra** and **8kHz – Hi-Hats**

Sample A was chosen for strong transient sounds from both a low frequency Kick drum, a Snare drum and hi frequency transient content from Hi-Hats (Figure 41) Figure 35 (from the main results – Section 2.5) show that the removal of *any* specific octave bands within the reflection did not result in a more *consistent* level of preference. However, median values for the removal of octave-bands 125Hz and 250Hz in the reflection loudspeaker, were much lower than other octave bands. Therefore, the removal of these octave-bands (where a predominant amount of energy lies for this sample) from the reflection loudspeaker in proportion to the output of the direct loudspeaker could lead to undesirable effects. Descriptions given for these octave bands comparisons are "*Decreased bass strength*", "*Decreased bass depth*", "*Decreased fullness*" and "*Increased treble strength – sharp (negative)"* which supports this suggestion.

FIGURE 41: SAMPLE A (DRUMS) SPECTROGRAM

Sample B (Guitar) had both transient and sustaining sounds from percussive slaps on a guitar body and sustained strings (Figure 42). Playback of this sample with octave band 500Hz removed from the reflection loudspeaker resulted in the highest median preference of *all* comparisons for this sample (Figure 40). Verbal descriptions for this frequency dependant paired comparison varied greatly across subjects, suggesting that no single descriptor can account for this result.

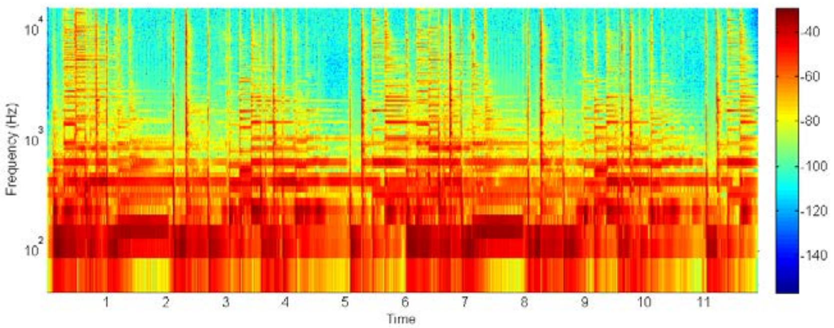


FIGURE 42: SAMPLE B (GUITAR) SPECTROGRAM

Sample C (Hi-Hats) was chosen for specific high frequency content in isolation (Figure 43). Although one may argue Hi-Hats are rarely heard in isolation, the experiment needed specific high spectral content to use for comparison and any instrument/sound of such high frequency will rarely be heard on its own. The results show that for playback with the reflection of octave band 8kHz removed, the quartile ranges of preference is reduced as well as producing the *lowest* median preference overall (Figure 40). This would suggest that subjects prefer the presence of this frequency content in the reflection for this sample. Descriptions for this frequency dependant playback were based on transparency - *"decreased naturalness" and "decreased detail"* **76%** of the time, with the inclusion of one timbral attribute "*decreased brilliance"*. The occurrence of transparency based attributes for *other* frequency dependant reflection of this sample ranged from **7% - 14%.** This indicates that the 8kHz octave band content in a reflection from above, could be providing naturalness and detail.

FIGURE 43: SAMPLE C (HI-HATS) SPECTROGRAM

Sample D (Orchestra – Figure 44) showed that a reflection with octave band 2kHz removed, resulted in the highest median value and the smallest range in subject's preference (Figure 40). Also, this is the only frequency dependant sample to *consistently* result in a positive preference across all subjects. Both these results suggest that for an orchestral sample, the *removal* of 2kHz from a reflection provides a beneficial effect. Subject's descriptions of this perceived effect were "*increased vertical spread*', "*increased frontal envelopment*" and "*increased precision*" showing that these beneficial effects, were all spatially based. This would suggest that for sustained orchestral stimuli, *increased* spatial attributes are possibly beneficial and that the absorption of octave-band 2kHz from a vertical reflection, will increase this beneficial sensation even further.

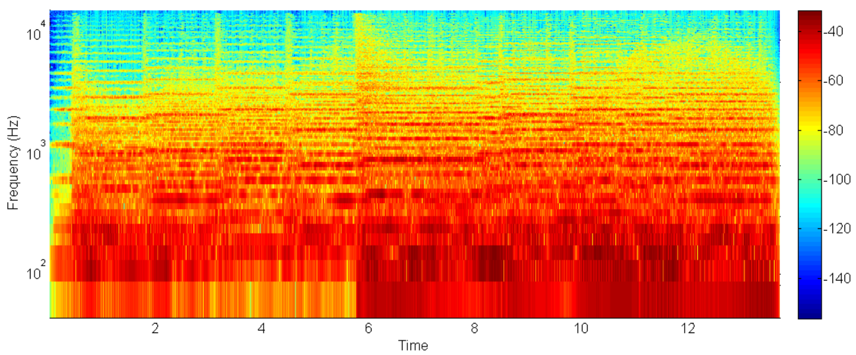


FIGURE 44: SAMPLE D (ORCHESTRA) SPECTROGRAM

Sample E was chosen for a low-fi piano effect (Figure 45). Playback with the reflection of octave-band 125Hz removed resulted in the *highest* median value for preference which is contrary to results for sample A (Drums). This frequency dependant comparison of sample E also resulted in the largest range of preference. This implies that where 125Hz is *not* the main spectral content of the sample (as in sample A) subjects may prefer its removal from the reflection. However, this suggestion is not observed where 125Hz is removed from other samples that 125Hz is not the main spectral content (i.e F – speech, D – orchestra). Subject's descriptive response for this comparison varied greatly across timbral, spatial and transparency based attributes suggesting no single attribute can account for this result. This may also be why this frequency dependant comparison resulted in the largest degree of spread for this sample.

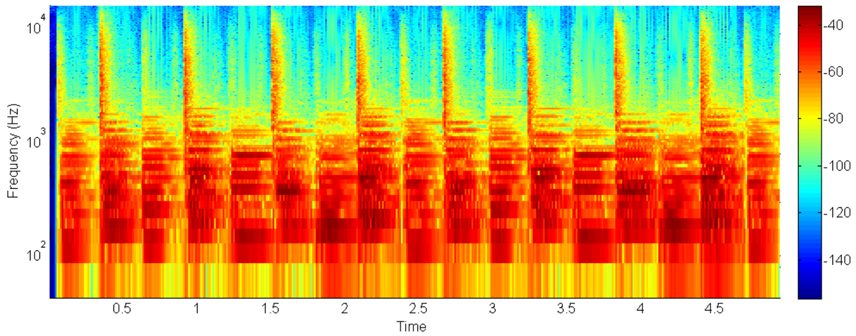


FIGURE 45: SAMPLE E (PIANO) SPECTROGRAM

Sample F was male speech, chosen for a comparison against musical signals (Figure 46). The only observation for this sample is that all frequency dependant comparisons resulted in a median below 'no preference' indicating that the presence of a reflection in general for speech stimuli has a negative effect.

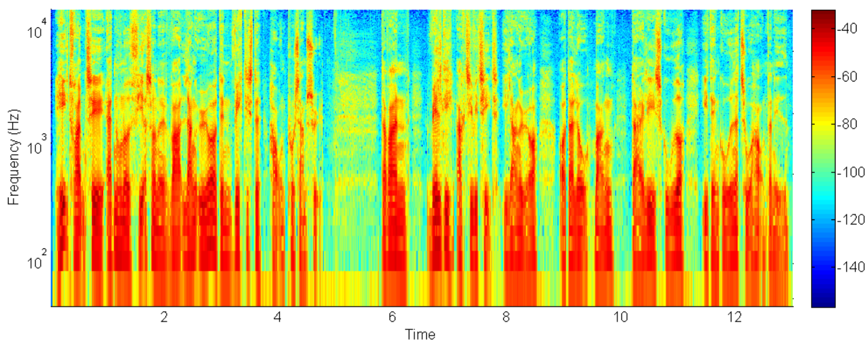


FIGURE 46: SAMPLE E (SPEECH) SPECTROGRAM

5.3 EXPERIMENT TWO SUMMARY

The aim of this experiment was investigate the effect of a frequency dependant reflection on subject's preference in a more realistic listening environment. By using a more detailed list of descriptive terms subjects were able to identify and describe their perceived auditory sensation more accurately and consistently. Playback of stimuli through the direct loudspeaker and the direct loudspeaker with its relative reflection, were presented to subjects at the same SPL so as to not bias perception through loudness and thus influence results.

The following points can be drawn from experiment two:

- Timbral auditory sensations were the basis of subject's preference 42% of the time.
- Increased/decreased '*vertical spread*' and '*vertical image shift'* did not correlate with positive and negative preference respectively. This suggests they *cannot* be categorised as a positive or negative attributes, unlike timbral attributes (or the spatial attribute '*precision'*) which can be categorised into positive and negative auditory sensations.
- Whilst some attributes are used to describe a listeners preference more than others ('*brilliance'* = +**1.9**), the average impact on preference may be much less than other attributes (nasal = **-2.56**).
- The alteration of frequency content from a vertical reflection *does* have a perceptual effect on what subjects hear and therefore, which descriptive terms they responded with.
- The removal of *specific* octave bands within a reflection for *particular* samples (where key frequency content in the sample corresponded to the frequency removed from the reflection), provided a much more concise response of subject's preference compared to playback with other removed octave bands. This was also supported by consistent use of certain descriptive attributes for that frequency dependant comparison.

CHAPTER 6

DISCUSSION

6.0 INTRODUCTION

The following chapter provides a discussion of both experiments conducted in this thesis. The comparison of results between both experiments will be investigated along with further discussion as to *why* these attributes may have been perceived and how this may be applied to the physical domain (Figure 1). Some of the limitations of this work have been mentioned, these will be discussed further along with the impact of this research.

6.1 INTER-EXPERIMENT OBSERVATIONS

Results from experiment one suggested that subjects may base preference in two different ways, depending on if their preference response was positive or negative (Section 4.2.3). No direct comparison can be drawn against experiment two, as perceived magnitude of timbral and spatial difference was not measured. However, results of preference did not infer a split above and below the '*no preference*' region as observed in experiment one. This suggests that listening in a more natural environment with more reflections, may reduce this divide. More research on subject's preferential split with musical stimuli would be required to support any hypothesis regarding this.

Highlighted in Section 4.3, experiment one outlined the possibility that subjects may base their preference on timbral characteristics, to a *greater* extent than perceived spatial attributes. The use of timbral descriptions that had positive and negative implications generally followed subject's level of preference. This can also be observed in experiment two, with timbral attributes being used to base listener preference for 311/480 paired comparisons. Whilst the use of spatial attributes such as "*vertical image shift*" and "*vertical spread*" were responded frequently in both experiments, the perception of these spatial sensations may have either: not impacted the listener to enough to influence preference (experiment one) or, impacted listener preference positively and negatively with no connection to increase or decrease of the sensation (experiment two). With the results of both these experiments, it is possible to suggest that the presence of a strong vertical reflection from above will impact subject's preference based on the timbral differences it provides, *greater* than the spatial differences.

Experiment one shows that the median preference value for playback with a reflection is generally above the '*no preference'* origin (5/6 comparisons). Interestingly, the opposite is observed in experiment two, showing that the median preference was generally *below* no preference, and only above for 12/48 comparisons. One possibility previously mentioned, is that a semi-anechoic chamber is not a *natural* environment and seldom used for listening to audio for entertainment. The inclusion of a single reflection in experiment one may therefore, always be perceived as beneficial placing the listener within a more natural environment. Section 2.1 discusses how studio design has aimed to *remove* us from a natural environment for critical listening, supports this suggestion. This therefore complies with a previous hypothesis from the results of experiment one (Section 4.3), that the inclusion of a vertical reflection in a semi-anechoic environment may have *consistently* provided a positive spatial contribution resulting in subjects basing their preference *more* on timbral qualities. Whereas in experiment two, listening conditions included a more natural sound field. The inclusion of a strong vertical reflection amongst this, seems to have contributed spatial attributes (*to a greater extent than previously*) to the cognitive process which subject's base their preference upon. This conclusion from both experiments *could* be applied in different ways depending on the room's current acoustic state. Implying that room's possessing a great amount of absorption resulting in minimal acoustical information to the listener, could benefit from the addition of a vertical reflection providing spatial attributes, and vice versa from a more '*live*' room, in the context of listening for entertainment reducing spatial attributes. However, as discussed in the following chapter, although a vertical reflection may increase spatial sensations, these attributes within these sensations may not always be beneficial.

6.2 INCREASED AND DECREASED AUDITORY SENSATIONS

6.2.1 SPATIAL ATTRIBUTES

An aim of experiment one, was to ascertain if any correlation can be found between perceived magnitude of change and subject's preference, between playback with and without a vertical reflection. Only when positive and negative preference values were seperated with respective perceived timbral and spatial values, could correlations be observed. However, experiment one *did* highlight that although people perceived a *spatial* change and described these changes throughout, the *magnitude* of these spatial changes may not matter. Experiment two on the other hand shows that increasing and decreasing *specific* spatial sensations ('*precision*' and '*distance*') could coincide with subject's preference. Also, that the response of some spatial attributes related to lateral energy is interesting to observe, given the specific monophonic layout of the experiment.

Consistent use of '*vertical image shift'* and '*vertical spread'* were observed throughout all samples in both experiments and remained present across the majority of subjects however, other spatial descriptions were also observed (Figure 38 – Section 5.2.1) such as *horizontal spread*. Monaural reproduction used in this experiment was to ensure that the listener processed no inter-aural time/level differences. However, auditory sensations generally provided by lateral energy were still given as a reason for preference. Increased '*horizontal spread'* was given for samples: C (Hi-Hats) and D (Orchestra), a total of eight times for the frequency *independent* test. Following this, experiment two using frequency *dependant* reflections also yielded spatial descriptions that would be associated with lateral reflections such as: increased/decreased '*horizontal spread*' (11 times) and increased '*frontal envelopment*' (12 times). Objective IACC measurements taken with a binaural head at the listener position for all frequency dependant tests showed that varying the frequency content from a vertical reflection, provided varying IACC values. It is acknowledged that given central speakers along the median plane and listener position, an IACC of 1 would be expected. However, room variances left and right of the listening position may account for discrepancies in IACC measurements such as distances from left to right walls, a listening position not symmetrically in line with permanently installed equipment and one sided computer monitors. Consequently, frequency dependant IACC measurements may have differed due to changing frequency content within reflections. Therefore, the only observation that can be made from this result is that a vertical reflection may provide properties that contribute to a sensory process (Figure 1), which makes us perceive greater '*horizontal spread'*.

The presence of a vertical reflection also provided spatial sensations such as increased/decreased *"precision"* and "*distance'*. The subject's perceived increased/decreased level of precision always followed the same direction as their positive and negative preference. The term '*precision*' was described to subjects as "*Individual instruments can be placed and separated within the sound image"* (Zacharov, 2015), therefore suggesting that the easier instruments are to localise, the more preferable the sound reproduction is. However in experiment two, playback with the simulated ceiling reflection was *not* always perceived to add precision by subjects, and therefore may not always be beneficial.

Finally for some frequency dependant comparisons, subject's additionally noted pitch and height seemed to be unbalanced, describing certain instruments containing low or high frequency content were moved to an incorrect vertical position. In particular, expressing that the kick drum (sample A) was too vertically high for its frequency content, and violins (sample D) *too* vertically separated from the rest of the ensemble. This may be due to the removal of particular octave bands in the reflection affecting subject's localisation through the '*pitch-height*' effect discussed in Section 2.5.2. Exploiting the '*pitch-height*' effect for localisation, Lee (2015) proposes a method for rendering vertical image spread known as '*Perceptual Band Allocation*' (PBA). Within this study, Lee demonstrates the non-linearity of the '*pitch-height*' effect from low to high frequency. In the context of this thesis, this means subjects who reported an unbalanced pitch-to-height, may have done so based on the additional frequency content of the reflection. This addition may have 'shifted' where specific instruments were localised due to their frequency content and that it may have only been perceived for particular instruments due to its non-linear behaviour. Supporting this may be the theory of directional bands discussed by Bleaurt (1997). His work showed that particular frequencies originating from loudspeakers *horizontally on-axis* to the listener may be localised at different angles to the listener along the median plane (Figure 14). Therefore, the presence/absence of particular frequencies within a reflection from above, may 'shift' the perceived origin of certain instruments containing specific frequency content. Bleaurt's work has since been refined (Wallis & Lee, 2015) however, the actual source location was still located horizontally on-axis to subjects. No work has been conducted regarding this directional band theory by adjusting source location along a radius of the median plane to see if this affects the localisation of frequencies.

Investigating what attribute would categorise this effect within the perceptual domain could be a topic for further study as subjects struggled to relate it to a provided attribute. An argument could be made that it regards precision – "*Individual instruments can be placed and separated within the sound image*", however subjects said that instruments *could* still be placed and separated, just the localisation of instruments *within* the image was exaggerated.

6.2.2 TIMBRAL ATTRIBUTES

Subject's use of timbral attributes throughout both experiments were very sample dependant. Experiment one demonstrated that the response of timbral attributes stated as positive or negative perceived change, corresponded with preference for 80% of comparisons for 10/13 subjects. This was then explored further in experiment two with the use of a more concise and controlled attribute list directly used to describe preference.

As shown in Figure 16 (Chapter 2.6.2), by Howard and Angus, (2013) it is possible that the perception of timbral characteristics ("*warm*" and "*sweet*") can be excited by both the increase of certain spectral content at one point, or the decrease of content at another. The verbalisation task in experiment one was a free response format allowing subjects to use any attribute to characterise their timbral perception[12]. Reflection content from experiment one was also frequency *independent* and therefore, subjects responses of timbre for the *same sample* was relatively consistent. Experiment two however, contained eight individual varying frequency *dependant* reflections of each sample. This resulted in subjects responding with auditory sensations that *could* be related to both the perceptual increase of frequencies at one spectral point, and/or the absence of other frequencies at another.

Playback of reflection with octave band 2kHz removed for stimuli B (guitar), D (orchestra) and E (piano) often resulted in subjects defining their preference with '*decreased brilliance*' and/or '*increased bass resonances*' both seen as negative. Given the descriptions of these (Appendix D), this suggests that the increase in one sensation and a decrease in another may actually be the perception of one phenomenon. These contrapuntal spectral descriptions may be the reason as to why descriptions were so varied for certain samples. As mentioned in Chapter 5, the intention of attenuating individual octaves, was in the context of an absorber with a high absorption coefficient across particular frequency points. It would be interesting to see if this effect was observed with the reverse of this experiment, with one specific target frequency was *reflected* and all others absorbed.

Finally, the use of an additional timbral descriptor '*phasey*' was noted for 29 of the comparisons made, which, given the addition of a delayed signal from a reflection, is not surprising. Whilst this was binaural listening (I.e. no ears were blocked during listening tests) each ear was receiving the same signal, therefore the summation of left and right may not have been able to disregard the comb filtering induced by the reflection (Blauert, 1997). This comb filtering may be the perceptual effect subjects were attempting to convey with the term '*phasey*', however with no description given for this attribute it is not possible to say that subjects are uniform in interpreting the attribute this way. Regardless of this, without the use of head shadowing (Toole, 2008) reducing any perceived comb filtering, it is likely that this timbral effect is noticeable to

[12] Although as stated, clarification of these descriptions were discussed between assessor and subject

a greater extent when induced by a vertical reflection. In addition to this, with 21 instances of subjects decreasing their preference by -2 (on average) (Table 10) when a '*phasey*' sensation is perceived, comb filtering along the median plane may be a negative effect.

6.3 THE PRECEDENCE EFFECT

Section 2.5.2 discussed the operation of the precedence effect along the vertical plane, evidence that a '*click*' signal with a specific time delays along the median plane *could* be located at the initial source (Hartman, 1993). However experiments by Lee (2011) demonstrate the precedence effect did not operate and sound sources were not fully located at the initial loudspeaker. Though these experiments are both different in terms of experimental set-up and context, they still assess vertical localisation and precedence effect. The reflection simulated in experiment's one and two had an ICTD of 1.48*ms* and 2.912*ms* respectively and therefore, lie within the fusion zone of the precedence effect (1ms – 5ms). As this context was assessing reflections, the delayed sound was attenuated by at least 2.5dB below the direct sound. For experiment one, subject's repeatedly responded with increased "*vertical image shift*", suggesting that the precedence effect did *not* operate, as the sound image was located above the direct loudspeaker with reflection playback. This is also observed in experiment two with a slightly weaker reflection. These results imply that vertical reflections impact our ability to fully localise a sound source at its origin and provides further evidence to suggest that the precedence effect *does not* operate along the median plane and requires ITD and ILD. More research on vertical reflections impact on the precedence effect is needed to further this hypothesis.

CHAPTER 7

THESIS SUMMARY

7.0 CONCLUSION

The aim of this thesis has been to investigate the perceptual effect of a vertical ceiling reflection on a listener's preference of monophonically reproduced sound. Specifically, to investigate: if subjects base their preference on timbral and/or spatial sensations, if any particular perceived attributes have the most influential effect in the perceptual domain and how these may be related to the physical domain (Figure 1).

Experiment one's paradigm was designed such that subject's level of preference could be mapped against the perceived magnitude of timbral and spatial difference. This magnitude of difference was not rated along a positive or negative rating scale, but rather from 'no perceived difference' to 'extreme perceived difference'. The descriptions provided by the subjects allowed the assessor to examine if these perceived differences were beneficial or detrimental by comparing them to their preference. The results implied that subjects based their preference on the timbral differences between the two playback cases to a greater extent than the spatial differences, in a semi-anechoic listening environment.

Following these results, experiment two was designed to explore these reflections further, using the subsequent adaptations: The assessment environment was in a more natural listening room complaint with ITU BS.1116 regulations, playback with and without the reflection were normalised to the same level, and frequency dependant reflections were played with removal of one of eight individual octave bands 125Hz – 16Khz.

The results of both experiments lead to the following novel conclusions:

- When expressing their preference with and without the presence of a reflection, the most frequently used attributes were timbral based.
- Whilst some timbral attributes were used infrequently, when these *are* perceived, they may have a greater impact on preference than other attributes more frequently used.

- The increase and decrease of timbral sensations with positive and negative connotations, all corresponded with individual subjects increased and decreased preference for playback with the reflection.
- Although the most commonly used, the increase of 'vertical image shift' and 'vertical spread' provided by a reflection cannot be used to increase preference.
- The removal of any of the individual octave bands from a vertical reflection does not *consistently* result in a more preferable listening environment than others across all musical stimuli and speech. However, the removal of *particular* octave bands in a reflection for *particular* stimuli, result in a more consistent level of preference along with a more consistent use of attributes to describe this sensation.

In addition to these results, the following suggestions can be made in the support of work in other fields of audio perception but require further investigation:

- The precedent effect may *not* operate along the median plane using musical or speech stimuli as '*vertical image shift'* was reported 131 times across both experiments, thus localisation was not fully at the direct loudspeaker sound source.
- The alteration of spectral content from a reflection, may be described in two contrapuntal ways. i.e. - A reduction in certain bass attributes, may also be perceived as an increase in certain treble attributes.

7.1 Limitations of this Research

One main consideration of this research has been to ground the results into applicable areas of audio reproduction. Unlike other research involving reflections, in these experiments, reflections were assessed in the context of listening for entertainment and using musical stimuli. Section 4.4 outlines the limitations of experiment one which are addressed in experiment two, such as, adapted rating scales and loudness discrepancy between playbacks with/without reflection.

However, it is acknowledged that the subject pool for both experiments was small, making it difficult to draw concrete conclusions, and may also not be representative of larger population sample. Increasing the subject pool may provide robust statistics, providing a more solid foundation to the conclusions. This increase in subject pool may also have resulted in normally distributed data and allowed for a different approach in statistical analysis.

Regardless of this, the author feels this thesis has contributed to knowledge into the cognitive process with which subjects base their preference of sound reproduction with the presence of a vertical reflection from above. The research was grounded in the applicability within small rooms which has become more relevant in recent years with growing complex sound reproduction systems readily available. This research may also provide an insight into the perception of height channels in multi-channel audio gaining more popularity. Although this simple vertical reflection was relatively strong with a low vertical boundary, the use of elevated sources in multi-channel audio (as well as being an elevated source) would in effect reproduce a stronger vertical reflection as the source moves closer to a reflection surface.

7.2 Further Research

The results from experiment one investigated the perceived magnitude of timbral and spatial differences, with a free verbalisation test describing these changes. The descriptive terms used in experiment two identified attributes at a level deeper regarding *what* timbral and spatial attributes could be heard with the addition of a frequency *dependant* vertical reflection and which ones may have a greater impact. Further study could investigate the level of difference heard from *each* individual comparison using direct attribute scaling and correlate it to subject's preference to ascertain which attributes are more preferable than others.

As discussed in the previous chapter, this thesis conducted experiments with a mono reproduction[13]. To progress results from this investigation and apply it to more commercial reproduction systems, vertical reflections should be assessed using stereophonic reproduction. The direct sound sources would then produce four reflections (one per ear, per source). Section 2.4.4 discusses work by Clark (1983) who demonstrates that the comb filtering provided through stereophonic reproduction can in some cases be a preferred response. Therefore inter-aural difference of reflections may also be preferable in terms of timbre.

Section 2.5 discussed the effect of elevated sound sources on subject's localisation both in terms of spectral content and the precedence effect. From the results of experiment two, the vertical reflection used provided us with information that resulted in the image shifting and sound source elevating and therefore showed us the precedence effect was not operating under conditions with a strong reflection from above. Further work should be conducted with varying

[13] Although stereo outputs were used to electro acoustically simulate a reflection, the direct sound was only produced through one channel.

reflection levels and delay times along the median plane to progress our understanding of vertical localisation regarding reflections impacting the precedence effect.

Reflections arriving from above were chosen for investigation in this thesis as this is where most acoustical treatment will be applied in small rooms. However reflections arriving from beneath us may still impact our preference. It is also apparent that sound reproduction systems are placed closer to the floor than the ceiling and therefore, will produce a stronger reflection which will contribute to our perception of timbre (Bech 1995). In conjunction with this, obstacles with surfaces may generally be placed in-between the reflection path and thus reduce delay time even more. Research into reflections arriving from a low angle could therefore, be important in understanding subject's preference along with timbral and spatial difference added. The correlation between pitch and height discussed in Chapter 2.7.2 show that localisation is impacted upon by frequency content, and that removal of frequency content alters our perception of spatial attributes (Chapter 5.2). Consequently a similar experiment conducted with floor reflections may yield interesting results for comparison.

RFERENCES

American National Standards Insitute. (1960). ANSI S1.1-1960 - Acoustical Terminology. American National Standards Institute, Inc.

American National Standards Insitute. (2010). ANSI S12.60-2010/Part 1 - Acoustical Performance Criterea, Design Requirements and Guidelines for Schools, Part 1: Permenant Schools. American National Standards Institute, Inc.

Ando, Y. (1977). Subjective preference in relation to objective parameters of music sound fields with a single echo. *The Journal of the Acoustical Society of America*, *62*(6), 1436–1441.

Angus, J. A. (1997). Controlling Early Reflections Using Diffusion. Presented at the Audio Engineering Society Convention 102, Audio Engineering Society.

Barron, M., & Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration*, *77*(2), 211–232. http://doi.org/10.1016/S0022-460X(81)80020-X

Bech, S. (1990). Electroacoustic Simulation of Listening Room Acoustics: Psychoacoustic Design Criteria. Presented at the Audio Engineering Society Convention 89, Audio Engineering Society.

Bech, S. (1994a). Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field. Presented at the Audio Engineering Society Convention 96, Audio Engineering Society.

Bech, S. (1994b). Perception of Timbre of Reproduced Sound in Small Rooms: Influence of Room and Loudspeaker Position. *Journal of the Audio Engineering Society*, *42*(12), 999–1007.

Bech, S. (1995a). Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field, II. Presented at the Audio Engineering Society Convention 99, Audio Engineering Society.

Bech, S. (1995b). Timbral aspects of reproduced sound in small rooms. I. *The Journal of the Acoustical Society of America*, *97*(3), 1717–1726. http://doi.org/10.1121/1.413047

Bech, S. (1996). Timbral aspects of reproduced sound in small rooms. II. *The Journal of the Acoustical Society of America*, *99*(6), 3539–3549. http://doi.org/10.1121/1.414952

Bech, S., & Zacharov, N. (2006). *Perceptual audio evaluation: theory, method and application*. Hoboken, NJ: John Wiley & Sons.

Beranek, L. L. (2008). Concert Hall Acoustics—2008. *Journal of the Audio Engineering Society*, *56*(7/8), 532–544.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization* (Rev. ed.). Cambridge, Mass: MIT Press.

Blesser, B., & Salter, L.-R. (2006). *Spaces Speak, Are You Listening?: Experiencing Aural Architecture*. MIT Press.

Bradley, J. ., & Soulodre, G. . (1995). Objective measures of listener envelopment. *Journal of the Acoustical Society of America*, *98*(5 I), 2590–2597. http://doi.org/10.1121/1.413225

Bradley, J. S., & Soulodre, G. A. (1995a). Subjective evaluation of new room acoustic measures. *The Journal of the Acoustical Society of America*, *98*(1), 294–301. http://doi.org/10.1121/1.413735

Bradley, J. S., & Soulodre, G. A. (1995b). The influence of late arriving energy on spatial impression. *The Journal of the Acoustical Society of America*, *97*(4), 2263–2271. http://doi.org/10.1121/1.411951

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound* (New edition edition). Cambridge, Mass.: MIT Press.

Brunner, S., Maempel, H.-J., & Weinzierl, S. (2007). On the Audibility of Comb Filter Distortions. Presented at the Audio Engineering Society Convention 122, Audio Engineering Society.

Clark, D. (1983). Measuring Audible Effects of Time Delays in Listening Rooms. Presented at the Audio Engineering Society Convention 74, Audio Engineering Society.

Cox, T. J., & D'Antonio, P. (2009). *Acoustic Absorbers and Diffusers: Theory, Design and Application*. CRC Press.

Davis, D. (1979). The Role of the Initial Time Delay Gap in the Acoustic Design of Control Rooms for Recording and Reinforcement Systems. Presented at the Audio Engineering Society Convention 64, Audio Engineering Society.

Davis, D., & Davis, C. (1980). The LEDE- Concept for the Control of Acoustic and Psychoacoustic Parameters in Recording Control Rooms. *Journal of the Audio Engineering Society*, *28*(9), 585–595.

Dunn, M., & Protheroe, D. (2014). Visualization of Early Reflections in Control Rooms. Presented at the Audio Engineering Society Convention 137, Audio Engineering Society.

Erickson, R. (1975). *Sound Structure in Music*. University of California Press.

Everest, F. A., & Pohlmann, K. C. (2009). *Master Handbook of Acoustics* (5 edition). New York: Tab Electronics.

Furuya, H., Fujimoto, K., Takechima, Y., & Nakamura, H. (1995). Effect of Early Reflections from Upside on Auditory Envelopment. *Journal of Acoustical Society of Japan (E).* Volume .16, (1995) No. 2 P97 - 104. http://doi.org/10.1250/ast.16.97

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033. http://doi.org/10.1121/1.382579

Griesinger, D. (1998). General Overview of Spatial Impression, Envelopment, Localization, and Externalization. Presented at the Audio Engineering Society Conference: 15th International Conference: Audio, Acoustics & Small Spaces, Audio Engineering Society.

Griesinger, D. (1999). Objective Measures of Spaciousness and Envelopment. Presented at the Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction, Audio Engineering Society.

Griesinger, D. (2009). The Importance of the Direct to Reverberant Ratio in the Perception of Distance, Localization, Clarity, and Envelopment. Presented at the Audio Engineering Society Convention 126, Audio Engineering Society.

Hartmann, W. M. (1993). Auditory Localization in Rooms. Presented at the Audio Engineering Society Conference: 12th International Conference: The Perception of Reproduced Sound, Audio Engineering Society.

Heyser, R. C. (1967). Acoustical Measurements by Time Delay Spectrometry. *Journal of the Audio Engineering Society*, *15*(4), 370–382.

Hidaka, T., Beranek, L. L., and Okano, T. (1997).) Some considerations of Interaural Cross Correlation and Lateral Fraction as Measures of Spaciousness and Concert Halls. Chapter 32 in Music and Concert Hall Acoustics, Ando,Y., and Noson, D. Academic Press, London.

Howard, D., & Angus, J. (2013). *Acoustics and Psychoacoustics*. Taylor & Francis.

Huang, P., & Abel, J. S. (2007). Aspects of Reverberation Echo Density. Presented at the Audio Engineering Society Convention 123, Audio Engineering Society.

Huang, P., Abel, J. S., Terasawa, H., & Berger, J. (2008). Reverberation Echo Density Psychoacoustics. Presented at the Audio Engineering Society Convention 125, Audio Engineering Society.

Imamura, H., Marui, A., Kamekawa, T., & Nakahara, M. (2013). Influence of First Reflections in Listening Room on Subjective Listener Impression of Reproduced Sound. Presented at the Audio Engineering Society Convention 134, Audio Engineering Society.

Imamura, H., Marui, A., Kamekawa, T., & Nakahara, M. (2014). Influence of Directional Differences of First Reflections in Small Spaces on Perceived Clarity. Presented at the Audio Engineering Society Convention 136, Audio Engineering Society.

International Organisation for Standards. (1994). ISO 8586-2:1994 - Sensory Analysis - General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors.

International Organisation for Standards. (2009a). ISO 3382-1:2009: Acoustics. Measurement of room acoustic parameters. Performance spaces, 36.

International Organisation for Standards. (2009b). ISO 3382-2:2009: Acoustics. Measurement of room acoustic parameters. Reverberation time in ordinary rooms, 36.

International Telecommunications Union Radiocommunication Assembley. (1997). ITU-R BS.1116-1:1997 - Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Telecommunications Union.

International Telecommunications Union Radiocommunication Assembley. (1998). ITU-R BS.6840-13:1998 - Sound system equipment. Listening tests on loudspeakers. International Telecommunications Union.

International Telecommunications Union Radiocommunication Assembley. (2015). ITU-R BS.1534-3:2015 - Method for the subjective assessment of intermediate quality levels of audio systems. International Telecommunications Union.

International Telecommunications Union, Telecommunications Standardization Sector. (1998). ITU-R P.800:1996 - Methods for subjective determination of transmission quality. International Telecommunications Union.

Kaplanis, N., Bech, S., Jensen, S. H., & van Waterschoot, T. (2014). Perception of Reverberation in Small Rooms: A Literature Study. Presented at the Audio Engineering Society Conference: 55th International Conference: Spatial Audio, Audio Engineering Society.

Katz, B. (2007). *Mastering Audio: The Art and the Science* (2 edition). New York: Focal Press.

King, R., Leonard, B., & Sikora, G. (2011). The Practical Effects of Lateral Energy in Critical Listening Environments. Presented at the Audio Engineering Society Convention 131, Audio Engineering Society.

Kishinaga, S., Shimizu, Y., Ando, S., & Yamaguchi, K. (1979). On the Room Acoustic Design of Listening Rooms. Presented at the Audio Engineering Society Convention 64, Audio Engineering Society.

Kuusinen, A., Pätynen, J., Tervo, S., & Lokki, T. (2014). Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls. *The Journal of the Acoustical Society of America*, *135*(1), 239–250. http://doi.org/10.1121/1.4836335

Lee, H. (2011). The Relationship Between Interchannel Time and Level Differences in Vertical Sound Localization and Masking. Presented at the Audio Engineering Society Convention 131, Audio Engineering Society.

Lee, H. (2015). Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array. Presented at the Audio Engineering Society Convention 138, Audio Engineering Society.

Lehnert, H. (1993). Auditory Spatial Impression. Presented at the Audio Engineering Society Conference: 12th International Conference: The Perception of Reproduced Sound, Audio Engineering Society.

Leonard, B., King, R., & Sikora, G. (2012). The Effect of Acoustic Environment on Reverberation Level Preference. Presented at the Audio Engineering Society Convention 133, Audio Engineering Society.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, *106*(4 Pt 1), 1633–1654. http://doi.org/10.1121/1.427914

Milner, G. (2010). *Perfecting Sound Forever: The Story of Recorded Music*. Granta Books.

Morimoto, M., & Iida, K. (2005). Appropriate frequency bandwidth in measuring interaural cross-correlation as a physical measure of auditory source width. *Acoustical Science and Technology*, *26*(2), 179–184. http://doi.org/10.1250/ast.26.179

Morimoto, M., Jinya, M., & Nakagawa, K. (2007). Effects of frequency characteristics of reverberation time on listener envelopment. *The Journal of the Acoustical Society of America*, *122*(3), 1611–1615. http://doi.org/10.1121/1.2756164

Newell, P. (2007). *Recording Studio Design* (2 edition). Oxford: Focal Press.

Niaounakis, T. I., & Davies, W. J. (2002). Perception of Reverberation Time in Small Listening Rooms. *Journal of the Audio Engineering Society*, *50*(5), 343–350.

Olive, S. E., & Toole, F. E. (1988). The Detection of Reflections in Typical Rooms. Presented at the Audio Engineering Society Convention 85, Audio Engineering Society.

Otto, N. C. (1997). Listening Test Methods for Automotive Sound Quality. Presented at the Audio Engineering Society Convention 103, Audio Engineering Society.

Ouis, D. (2003). Study on the Relationship between Some Room Acoustical Descriptors. *Journal of the Audio Engineering Society*, *51*(6), 518–533.

Pedersen, T. H. (2008). The Semantic Space of Sounds. *Lexicon of Sound-Describing Words-Version*, *1*.

Pedersen, T. H., & Zacharov, N. (2015). The Development of a Sound Wheel for Reproduced Sound. Presented at the Audio Engineering Society Convention 138, Audio Engineering Society.

Pike, C., Brookes, T., & Mason, R. (2013). Auditory Adaptation to Loudspeakers and Listening Room Acoustics. Presented at the Audio Engineering Society Convention 135, Audio Engineering Society.

Pike, C., Mason, R., & Brookes, T. (2014). The Effect of Auditory Memory on the Perception of Timbre. Presented at the Audio Engineering Society Convention 136, Audio Engineering Society.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*(3), 278–285. http://doi.org/10.1037/h0072651

Reiley, E., & Grimani, A. (2003). Room Mode Bass Absorption Through Combined Diaphragmatic & Helmholtz Resonance Techniques: ‘The Springzorber’. Presented at the Audio Engineering Society Convention 114, Audio Engineering Society.

Roffler, S. K., & Butler, R. A. (1968a). Factors That Influence the Localization of Sound in the Vertical Plane. *The Journal of the Acoustical Society of America*, *43*(6), 1255–1259. http://doi.org/10.1121/1.1910976

Roffler, S. K., & Butler, R. A. (1968b). Localization of Tonal Stimuli in the Vertical Plane. *The Journal of the Acoustical Society of America*, *43*(6), 1260–1266. http://doi.org/10.1121/1.1910977

Rubak, P., & Johansen, L. G. (2003). Coloration in Natural and Artificial Room Impulse Responses. Presented at the Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction, Audio Engineering Society.

Seetharaman, P., & Tarzia, S. P. (2012). The Hand Clap as an Impulse Source for Measuring Room Acoustics. Presented at the Audio Engineering Society Convention 132, Audio Engineering Society.

Schouten, J. F. (1968). The Perception of Timbre. Reports of the 6th International Congress on Acoustics. (Vol 76. p10)

Staffeldt, H. (1974). Correlation Between Subjective and Objective Data for Quality Loudspeakers. Presented at the Audio Engineering Society Convention 47, Audio Engineering Society.

Stephenson, M. (2012). Assessing the quality of low frequency audio reproduction in critical listening spaces. Doctoral Thesis, University of Salford

Toole, F. E. (2006). Loudspeakers and Rooms for Sound Reproduction—A Scientific Review. *Journal of the Audio Engineering Society*, *54*(6), 451–476.

Toole, F. E. (2008). *Sound Reproduction: Loudspeakers and Rooms*. Taylor & Francis.

Toole, F. E., & Olive, S. (1994). Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things. Presented at the Audio Engineering Society Convention 97, Audio Engineering Society.

Vickers, E. (2010). The Loudness War: Background, Speculation, and Recommendations. Presented at the Audio Engineering Society Convention 129, Audio Engineering Society.

Voetmann, J. (2007). 50 Years of Sound Control Room Design. Presented at the Audio Engineering Society Convention 122, Audio Engineering Society.

Wallis, R., & Lee, H. (2015). Directional Bands Revisited. Presented at the Audio Engineering Society Convention 138, Audio Engineering Society.

Wickelmaier, F., Umbach, N., Sering, K., & Choisel, S. (2009). Comparing Three Methods for Sound Quality Evaluation with Respect to Speed and Accuracy. Presented at the Audio Engineering Society Convention 126, Audio Engineering Society.

Zacharov, N., & Huopaniemi, J. (1999). Results of a Round Robin Subjective Evaluation of Virtual Home Theatre Sound Systems. Presented at the Audio Engineering Society Convention 107, Audio Engineering Society.

Zacharov, N., & Lorho, G. (2004). Subjective Evaluation of Virtual Home Theatre Sound Systems for Loudspeakers and Headphones. Presented at the Audio Engineering Society Convention 116, Audio Engineering Society.

Zacharov, N., & Mattila, V.-V. (2001). GLS - A generalised listener selection procedure. Presented at the Audio Engineering Society Convention 110, Audio Engineering Society.

Zielinski, S., Brooks, P., & Rumsey, F. (2007). On the Use of Graphic Scales in Modern Listening Tests. Presented at the Audio Engineering Society Convention 123, Audio Engineering Society.

Zielinski, S. K., Rumsey, F., Kassier, R., & Bech, S. (2005). Comparison of Basic Audio Quality and Timbral and Spatial Fidelity Changes Caused by Limitation of Bandwidth and by Down-mix Algorithms in 5.1 Surround Audio Systems. *Journal of the Audio Engineering Society*, *53*(3), 174–192.

APPENDICIES

A. SAMPLE DATA

FIGURE 47: SAMPLE A - DRUMS FFT 0HZ - 1600HZ

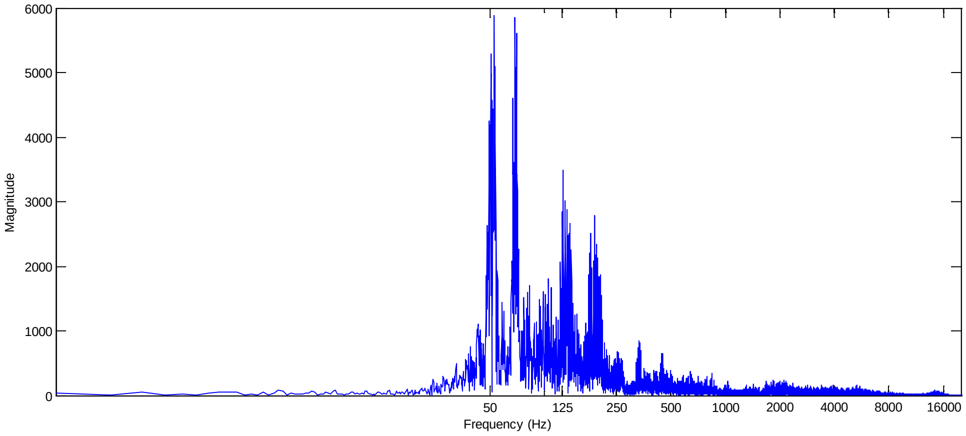


FIGURE 48: SAMPLE B - GUITAR FFT 50HZ – 1600HZ

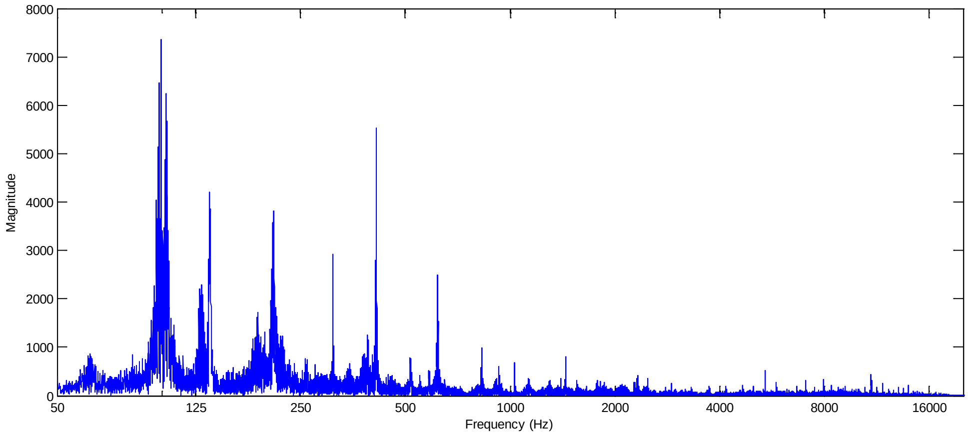


FIGURE 49: SAMPLE C - HI-HATS FFT 50HZ – 16000HZ

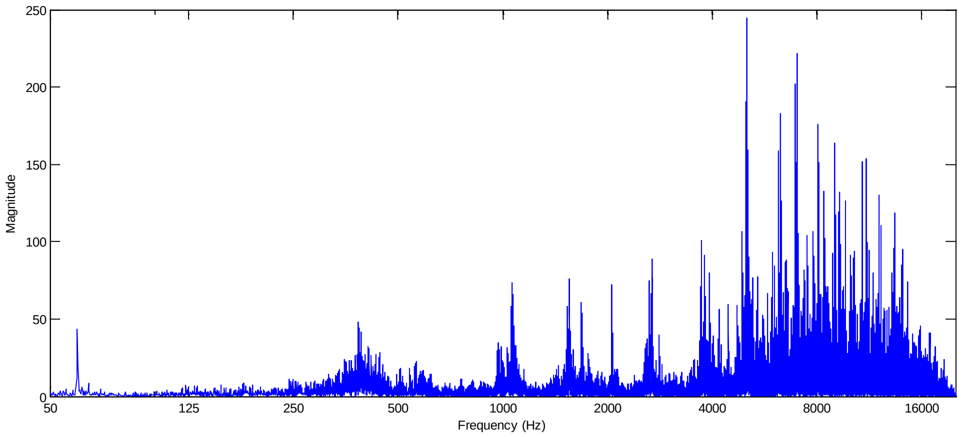


FIGURE 50: SAMPLE D - ORCHESTRA FFT 0HZ – 16000HZ

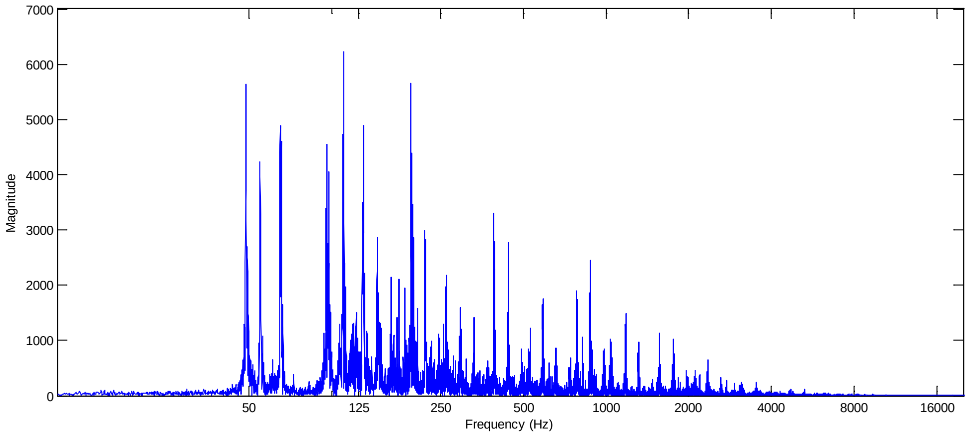


FIGURE 51: SAMPLE E - PIANO FFT 50HZ – 16000HZ

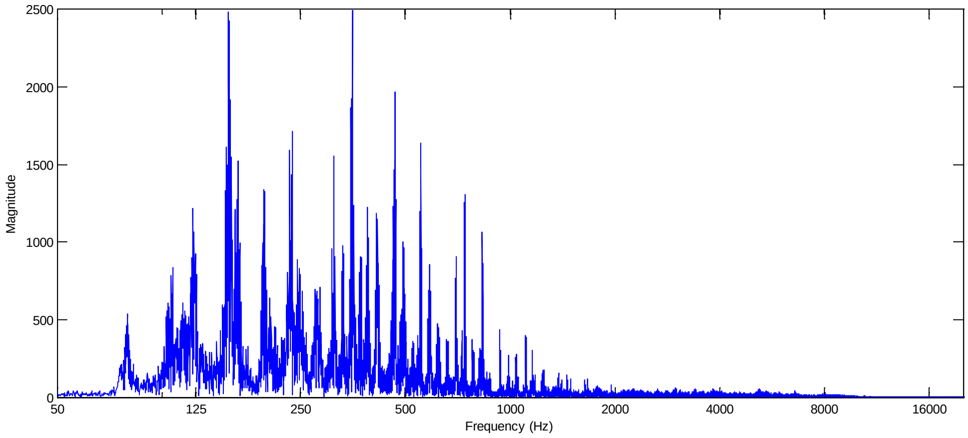
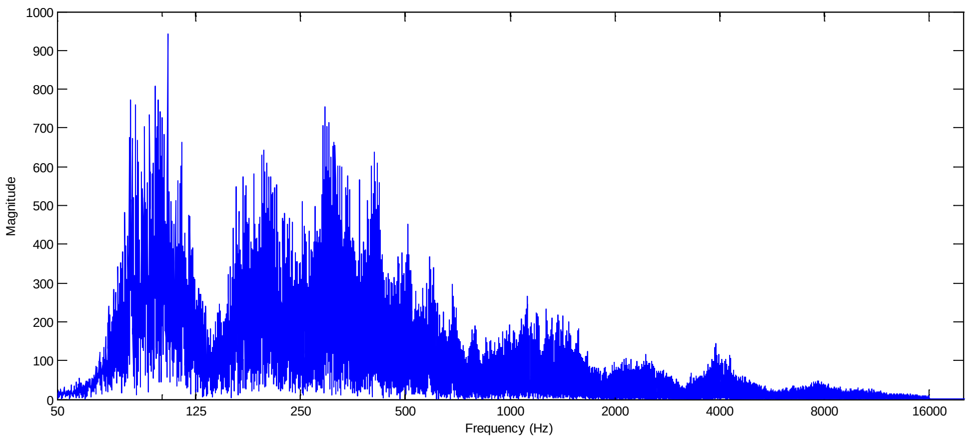
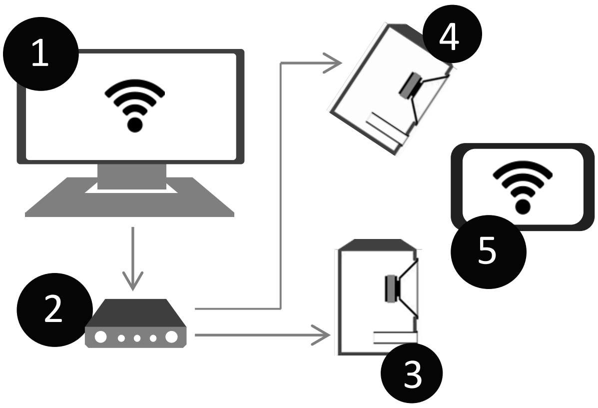


FIGURE 52: SAMPLE F - SPEECH FFT 50HZ – 16000HZ



B. Equipment & Interface Operation



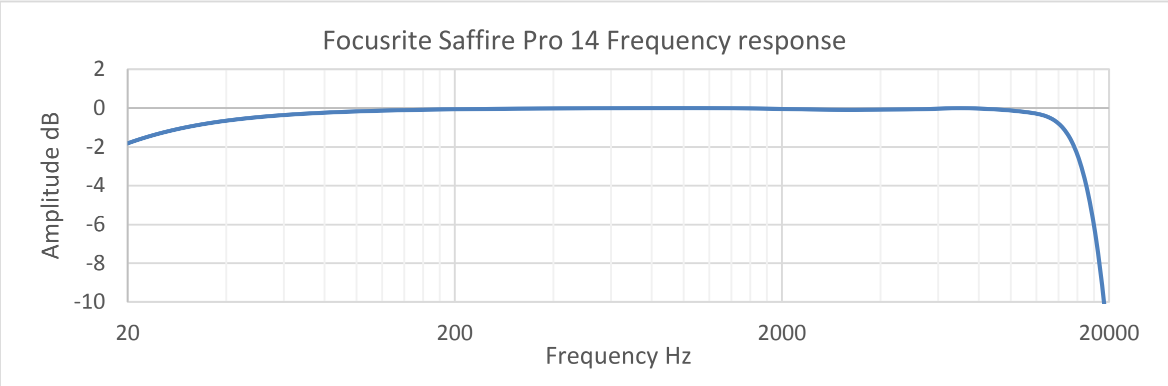
1 – Mac running MAX MSP programmed patch to wirelessly send and receive data from interface.

2 – Focusrite Saffire Pro 14 firewire soundcard.

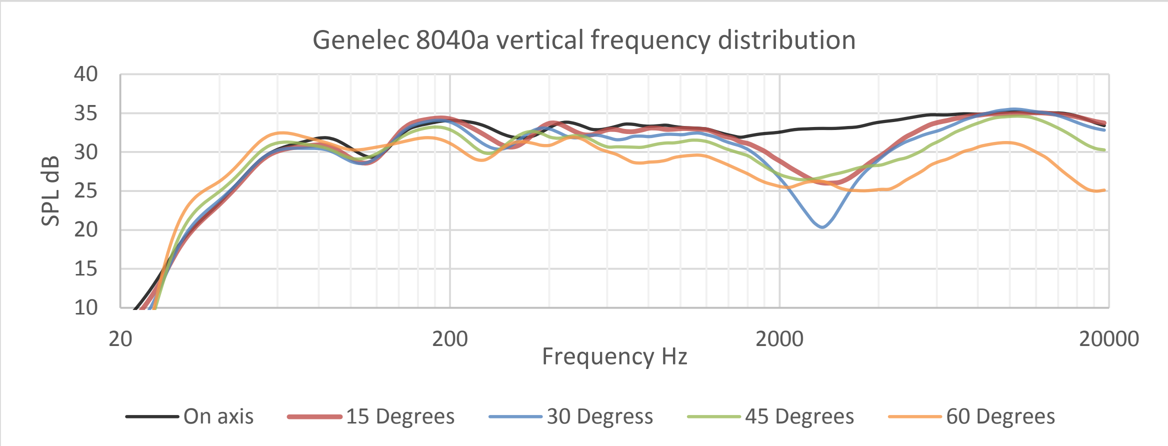
3 – Genelec 8040a Direct Loudspeaker.

4 – Genelec 8040a Reflection Loudspeaker

5 – Android table running Touch OCS software wirelessly linked to MAX MSP



Focusrite frequency response



Genelec vertical dispersion frequency responses measured at 1m radius

C. EXPERIMENTS ONE – SUBJECTS PROVIDED DESCRIPTIVE TERMS

Subjects were provided with a 'prompt sheet' with listings of popular descriptive terms for timbre used throughout audio assessment. As stated in Section 4.1.4, the description of timbre given to subjects was consistent with that of the ANSI standard, and that these terms were only used as a prompt to help. If the subjects felt a word better suited their auditory sensation then this may be used. The spatial information provided, were brief descriptions of sensations the author may feel be heard. Again, subjects were free to use other sensations as long as clarification could be given as to what this meant.

Timbral Definition

Timbre – *The description of timbre is the perceived difference between two similarly presented sounds of equal loudness and pitch sounding dissimilar.*

Timbral Descriptors

Bright	Rich	Nasal	Fat	Full	Tight
Muddy	Warm	Middy	Thin	Boxy	Harsh

Other: ____

Spatial Descriptors (Adapted from: Pederson & Zacharov (2015) and Kaplanis et. al (2014)

Image Shift – The sensation of the original image moving in location

Envelopment – The sensation of being in the middle of the sound rather than the sound arriving from a specific direction

Presence – A sense of space within the environment, as opposed to looking on through a window and no impression of being within the ambience

Other: ____

D. EXPERIMENT TWO – SUBJECTS RESPONSE SHEET

Timbral balance	Treble Strength	Covered, Un-sharp
		A ***soft*** sound without being dull
		Clearly distinguishable instruments
		Raised treble, ***sharp***, ***hard*** sounding
	Midrange Strength	Relative level of mid frequencies, ***low/high/balanced***
	Bass Strength	Relative level of bass, ***low/high/balanced***
	Fullness	Both low and high frequencies well represented with good extension
Treble	Brilliance (high frequency extension)	***Muffled***, blurred or dull
		Extended treble range, ***airy***, and ***open*** treble. ***Lightness***, ***purity*** and clarity with space for instruments Clear without being sharp or shrill.
	Tinny	Resonances or narrowband frequency prominence in the treble or high frequencies.
Mid	Nasal	A closed sound with pronounced midrange.
	Canny	Being played through a can or tube Prominent narrowband resonances in the midrange
Bass	Bass Depth	Low frequency extension
	Boomy	Prominent resonances in the low bass Tends to become muddy and imprecise
	Boxy	Hollow sounding Resonances in the upper midrange

Sound image	Distance	Perceived distance between listener and sound sources
	Vertical/Horizontal width	Vertical width of the sound image Greater spread across axis
	Depth	The depth of the sound source (Not to be confused with distance)
Localisation	Precise	Individual instruments can be placed and separated within the sound image.
	Envelopment	Are you surrounded by the reproduced sound? Does it give a space of sense around you? Frontal, Rear or total envelopment?
	Shift	Does the localisation of the sound source shift (Not to be confused with spread or width)

Dynamics	Punch	Strokes on drums or bass are reproduced with 'clout' as you can feel the blow.

Transparency	Presence	Is the sound present, distant or absent?
	Detailed	Sound rich in detail Details that cannot be measured that seem to give to music 'soul' or 'feel' such as audible nuances.
	Natural	Sound is reproduced with high fidelity Representative of the real sound without and timbral/spatial colouration or distortion.