Assessing the Quality of Low Frequency Audio Reproduction in Critical Listening Spaces

Matthew Stephenson

School of Computing, Science, & Engineering
University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy, May 2012

Abstract

The quality of low frequency audio reproduction in small spaces has always been problematic. For some time, methods have been suggested in order to optimise this reproduction. Many such methods have been based upon objective metrics which remain unproven from a subjective perspective. Whilst perception has been studied, this thesis identifies a research gap for more thorough testing.

A series of listening tests has been conducted, with virtual rooms auralised and presented over headphones in order to isolate specific modal parameters and allow efficient collection of subjective response from many listening environments.

The work presented searches for optimal values and perceptual thresholds of three parameters - modal spacing, density and decay. Results show that optimal spacings and densities may only be defined where assumptions are made which are not valid in realistic listening spaces.

Thresholds of modal decay[1] have been defined, which are considered valid regardless of stimuli or replay level. These are around 0.2 seconds for frequencies above 100Hz, and increase sharply below this point to around 0.85 seconds at 32Hz.

Through the testing of these parameters, it is shown that whilst discrimination between two rooms is usually a simple task, this does not reveal the underlying reproduction quality. The perceived quality of the room response is of great importance, and new experiments assess this quality using a paired comparison method which provides a simpler subjective task than direct scaling methods. A set of descriptors is elicited which can be used to evaluate low frequency audio. These descriptors - articulation, resonance and bass energy - are used to assess the impact of three room parameters on perceived reproduction quality. Room response metrics are also evaluated for perceived quality. Results reveal that modal decay is a primary indicator of quality, with shorter decays producing rooms with a higher perceived quality of reproduction.

[1] If a mode has a decay time less than a particular threshold, the individual effect of that mode on overall audio quality will be imperceptible.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I thank God for his constant love and faithfulness.

My wife Rachel (there is plenty of time for belly laughs now), and my parents for their support for many years.

Bruno Fazenda, for inspiration, encouragement and never being too busy to assist.

All my family and friends, too numerous to mention.

And finally, Keri Jones for his wisdom, gentle application of pressure and financial support.

Thank you.

The University of Huddersfield is acknowledged for the funding of this research project between September 2007 and 2009

Chapter 1

Introduction

1.1 Introduction

This chapter introduces the topic of research for this thesis - namely, the perception of sound reproduction at low frequencies in critical listening spaces. The research problem is introduced and a set of aims described. An overview of the methodology used in order to meet these aims is presented. The thesis offers a significant contribution to the current knowledge, and as such, the main findings are highlighted. Finally, the structure of this thesis is presented, allowing the reader to locate relevant chapters of interest.

1.2 Perception in Critical Listening Spaces

There are a number of situations where the correct reproduction of audio is critical. Examples are recording or mastering studios, and increasingly, within the context of a home theatre. Either critical decisions are to be made about audio content in these spaces, or a particularly high level of reproduction is required for greater enjoyment. Where a critical decision is required in a studio situation, the ability of the space to accurately reproduce audio is likely to impact directly upon the quality of the final recording. Much research has gone into the reproduction of audio in such spaces. Broadly speaking, such research has taken two forms. Firstly, seeking the improvement of reproduction equipment, such as a digital playback device or one or more loudspeakers. Secondly, the physical design of the space itself. It is this 'critical listening space' which is the focus of this thesis - and more specifically the *perception* of audio quality within the space. The literature review in the following chapter reveals much theory and a number of room optimisation techniques. However, no amount of theory as to why a room should be perceived as a 'good reproduction environment' will suffice if actually listening within it reveals that it is not.

As a result, researchers have been increasingly interested in understanding the aspects of room design which may affect the perceived quality of reproduction. It is important to clarify what is meant by potentially abstract terms such as 'accurate', 'ideal' and 'optimal' reproduction. An 'accurate' reproduction is defined as the ability to reproduce audio exactly (or as close as possible) to the original source. The signal received by the listener(s) should be a faithful representation of that fed to the playback device. However, a space will almost always exert its own specific influence upon the sound, and in certain cases this influence is actually considered desirable. Where a room does not have an influence on the source signal, such as an anechoic chamber, listeners often feel the space is unnatural and the reproduction undesirable. In other cases - for example, the opposite of the anechoic chamber, the reverberation chamber - the influence is also undesirable. Therefore, the term, 'ideal', refers to the case where the reproduction performance of the room is considered most desirable. Such a definition introduces a *subjective* element - what constitutes the most desirable? If there is no subjective agreement, any attempt to define the ideal would be fruitless. However, Toole (2008) argues that it a misconception that audio quality is a purely subjective quantity, and by treating it as such, we rule out the possibility of there being certain universal objective measures which could define the ideal reproduction. It is believed that such measures may also be defined for the subjective quality of listening spaces at low frequency. Therefore, the term 'optimal' is used in this thesis to define a measure or percept which represents the best case scenario for the majority of listeners. The thesis attempts to draw links between an objective understanding of the space and these optimal subjective measures.

1.3 Low Frequency Reproduction

Through extensive research, many challenging aspects in designing and optimising a space are well understood. For example, the concept of reverberation and reflections within a room, which can be controlled and treated with absorption and diffusion techniques, is well known. Popular recording studio designs implement many of the techniques suggested and see measurable and perceptible improvements (Newell, 2007). However, critical listening spaces are more often than not small in size. Due to the relationship between low frequency wavelengths and room dimensions, room resonances, known as *modes* are set up. These modes can cause audible problems in the sound field. Therefore, the low frequency region, typically below 250Hz, still presents a very real problem in many critical listening spaces (Fazenda and Davies, 2004). It is the perception of these problems which forms the basis of this thesis.

1.4 Research Problem

It is well known that within a room, the high frequency response can be considered a function of loudspeaker performance and absorption, while the low frequency response is highly related to the room geometry and position of source and receiver (Groh, 1974).

Even with the previous research into the perception of room modes (see Chapter 2), there remains a lack of subjectively relevant understanding in directly assessing the perceived quality of a listening space. Usable guidelines are also required for both room designers and those seeking to optimise existing facilities.

The *theory* of these room modes has been well defined, but the relative importance of different modal parameters, the extent to which we perceive the problems caused by each parameter and the overall quality of a room's reproduction are all less well understood.

1.5 Scope of This Thesis

This research relates to the *perception of audio reproduction at low frequencies in critical listening spaces*. Such a space is formally defined as "a space where an audio program is listened to in a way that allows the listener to evaluate and interpret its characteristics in depth and make decisions regarding any problematic features". Low frequency is generally considered to be below 250Hz throughout this thesis. Perception is investigated using modelled rooms presented over headphones. As such, the scope of this research does not include perception through loudspeakers within a real room, although it is acknowledged that such listening conditions may have an impact upon the perceived quality, due to factors such as whole body vibration. The main focus for listening tests is the assessment of typical program material such as popular music, within the context of a realistically modelled room. A number of investigations consider test tone stimuli and single resonances, although this is always with the intention of comparing these to similar results using musical stimuli. Speech transmission/reproduction is not considered here.

1.6 Aims of Research

Within the scope outlined, the aim of this research is to achieve a greater understanding of the perception of low frequencies in small rooms and to evaluate effective control of the sound field. In order to achieve this aim, the resulting objectives have been defined:

1. To determine *optimal values* for key modal parameters relating to the frequency distribution of modes.

2. To determine *perceptual thresholds* for low frequency room parameters such as modal decay.

3. To better understand and map the *relationship* between *room parameters*, *low frequency response* and *perceived quality*.

1.7 Methodology

The above aims are met through a series of subjective listening tests. These tests query groups of listeners regarding a set of stimuli to determine the optimal values, relevant thresholds or overall perceived quality. To facilitate this process, the low frequency sound-field is modelled and audio stimuli are processed through this model ready for playback over headphones. Various testing methods are then employed to ensure appropriate results can be obtained for the variable(s) under study.

1.7.1 Low Frequency Modelling

In order to overcome the difficulty of testing in a wide variety of spaces, and to retain the ability to isolate single modal parameters, the rooms under test are most often modelled using an auralisation technique, and replayed over high quality headphones. Similar methodology has been used previously (Fazenda et al., 2005; Weisser and Rindel, 2006; Avis et al., 2007; Stefanakis et al., 2008a). This work extends the methodology and modifies it where appropriate in order to meet the needs of the testing requirements for each experiment.

1.7.2 Testing Methods

Many test methodologies have been used by researchers in order to obtain the most relevant results. The exact choice of method is naturally dependent upon the individual experiment and associated data requirements. The work presented in the following chapters uses common methods such as the ABX comparison and Parameter Estimation by Sequential Testing (PEST), in addition to direct scaling methods such as MUSHRA (ITU, 1994). In later chapters, multidimensional scaling methods are also explored.

A number of statistical analysis techniques are also implemented. Again, techniques were selected on a test by test basis in order to extract the most pertinent results from the datasets. Examples include the common mean and standard deviation as well as ANOVA, Chi-Square, Principle Component Analysis and correlation techniques.

1.8 Contributions of the Author

The thesis presents a number of novel contributions to the field of study.

Firstly, an optimal subjective modal spacing is presented. Previous studies have attempted to optimise the spacing based on metrics such as an even distribution of modes. In this thesis, a new approach is taken - that of defining an optimal value to maintain the best quality of audio. This is shown to relate directly to the bandwidth of the spaced modes in question, rather than an absolute value.

Secondly, thresholds of modal decay are presented. A full investigation has been made into these thresholds, which incorporates new test methods, and provides evidence that stated thresholds are applicable not only for the worst case scenario of a single frequency decay, but remain valid within the context of a musical stimuli in a real listening space.

Thirdly, a set of low frequency descriptive terms have been produced and tested for validity. Such a term set is absent from literature focussing on low frequency perception. As the benefits of such term sets have been demonstrated in other areas of perceptual audio, such as spatial reproduction, it is seen as a great benefit to have a term set which can be used for low frequency quality assessment, not only in room acoustics but also other fields such as loudspeaker quality. Indeed, the terms have subsequently been used in a number of pieces of research conducted at the University of Salford.

Finally, the discussion in the final chapter provides a valuable contribution, highlighting a number of suggestions for ensuring the highest possible quality of low frequency reproduction, and defining perceptual quality as a combination of both ‘critical’ and ‘affective’ responses.

1.9 Thesis Overview

The thesis may be considered in three main parts. Firstly, Chapters 2 and 3 detail the research, background theory and details the auralisation method. Secondly, experimental work concerning the determination of optimal values and thresholds for individual modal parameters is presented in Chapters 4 and 5. The question of overall subjective audio quality is then addressed, with an experimental process undertaken in order to relate perceived quality to the objective parameters previously investigated. To conclude, the work is discussed in the context of the aims above, conclusions are drawn and further work suggested. A brief summary of each chapter is now presented.

A review of the relevant literature is presented in Chapter 2. This provides an overview of the wide body of research, covering early theory, suggested control methods, the perception of resonances and finally research conducted into the evaluation of audio quality. Throughout this process, a number of research gaps are identified which are addressed by this thesis.

Chapter 3 highlights the theory of low frequency sound-fields and their virtual reproduction. This includes the basic theory of room modes and builds on this to explain how rooms may be computationally modelled. The chapter focusses on the ‘modal decomposition’ model which is used extensively as the basis for listening tests presented in subsequent chapters. An explanation of the auralisation method used throughout the thesis to create ‘virtual rooms’ is given, focussing on high and low frequency alignment, headphone equalisation, loudness calibration and the suitability of audio test stimuli. Finally, an introduction into the perceptual framework which underpins the narrative of this thesis concludes the chapter.

Chapter 4 presents an investigation into the optimal modal spacing for the simplified scenario of two spaced resonances. Modal spacing is an underlying parameter of the modal distribution which in turn underpins much of the published research into room optimisation by aspect ratio (Bolt, 1939a; Louden, 1971; Walker, 1996). Results show that where the ‘optimal’ is defined as the spacing required to produce the shortest decay without the onset of audible beating, a result proportional to the modal bandwidth is observed. Furthermore, the chapter explores the possible validity of extrapolating these results to cases where a greater number of resonances exist. Here it is shown that although an optimal spacing can be found for simple cases such as two discrete resonances, such a method breaks down in more realistic scenarios. This is particularly important when considering both larger rooms, and higher frequencies, where a greater number of modes occur.

For this reason, Chapter 5 considers another commonly quoted parameter, also related to the distribution of modes - the modal density. It is widely implied that a higher density leads to a subjective improvement (Bonello, 1981; Everest, 2001; Howard and Angus, 2001), and yet no experimental results prove that this is the case. This chapter's investigation into the modal density includes three listening tests. The first considers an ideal scenario where both the source and receiver couple positively to each mode and with maximum magnitude. The second test highlights the difficulty in employing a similar methodology to realistic scenarios. The test reveals that when source and receiver position are accounted for, no specific threshold of modal density can be found. Rather, each room response can be distinguished from another unless the room conditions are very similar. As a result, a third exploratory test investigates whether the perceived reproduction *quality* increases with density.

With these two chapters highlighting the importance of an objective method more closely related to perceived quality, a third modal parameter was investigated - modal decay. This is perhaps the most intuitively satisfying parameter to determine a threshold for, as modal artefacts in rooms are often described as 'low frequency reverb', 'ringing' or 'decay'. Again a number of listening tests were conducted, firstly to determine an 'absolute' threshold, through the use of single frequency test tones, and secondly, thresholds using music stimuli and modelled room responses. Of particular interest is that the results reveal similar thresholds for both tones and music, suggesting that the threshold can be considered valid regardless of input material. Reduction of decay below this threshold is unlikely to be perceptible.

Whilst these results provide useful information regarding room optimisation, specifically in terms of decay thresholds, the conclusions drawn from the three chapters are that a) for a subjective assessment to be informative, realistic scenarios must be modelled, and preferably, musical samples used as test stimuli and b) it is the overall perception of low frequency in relation to a specific room response which is of the highest value in assessing the quality of reproduction. Listener responses may then be investigated in an attempt to determine features of the response which may account for their assessment. The remainder of the thesis therefore investigates this perceived quality. In order to do so effectively, Chapter 7 identifies a set of four descriptive terms by which a listener can rate the low frequency reproduction. These are determined using a consensus vocabulary approach known as 'descriptive analysis', and provide the subject with a set of rating scales more obviously linked with objective measurements, resulting in a simpler task when rating test stimuli.

Chapter 8 presents a listening test two-fold in purpose. Firstly, it is used to analyse and validate the descriptive terms elicited in the previous chapter. Secondly, it uses these terms and a paired comparison method to study the mapping of auditory sensations to both perceived quality and to three room parameters (decay, volume and source/receiver position).

With Chapter 8 investigating the relationship of quality and *room parameters*, Chapter 9 builds upon this technique to investigate the mappings between the quality and *response parameters* (measures calculated directly from the impulse or frequency response). These responses were chosen with careful consideration, based on the previous research, but without relation to any specific room parameter. As such, the findings of this chapter can be applied to any measured room response.

Finally, Chapter 10 reviews and discusses the main findings from the experimental work presented throughout the thesis and suggests a number of further studies.

Part I

Research and Theory

Chapter 2

Literature Review

2.1 Introduction

Chapter 1 presented an overview of this thesis, and the essential background to a study in the perception of low frequency reproduction quality in small rooms. Chapter 2 now focuses on the relevant literature which exists in this research area and is broken down into a number of sections.

Section 2.2 covers the necessary theory for studying the physical aspects of resonances within rooms. Much research has focussed on using this theoretical knowledge to model the sound-field. Whilst room modelling is a vast subject in its own right, beyond the scope of this thesis, a number of modelling methods are briefly introduced in Section 2.3 as their inherent advantages and disadvantages impact upon the model chosen for the implementation of auralisations for the subjective listening tests which underpin this thesis.

Section 2.4 then considers the body of research in the area of modal control. Differing control methods attempt to manipulate the sound-field in different ways, and an understanding of these provides essential knowledge of the underlying parameters which are to be further studied. Furthermore, it will be shown that many control methods have been proposed due to some *objective* criteria, and these then present ideal opportunities for new experimental work to be undertaken which will test their *subjective* validity.

Previous work has not however, been solely theoretical and objective in nature. A number of subjective studies have been carried out which have considered the human response to individual resonances. Section 2.5 draws together this work and discusses the importance of these studies as a stepping stone to the wider question of our subjective response to the overall low frequency sound-field.

Finally, with much of the previous work concentrating on individual modes and their associated parameters in isolation from the overall subjective impression in real listening scenarios, Section 2.6 provides a brief overview the research conducted into the subjective evaluation of an audio stimuli. Methodologies used for preference characterisation in other areas of audio reproduction, most notably surround sound perception are considered, and a research gap is highlighted for the use of these techniques to be exploited in the perception of low frequency reproduction quality.

2.2 Modal Theory

Modal theory refers to the method of defining a sound-field within an enclosed space, primarily at low frequencies. This sound-field is complex - characterised by the presence of individual resonances, or modes (Morse, 1936). The influence of these modes on the sound-field is particularly important at low frequencies, where the wavelengths involved are a similar size in comparison to the physical dimensions of the space. Modes are generally more sparsely distributed in the low frequency region and the perception of an individual mode is generally considered to degrade the audio signal (Howard and Angus, 2001). Indeed, resonances often add noticeable artefacts to audio reproduction. For example, resonances within loudspeaker units have been identified as a major source of colouration (Toole, 1986).

The sound-field at higher frequencies, whilst still containing modes, is rarely characterised by them, and does not suffer from the same perceptual effects. The distinction between high and low frequency perception is an important one. At higher frequencies, the sound-field is assumed to be diffuse - that is, homogeneous and isotropic (equal sound energy arriving equally from all directions) (Toole, 2006). Where a diffuse field is assumed, statistical methods are a useful way of characterising the sound field (Toole, 2006). Perception is often characterised by a response to these statistical measures, one such example being reverberation time (RT). However, at lower frequencies, the assumptions required in the calculation of RT no longer hold, as each mode does not decay at the same rate (Howard and Angus, 2001), and the energy of each mode is distributed spatially throughout the room. When a large number of modes occur within a given bandwidth, this spatial variation is reduced and the field becomes more diffuse. It is therefore this region which is considered in this thesis. This frequency is commonly referred to as the ‘modal region’ (Howard and Angus (2001) for example). Schroeder (1987) attempted to define the frequency at which a transition between these two regions of the sound-field occurs. His calculation is dependent upon the room volume and reverberation time

(RT_{60}). The original formula was based upon the assumption that a diffuse field may be assumed at the point where the average mode spacing is one tenth of the half power bandwidth. In a subsequent revisit to the subject, the author revised this figure to a more conservative three modes per bandwidth (Schroeder, 1996), thereby lowering the transition frequency for a room of given volume. This formula has been used by a number of researchers to justify a low frequency region within which to base their experiments (Cox et al., 2004; Fazenda et al., 2005). The discussion above should lead to caution in applying the Schroeder Frequency - it requires the statistical reverberation time, and yet this is a concept valid only where a diffuse field is assumed. Indeed, Baskind and Polack (2000) have shown that reverberation time cannot be reliably determined in small rooms. Nevertheless, Toole (2006) claims that the Schroeder Frequency is still a useful guide, and furthermore, Nelisse and Nicolas (1997), characterising a diffuse field using two different descriptors, find good agreement with Schroeder's equation. Davis and Patronis (2006) suggest an alternative formula for the high/low transition frequency, equal to three times the velocity of sound divided by the room's smallest dimension. In many cases, differing methods lead to a similar frequency. Howard and Angus (2001) point out that although a single frequency results from equations of the kind, this is very much a guide for the transition *region*, rather than a *fixed* crossover point.

As mentioned, a review of the literature highlights many authors referring to the transition frequency between a 'modal sound-field' and a 'statistical' or 'diffuse sound-field' which may provide a limiting point to their investigations (Blaszak, 2007). However, as there is no single point where the nature of the sound-field immediately changes, investigations in this thesis are not restricted only to those scenarios where the modal density is low, but includes those where the modal density is high. In cases where the density is high, and the transition frequency low, further understanding of our perception is required. That the room may by some standards be considered diffuse should not introduce a barrier preventing study of our perception in the low frequency region.

The question of what frequency range should be considered as 'low' therefore arises. The literature offers no consensus. Toole (2006), in his review paper, alludes to an upper bound of 300Hz. Other authors have used 250Hz (Osipov et al., 1997), whilst the most prominent cut-off for papers studying the low frequency region appears to be 200Hz (Darlington and Avis, 1996; Makivirta et al., 2003; Cox et al., 2004; Fazenda et al., 2005; Avis et al., 2007).

2.3 Describing the sound-field

Within the low frequency region, an acoustic response may be described by wave theory. An analytical solution, known as the modal decomposition is an infinite summation of elements consisting of geometric and frequency components which produces a complex pressure for a given source and listener position (Morse, 1936). There are a number of assumptions with this solution. Firstly, the solutions are only simple where the enclosed space is rectangular (although Kuttruff (1991) shows that other geometries may be calculated). Secondly, there is the assumption that the boundaries are perfectly rigid, damping is introduced as a loss of energy of each mode and the room is only lightly damped (Morse, 1936). Given these assumptions, the model has been shown to produce the general pressure response of a small room faithfully (Fazenda et al., 2005). This model is presented formally in Chapter 3 and may be considered sufficient for modelling the critical listening spaces under study.

The sound radiating source must also be modelled. The simplest case is that of a point source (Morse, 1936). The doctoral thesis of Fazenda (2004) shows this to be an accurate ideal source, although it becomes inexact when predicting the sound field close to the source.

The literature highlights a number of other methods for reproducing the modal sound field. These include the so called 'Image Source Model' (first implemented by Allen and Berkley (1979)) which calculates the room impulse response through a summation of boundary reflection 'images' in the time domain. Furthermore, numerical techniques such as Finite Element Analysis (FEA), Boundary Element Modelling (BEM) and Finite Difference Time Domain (FDTD) have all been shown to successfully model small rooms (Geddes and Porter, 1988; Ciskowski and Brebbia, 1991; Botteldooren, 1995). There are a number of advantages to these solutions, such as the ability to model complex geometries, boundary conditions and modelling the influence of objects within the space. However, significant disadvantages lie in the computational time required to process such models. Therefore these models become less attractive for subjective testing purposes, where changes to the room conditions are often required immediately for audition and subject response. Avis (2001) showed that BiQuad IIR filters can also be used to reproduce individual room modes, which may in be combined to form an approximate model. Once specific advantage of using this method is the ability to model the Q factor of each resonance, while the disadvantage is that modelling the modal interaction is difficult.

2.4 Modal Control

As an understanding of modal theory grew, researchers attempted to use the new knowledge as the basis for control methods aimed at improving the quality of reproduction at low frequencies. Therefore, this section presents a review of such attempts to control the sound-field. There have been a number of different approaches, including investigations into the physical design of a room, placement of loudspeakers, passive absorption, and a number of active control methods. An overview of the key literature in each of these methods is presented here, with reference to the modal parameters which they attempt to improve.

2.4.1 Physical Room Design

The first method considered is that of the physical design of the room. This is not strictly a passive or active control method, but rather an attempt to alleviate the problems caused by room modes through a specific *modal distribution* within the room. Work of this nature resulted in an attempt to define room geometries which achieve the best performance, and has been evident for many years.

Early work by Morse (1936) considers the frequency distribution of modes within rooms, forming a set of equations to determine the number of modes expected up to a certain frequency. Further expressions of this statistical spacing were published in 1939 (Bolt, 1939a; Maa, 1939). This understanding of the frequency spacing statistics formed the foundation for the theory of *aspect ratio design*. By defining the ratio of length, width and height prior to construction, control is gained over the centre frequencies of each mode. By considering these frequencies in all three dimensions, it is possible to determine cases of modal degeneracy - that is, where two or more modes share the same centre frequency. A number of authors have therefore sought to define a series of room aspect ratios having a modal distribution which, by some criteria, may be deemed ‘good’.

In revisiting the statistics of modes in rectangular rooms, Bolt (1946) became the first to publish a set of room aspect ratios. Here, ‘good’ rooms were considered to have little modal degeneracy. In other words those ratios which achieve an even average mode spacing. The results are commonly referred to as ‘Bolt’s blob’ (Toole, 2006), due to the graphical representation of those favourable ratios. Often left unquoted is the corresponding ‘range of validity’ graph which shows the frequency range over which these optimum ratios are valid, for a given room volume. Taking up this work, Louden (1971) used the newly available computational power of the microcomputer to derive a ranking of room ratios using a slightly different metric based upon the standard deviation of spacing between all calculated modal frequencies. The ratio 1:1.4:1.9 was hailed as the best example of a good distribution, having the lowest standard deviation across the range considered.

Such individual ratios may be considered impractical and overly restrictive. Walker (1996) took up this criticism, making the case instead for a *range* of acceptable ratios. A quality index based upon the mean-square spacing of modal frequencies was employed, and importantly, shown to give similar results to other spacing indexes. By fixing certain variables, such as the room volume and height, Walker produced a new contour map with criterion for a range of accepted length/height and width/height ratios. The criteria for 'acceptable' scores on this quality index was given with a fixed upper bound for the length/width ratio and a simple straight line equation for the width/height.

Bonello (1981), departing from the derivation of optimal ratios, used the spacing statistics to suggest a criteria which may be simply applied to determine if any room scenario is 'acceptable'. This was said to be developed through "practical experience of many rooms". The criterion is based upon the assumption that if a certain number of modes within a particular bandwidth is met, colouration will be less perceptible. It should be noted that this assumption is in agreement with the aforementioned theory of the Schroeder Frequency. In order to assess each room, the modal frequencies must first be calculated. Two conditions are then given which must be met. Firstly, that each successive third octave band should have more modes than the preceding one (or equal if the first band has only one mode), and secondly, that whilst there should ideally be no modal degeneration, it should only be tolerated where there are at least five modes in the bandwidth. As the density of modes increases with increased room volume, the criterion is not readily comparable with the fixed aspect ratios mentioned above. Indeed, according to Louden (1971), certain accepted ratios may fail the criterion at one room volume, and yet be validated by it at another.

Welti came to a similar conclusion, comparing a large number of modelled rooms and listening positions to the criteria and showing a lack of compatibility with some of his own metrics, such as the 'variance of spatial average' and the 'mean spatial variance'. These metrics both consider the dB level of a measured room response and are discussed further in Section 2.4.4 (Welti 2009).

In each of the methods described above, all possible modal frequencies within a given room are considered with equal weighting for any optimisation. This implies that all modes will be both excited by a source, and received by a listener at maximum amplitude. However, this will only occur with the source located in one tri-corner, and the listener in the corner diagonally opposite - a scenario Toole (2006) brands as "simply ridiculous". In an attempt to combat this inherent problem, that these metrics based on spacing statistics do not consider the coupling of source and receiver to modes within the room, Cox et al. (2004) based a new metric upon the magnitude frequency response. By using the frequency response, a figure of merit may be calculated which is not only dependent on aspect ratio, but also the position of the source and receiver, and the room volume. In fact, as it is simply a measure of the frequency response, unlike the previously mentioned methods, this method does not require the assumption of a regular shaped room. The optimisation procedure uses a cost parameter generated from the frequency response's deviation from a best fit line. Deviation above and below this line were given equal weighting. The algorithm generated room responses over the range of aspect ratios, and the results, when mapped in a similar manner to Walker (1996), show a dependence on both volume and source and receiver position. However, some comparison was drawn against the aspect ratios of Bolt and Louden, in both cases suggesting these ratios were not in agreement. Cox et al. (2004) cite that the early ratios not being robust to 'constructional variations' as a possible reason for this. They also suggest that 'subjective characterisation' of the cost parameter is needed in order to properly determine its relevance, as well as subjective thresholds for the detection of changes in the computed figures of merit.

To this point, the control of modes using specific room dimensions has been considered. The various methods are preventative in nature, in that they offer some element of control prior to construction. With the exception of Cox et al. (2004), they are based upon the center frequencies of the room modes, which become fixed quantities at the point of construction. The following methods each offer an element of subsequent control.

2.4.2 Absorption

Room modes are standing waves which, when the excitation source is removed, decay exponentially over a period of time. This time may be determined by the amount of damping present (Morse, 1936). This damping removes energy and therefore shortens the decay time. One of the simplest ways to remove this energy is to absorb it. Alongside Morse's theoretical work in the 1930s, the positive effects of absorption on the reduction of modal energy in a room were shown experimentally by Bhatt (1939), who measured the reduction of level for the steady state, and reverberation time for transient modes within a reverberation chamber with one absorbing wall.

It should be noted that the subject of absorbers and their design is large. For a thorough text on the subject, see Cox and D'Antonio (2009). A brief overview of the key points in relation to modal control are given here.

Many designs focus on controlling the objective reverberation time parameter within a space. This is an extensively quoted parameter in terms of perceptual relevance in the mid to high frequency region. However, primarily due to the large wavelengths involved, absorbers become less efficient for the control of low frequencies. Nonetheless, there has been important research directed at the use of absorbers for modal control, and a brief review of these is important. We may note the shortcomings of such devices, which have in turn driven continued research into further control methods.

The simplest form of absorber is the passive absorber, which may generally be considered in two classes - porous and resonant. Porous absorbers may typically be open cell foam, acoustic tiles, carpets and other soft furnishings (Cox and D'Antonio, 2009). Absorption coefficients may be determined when placed within the statistical, or diffuse, sound field and relate to the surface area of the absorber. However, within the low frequency sound-field, as we have seen, this concept of a diffuse field is invalid and as such, coefficients can often be unreliable. The use of porous absorbers is primarily undermined by the large wavelengths, which imposes on them the need to be extremely thick in order to provide adequate bass absorption - at least a tenth of a wavelength has been recommended for significant absorption (Ingard, 1994). At 30Hz, this would be over 1m thick. Furthermore, as porous absorbers are restrictive where particle velocity is high, they should be placed away from room boundaries (Newell, 2007). This further decreases their feasibility in small spaces. Finally, porous absorbers often act as wide-band devices, and this may result in unwanted absorption in the mid and high frequencies. For these reasons, a second type of passive absorber may be preferred for modal treatment - the resonant absorber.

Resonant absorbers have been shown to provide a better low frequency absorption performance since they are 'tuned' to resonate at a design frequency, where absorption efficiency is maximised (Cox and D'Antonio, 2009). Absorption tends to decrease rather drastically on either side of this frequency and this behaviour (Q-factor) may be varied. As a result, such resonant absorbers offer benefits in a number of ways. Firstly, their effectiveness increases at a room boundary, where pressure is high, meaning they can be placed there successfully. Secondly, it is possible to achieve significant absorption in a much smaller physical space. Resonant absorbers themselves may be classified in two broad categories, Helmhotz resonators and panel resonators. They work by utilising a 'mass-spring' arrangement - with the spring being an enclosure of air. Within Helmhotz resonators, the mass is a 'plug' of air, and a sheet of some material is the mass in a membrane design. Within the resonating cavity, porous absorption can be placed, which enables losses and further control of the Q-factor of the absorber. These types of devices have been used to control room modes in studio spaces (Zha et al., 1999).

As mass/spring systems, such absorbers may be tunable to the room. They can be used to treat problematic modes by removing energy at these frequencies. Cox and D'Antonio (2009) note that of the two types, the performance of membrane absorbers is more difficult to predict. They also point out that at low frequencies, the effect of any absorber is difficult to calculate, particularly because of the non-diffuse nature of the sound field.

Although resonant absorbers are generally more efficient at low frequencies when compared to porous absorbers, the total area required to effectively treat modal problems is comparable to mid-high frequency absorption and this a significant number of units are still required. Further control methods have therefore been considered.

2.4.3 Active Control

Whilst widely used, it can be seen that passive absorbers have their limitations, and with these in mind, researchers have explored the field of *active* absorption. Whilst potentially more expensive, this method a) extends the possible absorption to much lower frequencies, b) may be adaptively varied according to the requirements of the room in question, and c) may offer bass absorption with a relatively small footprint (Cox and D'Antonio, 2009). Olson and May (1953) were the first to propose active absorbers, which use a secondary radiator driven by a control signal. When implemented properly, these systems allow good control over modal decays, particularly for modes of high Q which are problematic for passive absorption (Cox and D'Antonio, 2009). A number of researchers have investigated active impedance systems, which may be classed as either fixed, or adaptive (Avis, 2000; Darlington and Avis, 1996; Herzog et al., 1995). The thesis of Avis (2000) provides a thorough summary of the relevant literature, and the 1995 book 'Active Control of Sound' is a particularly valuable resource (Elliott and Nelson, 1995).

In his thesis, Avis (2000) shows that within a one dimensional duct, modal decay could be significantly reduced. This work also reveals a steady state pressure reduction of a 44Hz mode by around 6dB when scaled up to three dimensions. However, using this method, each mode must be treated individually. In an attempt to combat this, Avis produced a single control filter which encompasses the whole modal sound field, effectively deconvolving the signal for a single source/receiver position. For his particular system, the room response was modelled using IIR biquad filters for the control filter (Avis, 2001). The advantage of such a system is that it incorporates control not only in the frequency domain, but also to the time response of the signal. Whilst this method is effective at a single listening position, it is less so at other points within the room, which, as previously discussed, will have a different response and therefore require a different filter.

In addition to the active absorption methods highlighted above which attempt to modify the impedance of the room boundaries, correction to the modal response may be attempted through a deconvolution of the transfer function, which is generally referred to as 'modal equalisation' (Makivirta et al., 2003). Here, secondary radiators are not employed, but rather, the transfer function is equalised before being fed to the primary radiators. A number of researchers have studied this field, and reported on the design of digital filters which allow a deconvolution (Mourjopoulos, 1985). Makivirta et al. (2003) refer to the method as Type I equalisation. The authors discuss the design of such a filter, which works below 200Hz and is an arrangement of the poles and zeros. Mourjopoulos and Paraskevas (1991) and Farnsworth et al. (1985) show that with a single point equalisation, while there is a flattening of the frequency response at the listening position, the audio is degraded at other locations.

In an attempt to overcome the problem of single point equalisation, Elliott and Nelson (1989) attempted to minimise the sum of squared errors between the equalised pressure and a flat response using adaptive digital filters. Whilst such an implementation produced better results than for a single point, the area of improvement is described by Santillan as "small spheres around each control point" (Santillan, 2001).

To improve upon the modal equalisation, more recent work by Santillan (2001) and Stefanakis et al. (2008b) consider the use of multiple sources for the simulation of a plane wave through the room. By creating such a wave, the physical propagation of modes may be corrected, leaving the equaliser the simpler task of correcting any remaining artefacts in the response.

With careful consideration, it is possible to create this plane wave with fewer sources. A method known as Controlled Acoustic Bass System (CABS) has been proposed by Celestinos and Nielsen (2008a). Their placement of four sources in specific room positions (two on each of the front and back walls, 1/4 and 3/4 on the horizontal plane and 1/2 on the vertical) eliminates axial modes in the height and width dimensions (up to the fourth order mode along the width), allowing only modes along the room length to be excited. By dealing with the propagation of the modes directly, it is possible to alleviate the problems of control being restricted to a single listener position. The secondary radiators, known as ‘sinks’ then absorb the sound energy at the rear of the room when delayed, attenuated and reversed in polarity. This set-up has been shown both theoretically and through measurement in two rooms to improve the frequency response (in terms of a cost parameter known as ‘mean sound field deviation’ and also the standard acoustical measure of clarity). Within such a source and sink system, the positioning of loudspeakers has been investigated by Stefanakis et al. (2008a), who incorporate an algorithm to extend optimal source placements to non-rectangular enclosures. Although a promising system, a number of practical issues should be considered, such as reliance (in the CABS system at least) on a rectangular room, the cost of additional subwoofers, less than convenient placement and the complexity of set-up (processing of delays and attenuation is necessary). In response Vanderkooy (2007) explored the possibility of using a simpler arrangement of sources and sinks in an irregular room. Assessment was made by way of a ‘room demerit’ figure which is based upon a similar principle to that of the deviation from an ideal frequency response seen earlier (Cox et al., 2004). The figure was applied to real measurements taken in rooms. Initial conclusions showed that, although in some positions the objective measure shows an improvement from multiple sources with appropriate delay and gain, others do not, and it is clear that further work is required.

2.4.4 Loudspeaker Placement

Yet another method of modal control revolves around the simple concept of loudspeaker placement. This affects the coupling between the loudspeaker and the mode.

Groh (1974), Allison (1976) Ballagh (1983) have each proposed methods to define optimal loudspeaker positions. Groh (1974) attempted to equalise two listening rooms through the placement of loudspeakers relative to a particular listener position. This method involved identifying problematic modes at each third octave band and moving the loudspeaker to either increase or reduce the coupling. Although an improvement was reported, compromises must often be made with this technique, as a movement to improve one frequency band may have the inverse effect on another band. Such a manual method is therefore labour intensive, and also room specific. Allison (1976) also searched for uniform frequency responses through the placement of sources, but considered the effects over a number of listening positions within the room. The placement was chosen with respect to the low frequency power output obtained. Three standard loudspeaker orientations were used, and a fourth was a direct radiator loudspeaker system, designed for uniform power loading. Five rooms, with a total of 14 listening positions were measured in third octave bands between

80 and 400Hz. Results showed that the direct radiator system produced the most uniform response throughout the rooms, with the third standard system consistently showing better results than the first two. Allison therefore concludes that a system either designed, or placed in respect of the boundaries, to produce a uniform power output will subsequently produce a more uniform sound pressure across multiple listening positions. Ballagh (1983) further investigated the uniformity of the total radiated power using direct radiator loudspeakers placed near the room tri-corners. Optimal placement was shown to be dependent upon the Q of the speaker, and a function of polar coordinates at fixed distances from the corner. The results also show that through optimal placement, the low frequency response of a loudspeaker may be extended by two thirds of an octave.

Further considering the coupling effects of subwoofers with the nodes and anti-nodes within the room, Welti and Devantier (2006) offer a multiple subwoofer optimisation method. Investigations show that some improvement may be made through the use of multiple subwoofers and minimal signal processing (simple EQ, delay and amplitude control). This approach differs from many previously seen, as a search algorithm is employed, which calculates a number of 'best case scenarios' through superposition of real measured responses at a number of seats from a number of subwoofers. The authors derive a metric known as the 'Mean Spatial Variance' (MSV) in order to assess the spatial uniformity at multiple positions from differing combinations of subwoofers, assuming linear superposition of each loudspeaker/seat measurement. Two set-ups are found to be the best practical configurations in respect to the MSV. These are four subwoofers at four wall mid-points, and two subwoofers at opposing wall mid-points. Brute force search results can be reviewed and implemented by individual users dependent on their specific room and needs. The implication is that no one-size-fits-all method is feasible in practice, and that any given room requires a tailored solution.

For each of these studies, we see that although results are presented, they are only valid for a particular loudspeaker, or room, or listening position. Furthermore, the 'optimisations' are again based on objective criteria, such as a uniform power response or the above mentioned MSV. The subjective effects of these optimisations are more often than not left unconsidered.

2.4.5 Further Control Methods

As shown, in addition to spectral differences, room modes exhibit highly spatial differences over the listening area. Many of the optimisation methods reviewed are limited, in practice if not in theory, to a small optimisation area. Recent research has proposed a psychoacoustic control method which may be applied to any system (Hill and Hawksford, 2010). Here, problematic modal frequencies may be removed through the use of parametric equalisation, and replaced with a 'virtual bass' system, which compensates by adding spectral components at higher frequencies which give the impression of the lower through the concept of the missing fundamental. Initial subjective testing suggests that the system may reduce perceived spatial variation.

2.5 Perception of Modes

The previous section highlights the current research in modal control techniques. It was shown that many of these have focussed upon alleviating the problems of room modes by aiming to achieve some target response based upon a specific metric. However, in the vast majority of cases, these metrics have not been proven to correlate with reliable subjective data. A number of authors acknowledge this, and explain that their objective metrics are those most likely to have subjective relevance (e.g. (Cox et al., 2004; Welti and Devantier, 2006; Vanderkooy, 2007). If a reliable subjective metric was available, it would be possible not only to assess current room treatments more accurately, but also guide those working on further optimisations.

The field of psychoacoustics studies the nature of our perception of sound, and covers a wide range of concepts. Of particular interest for this study of room modes is that of spectral and temporal perception. There are both physiological and cognitive considerations here, which are extensively covered by Moore (2003). Fletcher and Munson (1933) introduced the term critical band, which relates to the frequency response of the basilar membrane. This in turn relates to the frequency resolution of our hearing and is therefore underpins many of the assumptions in our understanding of the perception of low frequencies. This human frequency selectivity is evident in the concept of the aforementioned Schroeder Frequency. Although the Schroeder Frequency relates specifically to the frequency above which statistical methods may be used to analyse the sound-field, much research has used this frequency as a subjective limit, with the assumption that where enough modes overlap within a certain bandwidth, they will not be individually distinguishable (Howard and Angus, 2001).

Combining the theory of early psychoacoustic models, and the desire to better understand our perception of audio systems, be they transducers or a room response, research by Bucklein (1981) studied the audibility of individual resonances in the presence of music, speech and noise signals. The work was carried out in 1962 although first published in 1981. Although not directly linked to room resonances, Bucklein's study of ten subjects listening over headphones, showed that peaks (resonances) in the frequency response were more audible than dips (anti-resonances). Whilst the experiment was not designed specifically for low frequency, some examples were included, with frequencies studied as low as 85Hz. In addition to the increased audibility of peaks, it was reported that when in the presence of musical signals, difficulty was found in perceiving the resonance, unless it corresponded in frequency to the musical note being played. This has subsequently been confirmed by Avis et al. (2007). This finding provides a rationale for further testing of modal perception with musical stimuli, which is the basis of much of the experimental work in this thesis. Bucklein's study also reported that the detection of resonances was reduced as the Q-factor was increased. Qs of 1.8, 2.9 and 5 were considered. These findings are further supported by the work of Fryer (1975, 1977). In continuing to study the perception of resonances, Fryer also confirmed that resonances of low Q could be detected more easily than those of high Q. Whilst exact Qs under study were not reported, it is clear that the experiments included a range from at least 0 to 5. With each doubling of Q, an approximate 3dB decrease in detection threshold was observed. This increase in detectability of low Q resonances suggests that the temporal aspect (ringing) is not a reliable indicator of perceived degradation (Toole and Olive, 1988). Fryer's study also confirmed the increased difficulty in the task of resonance detection in musical signals compared to noise.

Addressing the issue of resonance perception being influenced by parameters other than simply Q and signal type, a comprehensive study was carried out by Toole and Olive (1988). Here, a number of parameters were varied which studied perception as a function of frequency, Q, time delay, presence of reverberation, loudspeaker directivity and listener performance. The study considers mostly higher frequency resonances, although 200Hz is included. A number of questions were asked. Firstly, how to detect a resonance. Secondly, if it will be audible by looking at its characteristics, and thirdly, to what extent it needs to be altered in order to become inaudible. Again, offering confirmation of Bucklein and Fryer's studies, low Q resonances were shown to be more audible than high Q (also approximately increasing 3dB for a doubling of Q). In most cases, subjects were more sensitive to resonances in the presence of noise signals.

Interestingly, the work also reveals that for discontinuous signals, which may be found in music, the listening environment has an important role. At low Q (1), the addition of reverberation lowers the threshold of detection, suggesting that in typical listening rooms, which may indeed have some reverberation, broadband resonances are even more likely to cause problems (Toole and Olive, 1988). For high Qs (50), detection is unrelated to the environment. It is interesting to note that resonances described as low-Q in this work would cover all three Qs reported by Bucklein.

Much of the above work focuses on resonances higher in frequency than would usually be considered to lie within the modal region of a critical listening room. The literature also focused primarily on peaks with little study into dips (anti-resonances). In order to shed light on these gaps, Olive et al. (1994) applied an Up Down Transform Rule approach to threshold detection tests with resonances of 63, 125, 250 and 500Hz and Qs of 1, 10 and 30 (considered low, medium and high). In line with previous work, the detection thresholds were obtained for resonances in the presence of both pink noise and transient signals (pulses). With pink noise signals, low Q resonances were found to have a threshold independent of frequency, at about 2dB for peaks and -2dB for dips. However, for peaks at medium and high Qs, as the frequency increased, the detection thresholds decreased, suggesting that resonances become more audible at the higher end of the low frequency range. With pulses, results differed, with the detection thresholds lower for higher values of Q and anti-resonances just as detectable as resonances.

Differences in absolute thresholds between this work and the previously mentioned studies are attributed to the methodology, which was here based upon the 70.7% detection threshold on the psychometric function, where as the other studies highlighted used methods which would yield results closer to the 50% threshold. However, similar trends across all these studies can indeed be shown, with each reporting low standard deviations in results. In summary, resonance perception is cleary complex and difficult to predict. However, it has been shown to be generally dependent upon frequency, Q and program material.

Where the resonance is a room mode, the difference in thresholds for differing signal types is of particular interest. Signals commonly found within popular music often contain combinations of both noise-like and pulse-like sounds. The differing perception of each type therefore leads to some confusion as to the true thresholds. With the importance of Q clearly shown, Avis et al. (2007) modelled a number of resonances across the low frequency range using bi-quad filters in order to determine the overall threshold of Q[1] within the presence of music signals. This method therefore moves away from single frequency resonances, towards the more realistic scenario of many resonant modes present in the response. Thirteen resonances were modelled between 34 and 198Hz using bi-quad filters, thereby enabling the Q factor to be systematically modified, and remain constant with frequency. Centre frequen-

[1] If a mode has a Q-factor less than this threshold then the individual effect of that mode on overall audio quality will be imperceptible

cies were chosen to correlate with a real room measurement, and the test signal was a musical sample chosen for its low frequency content. The Parameter Estimation by Sequential Testing (PEST) method was employed. Results indicate an absolute threshold for Q of 16 - below this level, further treatment of the room becomes unnecessary. Trends from this work agree with the conclusions of Olive et al. (1997) for detection where the signal is a square pulse, and is in contradiction with the findings for steady state noise signals, where we are better able to detect resonances of low Q. Avis et al. (2007) make a tentative suggestion that this is due to the importance of modal decay as a cue as opposed to frequency based cues in steady state signals. This suggestion contradicts Toole (2006), who has claimed that spectral cues can be considered a more important factor than temporal cues in hearing colouration.

Avis et al. (2007) also extrapolate their Q thresholds to decay times across three octave band frequencies of 32, 63 and 125Hz. A Q of 16 at these frequencies corresponds to decay thresholds[2] of 1.10, 0.56 and 0.28 seconds respectively. However, these times must be taken only as indicative, as no explicit frequency dependence was studied. In fact, there is also the suggestion that, due to increased sensitivity at higher frequencies, the Q threshold results may be biased towards them.

Considering the question of frequency dependent decay thresholds, Karjalainen et al. (2004) returned to using a single synthetic resonance in the presence of speech, noise, a drum hit and rock music, for frequencies ranging from 50Hz to 800Hz. The decay times reported incorporate all four signals, and show thresholds of around 0.2 - 0.3 seconds for all frequencies down to 100Hz. Below this, for the 50Hz test frequency, the threshold increases dramatically to around 2 seconds, suggesting that at such low frequencies, any reduction beyond that of a typical listening room would be essentially unnoticeable.

The aforementioned PEST methodology (Taylor and Creelman, 1967; Taylor et al., 1983) used before to obtain thresholds of Q may also be employed in the extraction of subjective thresholds of decay Goldberg (2005, 2006). In these studies, a new signal type is introduced - an upwards sine sweep from 10-1000Hz over 0.5 seconds with a single frequency temporal decay added using a simple resonator and gain stage. A small scale listening test revealed lower thresholds than those of Karjalainen et al. (2004). Exact thresholds were not reported with confidence due to low subject numbers and some issues with the equipment. The initial lower thresholds may however be explained by the difference in signals, with this test using the more analytical sweep. Goldberg (2006) also highlights issues in the test

[2]If a mode has a decay time less than a particular threshold, the individual effect of that mode on overall audio quality will be imperceptible.

signal and replay equipment, citing some low frequency distortion, and other non-modal cues as problematic. Subsequent work by the author introduces a new test signal, the windowed sine burst, created specifically for modal decay threshold tests (Goldberg, 2009). A similar test method was used to define the optimal fade in and out times for such a signal, revealing that for sine bursts below 100Hz, 1.5 cycles of the sine should be windowed as a fade out to prevent audible artefacts, while above 100Hz, a fade out of 15ms is required.

It has been shown that a number of studies have looked at our perception in terms of Q factor, signal type, peaks and dips and temporal decay. However, no current study has successfully measured thresholds of decay in a frequency dependent manner, related to both typical 'realistic' modal responses within rooms, and in the presence of musical signals. Chapter 6 addresses this area.

Considering the above discussion, it is suggested that a greater understanding of perception at low frequencies, where realistic rooms are considered, along with the effects of music stimuli is required. Such a study finds a basis within the realm of *overall* perceptual quality rather than the effects of *specific* resonances. It is to this area that we now turn.

2.6 Audio Quality Evaluation

A great many of the previous studies do not directly address the question of overall perceived audio quality within rooms at low frequency. Especially at higher modal densities, the interaction between modes, due to their complex nature may produce highly irregular responses which are difficult to relate to individual modes (Karjalainen et al., 2002). It is here that studies into the subjective evaluation of audio quality are more revealing.

Firstly, a number of objective models which attempt to predict low frequency quality are highlighted. This is followed with a review of more recent work which highlights the methodologies for the assessment of overall subjective quality. It is shown that there remains a gap in the research for implementing a similar methodology in the context of low frequency perception.

2.6.1 Objective Models

Throughout this review, it has been shown that a number of room optimisation and modal control methods have been based upon some objective criteria. Objective measurement is described by the Oxford English Dictionary as the measurement of a stimulus "not influenced by personal feelings or opinions" and must be repeatable independent of person, equipment or method. These objective criteria have variously been referred to as metrics, figures of merit, cost parameters or simply 'measures'. For clarity, the term 'metric' will be used henceforth.

Some work has been done in an attempt to study the subjective importance of these metrics. If any metric is to be applicable, either in an optimisation routine, or the assessment of sound quality, it must be shown to have a valid subjective relevance. As shall be shown, in many cases this has not been done.

As Section 2.4.1 showed, a number of metrics have been considered for the prediction of optimal room ratios. Two general forms have been used - those based upon modal spacing statistics (i.e. Louden (1971) (Bonello, 1981)), and those based upon the goodness of fit of the magnitude frequency response to a chosen ideal (i.e. Cox et al. (2004)). Fazenda et al. (2005) observed that in some cases, the two metrics appear to contradict each other and therefore deployed a listening test in an attempt to determine if either metric was preferable in terms of their ability to predict the 'quality' of the room. This was achieved by way of contrasting rooms which score highly according to a spacing metric, but low when considering goodness of fit, and vice versa. These contrasts were then compared subjectively using virtual rooms presented over headphones. The work concludes that as clear differences between stimuli are almost always detectable, neither metric should be considered a reliable predictor on a linear scale of quality. However, subjects were not asked to directly rate rooms for either of the individual metrics in terms of quality, and as such, it may be argued that one metric may still have a psychoacoustic relevance. Such an objective metric/overall subjective quality relationship remains to be seen, and Part III of this thesis addresses this important question.

It would seem that even with findings such as those of Fazenda et al. (2005), much of the research relies on an objective metric which would intuitively appear to have a subjective relevance when considering a particular optimisation. One such example is the 'Figure of Demerit' introduced by Vanderkooy (2007). Here, the metric is similar to that of Cox et al. (2004) - essentially a deviation from a target response. While Vanderkooy is careful to explain that the metric may not correlate to perception, it is nonetheless used to assess a multi-point equalisation control technique.

In a departure from purely frequency domain metrics, the Modulation Transfer Function (MTF) also incorporates time domain information, and has been suggested for assessment of room impulse responses (Fazenda et al., 2006a; Holland et al., 2004, 2006). These works highlight the MTF's usefulness in predicting the performance of low frequency room responses, with results showing that the MTF may produce a rough prediction of quality, particularly in comparison with subjective listening tests for thresholds of Q (Avis et al., 2007). Using signals with a Q of 16 (the previously determined threshold), the MFT rated the room with a score suggested to be 'good'. With an increased Q of 19, the ratings fall in a 'fair/good' category, while lower Qs of 11 are considered 'excellent'. Harris et al. (2006) shows further correlation between subjective responses to low frequency stimuli and the MFT's prediction.

The modulation frequencies within the algorithm for the above studies were originally developed with speech signals in mind. These may be considered less relevant in the context of musical stimuli, for which different frequencies may be more representative of the content. To determine these, Harris and Holland (2008) undertook a comprehensive analysis of modulation in 173 typical musical extracts, revealing that a reduction in the number of modulation frequencies to just six has no noticeable effect upon the result. The frequencies themselves however, are similar to those used in speech.

Finally it is necessary to consider a third category of objective metrics - spatial distribution. Such metrics aim to classify a room, or more often, a specified listening area, in relation to the uniformity of the sound-field within them. One such example of this in the literature is the Mean Sound Field Difference (MSFD) (Celestinos and Nielsen, 2008b). The MSFD is a combined measure of the spatial deviation of 25 measurement positions and the magnitude of deviations in each frequency response. Again, whilst no specific claims to subjective correlation are made, the implication is that a 'good' MSFD could be mapped successfully to a 'good' subjective impression over the corresponding listening area. A similar metric, calculating the SPL differences of virtual microphones throughout a listening space, the Mean Spatial Variance (MSV), is used by Welti and Devantier (2006). Here, the author explicitly states that "though it has not been proven that the metrics used here correlate exactly with listener preference, minimizing the seat-to-seat variation is a reasonable goal if it is assumed that the system is to be equalized.".

The purpose of this review is not to suggest that all metrics shown are invalid. Indeed, many have a logical psychoacoustic basis and are indeed considered a 'reasonable goal'. It should, however, be noted that a number of metrics which rate rooms according to some scale have no associated perception of magnitude. Cox et al. (2004) highlight this point. They report a figure of merit in dB, but explain that the smallest perceivable difference on that scale is not known. This is common to all such metrics, and can only be answered as our subjective model of hearing within the low frequency sound-field is improved and thoroughly tested.

2.6.2 Overall Quality Assessment

Whilst it is important to understand modal parameters, it must not be neglected that the perception of sound has been shown to be a multidimensional problem (Bech, 1999). This may be considered important at low frequency as there are many variables which make up the complex room response. Not only the construction of the room, but also its fittings and the positioning of loudspeakers and listeners all affect the overall room response. Furthermore, this response itself may influence a number of perceived aural sensations, which in turn are processed at a higher cognitive level before a decision as to ‘low frequency quality’ may be made. It is therefore important to ascertain the role which each parameter and each sensation play in the overall subjective evaluation, along with their relative importances.

The area of modal perception lags behind other aspects of audio perception, particularly in the areas of loudspeaker assessment and the perception of multichannel reproduction quality. This section summarises the advances made in quality assessment of audio with relation to these fields, and Part III uses this knowledge to characterise and assess the perception of low frequency reproduction.

Considering the room response as a whole, it is widely accepted that a ‘smooth’ frequency response provides the ideal listening situation, avoiding resonances, and indeed, spectral distortions of any kind (Toole and Olive, 1988; Toole, 1986; Gabrielsson, 1990, 1991). There appear to be very few studies however, which map frequency distortions to perceived quality. Gabrielsson (1990) applied three filters (low, mid and high frequency) to music recordings and comparisons were made between these and the original recordings (considered ‘flat’ response). Ratings were made on a number of descriptive scales, including fidelity, clarity, softness and fullness. Where the low frequency filter was applied, there was a 9dB boost below 200Hz (and a natural roll-off below 100Hz due to playback equipment limitations), and the music samples were generally rated lower in *fidelity* and *clarity*, whilst being considered *softer* and *fuller*. No attempt was made to map the contributions of these descriptions to an overall preference. Regarding the perceptual effects at even lower frequencies, where modal activity is likely to be most prevalent, an increased lower cut-off frequency results in perceptual effects (Bech, 2002). Specifically, lower cut-offs resulted in listeners reporting that ‘more bass’ was audible. Distortions in the form of passband ripple have also been shown to be degrading in terms of quality (Moore and Tan, 2003).

Techniques have been exploited in listening tests for many years, making use of simple judgements of quality, or producing ratings for parameters such as detectability or annoyance - so called ‘affective measurements’. However, in recent years, researchers have attempted to probe the notion of *perceived quality* further, to gain “an overall impression of the sound based on a combination of the individual attributes and the so-called cognitive factors” (Bech and Zacharov, 2006).

The first stage of such a process is to develop a language, or vocabulary to be used by the subjects (Rumsey, 2004). This process is no trivial matter, and a number of procedures for doing so exist. Within the food industry, a method known as Descriptive Analysis has been developed (Stone and Sidel, 2004). The method has been shown to be successful in the field of audio assessment, and is covered comprehensively by Bech (1999). Bech’s work originally took place in the context of spatial sound, but the method has been shown to work well across a number of audio related fields (Mattila, 2001; Martin and Bech, 2005). Further methods have been suggested for the generation of a descriptive language, notably, the Repertory Grid Technique, where each subject develops their own language and statistical analysis then refines the scales (Berg and Rumsey, 2006) and Free Choice Profiling, where each listener uses their own separate language, and ratings are given independent of any other subject’s language (Bech, 1999; Lorho, 2005).

The above work identifies methods which may be used to define a language, but it is noted that this can only be used for assessment within the scope of which the language was developed (Bech and Zacharov, 2006). At present, although a number of term sets contain descriptors seemingly related to low frequency concepts, such as Gabrielsson and Sjogren (1979)’s *muddy* or *rumbling*, there remains no language specifically for the characterisation of low frequency timbral quality. Interestingly, there is some evidence in the literature of terms being used for listening test scales, although these seem to have been decided on arbitrarily. Harris et al. (2006), for example, asks subjects to rate low frequency according to *clarity*, *bass extension and fullness*, and *relative level between instruments*;

Weisser and Rindel (2006), in studying a number of real listening rooms use the rating scales of *overall sound quality*, *boxiness* and *boominess*. Concepts such as these are again, both familiar and logical to experienced listeners, but there remains scope for a clearly defined and robust descriptive language at low frequencies. This work is undertaken in Chapter 7.

Where such a language is available, it can be shown that powerful statistical analysis, such as Principal Component Analysis and Multidimensional Scaling may be used to map both individual parameters and scores against proposed metrics to the perceptual space in order to build a model for the particular aspect of hearing under study. Such a study is reported in 8.

2.7 Summary

This chapter has reviewed the relevant literature when studying the perception of room modes. A brief theoretical review has provided sources for the basic understanding of the formation of standing waves, and the complex spectral and temporal responses occurring when considering a three dimensional rectangular enclosure. A number of modelling techniques have been introduced, whose implementation allows for detailed studies both objectively, in allowing new scenarios to be modelled, and subjectively, in that they form the basis of the auralisations used in listening tests. The modal decomposition model, with its ability to reproduce the general characteristics of a room and having efficient computation, was shown to have been used with success in low frequency psychoacoustic tests and is used in the testing throughout this thesis.

Methods by which researchers have attempted to control the low frequency sound-field have been presented systematically, from the physical design of the room dimensions, through simple absorption and more complex combinations of active control and equalisation. Importantly, it has been shown that the analysis of each of these control methods has relied upon some objective metric to determine performance. There has been very little significant study into the perceptual validity of these metrics, and this is identified as a clear research gap.

Finally, studies into the detection of individual modal parameters, such as resonance frequency, Q factor, and decay time have revealed some useful thresholds and trends. These include a threshold of Q of 16, and a number of exploratory studies reporting decay thresholds. However, it has been suggested that some of these are less valid in the context of real rooms with realistic music signals. It has been highlighted that overall perception, relating audible parameters and auditory sensation to an overall score, and the inclusion of perceived quality as opposed to simple affective judgements have been successfully employed to enhance the perceptual model in fields such as surround sound, but that these techniques have not yet been implemented for characterisation of the low frequency sound-field.

Chapter 3

Low Frequency Sound Fields

3.1 Introduction

As has been shown in Chapters 1 and 2, room modes are the cause of significant audible problems, such as an excessive bass level, audible distortion and ringing at specific frequencies. It will be shown that these problems manifest themselves particularly in small rooms, which this thesis considers to be between around $50m^3$ and $250m^3$. This is a typical volume for rooms used as studio control rooms, mastering studios and home listening rooms (Newell, 2007).

Chapter 2 highlighted research gaps in the area of the understanding of our perception of low frequency reproduction quality, in terms of both room and modal parameters and also overall quality. In order to understand the subjective aspects of listening within these critical listening spaces, where the sound-field is characterised by spectral, temporal and spatial elements, this chapter presents the basic theory governing the relationship between modal resonances, the physical space and the propagation of sound from a source to a receiver. Whilst this theory is not novel work, the key aspects are presented here in order to provide background and context to the interested reader.

The literature review identified a number of modelling techniques which may be used to simulate the sound-field within a room. One of these was the simple, computationally efficient analytical solution commonly known as the *modal decomposition*. This model is derived by considering the resonant frequencies within an enclosure, the propagation from a single point source and the resulting room transfer function at a listening position. The method has been used in successful psychoacoustic research by Fazenda et al. (2005), Stefanakis et al. (2008a) and Welti and Devantier (2006) amongst others, and will form the basis of many of the listening tests within this thesis. As such, the derivation is presented in Section 3.3. A number of assumptions are made when using this model which are discussed in Section 3.3.5. Section 3.4 introduces the concept of modal distribution within a room.

Finally, Section 3.5 details the use of the decomposition model in producing virtual room auralisations which may be presented to listeners over headphones during listening tests.

3.2 Modal Theory

In understanding the acoustics of enclosed spaces, it is not possible to consider sound radiation in the same terms as when radiated into free space. Therefore, such theoretical approaches as the inverse square law and the simple relationship between sound intensity and the power of a loudspeakers are no longer applicable. Rather, the sound-field is comprised of multiple *modes of vibration* of the air within the room. Because of this, the distribution of sound intensity may vary widely throughout the room, regardless of the distance from the radiating source.

These modes of vibration can be described as resonances, or standing waves, but are most commonly referred to simply as *room modes*. It is the combination of these modes which determines the unique characteristics of a room. As with all resonances, room modes are associated with a centre frequency, known as the modal frequency or *eigenfrequency*. The resonance also has temporal characteristics, decaying exponentially when the sound source is removed. Finally, each mode has an associated spatial distribution, and this is known as the mode shape or *eigenfunction*.

For a given enclosure, both the eigenfrequencies and associated eigenfunctions are derived in the following section. It must be noted that it is only possible to derive these for enclosures of certain geometries - examples being rectangular, cylindrical or spherical. This thesis considers only rectangular spaces, and in a further limitation imposed by the scope, the walls of this enclosure are considered to be smooth and perfectly rigid. This gives a boundary condition of zero particle velocity at each of the walls.

3.3 Resonances in a Rectangular Enclosure

3.3.1 Eigenfrequencies and Eigenfunctions

Morse (1936), Kuttruff (1991) and Kinsler et al. (2000) each thoroughly derive both the eigenfrequencies and eigenfunctions, beginning from the wave equation in Cartesian coordinates. Throughout this thesis, the point of origin (x,y,z) is (0,0,0).

Note that Morse (1936) uses the mid point within the room as the origin, which slightly alters the derivation of eigenfunctions. The room has coordinates from $x = 0$ to $x = L_x$ and similarly in the y and z dimensions, where L is the length of the dimension in meters.

The time invariant form of the wave equation, where a harmonic time law is assumed for pressure and particle velocity, using angular frequency ω is:

$$\Delta p + k^2 p = 0 \tag{3.1}$$

where Δ is the Laplacian operator, p the complex sound pressure and k the wave number.

In Cartesian coordinates this takes the form:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} + k^2 p = 0 \tag{3.2}$$

Where particle velocity is zero at each boundary ($x = 0$ and $x = L_x$ and similar in y and z dimensions):

$$\begin{aligned}
\left(\frac{\partial p}{\partial x}\right)_{x=0} &= \left(\frac{\partial p}{\partial x}\right)_{x=L_x} = 0 \\
\left(\frac{\partial p}{\partial y}\right)_{y=0} &= \left(\frac{\partial p}{\partial y}\right)_{y=L_y} = 0 \\
\left(\frac{\partial p}{\partial z}\right)_{z=0} &= \left(\frac{\partial p}{\partial z}\right)_{z=L_z} = 0
\end{aligned} \tag{3.3}$$

which, when substituted into a separated form of the wave equation:

$$p(x, y, z) = p_1(x) p_2(y) p_3(z) \tag{3.4}$$

becomes:

$$\begin{aligned}
\left(\frac{d^2}{dx^2} + k_x^2\right) p_1 &= 0 \\
\left(\frac{d^2}{dy^2} + k_y^2\right) p_2 &= 0 \\
\left(\frac{d^2}{dz^2} + k_z^2\right) p_3 &= 0
\end{aligned} \tag{3.5}$$

The constants in Equation 3.5 relate to k^2 by:

$$k^2 = k_x^2 + k_y^2 + k_z^2 \tag{3.6}$$

and therefore, for the x dimension, we observe the general solution:

$$p_1(x) = A_1 cos(k_x L_x) + B_1 sin(k_x L_x) \tag{3.7}$$

In order for the particle velocity to be zero at the boundaries, B_1 must be set to zero, as at the boundaries, a non zero sine term would result in a positive pressure. Furthermore, the $cos(k_x L_x)$ term must equal ± 1. Therefore, $k_x L_x$ must be a whole multiple of π, which is achievable where the following values are allowed:

$$\begin{aligned} k_x &= \frac{n_x \pi}{L_x} \\ k_y &= \frac{n_y \pi}{L_y} \\ k_z &= \frac{n_z \pi}{L_z} \end{aligned} \tag{3.8}$$

where n is a positive integer. Substituting these values into Equation 3.6 gives:

$$k_{n_x n_y n_z} = \left[\left(\frac{n_x \pi}{L_x}\right)^2 + \left(\frac{n_y \pi}{L_y}\right)^2 + \left(\frac{n_z \pi}{L_z}\right)^2 \right]^{\frac{1}{2}} \tag{3.9}$$

Remembering that k is the wave number in angular notation, using the relationship:

$$k = \frac{2\pi f}{c} \tag{3.10}$$

where f is frequency in Hertz and c the speed of sound in air in ms^{-1}, we can rearrange into the following, well known Equation 3.11 for obtaining the *eigenfrequencies* in Hertz of the enclosure. n_x, n_y and n_z may take zero values as well as a positive integer.

$$f_{xyz} = \frac{c}{2} \sqrt{ \left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2 } \tag{3.11}$$

Finally, the *eigenfunctions* associated with each eigenfrequency may be obtained by multiplying the cosine function in each dimension, with the k_x, k_y and k_x values of Equation 3.9 included:

$$p_{n_x n_y n_z} = C \times cos\left(\frac{n_x \pi x}{L_x}\right) \times cos\left(\frac{n_y \pi y}{L_y}\right) \times cos\left(\frac{n_z \pi z}{L_z}\right) \tag{3.12}$$

where C is an arbitrary constant and x, y and z positions within the room. $p_{n_x n_y n_z}$ now represents a three dimensional standing wave at a single point within the room. The inclusion of the time dependent factor $e^{i\omega t}$ would complete the expression, although as shall be shown, is not necessary in the following implementation of the room model in section 3.3.3.

The integer n refers to the 'mode order' in that particular dimension. Hence, where $n_x = 1$, there will be one nodal line in the x plane, occurring centrally. Where a single cosine term has a non zero integer value, i.e ($n_x = 1, n_y = 0, n_z = 0$), the mode is referred to as an *axial* mode. The plane wave visits only two parallel surfaces of the room, reflecting back and forth between each. Where two terms have an non-zero value, the mode visits four surfaces and is a *tangential* mode. Finally, where no terms are zero, the mode is *oblique*, visiting all six surfaces.

At any point r, where one of the cosine functions is equal to zero, the pressure also equals zero:

$$p_n(r) = 0 \begin{cases} L_x/2n_x \\ L_y/2n_y & \text{where } n \text{ is an odd integer} \\ L_z/2n_z \end{cases} \tag{3.13}$$

hence, wherever reflections of sound combine so as to oscillate in a manner specified by Equation 3.11, a standing wave is produced, with lines of zero pressure (nodal lines) wherever Equation 3.12 has cosine functions equal to zero. Any other frequency will not reflect in the same manner and therefore will not produce a standing wave resonance.

3.3.2 Excitation of the Enclosure

Having determined the unique set of eigenfrequencies and cosine eigenfunctions for a rectangular enclosure, it is now possible to formulate a model for the steady state point to point transfer function. These two points are the source and receiver, typically a subwoofer and listener respectively. In the frequency domain, this transfer function, assuming linearity of pressure superposition, may be seen as a summation of the contribution of each mode across the frequency range.

We begin by taking the wave equation in the form of Equation 3.1, which contains no source (equal to zero), and adding a source term ($q(\mathbf{r})$) on the right hand side:

$$\Delta p + k^2 p = -i\omega\rho_0 q(\boldsymbol{r}) \tag{3.14}$$

where $\mathbf{r}$ is an abbreviation of the x, y and z coordinates and ρ_0 is the density of air. It is possible to expand the source function $q(\mathbf{r})$ in a series of pressures p_n:

$$q(\boldsymbol{r}) = \sum_n C_n p_n(\boldsymbol{r}) \tag{3.15}$$

where:

$$C_n = \frac{1}{K_n} \int\int\int_V p_n(\boldsymbol{r}) q(\boldsymbol{r}) dV \tag{3.16}$$

where Kn is a scalar dependent on room volume and V is the room volume in m^3.

We note that this is only possible since the eigenfunctions form a complete and orthogonal set of functions:

$$\int\int\int_V p_n(\boldsymbol{r}) p_m(\boldsymbol{r}) dV = \begin{cases} K_n & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases} \tag{3.17}$$

3.3.3 Point Source Excitation

The source shall be modelled with the simple case of a point source. Whilst infinitely small in realistic physical terms, the point source may be considered valid where the dimensions of the source are small in comparison to the wavelength. At low frequency, this can be considered true (Pierce, 1989). A Dirac delta function is used to represent this source, with the form:

$$q(\boldsymbol{r}) = Q\delta(\boldsymbol{r} - \boldsymbol{r_0}) \tag{3.18}$$

where Q is the volume velocity. When substituted into Equation 3.16 this gives:

$$C_n = \frac{1}{K_n} \int\int\int_V p_n(\boldsymbol{r}) Q\delta(\boldsymbol{r} - \boldsymbol{r_0}) dV \tag{3.19}$$

A special property of the delta function allows us to apply the 'sifting property' where by:

$$\int f(x)\delta(x - x_0)dx = f(x_0) \tag{3.20}$$

resulting in:

$$C_n = \frac{1}{K_n} Q p_n(\boldsymbol{r_0}) \tag{3.21}$$

Ultimately we wish to find the steady state function - $p_\omega(\boldsymbol{r})$. We can expand this function in a similar manner to the expansion of the source function (Equation 3.15):

$$p_\omega(\boldsymbol{r}) = \sum_n D_n p_n(\boldsymbol{r}) \tag{3.22}$$

In order to obtain $p_\omega(\boldsymbol{r})$, we note that the unknown coefficients D_n must be determined. This is possible by defining it in terms of C_n, which can be achieved by substitution into Equation 3.14:

$$\sum D_n(\Delta p_n + k^2 p_n) = i\omega\rho_0 \sum C_n p_n \tag{3.23}$$

Using the relationship $\Delta p_n = -k^2 p_n$, this equation can be simplified and arranged in terms of D_n

$$D_n = i\omega\rho_0 \frac{C_n}{k^2 - k_n^2} \tag{3.24}$$

Finally, we may equate 3.22, giving us what is known as the 'Green's function' of the room, with a point source excitation at r and receiver at r_0.

$$p_\omega(\boldsymbol{r}) = iQ\omega\rho_0 \sum \frac{p_n(\boldsymbol{r}) p_n(\boldsymbol{r_0})}{K_n(k^2 - k_n^2)} \tag{3.25}$$

With complex boundary conditions, the k_n terms are usually complex quantities, therefore if:

$$k_n = \frac{\omega_n}{c} + i\frac{\delta_n}{c} \tag{3.26}$$

where the δ term is a damping constant then Equation 3.25 becomes:

$$p_\omega(\boldsymbol{r}) = iQc^2\omega\rho_0 \sum_n \frac{p_n(\boldsymbol{r}) p_n(\boldsymbol{r_0})}{K_n(\omega^2 - \omega_n^2 - 2i\delta_n\omega_n)} \tag{3.27}$$

where it is assumed that $\delta_n << \omega_n$. Here we have the transfer function of the room - the pressure at point r_0 as excited by a point source of volume velocity Q. r and r_0 are the eigenfunctions of 3.12.

3.3.4 Modal Damping

Within this thesis, the damping constant δ_n has been obtained in one of two ways. In the simple case of a rigid walled room, with an assumed uniform sound-field, the damping constant can also be assumed uniform, and therefore, reverberation time decreases linearly. In this case, there is a direct relationship between the weighted average δ_0 and the commonly used reverberation time RT_{60}

$$\delta_0 = \frac{6.91}{RT_{60}} \tag{3.28}$$

Damping is also affected by the mode type - axial, tangential or oblique (Howard and Angus, 2001). A wave travelling a greater distance in its reflective path is absorbed to a greater extent. The *grazing effect* (the absorption due to the angle of incidence as a wave travels along a surface) is assumed to be zero here. It follows that modes of the three different types are absorbed by differing amounts. We assume that, on average, each mode type is subject to the same additional absorption. Therefore, the constants ϵ are introduced. Where $n = 0, \epsilon = 1$ and where $n > 0, \epsilon = 2$.

To further increase the realism of the damping condition, we include both these mode specific ϵ values and the absorption coefficients α for the material on each wall. Morse (1948) shows that the approximate formula for the damping coefficient in this case is

$$\delta_0 = \frac{c}{8V}\left(\tfrac{1}{2}\epsilon_{n_x}\alpha_x + \tfrac{1}{2}\epsilon_{n_y}\alpha_y + \tfrac{1}{2}\epsilon_{n_z}\alpha_z\right) \tag{3.29}$$

3.3.5 Model Accuracy

The model for predicting the point to point low frequency sound-field is considered reliable for the investigations within this thesis, although it should be noted that its suitability has at times been questioned. Fazenda et al. (2006b) have argued that it is valid for subjective testing purposes even where damping is low, due to its ability to predict the *general* case response of a room (see also Fazenda (2004)). Specifically, as damping increases, the accuracy of the model reduces. For the purpose of the experimental work within this thesis, the approximate solution given by the model is considered acceptable. This is discussed further as necessary throughout the text.

It is acknowledged that other modelling techniques are available, should a more accurate representation of the sound-field be required. With current computer processing power however, such models are impractical for the real time, 'on the fly' simulations used in the testing methodology of this thesis.

3.4 Distribution of Eigenfrequencies

We may now consider the distribution of the modes within the room. It is possible to derive the average number of modes within a given enclosure. A number of equations can be found in the literature. It is beyond the scope of this thesis to formally derive these, but nonetheless, the ability to calculate the average number of modes within a given room is important. This has a number of implications, including calculation of the modal density at a specific frequency, and as will be seen, is important in understanding the quality metrics based on the distribution statistics which are investigated in Chapters 4 and 5. Equation 3.30 reveals the average number of modes up to frequency f (Bolt, 1939b).

$$N_f = \frac{4\pi}{3}V\left(\frac{f}{c}\right)^3 + \frac{\pi}{4}S\left(\frac{f}{c}\right)^2 + \frac{L}{8}\frac{f}{c} \tag{3.30}$$

where $S = 2(L_xL_y + L_xL_z + L_yL_z)$ and $L = 4(L_x + L_y + L_z)$.

This equation indicates that the number of modes increases with both frequency and volume. The distribution of modes was studied further by Schroeder, who derived the often quoted 'Schroeder Frequency':

$$F_c = 2000\sqrt{\frac{T_{60}}{V}} \tag{3.31}$$

where f_c is the crossover frequency, V the room volume in m^3 and T_{60} the decay time analogous to the RT_{60} reverberation time.

The literature review suggested that this frequency has been 'misrepresented' as the frequency above which we may no longer need to be concerned by modal frequencies. However, the frequency was originally intended to reveal the frequency above which a sound-field could be considered as statistically diffuse. Specifically, this is considered as the point where "the average frequency spacing between the maxima is 4/T" and "the average range of statistical fluctuations between maxima and minima is 10dB" (Schroeder, 1996).

The constant of 2000 ensures that the crossover frequency is met above the point where three modes are present within one modal bandwidth. As Skålevik (2011) points out, this figure is an arbitrary choice, based on Schroeder's experience measuring rooms. This concept is investigated further in Chapter 5.

3.5 Auralising a Modal Soundfield

Throughout this project, many listening tests studying modal sound-fields are performed. Key to these is the formation of *auralisations*, which allow the presentation of the modelled listening space to be accurately reproduced over headphones. This concept is not new, and has been successfully implemented by Fazenda (2004). However, the full method is presented here due to some differences from the referenced work, including new calibration, the high frequency content of the auralisation and the introduction of headphone equalisation.

3.5.1 Benefits of Auralisation

Auralisation of virtual listening rooms has a number of benefits. Firstly, the construction of a large number of listening environments to test in is simply not possible. Secondly, modelling enables the modification of individual modal parameters in isolation. In real rooms, the interaction between multiple parameters would not allow this. Thirdly, with auralisations, it is possible to switch a listener quickly between two rooms, eliminating the problems associated with our short 'acoustic memory' (Conrad, 1964; Crowder and Morton, 1969). Finally, using this technique, it is possible to modify the low frequency region whilst keeping the high frequency constant. In a real room, the high frequency reproduction would differ in each different room. When trying to assess the low frequency, this high frequency change would be a confounding factor.

Figure 3.5.1 shows the schematic for the auralisation of an audio test sample. Essentially, the auralisation is broken into two frequency ranges:

1. The high frequency presented is the original music signal high pass filtered at a chosen crossover frequency using a fourth order Butterworth filter. It therefore contains no specific 'room' content, although as the samples used are commercial recordings, the high frequency region naturally contains those room artefacts present in the original recording space. This method differs from the previously mentioned work of Fazenda (2004), where the high frequency content presented was the original recording convolved with a measured room impulse response. It may be argued that the addition step provides spatial and temporal cues to help create the impression that the listener is located in a real room. However, during the set-up of the listening tests for this thesis, it was concluded that no real benefit was gained through this additional step, and in fact, removing this stage was beneficial in that is eliminates the need

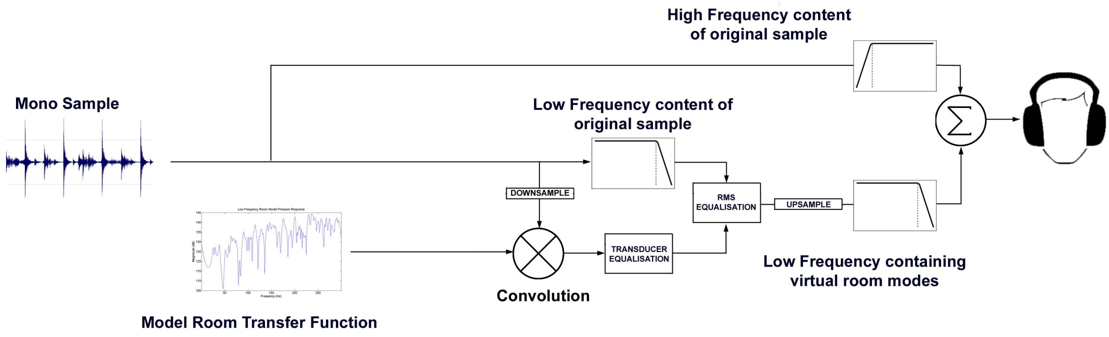


Figure 3.1: A schematic of the auralisation process used throughout testing

for a steep filter to remove the low frequency from the real room convolution. A steep filter can cause audible artefacts and a filter not steep enough leaves audible elements of the original measured room response at low frequency, thus interfering with the low frequency model under test.

2. The low frequency response is first modelled, and then convolved with a downsampled version of the input audio. After transducer equalisation and level calibration (see following sections), this is upsampled using MATLAB's *resample.m* function. Finally, this low frequency signal is low pass filtered at a frequency matching that of the high frequency, ensuring that the two regions cross over with an overall flat magnitude response. The high and low regions are then summed to produce the output signal.

A number of the elements within this process are explored in greater detail in the following sections.

3.5.2 Monaural Listening

Mono samples are used throughout the process. The low frequency model generated the room response at a single listening point. This results in a monaural sample being replayed through headphones to both ears. It is acknowledged that this is a possible source of error, although at the low frequencies tested, the typical distance between the two ears should not have a large effect.

3.5.3 High and Low Frequency Alignment

When splitting the audio into high and low frequency regions, and modelling the low frequency, it is important to ensure that the balance of levels between them,

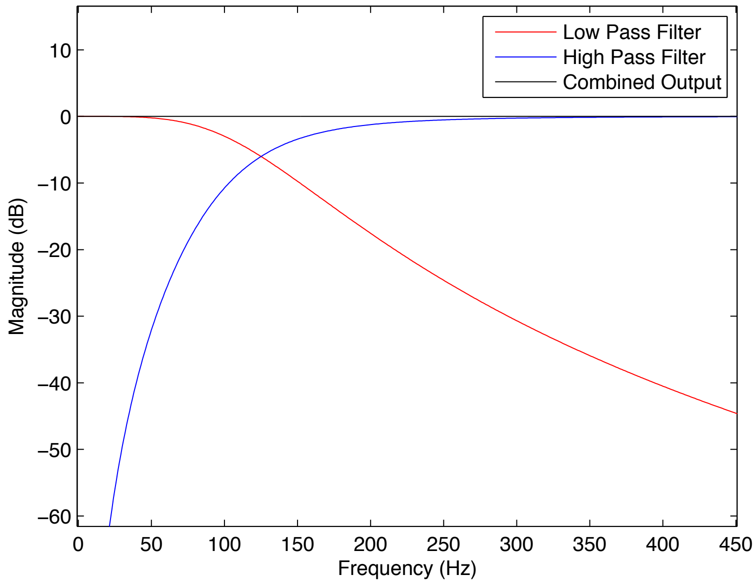


Figure 3.2: Low and high frequency crossover filters for auralisation

when recombined, produces a convincing output where audio reproduction is realistic and the low frequency content representative of that perceived in a real listening scenario. If the low frequency is too quiet, differences which are to be assessed may not be perceived. However, if the opposite is true, certain degrading effects may be exaggerated. In order to achieve the correct balance, the low frequency is subjected to an adjustment based upon the amount of bass energy present in the original sample. The original mono sample is low pass filtered at the crossover frequency (dependent upon the specific test) and a root mean squared (RMS) value taken to determine an overall energy level present in this frequency region over the duration of the sample. The RMS is also taken for the convolved low frequency model and mono sample. The difference between these two levels can then be used to obtain a gain factor which is applied to the modelled low frequency to ensure it is consistent with that of the original sample.

The filters used are also important in ensuring a natural overall response. Figure 3.2 shows the transfer functions of the high and low pass filters which are based upon a fourth order Linkwitz-Riley topology, resulting in a flat overall response. Figure 3.3 shows an example mono sample, filtered high and low regions and the combined response.

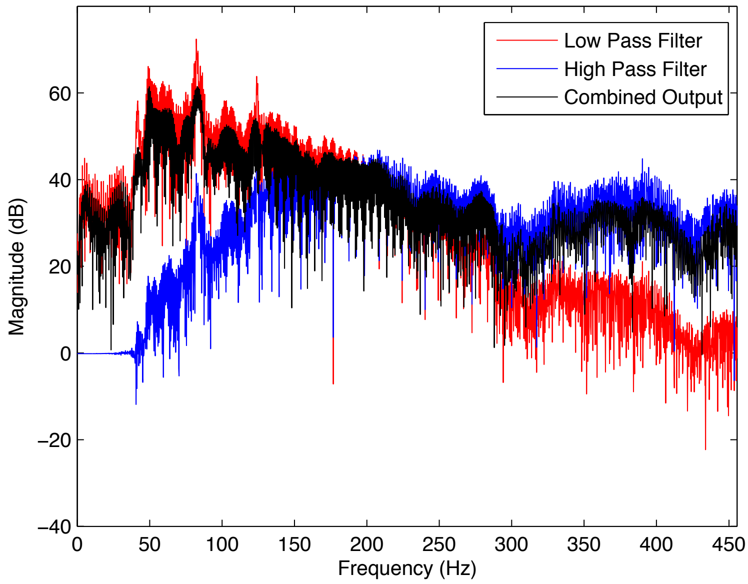


Figure 3.3: Crossover filters applied to a test sample

3.5.4 Transducer Equalisation

A pair of Sennheiser HD-650 headphones are used throughout the testing reported. A high end reference model, these headphones are an open back design and considered to have a good extension into the low frequency region. In order to ensure the flattest response possible, their transfer function was measured, using the sine sweep method and a Neumann K900 dummy head, whose microphones are located at the position of the ear drums. This response was then exported to MATLAB and a 3000 tap FIR equalisation filter created between 10 and 1000Hz (see Figure 3.5).

This equalisation was applied to the low frequency model after convolution with the stimulus and prior to level alignment between high and low frequency regions.

As mono samples were used during testing, equalisation was applied which corresponds only to the left side transducer measurement. Measurements of the two transducers show little variation and so this was considered appropriate (Figure 3.4).

The interface between computer and headphones was a professional M-Audio sound card and was, as such, considered a negligible source of replay error.

3.5.5 Playback Level

When playing back audio files during testing, it is important to know at what level sound is being received at the ears. Using a sound level meter it is possible to determine the sound pressure level (SPL) of pure tones. However, when using

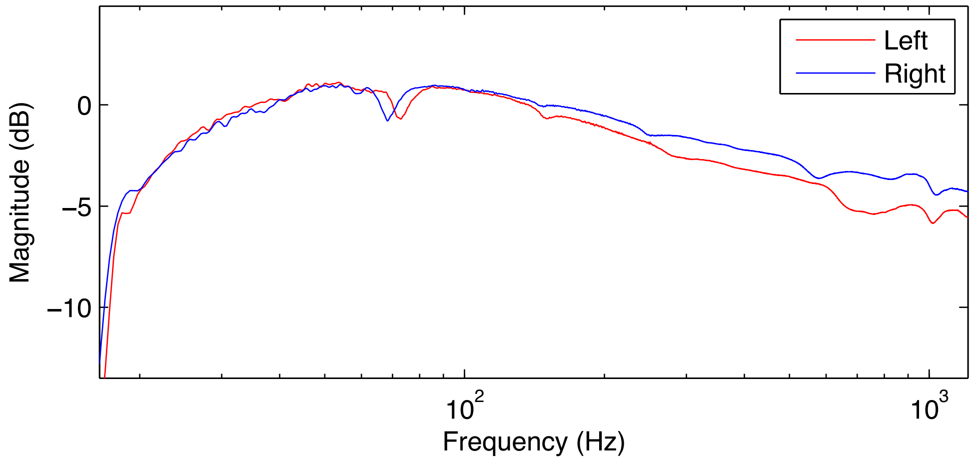


Figure 3.4: Comparison between measured responses of left and right headphone transducers

musical signals, it is much more challenging to report a 'loudness level'. The term loudness level is here defined as the equivalent sound pressure level of a complex, time varying signal.

The auralised music samples were therefore calibrated by measuring the loudness level as defined in ITU-R BS.1770-1 (ITU, 2006). This standard refers to an 'audio program loudness' which has units which are equivalent to a decibel level. In order to calibrate, a standard 1kHz sine tone is used as a reference. The tone is played through the headphones to a B&K Head and Torso Simulator whose microphones are placed at the ear drum, and were calibrated with a standard B&K calibrator. The level is measured with a sound analyser and output from the soundcard adjusted until the level is 85dB. This tone is then passed through the ITU loudness algorithm to obtain a reference loudness value for the 1kHz tone appearing at the ears at 85dB. Any auralised music sample can then also be passed through the algorithm and adjusted by the appropriate gain factor such that the overall sample loudness can be considered perceptually similar to the 1kHz tone.

3.5.6 Suitability of Auralisation Signals

When considering audio signals for testing purposes, it is essential that a suitable signal be chosen. ITU BS.1116 (ITU, 1994) asserts that only critical material should be used, which is defined as "that which stresses the signal under test".

Signals may be considered natural (speech, music, natural noise) or artificial (pure tones, sweeps, algorithmic noise). Within the scope of this thesis, only musical signals are considered. It is uncommon for speech to contain spectral content

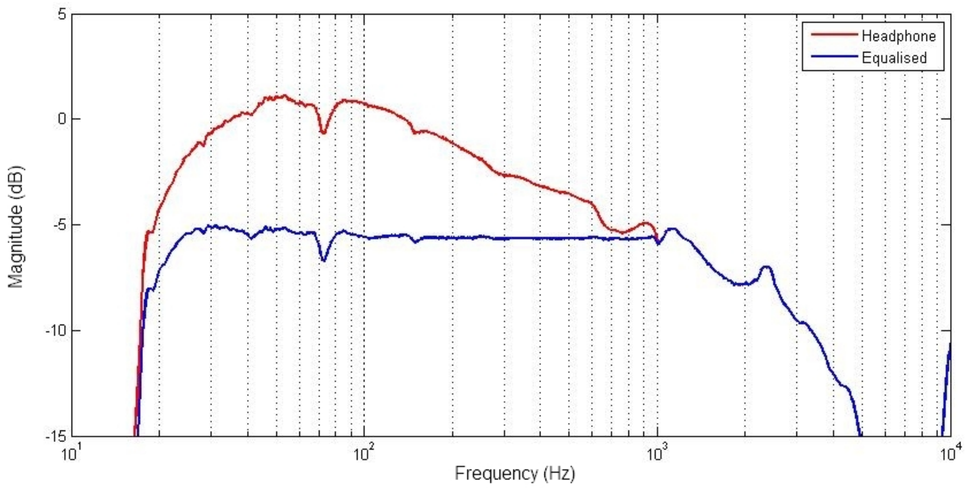


Figure 3.5: The original measurement of the headphone's left transducer and the resulting response after application of an equalising filter

at the low frequencies of interest. Auralisations using the room model described are produced using commercially recorded music samples extracted from CD. This places certain limitations upon them, such as the sampling rate (44100Hz) and bit depth (16). Anechoic recordings are not necessary for this type of auralisation, as the listening spaces under consideration would usually be used for the reproduction of either recorded audio, or that which is being performed within another non-anechoic space.

All sample selected were considered with respect to:

- Temporal characteristics: As has been shown in the work of Toole and Olive (1988), the audibility of resonances varies according to the temporal characteristics. Perception varies for continuous and impulsive sounds. Therefore, in order to test for the most realistic music cases, signals should be chosen which contain fast attacks and decays, and also sustained elements.

- Spectral characteristics: Naturally, it is particularly important to select program material which excites the audio range. Indeed, Zacharov et al. (1998) cite the problem of finding suitable material below 120Hz. Musical samples used within this thesis have each been checked for low frequency content.

- Sample familiarity: Listeners should not be overly familiar with a particular sample. For example, if an excerpt has been a commercial success, biasing can be introduced, particularly when asking subjects to evaluate audio quality.

3.6 Musical Excerpts

Throughout the following listening tests, four different music excerpts have been used.

1. 'LEN' - may be considered to be in the 'pop' genre. It is around six seconds in length and contains a number of sparse bass notes with short attack and decays, and a sparse kick drum rhythm. This sample has been used in previous research (Avis et al., 2007) and is considered to be particularly good at highlighting modal artefacts whilst remaining clearly 'musical' in nature.

2. 'HC' - is of similar length and may be considered to be in the 'jazz' genre. It is a solo double bass refrain, with a greater number of notes. It is therefore likely to excite a wider band of frequencies in the experimental region. The natural resonant behaviour of the double bass results in longer decay times of the notes themselves, and therefore many of the notes can be heard to blend into one another.

3. 'FW' - which consisted of a single bass line with a number of notes, but less resonant in nature than the HC sample.

4. 'FG' - and can also be considered in the 'pop' genre, this time with a synthesised bass-line with very short decay envelope.

Spectrograms of the four samples are shown in Figure 3.6.

3.7 Summary

This chapter has introduced the basic theory necessary for an understanding of low frequency sound-fields in enclosed spaces. As these spaces are to be modelled as 'virtual rooms' in the following chapters, the derivation of the low frequency room response was presented, assuming a lightly damped rectangular room with rigid walls and point source radiation.

The use of this model in producing auralisations for presentation to subjects has been shown, including the choice of stimuli, alignment of high and low frequency, replay levels, calibration and headphone equalisation.

Building on both the theory presented, and the review of the literature in Chapter 2, the following part presents the search for optimal modal distribution parameters and perceptual thresholds for decay.

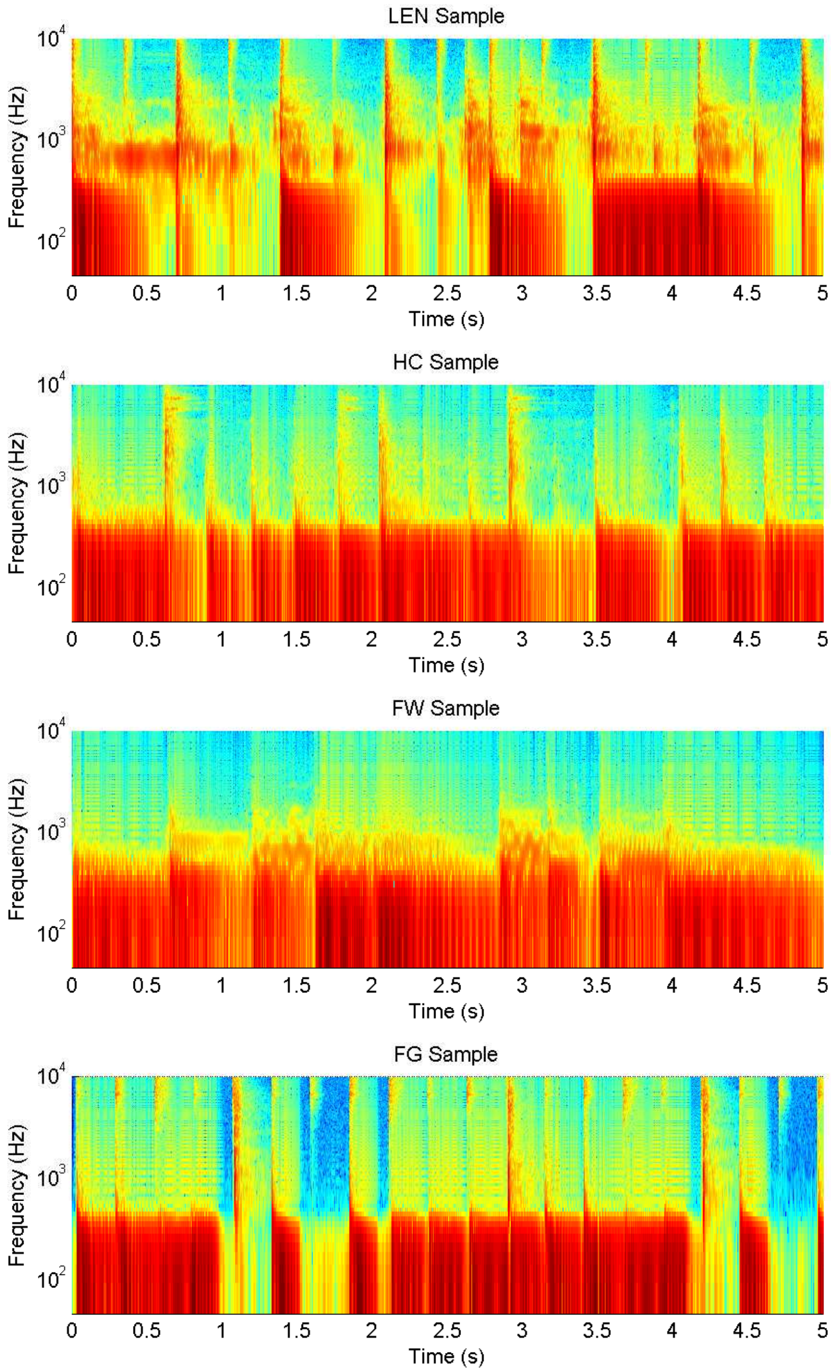


Figure 3.6: Spectrograms of the four music samples used throughout testing

Part II

Modal Parameters

Chapter 4

Optimal Modal Spacing

4.1 Introduction

Chapter 2 presented previous research in the field of modal control. It was shown that one of the earliest attempts to control the low frequency region was through modification of the modal distribution. For the simple case of a rectangular enclosure, Equation 3.11 showed that each individual modal frequency may be calculated, allowing us to know the *spacing* between each. It was also shown that, using statistical methods, we may estimate the average number of modes present within a given frequency region (Equation 3.30). We therefore have two related concepts which we may use in an attempt to better understand the modal distribution - *modal spacing* and *modal density*. This chapter focuses on the concept of spacing. The perception with reference to density is explored further in Chapter 5.

The difference in perception due to different spacing between modes is important to understand. If there is a particular spacing which affords a clear subjective improvement, it is possible to use this as a design goal for new construction, or even to attempt to modify the distribution within an existing room.

The interest in finding an 'optimal spacing' can be seen through a review of the literature, a number of metrics having been defined which use the statistics of typical spacing between modes as a criteria for determining good quality rooms. Section 4.2 briefly covers the theory behind such metrics, and presents a critical analysis into their validity. A number of underlying problems with these objective metrics are discussed, revealing the necessity for a better *subjective* understanding. In light of this need, Section 4.4 details experimental work which attempts to find a subjective answer to the question 'is there an optimal spacing between two individual modes?'. Results are presented in Section 4.4.3 and the following sections discuss how the results of this simple test can relate to real room scenarios, and present a number of issues which arise when attempting this. Finally, this investigation concludes with a discussion of the implications of subjective findings for room design and treatment.

4.2 Spacing Theory

4.2.1 Modal Degeneracy

As shown in Chapter 3, given a set of room dimensions, assuming a rectangular shape, it is possible to calculate all modal frequencies. With this dependency on the room dimensions, many attempts have been made to optimise the room by altering these dimensions, and thus the modal distribution in such a way as to give the 'best' subjective response. In cases of new constructions, room designers may be afforded the luxury of having control over some or all of these dimensions. It was originally suggested that good and bad *ratios* exist - the rationale being that a carefully selected ratio will allow the problem of modal degeneracy to be avoided. Why then, is such weight put upon this avoidance of degeneracy?

Firstly, two modes occurring at the same frequency will add together and the magnitude of reinforcement of that frequency will be increased. Bucklein (1981) and Fryer (1975) showed that the detection of resonances increases with increased magnitude. It is implied that a 'detected' resonance is undesirable as it alters the perception of the original signal. A second, related issue is that there is a reduced chance of an evenly spaced alignment of modes. As eigenfrequencies have a relationship proportional to the dimensions, if two of these are similar, modes associated with the similar dimension will occur at similar frequencies. Degeneracy is therefore considered problematic. Multiple strong resonances and uneven spacing are also unconducive to obtaining a *flat frequency response*.

4.2.2 The Flat Response

When a response is considered flat, within the frequency domain, it will contain no peaks or dips (magnitude deviations). Where such irregularities do exist, they modify the overall sound for the listener by altering the amplitude at certain frequencies. Furthermore, where such peaks and dips are present, they are associated with decay times at that particular frequency. These decays are related to the parameter known as the Q-factor, where Q refers to the 'quality'.

The system exhibiting a flat frequency response is essentially considered desirable because of its ability to reproduce the original audio signal. This is the overwhelming assumption noted within the literature (see (Cox et al., 2004) and (Fazenda

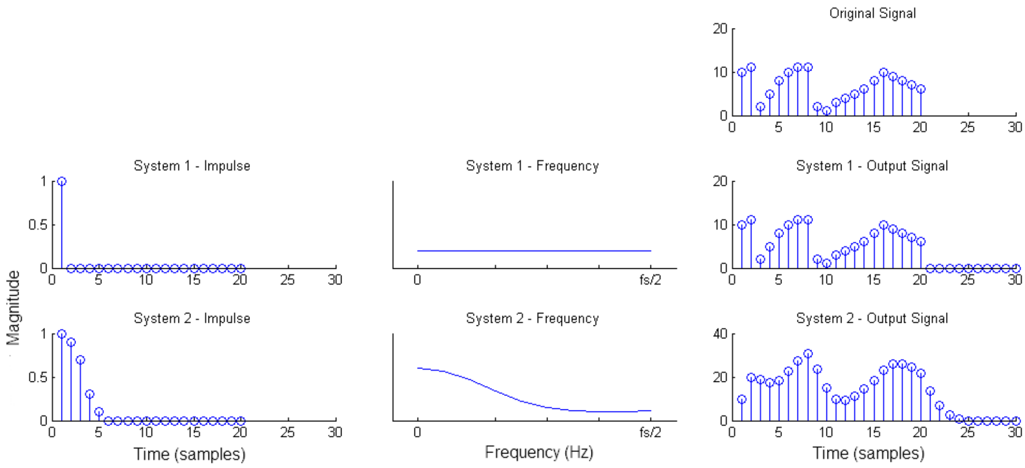


Figure 4.1: Comparison of convolution with a Dirac impulse and a decaying impulse

and Davies, 2004)) and also through conversation with studio designers, end users and discussion on numerous websites. A completely flat response has a number of properties which lead to this assumption. When an audio signal is passed through any system, the mathematical process of convolution takes place between the signal and the transfer function of the system. In this case, the room itself is that system.

The flat response can be shown to be equivalent to an impulse. This is an infinitely loud, infinitely short signal at some point in time. In the digital domain, this can be effectively expressed by the Kronecker delta:

$$\delta = \begin{cases} 0 & n < 0 \\ 1 & n = 0 \\ 0 & n > 0 \end{cases} \tag{4.1}$$

If we produce this pulse, and take the FFT, a flat frequency response between DC and half the sampling frequency is observed. The inverse is also true, if a room were to exhibit a completely flat response, its time response will be an impulse. In reality, this is never fully achieved, although a response that tends towards flat, or even smooth, will have a short time response. This short response is often associated with high quality reproduction (Fazenda and Davies, 2004). This would be our 'ideal' case, where we define a high quality system as one which will faithfully reproduce an audio signal fed to its input. Figure 4.1 shows that due to the convolution process, a short response is better at maintaining the original signal than that of a longer one. Avis et al. (2007) suggest that the reduction in decay time may be one of the most important factors in the perception of modes.

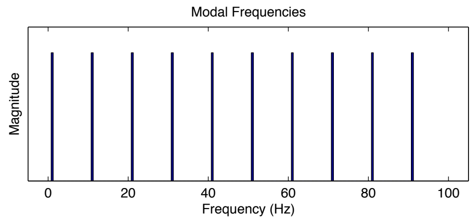


Figure 4.2: Example scenario of ten modes evenly spaced with equal magnitude

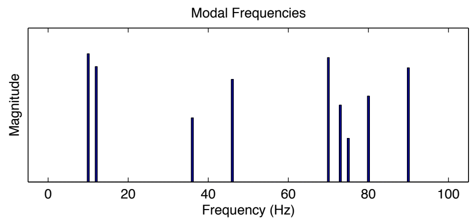


Figure 4.3: Example scenario of ten modes randomly spaced with differing magnitudes

4.2.3 Aspect Ratio

If we return to the original idea that specific aspect ratios may help us distribute the modes, we see that such a ratio may help ensure a flat response and therefore short decays of resonances, and a faithful representation of the input signal. It is easy to see why this theory might be appealing if we glance at Figures 4.2 and 4.3, one homogeneously spaced with ten modes, and one with randomly spaced modes. Our eyes 'join up' the modes in the frequency response. It is easy to see why the evenly spaced modes look best. It follows that an arrangement of the modal frequencies corresponding to a more homogeneous frequency response will result in an improvement in the perceived audio reproduction quality as decay times reduce and the amplitude effects of peaks and dips become less noticeable. This then becomes our 'optimal' condition - where the decay time of each mode is as short as possible.

This idea underpins much of the study of aspect ratios. It follows that, if the length, width and height of the room are carefully chosen, the modes will be distributed in such a way that the overall effect is a smoothing out of the frequency response. Bolt (1939a) was the first to publish this idea, producing his famous 'blob', which defines theoretically good aspect ratios. In another attempt, Louden (1971) published a table of ratios in descending order, best to worst. Chapter 2 reviewed further attempts to define optimal ratios.

4.2.4 Problems with Spacing Metrics

There are a number of problems associated with the criteria for 'good' spacing used in much of the literature:

1. Methods such as that of Louden (1971), consider only the attempt to avoid modal degeneracy. An algorithm searches through a set of ratios, and selects the ones with the smallest standard deviation between modal frequencies. This method however, fails to consider the possibility that, even though modes may uniformly spaced, there may be no subjective improvement. Is the smallest standard deviation between eigenfrequencies the most perceptually relevant measure?

2. The metrics do not account for changes in perception as the frequency increases. It has been shown in Chapter 3 that a greater number of modes occur as the frequency increases. Over what range should we look at the uniformity of spacing? There is likely a difference in the perception of spaced resonances as a function of frequency. Indeed, it has been reported that we are more sensitive to frequency irregularity at higher frequencies (Avis et al., 2007; Karjalainen et al., 2004). Current aspect ratio metrics apply no specific weighting factor dependent on frequency.

3. A third problem with these criteria is the lack of account for positioning of the loudspeaker and listener in the room. It is assumed that each mode contributes equally, and with the same phase, to the overall response. In order for excitation and reception of every mode, the source must be placed in one corner, and the listener diagonally opposite. Furthermore, the response is significantly altered where two modes are out of phase with each other, as a result of the eigenfunctions, or 'mode-shapes' (Equation 3.12). The importance of these mode shapes will be fully explored in Chapter 5.

4. Figure 4.2 shows all modes having equal amplitude, which makes it tempting give them equal importance. However, in reality, modes of different types, axial, tangential and oblique will likely have different amplitudes, due to different levels of damping. This is also unaccounted for by the spacing metrics. It is likely that the perception of two spaced resonances will differ under differing damping conditions.

5. Finally, the volume of the room is often not considered. While Bolt (1946) does give a range of validity, Toole (2006) explains that it is rarely considered

in practice. A change in volume, whilst retaining the *relative* spacing, will alter the *absolute* spacing between modes.

These omissions form the rationale for the experiment in Section 4.4, investigating the subjective response to the interaction between two modes. The experimental work addresses a number of the issues above, and the findings are extrapolated in order to discuss the results with regard to each of the concerns highlighted here. Before the test method is described, the interaction between two modes is further explored.

4.3 The Spacing of Two Resonances

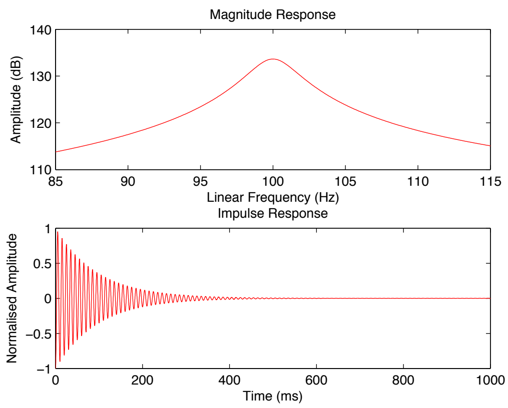
In order to determine if a frequency and damping dependent optimal spacing can be defined, the simple case of two adjacent modes is studied, both analytically and subjectively. In the case of a single mode, the decay, analogous to the RT_{60} reverberation time concept in the diffuse field, can be related to its Q-factor by the approximate equation:

$$RT_{MODAL} = \frac{2.2Q}{f_0} \tag{4.2}$$

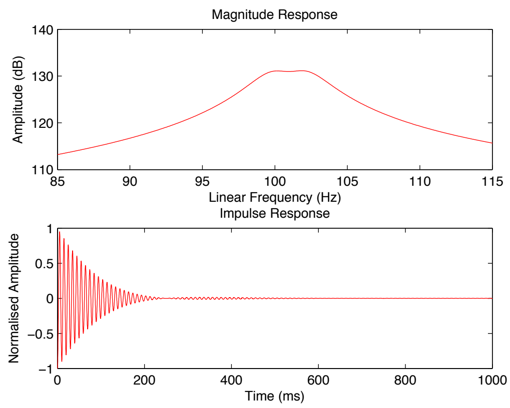
where Q is the quality factor of the mode and f_0 the frequency.

We see that as the Q increases, the decay time, often described as ringing, increases. We observe the effect of adding a second resonance by simply summing them. Figure 4.4 shows a second resonance, first at 102Hz and then at 105Hz, where the Q factor is 30. Secondly, Figure 4.5 shows the same resonant frequencies, but where each has a Q factor of 50. The differences in these figures highlight the fact that we must not limit our understanding of a ‘good’ spacing simply to the frequencies at which two modes occur.

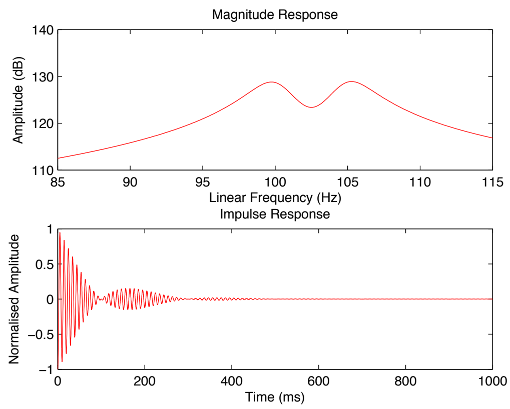
A simple visual investigation of the effect of altering the spacing between the two individual resonances reveals a clear impact on the temporal response. As the second frequency moves away from the first, the frequency response appears to flatten at the peak, and results in an apparent reduction of the overall decay time from the case where both resonances were at 100Hz. However, as the shift continues, the magnitude frequency response reveals a large dip and the resulting impulse response begins to show distinctive amplitude modulation. This is associated with the interaction between the two resonances and at these frequency differences, sound identical to first order beats as described in many psychoacoustic textbooks (eg. Howard and Angus (2001); Kuttruff (2007)). Where the Q is increased to 50 in



(a) 100Hz 100Hz

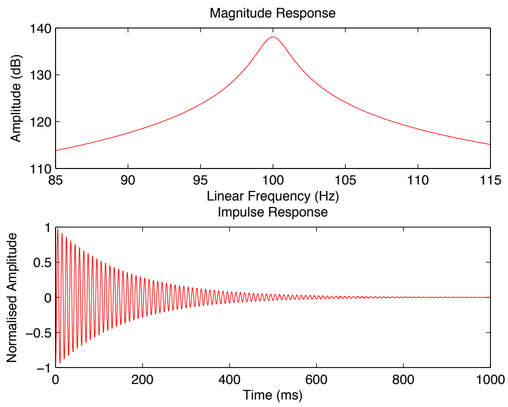


(b) 100Hz 102Hz

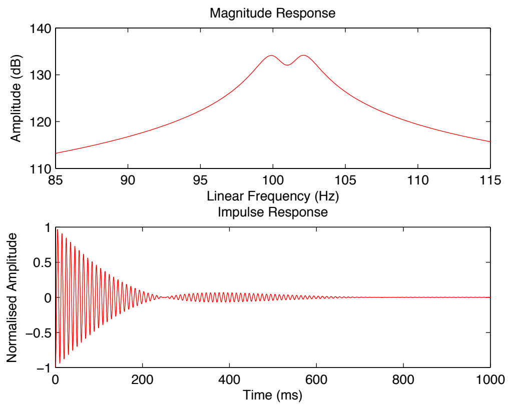


(c) 100Hz 105Hz

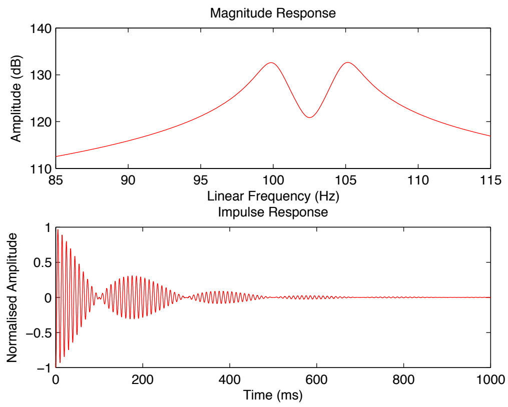
Figure 4.4: The spacing of two resonances - Q = 30



(a) 100Hz 100Hz



(b) 100Hz 102Hz



(c) 100Hz 105Hz

Figure 4.5: The spacing of two resonances - Q = 50

Figure 4.5, the onset of these beats appears at a closer spacing. Between two sinusoids, the beating frequency is equal to the frequency between them. Therefore, the greater the spacing, the faster the beating effect.

Once again it is possible to fall into the trap of making assumptions, based upon a visual inspection of the figures, as to the perceived quality of an audio stimulus when passed through these resonant systems (assuming the audio material were to excite the corresponding frequency range). If a perceptual improvement is associated with a shorter decay, then moving the second frequency away from the first is preferable. However, the introduction of beating is likely to be highly detectable to the listener and perhaps undesirable (Rasch and Plomp, 1999). Therefore, by taking into consideration both the modal frequencies *and* Q-factors, we see that the search for an 'optimal' spacing involves more than simply aligning the frequencies with uniform spacing.

4.3.1 Definition of Optimal

As this chapter presents an investigation into defining optimal spacing of modes, it is important to remember the definition of optimal within the scope of this work as outlined in Chapter 1. In a general sense, the 'optimal' is the value of that parameter which produces the best overall subjective response. Within this chapter therefore, 'optimal' refers to *the situation in which a system imposes the least degrading artefact onto the original stimuli*. In terms of two spaced resonances this is considered to be the shortest audible decay time, but where any perceived beating is not considered degrading. Such a definition allows the following listening test to be performed.

4.4 Subjective Testing

In order to answer the question posed, the two spaced resonances were artificially modelled. Resonances were generated using the Green's Function described in Section 3.3.3. It is common to supply this equation with an array of modal frequencies for a full room model. However, here, a fixed array (eg. $\omega_n[100, 102]$) of the two modes was fed into the equation, with the source and receiver eigenfunctions as unity, to obtain the system's response.

Equation 3.27 produces a complex output in the frequency domain. This response was transformed to the time domain, giving the impulse response of the system in question. It is common practice in psychoacoustic testing of this nature to convolve the system response with an input stimulus such as a test tone or musical refrain to determine the perception of that stimulus to that system response.

However, in this case, the impulse itself is played to the subject as the test stimulus, in order to obtain absolute thresholds between decay reduction and the onset of beating. A number of stimuli were considered during pilot testing, such as noise, and also sine bursts. However, these were rejected on the grounds that there would be unnecessary masking effects. Single frequency decaying sine tones were considered, but the decay length of the tone would in some cases be responsible for masking the decay of the resonance itself. Furthermore, a tone of increasing bandwidth would be needed to excite the full frequency range under test as the spacing increased. Using a musical stimulus, whilst realistic in a listening scenario, does not allow absolute thresholds to be detected. As such, an optimal value measurement from the impulse itself, corresponding to the 'worst case scenario', was found to be appropriate.

4.4.1 Method

As mentioned, both the frequency and the decay must be considered to better understand spacing. Therefore, two independent variables were chosen - the frequency of the first mode (63, 125 and 250Hz) and Q-factor. Four Q-factors (10, 20, 30, and 50) were chosen to represent a broad range, typical in listening conditions. The spacing of the second resonance was adjusted by way of a slider on a graphical user interface (Figure 4.6). Samples were generated immediately each time the slider was moved, reducing the resolution error which would be inherent if playing back pre-processed samples with a set interval. All programming was carried out in MATLAB®. Subjects were asked to adjust the slider to that point where the shortest decay time before the audible degradation of beats occurred, thereby corresponding to the aforementioned definition of 'optimal spacing'. Prior to the test, explanation of the differences in presentation sounds (long decay, shorter decay, and beating effect) were explained and demonstrated, along with images in the time domain. No time domain images were displayed during the actual tests to avoid bias.

The dependent variable was therefore the frequency spacing required by the subject. Eleven subjects were tested, in quiet studio conditions. Each subject was a student or lecturer of Music Technology at the University of Huddersfield and was given time to practice before the test commenced. The block of twelve tests was randomised, and repeated three times by each subject.

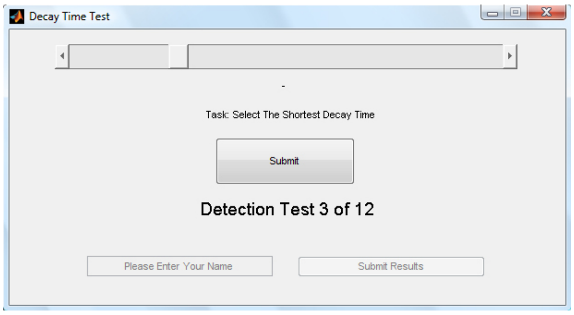


Figure 4.6: Screenshot showing the Graphical User Interface for the optimal spacing test

4.4.2 Calibration

A laptop computer was separated from the monitor and mouse by way of an acoustic screen. Stimuli were auditioned over a pair of Sennheiser HD-650 headphones. The presentation levels of the three frequencies were weighted to ensure that the perceived level of each sample was the same. Each was presented according to the 90dB equal loudness contour (Robinson and Dadson, 1956). The calibration was achieved by placing the headphones on a Neumann K900 dummy head and measuring the output level of full amplitude sine tones at each of the test frequencies. Scaling values were then applied to ensure correct playback level. These were: 63Hz - 0.551, 125Hz - 0.299, 250Hz - 0191.

Each impulse was filtered with a third order Butterworth low pass filter at 1kHz, and windowed with a Tukey window providing a 12 millisecond opening and closing time in order to remove any spectral artefacts from the beginning and end of the signal which may give audible clues to the listener. The sound card used was a professional M-Audio Firewire 410 and was considered to be a negligible source of replay error.

4.4.3 Results

Figure 4.7 shows the mean spacing across 11 subjects and reveals a clear trend. As the Q-factor increases, the optimal spacing needed to provide the shortest decay reduces. This result makes good sense, as the higher Q-factors are associated with a narrower bandwidth. Therefore they must be spaced closer together for the interaction of the two to produce a smooth response. Furthermore, being ‘more discrete’, the further apart they move, the more obvious the beating effect as each beat has a greater amplitude.

When comparing the test frequencies, it is seen that higher frequencies require a greater spacing between the two resonances. Interestingly, this puts the perceived

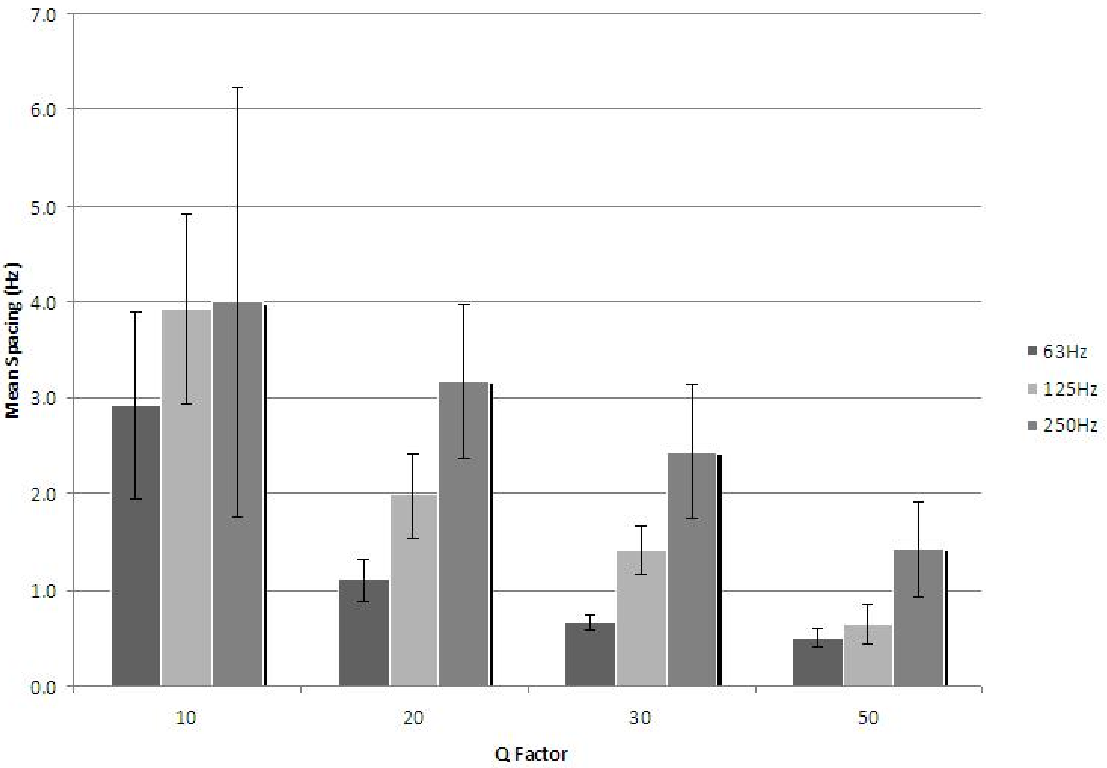


Figure 4.7: Subjective optimal mean spacing across Q-Factor and frequency

optimum in direct contradiction to the natural effect within a room, as the actual spacing of modes decreases with increased frequency. Somewhat counteracting these results, we see that the level of uncertainty shown by the standard deviation error bars also increases with frequency, suggesting that the concept of an optimal spacing becomes less meaningful at higher frequencies. An increase in standard deviation is also observed with decreasing Q, suggesting increasing difficulty in hearing modal effects as Q decreases. Again, this result makes sense - the reduced decay times of lower Q tests make it more difficult to accurately perceive the length of that decay compared to another or to notice the beating effect. It is noted that no direct comparisons of decay were made between cases. The 'optimal' spacing only refers to that frequency and that Q. It is not the case that with the spacings obtained, all impulses are equal in their perceived *quality*. For example, at 63Hz and Q=50, even with the optimal spacing of 0.5Hz, the perceived ringing is longer than 250Hz at Q=10.

Analysis of variance was carried out to ascertain the level of significance across the variable parameters. ANOVA shows that both Q Factor and modal frequency are highly significant at the 1% level ($p<0.01$), which indicates the success of systematic testing. Figure 4.7 reveals that there is less difference across the three frequencies at

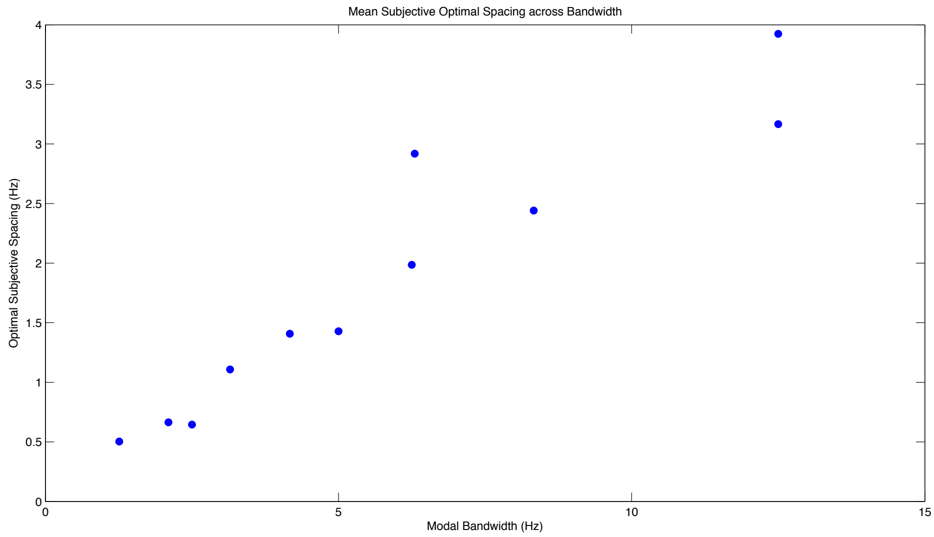


Figure 4.8: Mean subjective optimal spacing presented in ascending modal bandwidth

low Q (shorter decay times). The large overlap between frequencies suggests there is no significant difference between frequencies at these levels of Q.

Although both factors are themselves highly significant, it is useful at this point to note that they can be related in terms of a single value - modal bandwidth. The relationship is shown in Equation 4.3

$$B_w = \frac{f}{Q} \tag{4.3}$$

Table 4.1 and Figure 4.8 consider each of the 12 test scenarios in ascending bandwidth (the bandwidth of the first mode is reported, with the second mode's bandwidth assumed to be the same due to the relatively small spacing). The results again show a clear trend.

Bandwidth (Hz)	1.26	2.10	2.50	3.15	4.17	5.00
Modal Spacing (Hz)	0.50	0.66	0.65	1.11	1.41	1.43
Bandwidth (Hz)	6.25	6.30	8.33	12.50	12.50	25.00
Modal Spacing (Hz)	1.99	2.92	2.44	3.17	3.92	4.00

Table 4.1: Mean subjective optimal spacing presented in ascending modal bandwidth

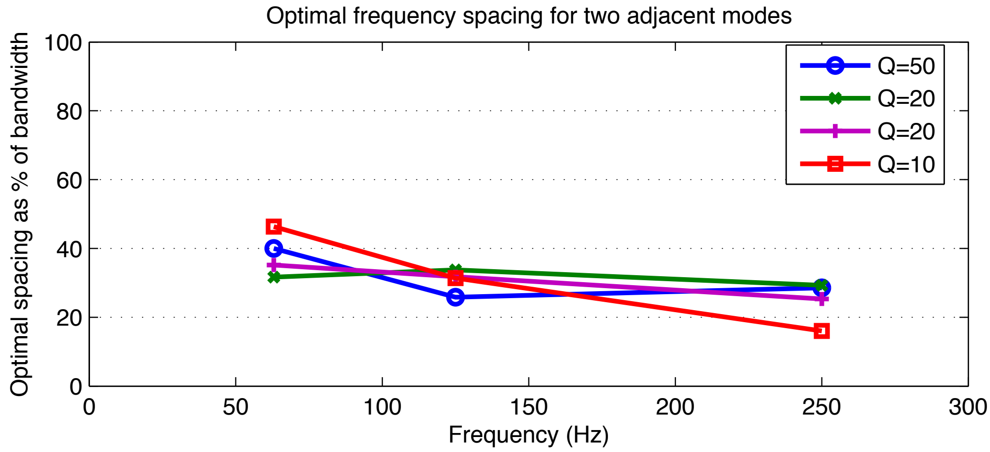


Figure 4.9: Optimal Spacing across ascending bandwidth for the four different Q Factors tested

4.5 Predicting the Subjective Result

With such trends revealed by the subjective testing, it is now possible to consider how we may predict this result, before expanding the study to more complex spacing scenarios.

It can be shown that the modal bandwidth is a significant indicator of the optimal spacing. Figure 4.9 shows the subjective results as a percentage of the modal bandwidth. This figure reveals that, for Qs of 20, 30 and 50, regardless of frequency or Q, the optimal spacing lies between 25 and 40%. At lower Qs, the standard deviation becomes higher (see Figure 4.7), and results are less reliable. Nevertheless, they appear in a similar region by and large. These results were confirmed by comments from subjects who each stated that the shortest impulses (lower Qs) were significantly harder to judge than those of longer length.

If we take an average of these results lying between 25 and 40%, we arrive at an optimal spacing for a second resonance of one third the bandwidth of the first. The result can be further confirmed by taking an approximate regression line through Figure 4.8, resulting in a y intercept at the origin, and therefore an equation of $y = 0.33x$. Taking this one third bandwidth spacing, we observe good correlation with the subjective results, independent of both frequency and Q. Whilst clear correlation is evident, it is of greater importance to understand the underlying principle behind this if the results are to be useful in cases more complex than two simply spaced modes. The question must be asked: why is this the optimal spacing?

To answer this, it is useful to observe the effects in the time domain. With the impulse response plotted with absolute values and on a decibel scale, an interesting observation is made. The relationship between two modes spaced according to the

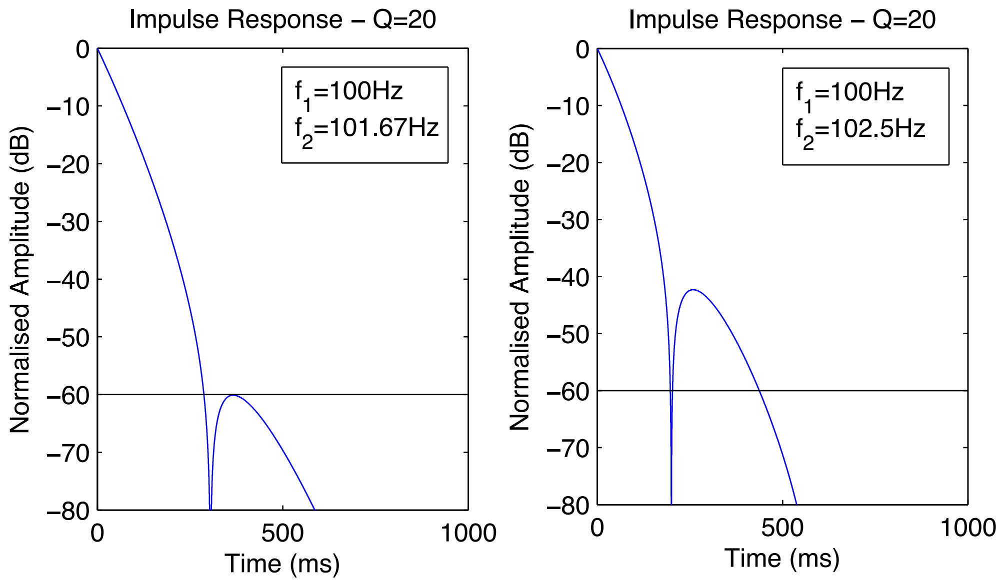


Figure 4.10: Absolute decay plotted in dB for a) optimally spaced modes and b) spacing greater than the optimal

one third bandwidth rule reveals that the decay initially falls below -60dB quicker than for a single mode, and that the first beat between the two frequencies rises to a peak amplitude no greater than -60dB. Where the spacing is greater, the initial drop to -60dB occurs quicker, but the peak amplitude of the first beat is then increased, becoming audible to the listener, who hears this as both degrading, and an additional part of the overall decay.

It appears that the answer to the original question of where the optimal point lies between the two degrading effects of excessive decay and obvious presence of beating can be answered objectively - where the peak of the first beat first becomes audible (Figure 4.11). It could be argued that the listening level is therefore important. However, it is suggested that this is not the case. If we increase the replay level, the beats may indeed become more audible. The natural response therefore is to decrease the spacing between resonances. This then results in a longer initial decay, which is also heard to a greater extent, and so the spacing will need to be increased. The one third bandwidth resulting in a peak amplitude of the first beat at -60dB provides the optimal solution.

It is now necessary to determine the improvement in decay time by spacing in this way. However, standard methods of determining the decay of an impulse response, such as the measure of RT_{60}, rely on the assumption that we are within the statistical region, where the contribution of all resonances are considered to be equal. As we have seen, this does not hold at the lowest frequencies, and especially

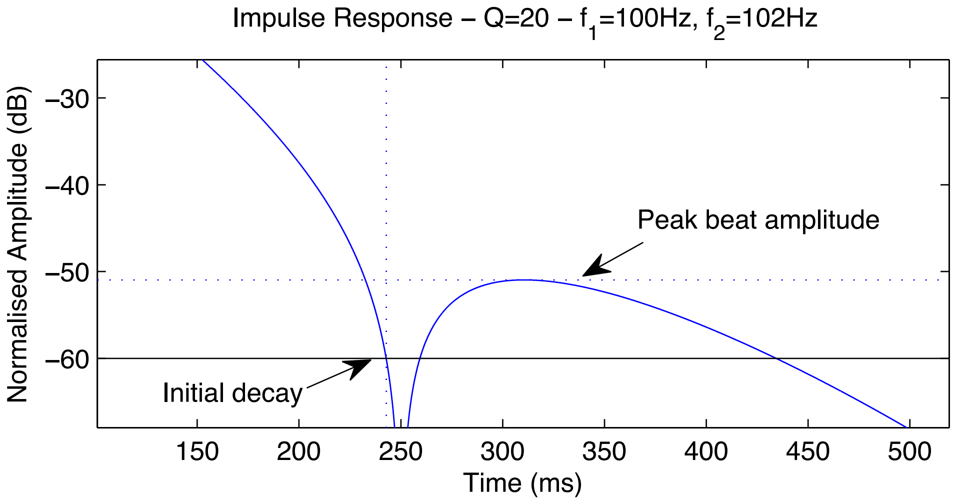


Figure 4.11: Concepts of ‘peak beat amplitude’ and ‘initial decay’

not when dealing with such isolated cases as two artificially modelled modes. It is therefore proposed that the results from this study are used, with the T_{60} equal to the time taken for the *initial* decay to drop -60dB, providing the peak amplitude of the the first beat does not rise above this level.

Based upon these concepts, the following section considers the use of optimal spacing to achieve a reduction in *virtual Q*, to an appropriate level as per the subjective thresholds defined by Avis et al. (2007).

4.6 Virtual Qs

It has been shown by Kuttruff (1991) that the decay time and Q factor of an individual resonance can be related with the simple relationship:

$$RT_{60} = \frac{2.2Q}{f} \tag{4.4}$$

Avis et al. (2007) have defined absolute thresholds of modal Q to be 16. A reduction below this level was shown to render the mode imperceptible. If the initial decay can be reduced, and beating not perceived, it is therefore possible determine how much an optimally spaced second resonance may reduce the ‘perceived Q’. The spacings revealed by this test are low enough that the perception is of a single tone (Moore, 2003). With this in mind, the concept of ‘virtual Q’ is introduced. This is defined as *the equivalent Q factor perceived from two ideally spaced resonances.*

Let us first imagine a room with damping sufficient to give a Q of 20 at 100Hz. We see from Equation 4.4 that this relates to a RT_{60} for that mode of 0.44 seconds.

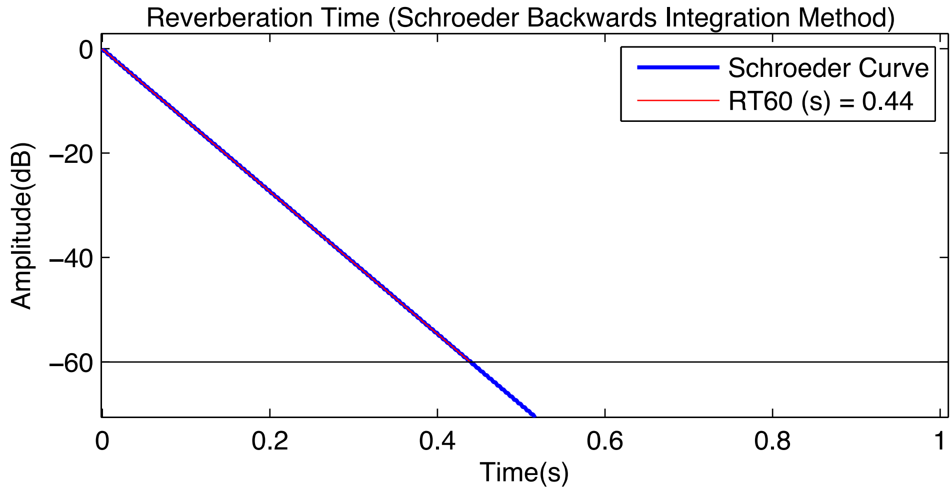


Figure 4.12: RT_{60} of a single mode using the backwards integration method

This can be verified with the commonly used 'backwards integration' method for determining a room's RT_{60} (Schroeder, 1965). Whilst we are not listening at frequencies within the statistical sound-field where this calculation is usually considered valid, in the case of our single exponential decay, the method can be used. Figure 4.12 shows that the method reveals the expected decay time of 0.44 seconds.

According to the subjective study, Equation 4.5 gives the optimal frequency for a second resonance:

$$f_2 = f_1 + \frac{f_1}{3Q} \tag{4.5}$$

Given this equation, when $f_1 = 100Hz$ and $Q = 20$, the optimal spacing for the second resonance is 101.67Hz. Figure 4.13 shows the impulse response in decibels of this system. The backward integration method now reveals a decay time of 0.39 seconds, and we can see from the curve that the method is no longer valid due to there being no 'statistical average' decay rate. However, by using the time taken for the *initial* 60dB decrease, we obtain a time of 0.29 seconds.

Where this 0.29 seconds is taken to be the perceived modal decay time, Equation 4.6 gives a 'virtual Q' for the system of 13.

$$Q_{virtual} = \frac{f_1 \times T_{60(1)}}{2.2} \tag{4.6}$$

Using this calculation, it can be shown that the relationship between original and virtual Q remains the same regardless of frequency. Any resonance with a Q of 20 is reduced to a virtual perceived Q of 13 by spacing a second 1/3rd bandwidth apart. This is the maximum reduction which can be made, as spacing the second resonance

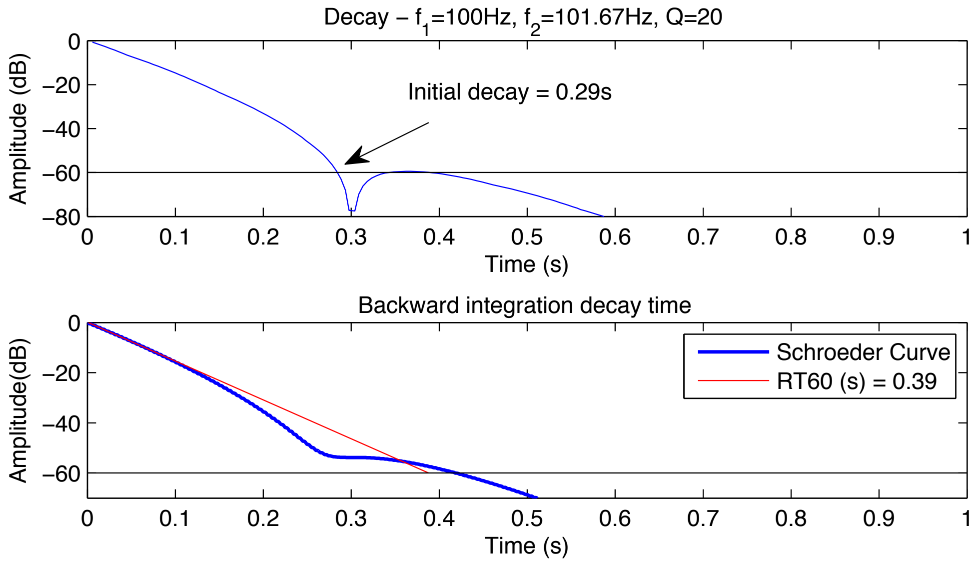


Figure 4.13: Calculation of RT_{60} using a) optimal spacing criteria and b) backwards integration method

Frequency (Hz)	32	63	125	250
Decay Time (s) - Q=16	1.10	0.56	0.28	0.14
Decay Time (s) - Q=24.6	1.69	0.86	0.43	0.22

Table 4.2: Decay times needed to achieve single resonance Qs of 16 and 24.6

further apart in order to reduce the decay time would increase the beat amplitude and negate the effect of the reduction. With all this in mind, if we are intending to reduce the virtual Q to below the threshold of 16 (Avis et al., 2007) we must begin with a single resonance Q-factor *no higher than 24.6*. An optimally spaced second mode will then provide the necessary decay reduction for the perception of a virtual Q of 16.

Table 4.2 shows how a Q of 24.6 corresponds to equivalent decay times of single modes at low frequencies. Such decay times are certainly realisable in critical listening room scenarios. This is opposed to aiming initially for Q=16, which gives higher damping requirements, adding complexity, physical size and cost.

4.7 Multiple Resonances

To this point, only simple cases of two adjacent modes have been considered. In realistic room scenarios, there may be a larger number of resonances which interact to form the overall response. However, as previously shown, at the lowest frequencies, these are often widely spaced, and therefore it is justifiable to study them discreetly,

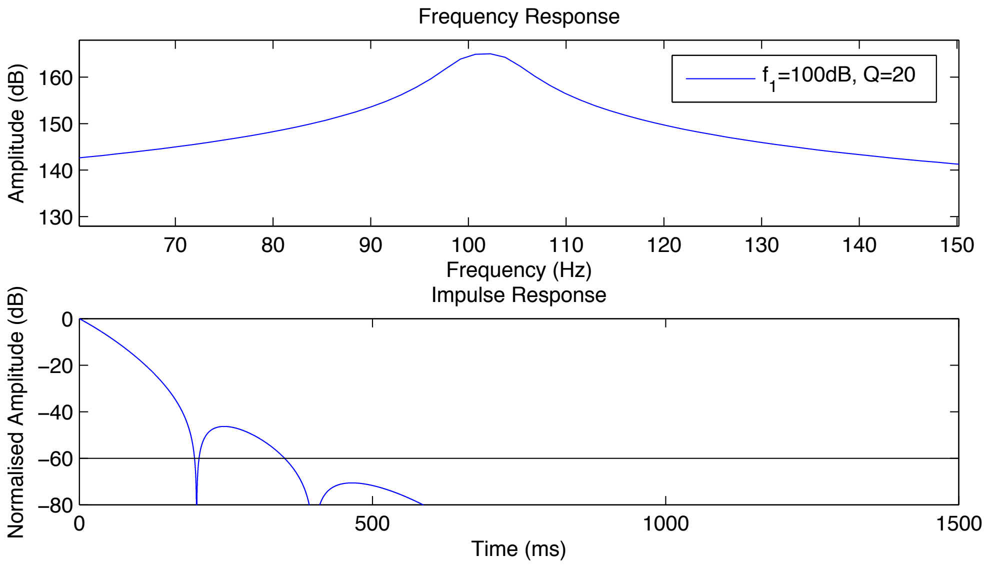


Figure 4.14: Frequency and time response where three modes are each spaced according to the one third bandwidth rule

or where two or three modes share a similar frequency band.

It has been shown that the addition of a second resonance needs, in most cases, to be within a few Hertz of the first, and the spacing is particularly sensitive to small fluctuations. A flattening of the response was observed, and it follows that the addition of a third resonance may 'continue' this flattening, providing a greater subjective improvement. With the optimal spacing defined at one third of the modal bandwidth, it is logical to attempt to place the next mode another third bandwidth apart. Here we assume that the bandwidths of all three are the same, as they are closely spaced in frequency. It is possible, although unlikely, that the absorption at the different frequencies would differ, therefore changing the bandwidths slightly. Figure 4.7 shows the frequency and time response to such a system, where the Q is 20 and the first frequency 100Hz.

Whilst the frequency response does indeed appear to be smoothed out, and the initial decay reduced further than in the case of only two modes, the peak amplitude of the first beat is now above the -60dB threshold. The subjective results suggest this will be audible and degrading. If we are to space the modes so that once again the first beat falls below this threshold, the spacing (when equal between all three modes) is required to be 23% of the modal bandwidth. Such a case may only be obtained with near cubic rooms and leaves practically no room for tolerance errors. With a first mode at 100Hz and a Q of 20, the second and third modes must be at 101.19 and 102.38Hz respectively. Such spacing yields an initial decay time of

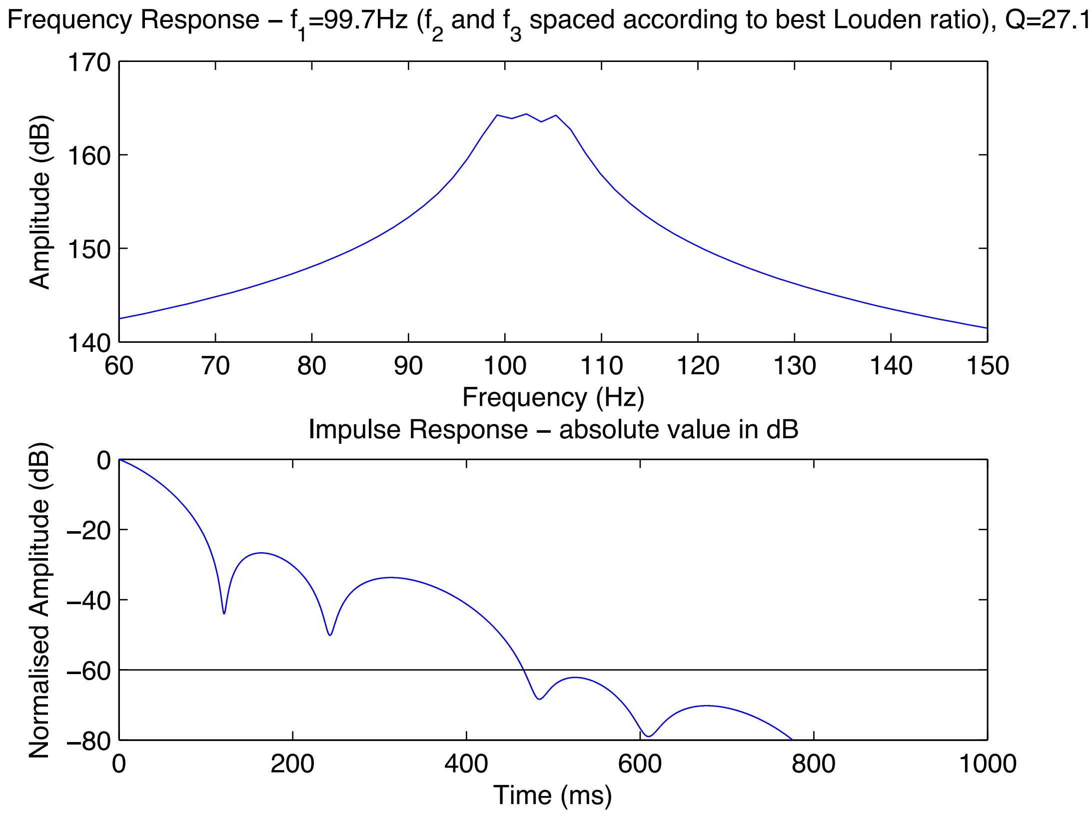


Figure 4.15: Impulse response for three modes spaced according to the best Louden ratio in a $75m^3$ room

0.27 seconds, and a virtual Q of 12.14. This is in comparison to the case with 33% spacing with two modes at the same frequency and Q, giving initial decay of 0.29 seconds and virtual Q of 13. Furthermore, there has been no significant change in the peak amplitude given three modes over two. It is therefore apparent that the addition of the third mode has very little benefit perceptually over the addition of a second.

With a greater understanding of the nature of these spaced modes, we may compare the effects caused by these very close spacings as suggested by the subjective testing, and a case where the modes are more uniformly spaced, such as by applying one of the aspect ratios previously mentioned. Let us consider the best ratio according to Louden (1971), of 1:1.4:1.9. In a $75m^3$ room, these dimensions would be $x = 4.36m, y = 5.78m, z = 3.04m$. At around 100Hz, there are three modes of 99.7, 102.35 and 105.02Hz. Given a typical decay time of 0.6 seconds, the Q will be 27.19, leading to the response shown in Figure 4.15. It can be seen that the frequency response looks smoother, but there is clear evidence of beating.

The problem here is that these calculations continue to fall foul of a number of the

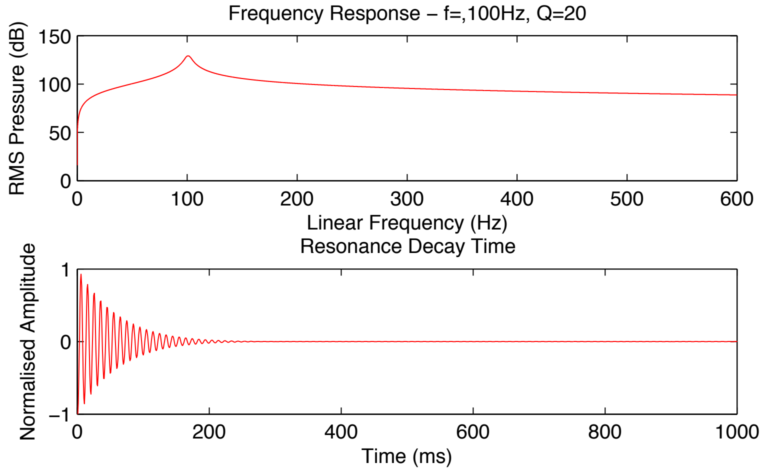


Figure 4.16: Two 'ideally' spaced modes in phase

problems highlighted earlier. Firstly, that we are looking at these isolated frequency ranges only. This removes the opportunity for further modes within the region to impact upon the response. This is considered further in Section 4.9, regarding the implications of spacing in room design. A second assumption, to which now turn, is that all modes occur with a positive phase and at maximum amplitude.

4.8 Eigenfunction Relationship

The eigenfunction relationship between two modes is one which can be shown to significantly affect the overall response. As has been shown, many metrics consider only the eigenfrequencies, that is, the centre frequency of each mode within the room, according to the dimensions. Metrics based upon such an approach suffer due to their neglect of the contribution and phase relationship of each mode. Section 3.3 shows that the contribution of each mode to the pressure sum at a specific listening point can be defined with a cosine function. Our assumption to this point has been that each mode will sum together with a maximum positive pressure amplitude. However, this is not necessarily the case within a real room. It is possible, due to the location of the source and receiver, and the order of each mode, that the reception of two modes, even if spaced at our optimal one third bandwidth spacing, may sum together to form a response which will be perceived very differently to the cases we have seen.

Figures 4.16 and 4.17 show the frequency and time domain representations of two optimally spaced modes, where the second is firstly in phase and then in anti-phase.

We can see that the implications are twofold. Firstly, there is a clear increase

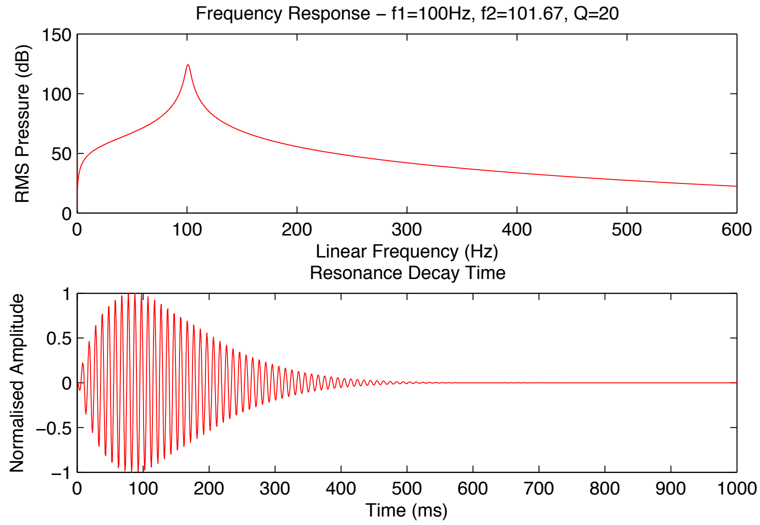


Figure 4.17: Two 'ideally' spaced modes in antiphase

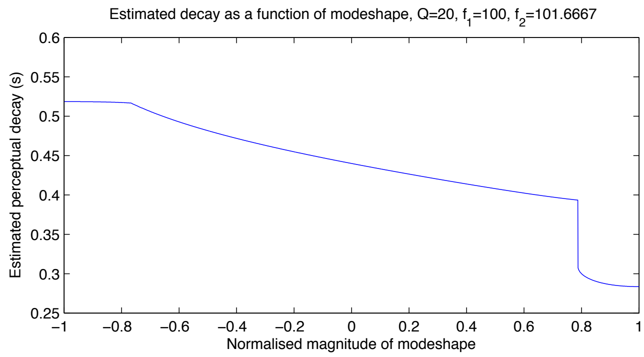


Figure 4.18: Estimated perceptual decay from two 'optimally' spaced resonances as a function of mode-shape of the second

in the decay time with the out of phase case. Secondly, there is a smearing of the beginning of the time response - the *attack* period in musical terms. In the case of modal resonances here, we see that the response to transient sounds, such as bass drum hits, or even bass guitar notes will be sluggish for the out of phase case. Figure 4.18 shows the estimated decay time, according to the first drop to -60dB criteria previously discussed, as a function of the second mode's phase, from -1 through to 1. This is equivalent to moving the receiver in Figure 4.19 from Position 1 to Position 2. It may be observed that the effective reduction in decay time by the one third bandwidth spacing, from 0.44 seconds of a single mode at 100Hz, to 0.28 seconds where both modes superimpose with positive phase, is 'undone' as the second mode moves out of phase with the first, ending up at 0.52 seconds, an increase in decay of nearly 0.1 seconds compared to the single mode at 100Hz.

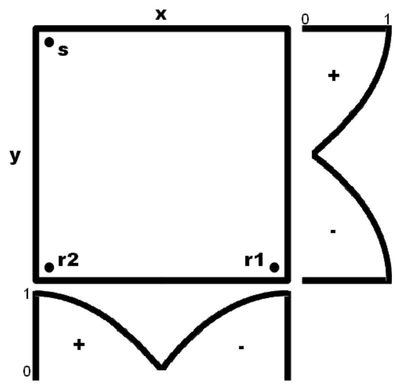


Figure 4.19: A source and two receiver positions with first order axial modes and their associated magnitude in the x and y dimensions

As a greater number of modes are added, the situation increases in complexity. The greater the number of modes, the greater the chances of these occurring with magnitude and phase mis-matches. As a final example, Figures 4.20 and 4.21 show how the case of three optimally spaced modes may differ, this time with the impulse response plotted in dB to highlight the beating effect. In Figure 4.20, all three modes are of maximum phase, while Figure 4.21 shows the same modes, this time with the central one in anti-phase. The results of the subjective testing show that this case will clearly be perceived as a less desirable response.

As we can see then, our 'optimal' spacing becomes less than optimal where a more complex system such as a realistic room is taken into consideration. In fact, not only is it less than optimal, but in some cases, a one third bandwidth spacing may be more degrading than any other arbitrary spacing.

4.9 Implications for Room Design

Finally then, we consider the results of both the subjective study along with the subsequent findings and assess the implications of modal spacing on room design and treatment.

We have seen that there is an optimal spacing, where two individual modes are concerned, which occurs where a second mode lies 33% of the bandwidth away from the first. The case occurs only when both modes add constructively in terms of their phase relationship. A deviation from this ideal scenario adversely affects the response, and extended decays and smearing of the 'attack' time of the resulting response occurs.

When considering the practicality of this optimum for room design, we note firstly that such cases of two isolated modes are relatively rare. It is only at the

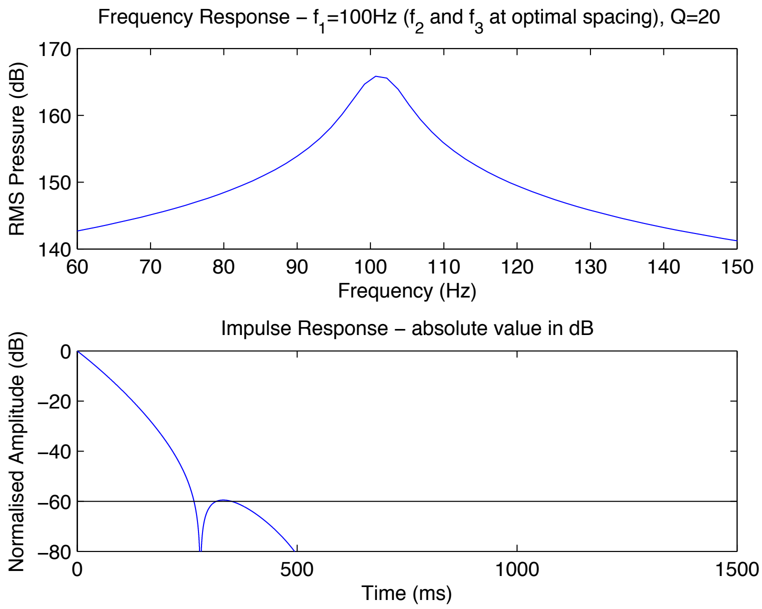


Figure 4.20: Three modes, spaced optimally, all in phase

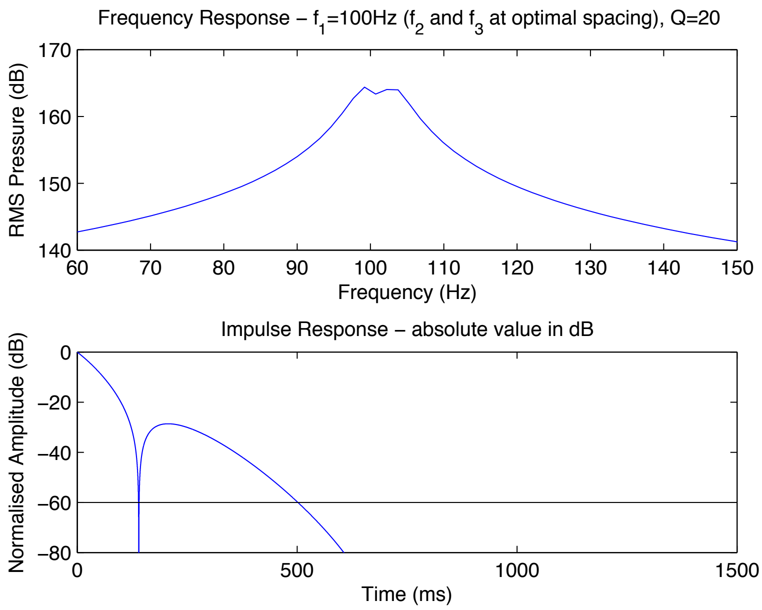


Figure 4.21: Three modes, spaced optimally, center mode 180° out of phase

Freq (Hz)	27.2	33.6	43.3	54.4	63.5	64.0	67.3	69.1	71.9
Type	A	A	T	A	A	T	A	T	T
Freq (Hz)	72.6	76.9	81.7	83.7	86.5	88.3	90.2	92.5	96.4
Type	T	O	A	T	T	T	O	T	O

Table 4.3: Distribution of modes in example room - Axial, Tangential and Oblique modes are given

lowest frequencies where few modes exist that this may actually happen. However, at these lower frequencies, there are a number of reasons which suggest that achieving the optimum may be beneficial.

Firstly, the standard deviation of the measured data is generally lower at low frequency, suggesting that subjects were more able to determine the reduction in decay time that a good spacing affords. Secondly, as Chapter 2 highlighted, absorption at these frequencies is difficult, and rooms often exhibit higher Qs in this region. Results show that there was good agreement of a beneficial optimal spacing where Q was higher, and decays longer. If these longer decays can be successfully reduced through carefully chosen room geometry, this is worth considering.

An example here is useful. In a room of 2.7m x 5.1m x 6.3m, which is $87m^3$ in volume, the modal distribution is shown in Table 4.3. It could be argued that from around 60Hz, the frequencies become closely spaced, and energy will spread to a number of surrounding modes. However, at lower frequencies, there are a few modes which can be considered as individual.

If one is to attempt to optimise the response using the first drop to -60dB criteria, it is first of all necessary to know the decay time for the frequency in question. If we consider a room of similar volume ($85m^3$), and assume a uniform decay time from 20-80Hz of 0.8 seconds, we can calculate that dimensions of 4.33m, 4.40m and 4.47m are required to produce the optimal 23% bandwidth spacing required where there are *three* modes interacting. It is convenient to use the optimal spacing of three modes, rather than two, simply to align with the three room dimensions. The bandwidth of a mode for a decay of 0.8 seconds is 2.75Hz - 23% of this, 0.63Hz, is therefore the spacing which will produce the shortest decay time. The first order axial modes will then occur at 38.35, 38.98 and 39.61Hz.

It is noted that even for a single mode at this frequency (38Hz) and decay time (0.8s), the Q-factor is around 14, which is already below the threshold of 16. According to Avis et al. (2007), further reduction is unlikely to result in a perceptual improvement. However, the second order axial modes will occur at twice the frequencies of the first (76.7, 77.96 and 79.22Hz), giving a Q of about 29 (assuming the decay time remains at 0.8 seconds). As a linear relationship exists, these three modal frequencies will also exhibit the 23% optimal spacing. The first drop to -60dB is at 0.43 seconds, giving a virtual Q of 15 - a perceptual improvement from 29, and now below the threshold of 16.

It must however, be remembered that at these 'harmonics' of the first set of axial modes, there is a much greater chance of other mode types (tangential, oblique) occurring. These are likely to interfere and produce artefacts which negate the decay reduction gained through the optimal spacing. Furthermore, it is unlikely that the assumption of a constant 0.8 second decay across the frequency range will be valid.

There are clear limitations with this theory, the most obvious being that the rooms dimensions must be very similar. There are inevitable tolerances in any such construction, and it is highly unlikely that the modal frequencies will occur as precisely as desired. The theory also assumes that low frequency sources are point sources and placed exactly. As Figure 4.18 shows, where the eigenfunction is not exact, a decay time above that of the optimum is always observed. Finally, as it has been shown that the calculations are particularly sensitive to decay times, it is likely to be infeasible to achieve the precise decay times/modal frequencies required for this type of optimisation.

4.10 Summary

In this chapter a subjectively defined optimal modal spacing has been measured. This is shown to increase with frequency and decrease with Q-factor. When specified in terms of percentage of modal bandwidth, the optimal spacing lies between 25% and 40% of modal bandwidth regardless of frequency and Q (with exception to a Q value of 10). It has been noted that this result seems to show some agreement with the Schroeder Frequency's prediction of the transition from the soundfield being classified as modal to statistical. If the optimal spacing is 33% of the modal bandwidth, this would allow three modes to fall within one bandwidth. However, Schroeder suggests a *minimum* of three. Results here suggest that spacing any closer together may result in degradation as the decay time increases.

The reliability of subject's responses also show that modal spacing is important at the lowest modes but its significance decreases with increasing frequency. A smaller spacing than optimal leads to longer but homogeneous resonant decays. This has been shown to be problematic for sound reproduction (Fazenda et al., 2005; Toole, 2006). However, larger spacing than optimal leads to audible beating in the decay. The relative importances of these two factors (long single decays versus perception of beats) has not been measured and it stands out as an interesting avenue for future research. It should be noted that this applies mainly to case where only two resonances share a very narrow band of frequencies which is representative of the lowest modes in a room where two or more dimensions are similar.

Using these subjective results as a basis, a number of further scenarios have been investigated. The concept of ‘virtual Q’ has been introduced in order to determine the perceived Q where two resonances are optimally spaced. The finding that the spacing resulting in the shortest decay before the onset of beats reached a -60dB threshold has been extrapolated to scenarios considering multiple modes and also where the two modes do not share a positive phase relationship. Both of these cases are problematic where the modes are spaced according to the one third bandwidth rule.

It is suggested that whilst there remains an opportunity for further investigation in this area, such as looking at how we might best resolve phase relationships, what our upper frequency limit for spacing optimisation and the psychoacoustic effects of beating type effects where many modes exist, it is beyond the scope of this thesis, and a greater advance in understanding perception lies in other parameters and optimisations. It is therefore concluded that it is unlikely that a particular arrangement of modes which aim for an ‘optimal spacing’ will be of great practical use for room designers.

Chapter 5

Thresholds of Modal Density

5.1 Introduction

Chapter 4 introduced the concept of the modal distribution, and focussed upon the parameter of spacing, in search of an optimum. This chapter considers a related concept, that of the modal density. This parameter is inherently linked to the spacing, and is defined as 'the number of modal frequencies within a given bandwidth'. Density as a parameter is perhaps more intuitively related to real room scenarios than spacing, due to the fact that as frequency increases, the discrete nature of single spaced modes disappears.

With this parameter more naturally correlated to the realities of a real room, researchers have considered objective quantities based upon the density in further attempts to classify the quality of a room in subjective terms.

As shall be shown, this research has typically led to the implication that a high modal density alleviates many of the perceived problems. If this is indeed the case, the obvious question is, how high is high? This chapter therefore attempts to determine if a threshold of modal density exists - and if so, define it and consider any limitations.

5.2 Modal Density

The modal density is essentially a measure of the number of modes occurring within a given bandwidth. This can be obtained by simply calculating all eigenfrequencies and summing over the given range. Furthermore, it has been shown in Chapter 3 that Equation 3.30 represents a statistical average for the expected number of modes at a specific frequency. It is this equation which is used in this chapter to calculate the room parameter known as modal density. The concept of metrics, used in scoring rooms based upon one or more objective measures was introduced in the previous chapter. Unsurprisingly, there have also been a number of metrics suggested in terms of the modal density. Two such examples are widely quoted. Firstly, the 'Bonello Criterion' (Bonello, 1981), which aims to guide room design through a set of criteria including that where modal degeneracy occurs, there should be at least five modal frequencies within that third octave band. The implication here is that a greater density may offer a subjective improvement to counteract the effects of degeneracy. Bonello explains that it is through personal experience that he comes to his conclusions. In a study 28 years later, Welti highlights that the criterion assumes certain conditions which are almost never met, though he does call the method 'intuitively satisfying' (Welti, 2009). It is for this reason – that such a typical metric may not offer a subjective improvement and yet remains enticing for use due to its simplicity – that a thorough subjective study of the modal density is necessary at this time.

The second widely quoted objective quantity derived from the density is that of the 'Schroeder Frequency' (Schroeder, 1996). This frequency indicates the point where sufficient modes exist, within the bandwidth of a single mode, that the frequency response can be assumed to be statistical in nature. Specifically, above this frequency, the average spacing between adjacent maxima is equal to $6.7/RT_{60}$, with the same result regardless of the room. The frequency is therefore traditionally said to define a transition between the 'modal' and 'statistical' regions in a given room (Toole, 2008). This transition frequency is most often determined by Equation 3.31. Note that the constant 2000 was changed in Schroeder's 1996 paper from his original stated value of 4000. This has the effect of lowering the crossover frequency to the point where just *three* modes are present within the bandwidth of one mode as opposed to *ten*. This change is interesting in itself and no real justification is given. Again, we are left with only an *implication* that this may have a subjective relevance.

It can be argued that it is widely believed that when above this transition frequency (f_c), we are listening within 'diffuse sound-field' conditions, and therefore individual effects associated with discrete resonances are no longer perceived. It must be stressed here that this objective measure reveals the transition to a room where the frequency response can be considered statistical in nature. Schroeder's papers make no reference to any subjective effects of reaching this frequency. However, tracing the history of room acoustics, one can see that the use of the Schroeder Frequency has led to an ever increasing polarisation of modal and diffuse sound fields despite some warnings (e.g. Toole (2006)). As this gap widens, the assumption that *perception* of audio falling within these two regions is also polarised seems to have proliferated. Evidence for this can be seen from attendance at major audio and acoustic conferences and lectures, but also within the literature. For example, many research papers use this crossover frequency as a limiting point for their investigations into the effects of low frequency resonances. The work of Avis et al. (2007) which investigates the perception of room modes uses the Schroeder Frequency as the point of transition when forming binaural room models. In their 'Room Sizing and Optimization' paper, Cox et al. (2004) also state that the frequency range under investigation can be "guided by the Schroeder Frequency". A further example of its use in this way can be seen in Blaszak (2007). Finally, Toole (2006) states the importance of the crossover region as a "real phenomenon" which needs to be better understood.

As the size of an enclosure increases, the Schroeder Frequency (f_c) decreases. Therefore, in large rooms such as concert halls, f_c is typically very low, often below the 20Hz threshold of our hearing. However, spaces such as control rooms, with typically small volumes (i.e. 100m^3), are classified as having modal regions at frequencies not only above 20Hz, but well into the range of most musical situations (i.e. $RT_{60} = 1.28s$, $V = 75m^3$, $f_c = 261Hz$ - middle C).

Regardless of room size, the modal density naturally increases with frequency. Eventually many hundreds of modes exist within just a few Hertz. It is this increase in modal density that underpins the definition of the Schroeder Frequency. It is argued here that, just as we have seen that we should not rely on solely objective measures of modal spacing, we should not rely on them for density. Rather, we must ask the question: is there a threshold above which enough modes exist such that we no longer perceive any degradation? Is it in fact possible to determine a 'subjective counterpart to the Schroeder Frequency?

To further emphasise a misunderstanding of the density parameter, it is noted that it is often either assumed or implied (e.g. in diagrams such as Figure 5.1) that as a large number of modes are concentrated in a given frequency range, as occurs with an increase in volume and/or frequency, the overall magnitude frequency response becomes 'flatter' and thus is commonly associated with better quality reproduction. We have seen from the spacing work that an addition of multiple modes may in some cases appear to smooth the response. With many modes sharing a similar bandwidth, any energy present through say, a bass note, will excite each of these modes, and dissipate through them. However, the previous chapter also revealed that the interactions between modes must not be considered simplistically. Phase interactions between closely spaced modes may in fact result in a highly irregular

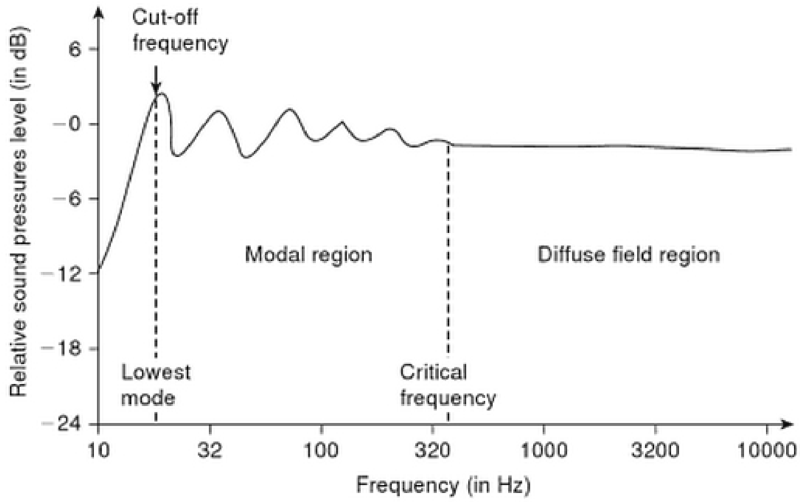


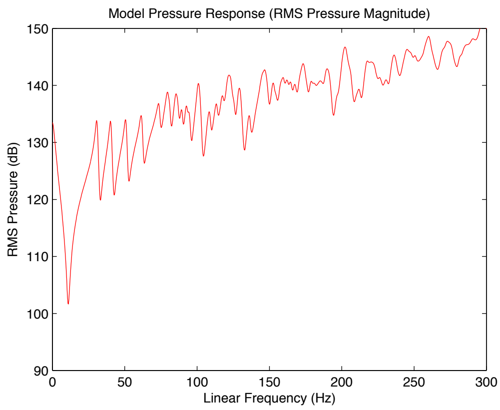
Figure 5.1: A typical representation of the transition frequency, taken from Howard and Angus (2001)

frequency response and actually degrade the audio. Therefore, a higher density may in fact compound such problems.

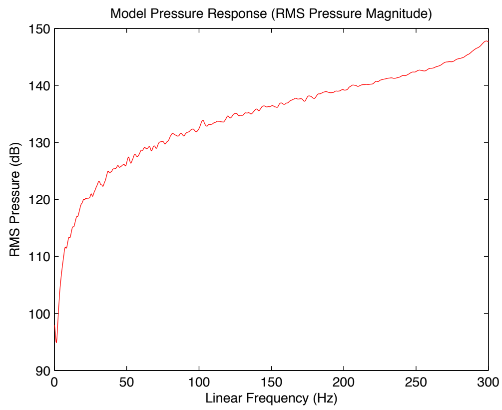
5.3 Density and Shape Functions

As shown in Chapter 3, each eigenfrequency has an associated eigenfunction. A common representation of this is the cosine function given in Equation (3.12). The contribution to the overall pressure response is therefore dependent upon the coupling of a source or receiver to the mode, which therefore differs as the source/receiver moves throughout the room. These spatial distributions are known as the 'shape functions', and are declared as $p_n(r)$ and $p_n(r0)$ within the modal decomposition room model (Equation (3.27)), for the source and receiver coupling respectively. These are essential for the realistic modelling of a room.

However, as has been mentioned, by studying a room and accounting only for the modal frequencies, the impact of these shape functions is ignored. This is mathematically equivalent to assuming that all modes contribute equally to the pressure response, and do so each with a positive phase. This scenario represents the conditions assumed for many objective measures, such as the room ratio metrics suggested by Bolt (1939b) and Louden (1971). When such scenarios are replicated using the decomposition model with the shape functions omitted, we witness an apparent 'smoothing' of the frequency response as density increases, which may add to the impression that a perceptual improvement is observed as the density rises. Furthermore, with the omission of shape functions, as the density increases, we observe a high frequency amplitude rise, as more and more modes sum together with maximum positive phase. This is a mirepresentation of the physical reality of a room, where the lowest frequencies are often higher in magnitude. In practice, such conditions are never actually attained in rooms. Figures 5.2 and 5.3 show the differences in frequency response where shape functions are both included and omitted. With a low density (simulated with a $50m^3$ room), both responses show peaks and dips at modal frequencies, although the 'all positive' nature of omitting the mode shapes is apparent at the lowest frequencies. Clearly, with a much larger density, the differences between the models are stark, and it is also clear that our perception of the two responses will differ, even though all room conditions are identical and both have the same high density. Even with a high density as is the case in a 3000m^3 room, the interaction of the source and receiver shape counteracts any smoothing which may occur simply due to the density increase. At this volume, the frequency band within just 5Hz between 125 and 130Hz already contains around 95 modes, corresponding to a spacing of 0.05Hz.



(a) 50m^3 without mode shapes

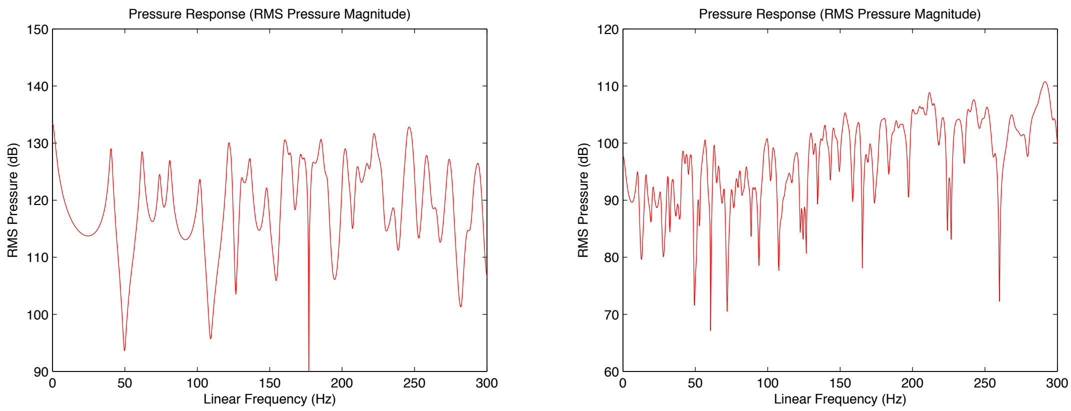


(b) 3000m^3 without mode shapes

Figure 5.2: Comparison of high and low density room responses when modelled without mode shapes

5.4 Density Threshold Omitting Shape Functions

As the omission of mode shapes does give a clear smoothing with density increase, it should be possible to define a threshold for the point where it is sufficiently high that reproduction of audio in this environment can no longer be distinguished from a smooth reference case. Any such threshold obtained may be treated similarly to



(a) 50m^3 with mode shapes (b) 3000m^3 with mode shapes

Figure 5.3: Comparison of high and low density room responses when modelled with mode shapes

other 'absolute' thresholds, that is, thresholds which relate to the absolute detection of human perception given an ideal condition. Examples include the detection of pure tones in equal loudness curves, and will be met again in the following chapter on modal decay thresholds. Although often simulating unrealistic scenarios such absolute results are still considered valid as a first point of reference in investigating thresholds, as they provide an interesting comparison with any thresholds obtained from more realistic scenarios where mode shapes are included (see Section 5.5).

5.4.1 Obtaining a Threshold

Providing a linear scale exists, with a perceptual improvement due to increasing modal density, it is possible to determine a threshold above which no further differences are perceived. Therefore, subjects were asked if they could perceive a difference between a high density reference and that of a variable case. Two samples were therefore necessary, which were produced by convolving a test stimulus with a reference room response and also the response where the modal density was that under test. Density itself was controlled through modification of the room volume within the decomposition model. The reference was modelled in the same way, but with a very large volume, and therefore density, effectively producing a completely smooth response. This test therefore identifies the detection threshold where the modal density is sufficient to produce a room whose audio reproduction is perceived as the same as that of the reference. The density can be extrapolated from the required volume using an expression describing the typical relationship in rectangular rooms (Bolt, 1939b).

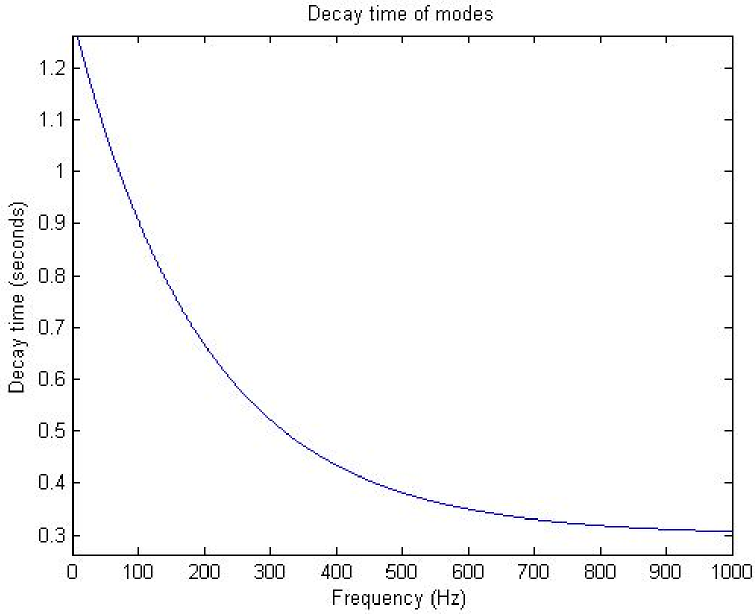


Figure 5.4: Decay time across frequency used in the modelling of room responses

5.4.2 Considering Stimuli

Concepts such as the Schroeder Frequency imply that the same density is required regardless of frequency - or to put this another way, whatever frequency the 'three modes per bandwidth' criterion is reached at, those three modes per bandwidth are sufficient that the response is uniform. In order to determine if this is indeed the case or if, in fact, there is a frequency dependency to the density threshold, it was necessary to study a number of different frequencies separately. To achieve this, the input stimuli to the modelled rooms were single frequency test tones. These were 0.4 second decaying sine tones, representative of single frequency bass notes. The same three frequencies as in the modal spacing test - 63Hz, 125Hz and 250Hz were investigated. These test tones were considered appropriate as it was not specifically the decay which was being tested, but the effects of the overall room response upon a stimuli. The tones were therefore convolved with the modelled room response. Damping within the room was modelled as frequency dependent, according to a simple exponential curve (Figure 5.4). This was done in order to keep the model results as realistic as possible (with the exception of the omission of the shape functions). Therefore, in assessing thresholds, each mode which does impose a characteristic should have a decay analogous to that of a typical critical listening room. As figure 5.4 shows, this has a decay of 1.01, 0.84 and 0.58 seconds at 63, 125 and 250Hz respectively.

As in Chapter 4, samples were calibrated in order to be presented according to the 90dB equal loudness contour. Eight subjects were tested, under the same conditions as the spacing test.

5.4.3 PEST/ABX Hybrid Method

In order to home in on the subject's threshold of detection between the reference and variable sample, the PEST (Parameter Estimation by Sequential Testing) methodology (Taylor and Creelman, 1967; Taylor et al., 1983) was considered. This is an adaptive method, similar to the 2-alternative forced choice, which forces a response to a simple question, "is there a difference between the reference and test stimuli?". The PEST produces results directly through its adaptive method. It is maximally efficient, resulting in fewer trials per listener to output the threshold. Each reference and variable sample pairing is known as a *trial*, with a set of trials completing a PEST *run* until the rules state that the PEST sequence is to be terminated.

An initial difference between the reference and variable volume was set which could be easily discernible by the subject, and following this first trial, a set of PEST 'rules' govern the determination of new variable room volume levels. In this test, the initial volume was $100m^3$, a relatively low modal density at each of the three test frequencies, where differences due to the discrete modes are clearly audible. Wherever a trial has been completed, a Wald sequential likelihood test (Wald, 1947) determines whether a new level should be set or the previous repeated. If a new level is to be set, the rules can be summarised as follows:

1. If a reversal is made, halve the step size.

2. A second step in the same direction requires the same step size as the first.

3. Fourth and further steps in the same direction require a doubling of the step size.

4. Step size on the third trial in the same direction depends on the step prior to the last reversal. If, on this step, no doubling occurred, the step size should now double, while if a doubling did occur, no doubling should take place here.

The original PEST specification suggests a logarithmic change to step size, although after initial experimentation, a linear change proved most appropriate. The end of a PEST run occurs when a new step size is required which falls below a defined minimum. The selection of this minimum step size is key to the success of the PEST run, as it determines the accuracy by which the final threshold can be defined. It is therefore necessary to make this small enough to output thresholds of reasonable accuracy, but not so small that it becomes difficult for the PEST process to converge. The selection of this value also directly affects the time taken to test, and therefore has a bearing on factors such as subject fatigue and boredom, which can be a significant source of error (see Goldberg (2006)). Pilot testing revealed that a reasonable minimum step was $100m^3$.

In such a test, it remains a possibility that the subject simply responds that they can hear a difference between the reference and variable samples even if they may perhaps not (most likely due to the belief that they are performing better if they can consistently hear a difference). To ensure that they could not simply claim to hear a difference, the PEST process was augmented with a standard ABX procedure whose output informed the PEST routine of success or failure.

At each room volume a maximum of three comparisons could be made. Sample A was the reference, B the varying volume (dependent on the PEST routine) and X either A or B, randomly chosen. If the X sample was correctly identified three times consecutively, the room volume and hence modal density was increased. However, a single incorrect answer would immediately register a failure to detect a difference and therefore the volume would decrease. An incorrect answer reveals that there is no difference evident between the current varying modal density and the reference case, an indication that a sufficient density has been reached in order to smooth out any audible artifacts arising from the modal response. The requirement of three consecutive correct answers reduces the probability of the subject guessing to 12.5%, and while this is not at the typical statistical threshold (<5%), it was considered sufficient given the association with the PEST methodology, which would bring the volume back down at the next comparison unless another three guesses were correct - totalling six consecutive guesses - a probability of just 1.6%.

5.4.4 Results

Figure 5.5 shows the mean room volume threshold and standard deviation where no detectable difference existed between that volume and the reference. In practice, the results provide the *volume* threshold for a particular frequency. However, to extract the *density* threshold, a modal bandwidth for the corresponding frequency has to be obtained. Modal density is therefore calculated using Bolt's equation, with modal bandwidth obtained using:

$$Bw = \frac{2.2}{RT} \tag{5.1}$$

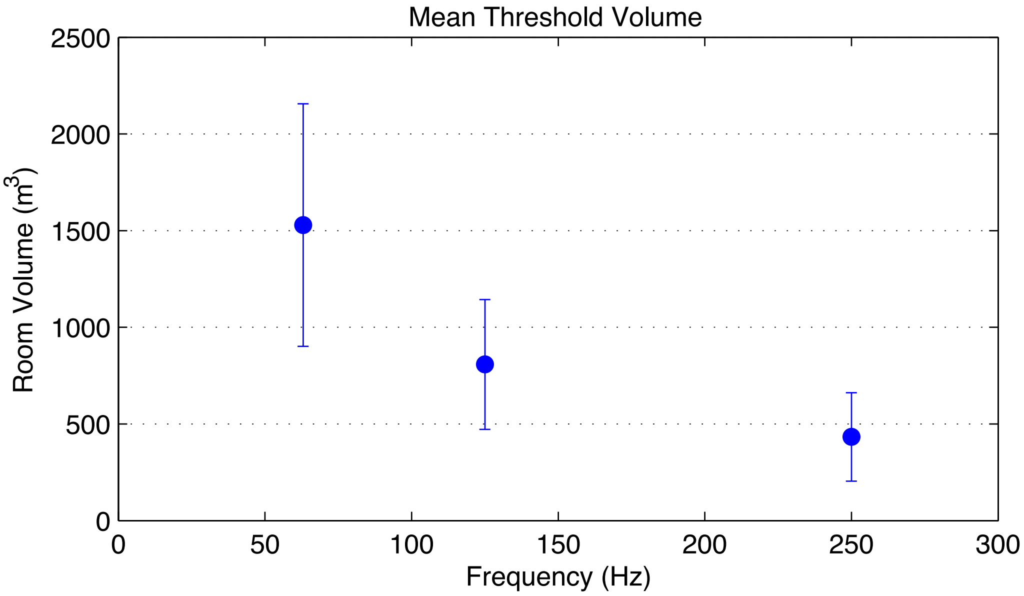


Figure 5.5: Mean room volume threshold for the detection of difference over three test frequencies

	Frequency (Hz)		
	63	125	250
Modal decay (seconds)	1.01	0.84	0.58
Required Volume (m^3)	1529	803	433
Required Modal Density (modes per bandwidth)	4.1	10.3	31.6

Table 5.1: Modal density according to bandwidth taken from reverb conditions in modal decomposition model

$$\delta N = \frac{8.8\pi F^2 V}{c^3 RT_{modal}} \tag{5.2}$$

where N is the modal density, F the frequency, V room volume, c the speed of sound in air and RT is taken from Figure 5.4 at any given frequency. This density is indicated in Table 5.1.

The results show that at 63Hz a subject would require around four modes per modal bandwidth to even out degrading effects. Furthermore, under these test conditions, subjects require an increasing modal density as frequency rises. This is shown in Table 5.1, where each volume is associated with a density, giving the required density threshold for perceiving a smooth response at a given frequency. Consequently, no definition of a generic modal density across frequency is possible from these results. It is also seen that the application of Schroeder's transition to a statistical region is not replicated subjectively, and it would be unwise to do so. Although at the very low frequencies a modal density of about four is sufficient and

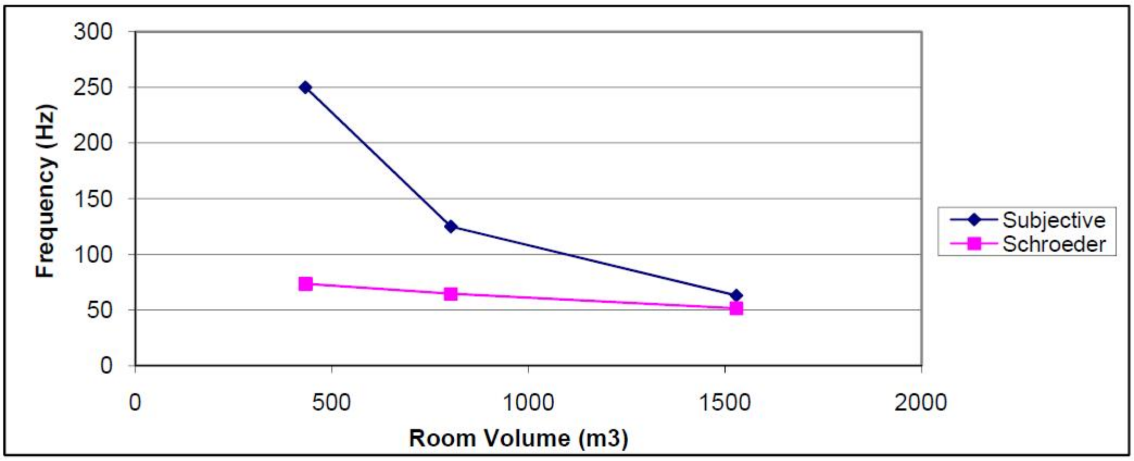


Figure 5.6: Schroeder and subjective 'cut on' frequencies across room volume

in accordance with the definition of three for the Schroeder Frequency, as frequency increases, subjects require a much higher density if degradation of the stimuli is to be inaudible.

Another way of representing these results is shown in Figure 5.6. A subjective 'cut-on' frequency above which modal effects are negligible is indicated both with regards to these subjective tests and the Schroeder Frequency. It is clear that, for smaller rooms the Schroeder Frequency underestimates the subjective 'cut-on' frequency - subjects still detect differences in modal sound fields above it. For larger volume rooms, the subjective results converge to towards it. For example, at $433m^3$, f_c is 66Hz, while subjectively, this was the required volume at 250Hz.

At lower room volumes, the Schroeder Frequency may well accurately predict where a *statistical soundfield begins*, but we do not begin to *perceive* a uniform, smooth response until we reach a much higher frequency. Therefore, at low frequencies, even in rooms large enough for us to expect that we won't suffer from audible modal effects, it has been shown that we can perceive differences from the ideal smooth case.

5.5 Density Threshold Including Shape Functions

Whilst the density thresholds observed may be interesting, it remains uncertain how they translate in real rooms. As previously discussed, any realistic scenario should include the effects of the mode-shapes as these carry crucial information about the way in which the source and receiver position couple with the modes - a scenario to which we now turn.

5.5.1 Obtaining a Threshold

In order to obtain similar thresholds, but with the inclusion of the shape functions, it seems logical to employ the same PEST methodology. However, pilot testing revealed that such an experiment is problematic; a convergence of the PEST routine was not observed. It will only do so where a true threshold exists, and where there is a linear change in perception in accordance with the changing of an independent variable. This is not the case where shape functions are included. During pilot testing, the subjects were able to perceive differences even at very high modelled densities. With reference to Figures 5.2 and 5.3 we see why this may be the case. The peaks and dips in the frequency response remain even at high volumes/densities. By changing the room volume we simply observe a *shift* in these peaks and dips, rather than the gradual smoothing towards and ideal case where degradation of the stimuli will be reduced, as was seen in the case with no mode shapes. This also means that no reference density can be set. Logically, any non linear parameter change cannot be tested for a threshold using adaptive methods such as PEST. Therefore, in order to investigate the effects of density increase when including the effects of modal coupling, a different approach is needed.

5.5.2 Considering Stimuli

We must not forget that the interaction of the test stimuli with the room response is of great importance, as shown by Fazenda (2004). In the case omitting shape functions, where the response shows a smoothing with increase in density, it was appropriate to use single frequency test tones as the stimuli. Although as volume increased, different combinations of peaks and dips were excited by these single frequency bursts, there was nonetheless a smoothing with each incremental volume increase. However, with the inclusion of shape functions, this is no longer the case. As the occurrence of peaks/dips can be considered random at a given frequency, when the density is increased, these peaks and dips may change in frequency and the artefacts heard will be a product of these interactions, rather than just the density itself. To test the effect of density upon the full frequency response, a full range stimuli is required. As ultimately, this investigation should reveal results applicable to realistic application in rooms, it is preferable to consider musical stimuli here.

As it is not possible to determine an absolute threshold, a new approach is considered. One of the key motivating factors in looking for a threshold was the widely held notion that at low modal densities, we hear clear differences due to the discrete nature of the modes, and then with a high enough density, such differences

	Dimensions (m)			Volume (m^3)
Small Room	11.9	9.1	4.6	500
Large Room	33.3	24.7	12.5	10000

Table 5.2: Reference room dimensions

are 'ironed out' as any room will have a statistically similar response. The following section attempts to investigate this theory.

5.5.3 High and Low Density Comparison

The following hypothesis is tested:

> The differences between two rooms where the modal density is low will be clearly audible. Conversely, when the density is sufficiently high, each room will be perceived similarly.

In order to test this hypothesis, a simple ABX methodology was employed. Twelve rooms were auralised using the procedure outlined in Chapter 3 (with the omission of the headphone equalisation stage). The LEN musical samples introduced in Chapter 3 was chosen for auralisation. In order to ascertain if there is a subjective difference between higher and lower density cases, two room categories were tested - typically small and typically large rooms. A reference sample at the larger end of these categories was created ($500m^3$ with typical dimensions for the small case, and $10000m^3$ for the large). Table 5.2 shows the actual dimensions for these rooms. Damping was modelled as frequency dependent, decaying exponentially, in the same manner as in the test with no mode shapes (Figure 5.4).

For each case, five additional rooms, of volumes increasing towards that of the reference were auralised. The small room volumes were 100, 250, 400, 450 and $490m^3$, with the large rooms 1000, 5000, 9000, 9500, $9990m^3$.

A series of ABX tests were conducted between pairs of each reference (Sample A) and the variable volume rooms (Sample B). Sample X was the unknown which was a random selection between the current variable and the reference sample. Subjects made ten comparisons of each pair to ensure statistical validity.

The same eight subjects participated as in the previous test with no mode shapes, during the same listening session. A short break was enforced between the two tests, and an explanation of the interface was given. Subjects were instructed to take further breaks whenever they felt it necessary.

Volume (m^3)	% of ref volume	Mean correct identifications	p
100	20	9.22	0.00
250	50	8.56	0.00
400	80	8.33	0.00
450	90	8.11	0.00
490	98	6.56	0.15

Table 5.3: Results for ABX testing of small room volumes

Volume (m^3)	% of ref volume	Mean correct identifications	p
1000	10	9.11	0.00
5000	50	8.56	0.00
9000	90	7.67	0.02
9500	95	5.89	0.13
9900	99	5.89	0.92

Table 5.4: Results for ABX testing of large room volumes

5.5.4 Results

Tables 5.3 and 5.4 present the results split according to the two different cases, small and large rooms. The mean correct identifications reveals the number of times it was possible for the subjects to identify sample X. For example, out of ten trials of the reference room $500m^3$ and the variable room of $250m^3$, 8.56 times out of ten the X sample can be identified.

A chi-square test was performed for each sample pair to determine the significance of the identifications (see p value in Tables 5.3 and 5.4). The expected value, if a subject is guessing at random is five out of ten, and this is used in the statistical test giving a p value. If this is below 0.01, it is an indication that there is a significant difference (at the 1% level) between the two samples. Above this p value, statistically, the subjects could not reliably distinguish between the samples.

Results show that similar trends are visible for both the typically small and large rooms (see Figure 5.7). It would appear that, regardless of whether the modal density is low or high, if the compared rooms differ in volume, differences in audio reproduction at low frequencies can be heard. Essentially, we can observe that a volume difference results in a difference in the room response. We can therefore extrapolate that, *regardless of density*, a difference in *room response* is perceptible. The ABX task remained simple until the differences in the response were small (volumes within 10% of each other). The results tables therefore also show the percentage difference between the reference and variable rooms. As can be seen, when this rises above 90%, no significant differences between the rooms can be heard, regardless of modal density. This is shown clearly in Figure 5.7. With such small

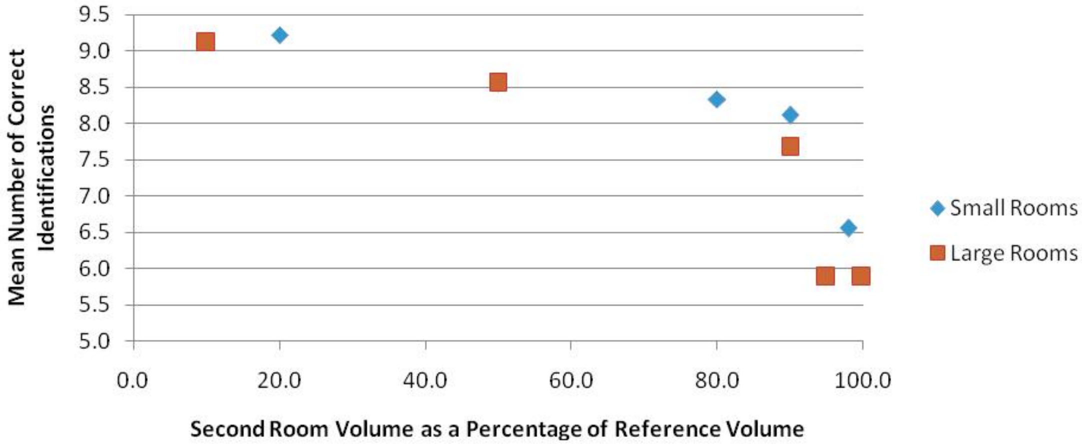


Figure 5.7: Correct answers in the identification of two room volumes

differences, the frequency response becomes similar, allowing individual frequencies within the stimuli to interact in a perceptually similar way with the rooms transfer function (Figure 5.8).

This result adds further weight to the argument that it is the individual room response which is responsible for the listener's perception of reproduction quality, and should not simply be taken as a constant above the Schroeder Frequency. In other words, whilst the statistical properties of these large rooms may well be the same, it is clear that the perceptual properties are not.

5.6 Perception of Audio Quality

Whilst it has been shown that listeners are able to perceive an absolute difference where the room transfer function differs, this gives no indication of the perceived quality of reproduction of that room. After all, it is logical to see that a different frequency response, even at a high modal density will be perceived differently. One only needs to consider our perception of the higher frequency range. Although this is often considered statistically, different rooms clearly produce differing auditory sensations. An example would be the reverberation time.

The question therefore remains, is there a perceptual improvement to the quality of low frequency reproduction as the density increases? A listening test has been designed in order to explore this, again using comparisons of an auralised music sample in modelled rooms of differing density.

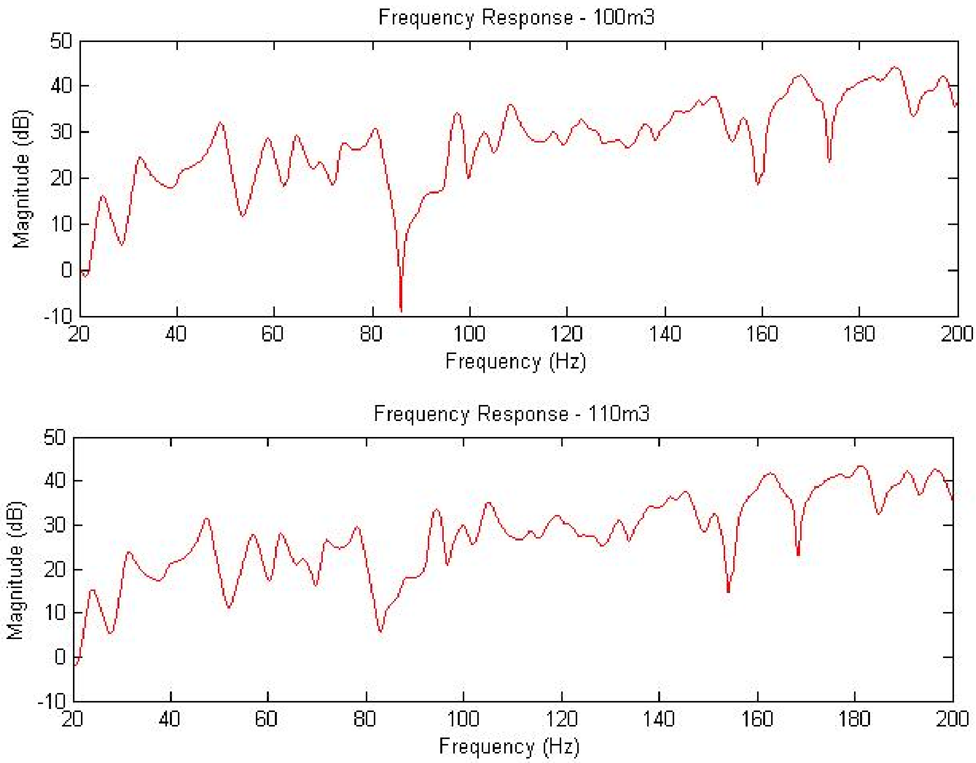
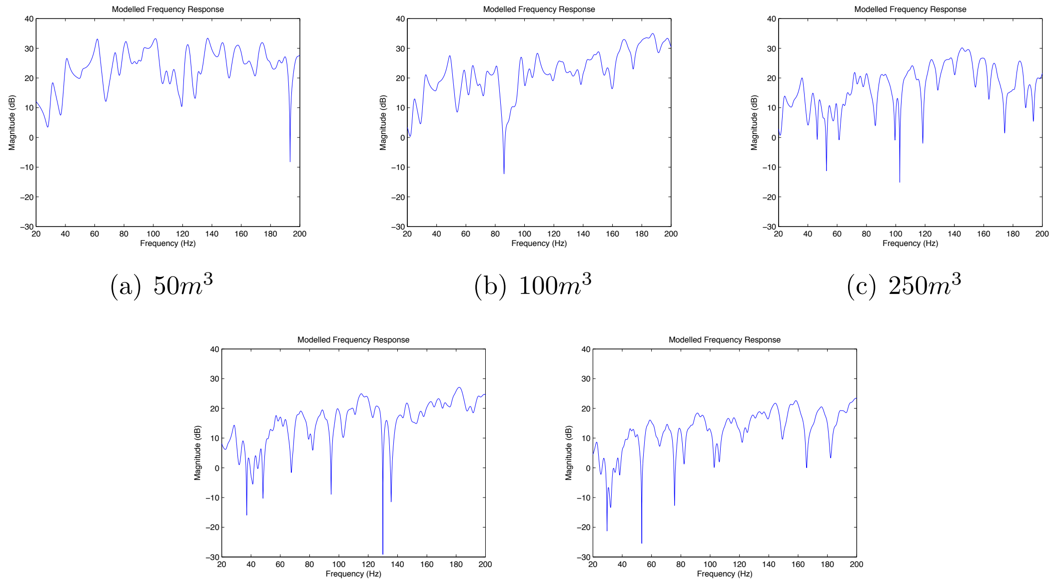


Figure 5.8: A comparison of two frequency responses in rooms with a volume difference of just 10%



(a) $50m^3$ (b) $100m^3$ (c) $250m^3$

(d) $500m^3$ (e) $1000m^3$

Figure 5.9: Frequency responses of all five rooms in the paired comparison perceived quality listening test.

5.6.1 Test Samples

The test was conducted with the same LEN sample auralised with five differing room volumes. The damping was kept constant in order to eliminate decay time differences as a confounding factor, and this followed the same typical decay time curve for small rooms as previous used (Figure 5.4).

For the purposes of testing perceived quality, rooms were modelled at $50m^3$, $100m^3$, $250m^3$, $500m^3$ and $1000m^3$ in order to study densities which may be considered typical across the range of standard critical listening rooms, rather than the artificially large rooms considered previously. As before, a point source was assumed to lie in a tri-corner. The room ratio was kept constant ($2.58 : 1.97 : 1$). The receiver was located at a distance varying with room size in the x and y planes but at a constant height of 1.3m.

The five frequency responses can be seen in Figure 5.9. Their densities varied considerably. Between 0Hz and 100Hz, the $50m^3$ room contains 13 eigenfrequencies, while the $1000m^3$ room has 154.

As both the literature and previous tests have shown, the frequency range over which we perceive modal effects is an important consideration. It is possible to alter this range in testing by limiting the crossover frequency between the modelled low frequency and the original music sample (see Figure 3.2 in Chapter 3). The test was therefore conducted with the crossover at two frequencies, 100Hz and 200Hz in order to draw a comparison between the two.

5.6.2 Test Procedure

During preparation of the test samples and pilot testing, it was noted that although absolute differences could be heard, these were considered small in terms of perceived quality. One explanation for this is the constant decay time, which, as will be discussed further in Chapter 6 is an important perceptual factor. Furthermore, particularly when considering the cut-off frequency of 100Hz, differences are harder to perceive due to our reduced perception at the lowest frequencies.

A task requiring subjects to rate all samples using a direct scaling method is difficult under these circumstances. Such challenges are noted by Bech and Zacharov (2006). For these reasons, a pair-wise comparison method was chosen, with subjects asked whether the quality of one sample was worse/same/better than another. The 'same' option was included due to having a small sample set of five samples. In subsequent testing of perceived quality, another method of paired comparison has been used, simply asking the subject to choose between two samples (see Chapters 8 and 9). Here however, the method allowing the three options was chosen. This has been used successfully by Huang et al. (2008). The method also allows for a comparison of identical samples which can help determine subject accuracy. Each sample was rated against each other including reversals. For example Sample A = $50m^3$, Sample B = $100m^3$ and also Sample A = $100m^3$, Sample B = $50m^3$. Listeners were instructed to audition Samples A and B, and then make a decision based upon the '*overall quality of low frequency reproduction*'. They were encouraged not to spend a great deal of time on each comparison - if they could not detect a noticeable difference, it was to be assumed that the quality was the same.

The interface was created in the MATLAB environment as shown in Figure 5.10. Before undertaking the test, subjects were given a short training phase where they were played the music sample that would be used in the test, with a variety of differing modal artefacts, including some which were exaggerated beyond the levels present in the test samples. This stage was designed in order for the subjects to become accustomed with the sample and the likely degradation effects. Furthermore, many of the subjects had taken previous tests and were becoming increasingly familiar with the music sample used.

Seven subjects were tested, with all but one having had prior listening test experience. All had experience mixing music in a number of listening environments.

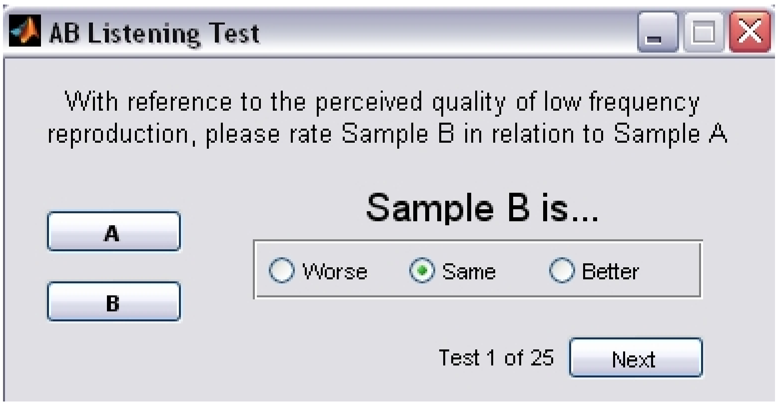


Figure 5.10: Graphical user interface for testing the perceived quality of modal density

Furthermore, five subjects had been through a listening panel screening test including audiometry and an introduction to critical listening comparisons. One subject reported tiredness before taking the test, although analysis shows their results to be similar to the others and so are included herein.

5.6.3 Results

The results for each subject's 25 comparisons were placed in a matrix, using -1, 0 and 1 for the ratings 'worse', 'same' and 'better' respectively. The resulting matrix was analysed not only to determine the quality rating of each sample, but also for 'judgement errors' which allow the validity of the scaling to be considered. As mentioned, this technique has been used successfully in the work of Huang et al. (2008) in rating the annoyance of noise samples and the analysis here follows a similar process.

An example matrix for Subject 1 is shown in Table 5.5. As can be seen from the table, the subject correctly identified each identical pair as the same (deep shaded cells scored 0). The scores in each row relate to the perceived quality of B against the column A. For example, the scores show that Sample 4 was rated as better than Sample 2 (score of 1). When this pair were rated the other way around, we can see that Sample 2 was rated worse than Sample 4. If this had not been the case, there would be an inconsistency in judgement.

In order to analyse the results, the first step is to determine the level of judgement errors for each subject. Two types of error are evaluated here - *self comparison errors* (sc) and *difference comparisons* (comp). Self comparisons refer to the case where Sample A and B are identical, while the difference comparisons are those where the reversal of the rooms for samples A and B are considered (see Table 5.6).

		A					
		1	2	3	4	5	R_{row}
	1	0	0	0	-1	-1	-2
	2	-1	0	0	-1	0	-2
B	3	0	1	0	-1	-1	-2
	4	1	1	1	0	0	3
	5	0	1	1	-1	0	-1
	R_{col}	0	3	2	-4	-2	

Table 5.5: Example matrix of a subject's responses to the five room volumes

Error Type	Error Recorded When:
sc	$R_{ij} \neq 0$ when $i = j$
comp	$R_{ij} \neq -R_{ji}$

Table 5.6: The calculation of judgement errors

For each matrix there are a total of five possible self comparisons and ten possible sources of error between comparisons of two samples. A total of 15 errors were therefore possible. The error rate was calculated for each subject as a percentage of this total (Table 5.7).

It is clear that the level of judgement errors was very high. The average error across subjects and the two test crossover frequencies was 49.9%. Therefore the quality ratings revealed in the following section are considered indicative rather than conclusive.

A number of comments from listeners suggested that the definition of quality was confounded by the differing effects of the bass guitar and kick drum parts.

5.6.4 Perceived Quality Matrix Analysis

The matrices were also analysed for perceived quality in order to further inform our discussion on modal density, although this is undertaken with some caution due to the high level of error. An overall quality score may be extracted by using Equation 5.3. Note that the value of R_{col} reveals the opposite quality result to R_{row}, hence the minus sign.

	Subject						
	1	2	3	4	5	6	7
100Hz	33	53	13	87	33	60	60
200Hz	40	53	67	33	27	80	60

Table 5.7: Percentage of judgement errors made by each subject

	Subject							
Volume (m^3)	1	2	3	4	5	6	7	Mean
50	-1.0	-1.0	0.0	-0.5	-2.0	-2.0	0.5	-0.86
100	-2.5	-3.0	-4.0	2.0	1.0	-1.5	-1.0	-1.86
250	-1.5	2.0	-1.0	0.0	-4.0	1.5	-2.0	0.71
500	3.5	1.0	2.0	0.5	3.5	0.0	2.5	1.86
1000	1.5	1.0	3.0	2.0	1.5	2.0	0.0	1.57

Table 5.8: Subjects quality ratings of room volumes - 100Hz crossover frequency

	Subject							
Volume (m^3)	1	2	3	4	5	6	7	Mean
50	-1.0	1.5	-2.0	-2.5	-3.0	-1.0	-2.5	-1.50
100	-2.5	-4.0	-2.0	-2.5	2.0	-0.5	-1.5	-1.57
250	-1.5	3.0	0.0	0.0	-2.5	0.0	-1.5	-0.36
500	3.0	-0.5	2.5	2.5	3.5	1.0	3.0	2.14
1000	2.0	0.0	1.5	2.5	0.0	0.5	2.5	1.29

Table 5.9: Subjects quality ratings of room volumes - 200Hz crossover frequency

$$\bar{R} = \frac{R_{row} + (-R_{col})}{2} \tag{5.3}$$

Tables 5.8 and 5.9 show the average score determined by Equation 5.3 for each subject and each room volume.

Such analysis produces a subjective scale. The mean across all the sample set occurs at zero, hence both positive and negative ratings. The greater the spread (where 10 is the maximum, between -5 and 5) the greater the perceived difference. Viewing Tables 5.8 and 5.9, we are able to see that the scores do appear to differ across the five rooms, although the spread is not particularly wide. For the rooms cut off at 100Hz, this spread is 2.72, which is equivalent to using around one quarter of the possible scale in a standard direct rating test. The low spread suggests the samples are indeed perceived as having a similar quality regardless of their differing densities. Interestingly, whilst the individual scores are not identical for the two crossover frequencies tested, the *order* of perceived quality of each room volume is preserved.

It is also noted that we do not see a linear trend of increased quality as modal density increases. However, the analysis does suggest some similarities at both crossover frequencies, possibly suggesting dominant artefacts in the lowest frequency region. The best and worst volumes are $500m^3$ and $100m^3$ respectively.

Through a visual inspection of the frequency responses (Figure 5.9), we may speculate as to why certain rooms score as they do. There are a number of ‘features’

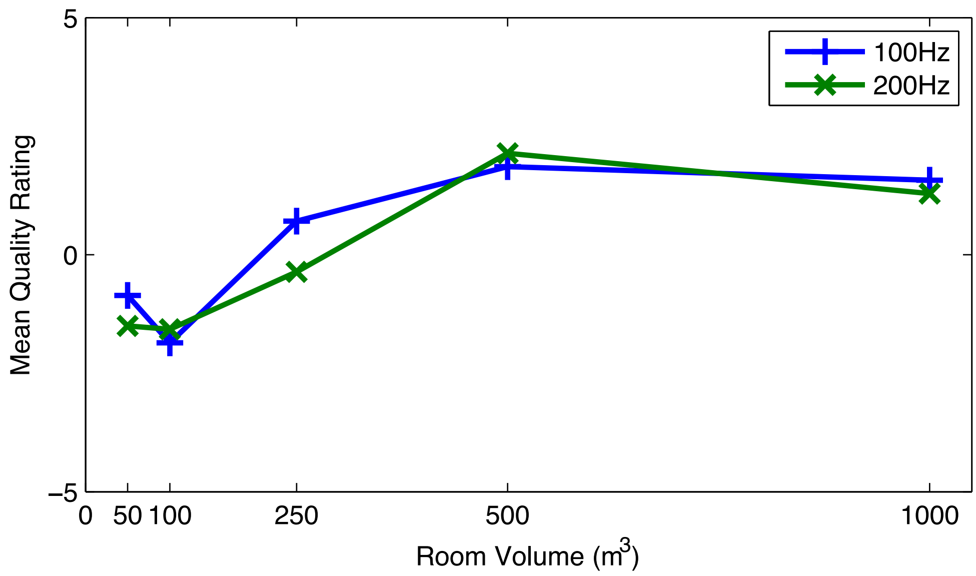


Figure 5.11: Mean perceived quality at each room volume for the two crossover frequencies

of the response, such as large notches or dips at particular frequencies, obvious peaks, or even an overall smoothness which may be responsible for their ratings. However, such speculation can be problematic. What is clear is that a greater understanding of the perceived audio quality in relation to a specific room response is required. This is considered in Part III.

5.7 Modal Density Discussion

The investigation of the modal density has been based around three listening tests, revealing an increasing understanding of the perception of this parameter in a variety of scenarios, from the absolute threshold omitting mode shapes with artificial stimuli to realistic models and stimuli and finally perceived quality.

As Chapter 4 showed, there has been widespread use of room metrics based upon the modal distribution which solely take account of the room dimensions, without consideration of the position of either the listener or the loudspeakers. Through the modelling of such a scenario, it is possible to observe the smoothing of the room response as the density increases. Obtaining a threshold in such a scenario, where it is no longer possible to distinguish between the room's response to two differing densities, allows us to determine something along the lines of a 'subjective counterpart' to the Schroeder Frequency. Using test tones which interact with the room only at a their specific frequency and a narrow band around it, an absolute threshold has been obtained.

It is noted that this threshold at the lowest frequencies, is around four modes per bandwidth. This correlates closely with the Schroeder Frequency, which suggests three modes per bandwidth is sufficient to move into the statistical region. The important issue to be wary of here is the linking of the transition into a statistical region with the subjective case of a density having been met where reproduction can no longer be distinguished from a very high reference case. Furthermore, this result appears to be in agreement with the optimal spacing revealed in Chapter 4. Here, the optimal spacing which would result in the shortest decay, which also relates to the smoothest response, is 1/3rd of a bandwidth of one mode, resulting, of course, in an average of three modes per bandwidth.

We may also observe that the spacing investigation revealed that sensitivity to decays was greater at higher frequencies. This is confirmed by the threshold of density results - a much higher number of modes per bandwidth are required at 125Hz and 250Hz. It is suggested that the individual peaks and dips at higher frequencies are perceived as more obvious degradation, and therefore a much smoother overall response (obtained only when mode-shapes are omitted) is required before we become unaware of the degradation.

Whilst a useful starting point in our investigation, one should carefully consider the use of thresholds based upon scenarios unlikely to occur in real rooms. More accurate modelling of the room, including the source and receiver positions, is considered more revealing in the search for a subjectively relevant density threshold. The interaction of mode-shapes results in irregularities in the frequency response. There is no longer a direct correlation between the modal density and the smoothness of the response. In large rooms, the typical definition of statistical region is reached at lower frequencies than in smaller rooms. It may therefore be assumed that low frequency treatment is less necessary where a large space exists. The listening test undertaken here reveals that the subjective response to a range of small and large rooms show almost identical patterns. It is clear that there is no linear relationship between density and the perception of modal artefacts.

It would seem natural that detection of differences between two modelled rooms is based upon their differences in terms of transfer function. It is for this reason that the final listening test was undertaken, in order to determine if there is an increase in perceived quality in rooms of higher density. The results presented show large judgement errors, and it is suggested that this is due to a) the complexity of the subjective task, asking subjects to evaluate ‘perceived quality’ and b) that the modal density does not sufficiently affect the quality of reproduction so as to be noticeable.

To conclude, it appears that we have too often considered modal density as a relevant subjective parameter. We should instead be focussing on the individual frequency response. Part III of this thesis is motivated by this finding.

5.8 Summary

Modal density thresholds have been investigated in the pursuit of perceptually relevant objective room response parameters. The importance of considering the interaction between room, source and receiver has been shown. When this is accounted for, and realistic stimuli are auralised, it is not possible to obtain thresholds. Instead, the specific frequency response becomes the dominant factor. It is suggested that it is the unique frequency response and not specifically the density which is responsible for the quality perceived by a listener. If these responses are to be rated in terms of quality, further detailed testing is necessary.

The results of these tests reveal two main conclusions. Firstly, it is the overall evaluation of a room which reveals its quality, rather than the insistence on whether a difference between two samples can be heard or not. Secondly, we can conclude that modal density is not a particularly relevant parameter in terms of understanding perception of low frequency reproduction quality. Based on findings in the previous chapter and previous research, we consider the perception of modal decay from here on. Therefore, the following chapter will further study this decay time in an attempt to define new thresholds, and later, Part III will consider the question of perceived quality, rather than absolute detection of differences.

Chapter 6

Thresholds of Modal Decay

6.1 Introduction

The previous two chapters have investigated parameters which are related to the modal distribution. In attempts to define optimal values, the spacing experiment looked for the shortest perceived decay and the density found a threshold with reference to a smoothing of the response - thereby reducing the decay time. The common factor is the importance of a short decay time for audio reproduction to be perceived as high quality. In response to this, and with the research gap identified in the literature review for further investigation, this chapter attempts to define a subjective threshold for modal decay.

6.2 Modal Decay

An individual mode has a decay characteristic which is related to both the mode type (axial, tangential and oblique) and the absorption present within the room. At higher frequencies, it has been noted that it is acceptable to consider a room as a diffuse field. Statistical calculations can therefore be applied and it is possible to derive a *reverberation time* which is applicable across a particular frequency band. At lower frequencies, this concept breaks down - it is possible for individual modal frequencies to decay at different rates (Howard and Angus, 2001). The uniform decay necessary to define a quantity such as reverberation time (RT_{60}) may not be present at low frequencies.

6.2.1 The Importance of Decay

The previous chapters have shown that the decay time of modes is of importance when assessing a room's low frequency reproduction. Indeed, this is in line with previous research, which has identified modal decay as possibly the principal element in our subjective assessment of low frequency audio (Avis et al., 2007). If this is the case, optimisation targets are of great importance to the industry. By defining thresholds below which no further reduction of decay is perceptible, both financial and physical space savings may be made.

The literature review highlighted attempts to use modal equalisation to control the room response. Such equalisation methods, whilst potentially reducing the decay time of individual modes, are based upon a modification of the signal input to the loudspeakers. They are also prone to introducing large distortions into the electroacoustic devices or a high cost into digital filters when trying to deal with frequency peaks or dips that are caused by modal interaction. This is particularly undesirable if the decay present is unlikely to cause a perceptual degradation of audio. As Goldberg explains:

> "objective system equalization is no longer the limiting factor in improving the listening experience, so the focus has moved towards a listener's ability to perceive an audible sound quality improvement" Goldberg (2009)

6.2.2 Earlier Thresholds

Each mode has a number of parameters, the amplitude, frequency, and also the so called 'quality factor' (Q factor). As shown in Chapter 4, this Q factor is related to the decay time. The higher the Q, the longer the ringing in the time domain. In order to study the perceptual response to these Qs, Avis et al. (2007) attempted to define Q thresholds through the subjective testing of modelled critical listening rooms. By modelling modes using BiQuad filters, the Q factors could be dynamically varied. Thirteen modal frequencies were produced using these filters, and the summation produced the frequency domain room model. These listening tests showed an absolute threshold of Q=16, below which, resonances were inaudible. The Q factors were kept constant across the frequency range (20 - 200Hz), meaning that the decay time for any frequency where Q=16 can be estimated. These are 1.1, 0.56 and 0.28 seconds at 32, 63 and 125Hz respectively. As the Q/decay relationship is frequency dependent, the decay result can only be extrapolated and the suggested decay thresholds are indicative only. Furthermore, during this testing, full room models were not considered, in order to maintain control over the Q factors, and the model took no account of the interaction of mode-shapes.

In a similar experiment, Karjalainen et al. (2004) studied perception of decay time directly at a number of frequencies. It was found that at typical listening levels down to 100Hz, the threshold increased from about 0.3 to 0.4. However, when testing at 50Hz, subjects observed no noticeable differences for decay times of up to two seconds.

Interest in the perception of modal decay has remained since these tests. Goldberg (2005, 2006, 2009), in a series of papers aimed at defining such thresholds produced a number of preliminary results. In his original work, modal decays were to be detected in the presence of sine sweeps and noise. In later work, he defined more suitable sine burst test signals in order to improve the methodology for revealing thresholds (Goldberg, 2009). In the light of these previous studies, this chapter attempts to continue the investigations begun, and produce valid, frequency dependent thresholds.

6.3 Determining Thresholds

In search of such thresholds, this chapter follows a similar structure to Chapter 5. Firstly, absolute thresholds of decay will be studied. These are achieved by testing using specifically chosen artificial stimuli. In much the same way as the previous chapter investigating optimal density, using artificial stimuli is seen as an important starting point as it is useful to compare such absolute thresholds with those found under more realistic scenarios. Furthermore, they also provide ‘worst case’ guidelines. Additionally, if the absolute thresholds are found to correlate with those of a more realistic nature, it may be concluded that a valid decay threshold exists regardless of factors such as musical genre. Following the absolute threshold testing, an investigation into obtaining frequency dependent thresholds for musical stimuli in realistically modelled rooms is presented. This presents a considerable challenge due to the interaction of modes and musical stimuli. A number of methodologies and listening tests are reported, culminating in a set of final thresholds and a discussion of their importance in the wider context of this thesis.

6.4 Artificial Stimuli

6.4.1 Test Stimuli

A suitable test signal must be generated if accurate thresholds are to be found. There are a number of signals commonly used in psychoacoustic testing, such as white and pink noise, pulses, logarithmic and linear sweeps and pure tones. These signals can be seen to represent differing components of a natural signal. For example, pulses have been used by Olive et al. (1997), where it is suggested that they are helpful in revealing the audibility when a musical signal contains transient sounds. Single tones are representational of harmonic elements while noise can be used as a controlled artificial stimulus to reveal perceptual responses to non melodic musical elements.

Goldberg (2009) has reported that one of the most useful signals for highlighting modal decays is the pure tone sine burst. A simple sine tone can be amplitude modulated with the exponential decay curve of the single frequency room mode in question. This sine burst was chosen as the most appropriate stimulus. The decaying sines simulate the resonant system directly, and therefore a convolution of stimulus and system is not required. Rather, each test can be a single burst at the exact frequency under test. It was decided that thresholds should not be obtained for other artificial test stimuli. This would increase the demand on subjects and, as the goal is to obtain absolute thresholds before continuing to determine thresholds in realistic music and room cases, these other stimuli which may represent elements of that music were considered unnecessary.

The construction of an ideal sine burst is discussed in detail by Goldberg (2009). The method is briefly presented here for clarity. Firstly, the Hann window was identified by Goldberg as the ideal window shape to reduce spectral spreading. This is important as when dealing with very short thresholds, any spurious effects introduced by either the playback equipment or signal processing methods have the ability to interfere with the results. A subjective test was then performed to determine the minimum required Hann window length independently at both the beginning and end of the burst to ensure that spectral leakage is not audible. The results showed that above 100Hz, there should be 15ms window, which he recommends to be extended to 30ms in order to be certain that no leakage is audible. Below 100Hz, three cycles of the sine tone were found to be a sufficient fade out. These envelopes are well below the decay times observed within rooms and also much shorter than the thresholds suggested by previous studies, and may therefore be used as the fade out times for the reference samples with which to compare longer decays to. Figure

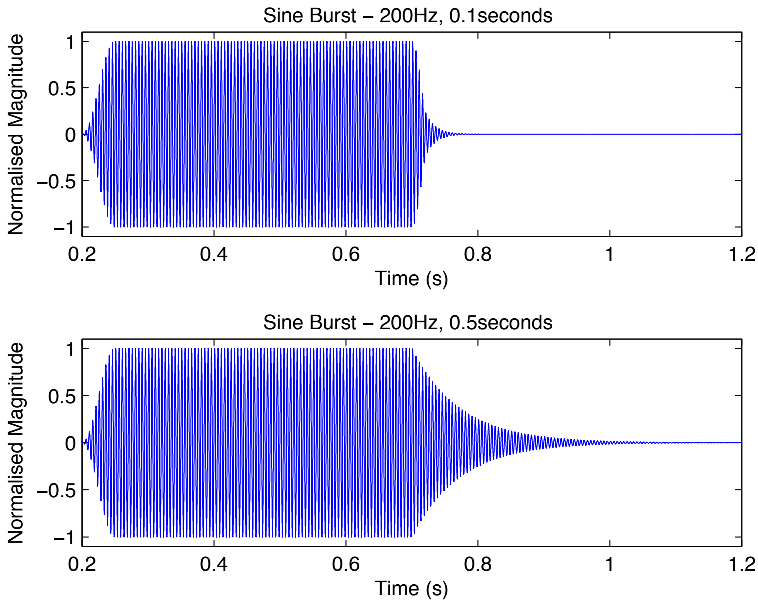


Figure 6.1: Sine bursts at 200Hz with decays of 0.1 and 0.5 seconds

6.1 shows the time domain representation of two sine bursts with different decay times.

6.4.2 Independent Variables

As with many psychoacoustic phenomena, thresholds of decay are expected to be frequency dependent. In order to test this variable systematically, and also compare results to those of previous preliminary tests, five test frequencies were chosen - 32Hz, 63Hz, 100Hz, 150Hz and 200Hz. It is also likely that thresholds will be dependent upon replay level. Therefore, each frequency was tested with SPL levels at the ear of 85dB and 70dB. Again, these were chosen as they correlate with previous testing (Goldberg, 2006) and can be seen to represent a range of expected signal levels in critical listening rooms.

6.4.3 Test Methodology

The hybrid PEST/ABX method previously presented in Section 5.4.3 was again employed, with the adaptive PEST rules allowing a direct threshold to be obtained in as short a time as possible and the additional ABX section ensuring subjects could not simply report that they hear differences when in reality they do not.

The testing therefore used fixed reference sine bursts with no decay (windowed as previously discussed in order to remove any audible spurious effects), and a second

Frequency (Hz)	32	63	100	150	200
Calibration Coefficient	0.18	0.15	0.13	0.13	0.13

Table 6.1: Calibration coefficients for 85dB playback of tone bursts at each test frequency

sample whose decay time was systematically varied using the PEST rules. A number of additional variables were logged, including the time taken for each PEST run. As the testing can cause fatigue, and training effects have also been witnessed in previous work of this nature, the first test was repeated at the end. The presentation order of all other tests was randomised across subjects. The repeated test was the 100Hz tone replayed at 85dB (linear). This case was chosen as the control as pilot testing showed that it was one of the most revealing in terms of decay. Therefore a total of 11 tests were conducted - five frequencies at two replay levels, with the first test repeated once at the end.

Breaks were enforced every 15 minutes by the test operator, and subjects were instructed to take additional breaks whenever required. Testing took place within the listening room at the University of Salford, and the host computer was located outside the room, leaving only headphones, mouse and monitor within the test environment to eliminate noise from the computer.

6.4.4 Calibration

With five discrete frequencies tested, it was possible to calibrate the replay levels at each individual frequency. This was achieved by calibrating the headphone output with a B&K HATS system and a Norsonic Sound Analyser. The right ear of the HATS was removed allowing access to the microphone, which was calibrated using a standard B&K calibrator. The ear was replaced and the level of normalised sine tones replayed at each frequency through the Sennheiser HD-650 headphones was recorded with the Sound Analyser. A correct attenuation coefficient for replay at both 85dB and 70dB could then be calculated according to the measured level of the sine tone and the SPL recorded at the ear. All measurements were the unweighted, linear frequency response. The analyser was unable to produce a direct measurement at 150Hz, and the coefficient was therefore calculated through interpolation based on the attenuation for other frequencies. Playback was through the same M-Audio Firewire 410 interface as in previous chapters. The reproduction chain was unable to produce a 32Hz burst at 70dB and this lowest frequency was therefore only tested at the 85dB replay level.

6.4.5 Results

This section presents the results of the subjective data that were gathered, beginning with an analysis of subject performance.

Subject Performance

Two measures of subject performance have been calculated. Firstly, the number of occasions where the PEST routine failed to converge, and secondly, the consistency of threshold between the repeated first and last tests. Large differences in this value may be as a result of either fatigue or training effects. It is not possible to directly measure this, as it remains a possibility that both effects occur and offset each other. Table 6.2 shows the number of occurrences for each subject where the PEST routine failed to converge, and the reason for this failure. Exceeding the number of allowed trials (20) is a possible indication of subject fatigue. A flagging of the maximum reversals exceeded can also suggest fatigue, showing the subject oscillating above and below the probable threshold. This may occur when a short decay is not detected and the PEST rules increase it to an obvious level. When this is detected, the decay reduces, at which point the fatigued subject may again fail to detect the decay. If a large number of subjects were to exceed the maximum number of reversals, this would indicate that the minimum step size in the PEST logic is too small. This was not the case, with only 4.7% of PEST runs resulting in this error. Finally, if the maximum decay of two seconds is exceeded, it is an indication that the subject is unaware of the artefact they are being asked to identify. Herein it is noted that if any of these flags were identified, and no threshold level recorded, subsequent analysis ignores these trials on an individual basis.

The scores given for the first and last tests show how consistently subjects performed, and is also reported in Table 6.2. In all but two cases, subjects obtained a higher threshold for the first test than the last (positive difference). Where this difference is large, there is some suggestion that training effects may have occurred. Subject 16 displayed an abnormally large difference in first and last test thresholds of 0.375 seconds in addition to having two tests fail to produce a threshold due to exceeding the maximum number of reversals. This subject's data were removed before further analysis.

The number of trials taken for each PEST run can also be seen as an indicator of subject performance. For all subjects, across all frequency and level combinations this was 10. In detail, we see the average number of trials in Table 6.3. There is no significant difference across the 11 tests, which suggests that there was no one frequency/level combination which was easier or more difficult than the others.

Subject	Max Trials	Reversals	Max Decay	Difference (s)	Time Taken (mins)
1	0	1	0	0.100	24
2	0	0	1	0.000	22
3	0	0	0	0.150	24
4	0	0	0	0.050	27
5	0	0	0	0.125	37
6	0	0	0	0.025	27
7	0	0	0	0.299	18
8	0	0	0	0.250	24
9	0	0	1	0.000	27
10	0	0	0	0.000	35
11	0	0	0	-0.050	26
12	0	0	0	-0.025	21
13	0	2	0	0.125	34
14	0	2	0	0.074	32
15	0	1	0	0.125	46
16	0	2	0	0.375	18
17	0	0	0	-0.001	35

Table 6.2: Analysis of the subject performance during testing of the absolute thresholds of modal decay

Frequency (Hz)	32	63	63	100	100	100	150	150	200	200
Replay Level (dB)	85	70	85	70	85	85	70	85	70	85
Average Trials	8	11	10	9	11	10	10	9	8	9

Table 6.3: Average number of trials taken in each test for the PEST run to converge

The mean time taken was 29 minutes, which included time taken for breaks.

Thresholds

Figure 6.2 shows the mean detection thresholds at each frequency and replay level. The figure also shows the standard deviation across the sixteen subjects. Overall, there is a clear trend, with threshold increasing with decreasing frequency. The threshold appears to be asymptotic to about 0.2 seconds at the higher frequencies, regardless of replay level. However, with a quieter stimulus, at 63Hz, the threshold is higher. This behaviour can be compared to the hearing threshold curves, which show that our sensitivity reduces at lower frequencies (Fletcher and Munson, 1933). Reproduction at 70dB at 63Hz is perceived substantially quieter than the same level at 125Hz. With reference to the perceived level at 1kHz, at 63Hz the 70dB burst will be perceived as approximately equivalent to 40dB. At 125Hz this will be approximately 60dB.

The significance of these results is analysed in Section 6.8, which also draws a comparison to thresholds obtained with musical stimuli.

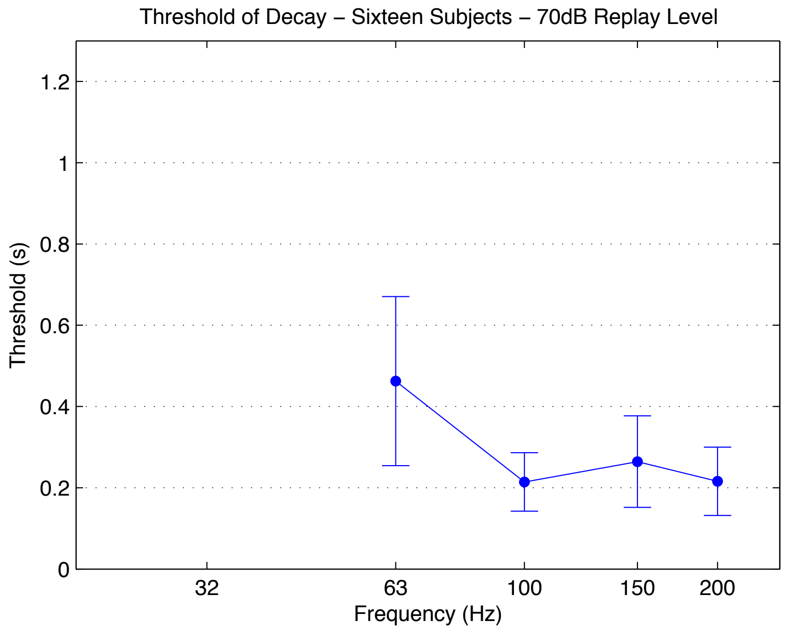
Comparison with previous studies

It is interesting to compare these results with those of previous tests. Firstly, there are the initial results from Goldberg's small scale threshold testing (Goldberg, 2005), which was primarily conducted to confirm the success of the PEST method for obtaining such decays. Precise threshold data was not reported, although a graph containing the median value across six subjects appears in the publication. At 32Hz and 85dB reproduction, the threshold is around 1.35 seconds, dropping rapidly to around 0.5 seconds at 63Hz, with values of around 0.1 seconds at 80, 125 and 200Hz. These results, while not identical, are of a similar order to the results obtained in this chapter and differences are attributed to the small size of Goldberg's test and inherent problems with test stimuli and methodology. Karjalainen et al. (2004) have also reported low frequency decay thresholds. Their method differers somewhat, with testing carried out in a real room over loudspeakers. A single synthetic mode was added to a number of different stimuli (music, speech, noise and percussive sound). When accounting for the reverberation time present within the room, they report a fairly constant 0.2 - 0.3 second decay time threshold down to 100Hz, at which point the threshold increases rapidly. The thresholds revealed in Figure 6.2 show good correlation with Karjalainen's study.

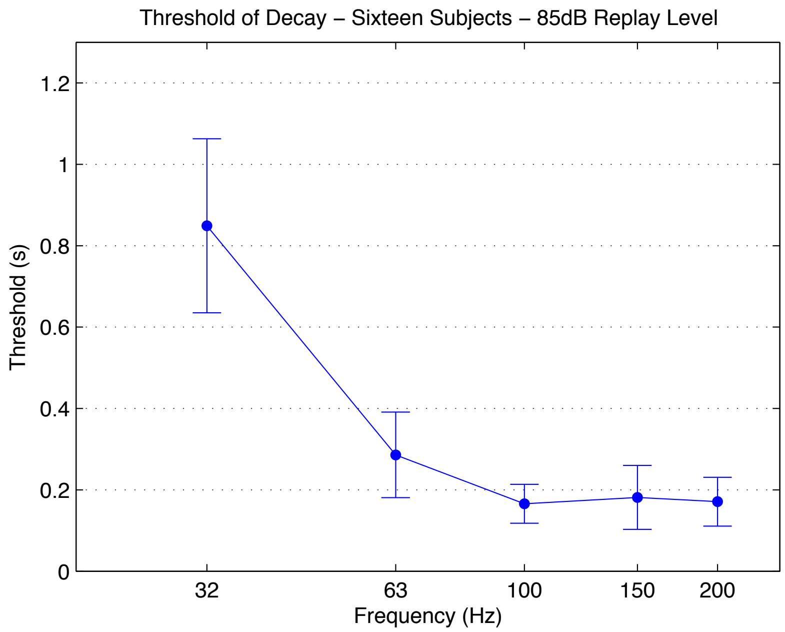
Finally, assuming the thresholds of Q reported by Avis et al. (2007) can reliably be used to extrapolate approximate decay thresholds, values of 1.10 seconds at 32Hz, 0.56 seconds at 63Hz and 0.28 seconds at 125Hz were found. Avis et al.'s results are slightly higher than the absolute thresholds observed in this chapter, but it should be remembered that they were using musical stimuli. Thresholds established with musical stimuli are therefore expected to be higher, and will be more directly comparable to the subsequent tests (Section 6.5) using musical stimuli.

6.4.6 Absolute Thresholds Summary

These results highlight the absolute thresholds of decay. A threshold between 0.2 and 0.3 seconds above 100Hz has been observed and is consistent with previous work in this area. The result can be seen as a 'worst case scenario' - that is, a target to be aimed at if no further decay reduction is to be perceived regardless of input stimuli in the room.



(a)



(b)

Figure 6.2: Decay time thresholds for sine bursts, a) 70dB and b) 85dB replay levels

6.5 Natural Stimulus Thresholds

Thresholds of decay will be most useful to studio designers and those wishing to treat existing facilities when they account for the room environment and the typical audio stimuli which will be replayed there. The absolute thresholds reported above do not take into account the possible effects of the presence of natural signals, such as music or speech, or the interaction of modes within the room. It is therefore necessary to test for similar thresholds whilst also accounting for these more realistic effects and stimuli.

6.5.1 Test Stimuli

Two music samples were used for the determination of decay thresholds. These were the LEN and HC samples introduced in Section 3.6. These represent samples with both short, punchy low frequency content, and also a resonant acoustic bass notes with a naturally longer decay envelope.

6.5.2 Decay Model

The modal decomposition model was again implemented to generate room impulse responses with varying decay times. The input parameters to the model, other than the decay time, were kept constant, as follows:

- Room volume: $100m^3$
- Dimensions: width=6.97m, length=5.32m, height=2.69m
- Source position: front-left-bottom tri-corner (modelled as point source)
- Receiver position: width=3.16m, length=1.97m, height=1.3m
- Model frequency resolution: 0.12Hz

In the previous chapter, damping was modelled as frequency dependent, following an exponential curve to reduce the decay with increasing frequency. In this investigation which studies the decay directly, the decay was kept constant across frequency. This differs from the experimental work of Avis et al. (2007), in their investigation of the threshold of modal Q. That work modelled resonances with 13 Bi-Quad filters, and the independent variable was Q, which was therefore kept constant, resulting in differing decays across frequency.

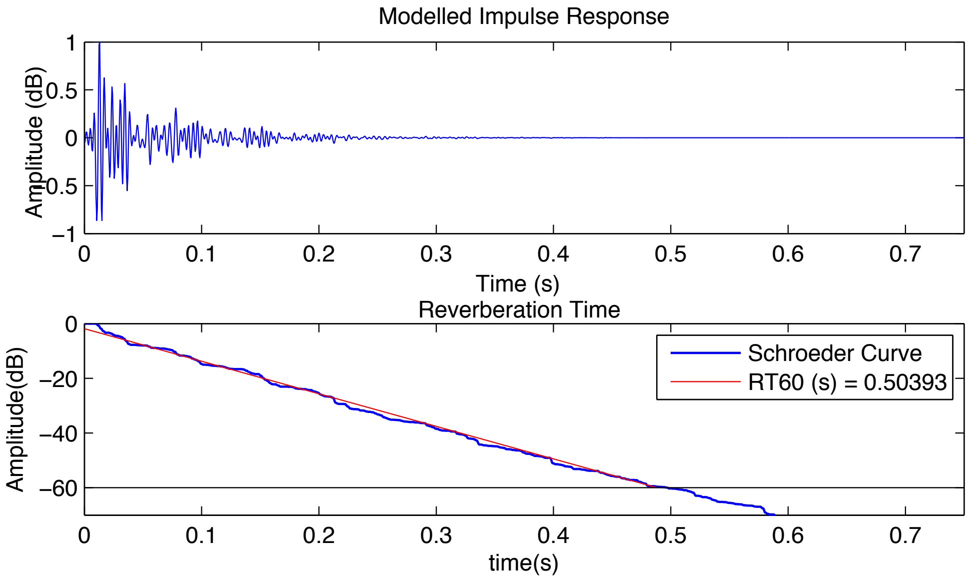


Figure 6.3: Model generating a 0.5 second decay time across frequency, and Schroeder backwards integration method confirming the RT_{60} of 0.5 seconds

In the test presented here, the decay time was controlled through the analytical model's damping parameter δ in Equation 3.27. The required alpha (α) for a given decay time was obtained through the use of Sabine's equation relating reverberation time (T_{60}) to the absorption coefficient, α (see Morse (1948)). It is therefore possible for decay to be modelled dependent on both frequency and boundary absorption, although this was simplified in this model by attributing a single α for all surfaces. The impulse responses produced from an inverse Fourier transform of the resultant complex pressure array may be verified using a Schroeder backward integration plot. Figure 6.3 shows the impulse and integration plot for an input RT of 0.5 seconds. We can see that the model produces a time response with a reliable decay time, whilst still retaining the typical modal interactions that are likely to occur in real rooms.

It is also important here to note the validity of the modal decomposition model. It is accepted that the model is usually considered valid only where damping is low and room geometry simple (Morse, 1948; Kuttruff, 1991). Its ability to accurately model a real room response breaks down when the damping becomes high. It could therefore be argued that it is unsuitable to use the model to test for decay thresholds, as the assumptions of low damping will be violated as this parameter is increased towards a perceptual threshold. However, it is argued that the model does continue to provide a *general case* of the room response (Fazenda, 2004). It is this general case, as opposed to a highly accurate room model which is required here. In other words, it is the decay time produced by the model, rather than an exact representation of the modal sound field in the room which is needed in order to reveal the desired decay thresholds.

Frequency Dependency

As observed from the tests run using artificial stimuli, the thresholds of decay are frequency dependent. It is simple to test for this dependency where the stimulus is a single tone, but this becomes more complex when considering a musical signal. In order to achieve frequency dependent results, the auralisation model was implemented with a variable 'cut-off' frequency. This corresponded to the crossover between the model convolved with the low frequency region of the music sample, and the original sample (see Figure 3.5.1). Three crossover frequencies were tested, 63Hz, 125Hz and 250Hz. This allows a 'cumulative frequency dependency' to be observed. For example, where the cut-off is 63Hz, the threshold revealed is applicable up to this frequency. The auralised sample above this simply reproduces the sample in its original form. This will mean that with 125Hz cut-off frequency, the samples will also include the modal decays below 63Hz. If the thresholds for a music signal follow those of the artificial stimuli, this cumulative effect is not seen as problematic as the thresholds are likely to be higher at the lower frequencies. For example, if the threshold was found to be 2 seconds at 63Hz, and 1 second at 125Hz, we could attribute the 1 second threshold specifically to the region around 125Hz as this decay will not have been perceived at 63Hz.

6.5.3 Methodology

An initial pilot test was run using the same PEST/ABX methodology as with the artificial stimuli, in an attempt to reveal corresponding decay thresholds. However, it became apparent that, as seen in Chapter 5 in the study of optimal modal density - when testing using music samples and the decomposition room model, the PEST failed to converge. The reasons for this were similar to those observed in Chapter 5. Where a difference in the room response is apparent, the convolution of that response with an audio signal will always produce a sample where differences can be perceived when compared to a reference case. A particular difficulty here lies in the selection of a reference sample. If a very short decay is used as a reference, say, 0.05 seconds, trained subjects are likely to perceive differences between most samples right down to a point very close to 0.05 seconds.

It may be suggested that the subjects could be asked a slightly different question, such as, "is there a clear difference in the decay times of these two samples?" However, this substantially increases the complexity of the task, and adds a further subjective dimension - what should be considered a 'clear difference'? The problem can be highlighted further if we were to change the reference sample, say to 0.1 seconds. Pilot testing showed that the variable sample's decay was reduced by the PEST algorithm as subjects again perceived a difference right down to 0.1 seconds. If the specific PEST run created a variable sample very close to 0.1 seconds, no difference was perceived, and the decay time increased. Furthermore, if the PEST rules resulted in a drop *below* 0.1 seconds, once again, a difference would be perceived - leading the PEST to further reduce the decay! When the PEST did converge, it would consistently do so around that of the reference, regardless of what that reference was. Clearly, this does not reveal a decay threshold but rather, indicates where an audible difference is perceived.

Two particular observations are considered in the light of this pilot, each leading to the formation of a separate test. Firstly, differences in samples with differing decay times *are* perceived, leading to the conclusion that there may be a 'quality threshold' - a point at which further reduction in decay is *noticeable*, but does not impact the perceived *quality*. Secondly, it was observed that perceiving the differences within the context of the ABX test, although possible, was not simple. After a number of auditions of the reference and variable samples, an accurate decision could be made, but it would appear that on first impression the samples appeared similar. The audible differences could often be described as timbral, rather than due directly to decay time. It is suggested that if subjects are aware that they should focus specifically on the audibility of decays, and reduce the number of auditions permissible, rather than allowing them to constantly replay the samples for timbral differences, then a true threshold of decay may be obtainable.

In the light of these observations, the two tests were devised which focus firstly on the definition of a 'quality threshold' and secondly, on a 'first impression' threshold.

6.6 A Threshold of Perceived Quality

The following test was devised in an attempt to define a threshold of modal decay below which, no further improvement in reproduction quality is observed. The two music samples introduced in Section 6.5.1, and three cut off frequencies discussed above were tested, at replay levels of 75dB and 85dB (calibrated as per Section 6.4.4).

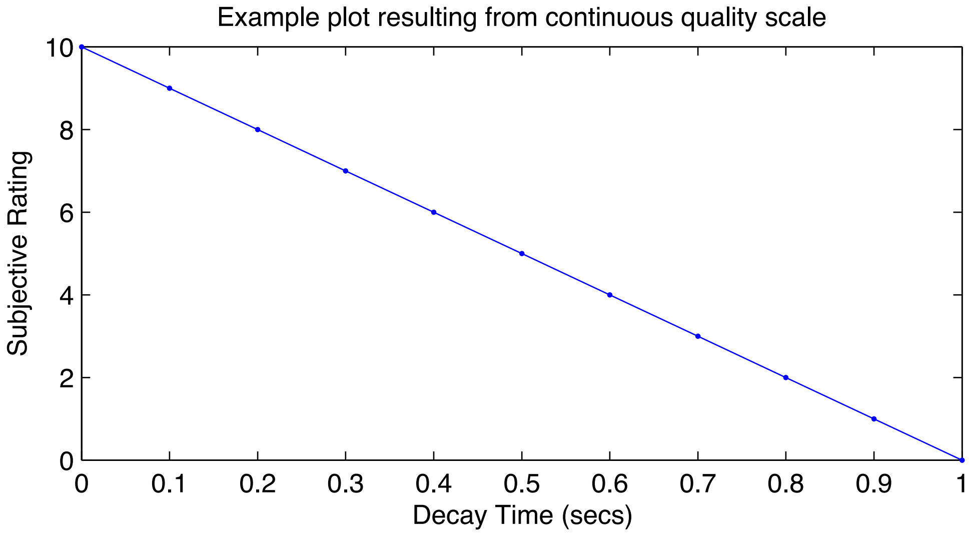


Figure 6.4: Example of the possible bias resulting from continuous scaling

6.6.1 Method

With three independent variables (frequency, music sample and replay level), each with multiple levels of decay, it was decided that the audio quality testing was most efficiently implemented using a direct scaling method (Bech and Zacharov, 2006). Indirect methods such as paired comparison (discussed further in Part III) are often unsuitable for tests with large numbers of independent variables, as the number of comparisons required rises exponentially.

It was therefore decided that, at each combination of sample, frequency and replay level (of which there were 12), a total of 11 auralised samples would be auditioned and rated. These would include auralisations of ten incremental decay times and a 'low anchor', which was to have a noticeably longer decay time. Each of the 11 auralisations were visible on a single GUI. The rating scale used was a forced choice, five level interval scale, chosen for a number of reasons. Firstly, if a sliding scale with a fine resolution was used, bias may be introduced due to the natural temptation to rate each sample differently, from best to worst. Therefore, even though two samples are perceived as the same quality, they are unnecessarily separated (Zielinski et al., 2008). As it is expected that a longer decay will result in a lower perceived quality of reproduction, subjects are likely to produce a data spread such as the one in Figure 6.4.

A similar effect may be seen even with lower resolution scaling, such as a fixed integer interval between 0 and 10. In this case, with 11 samples and 11 possible rating levels, the likelihood of subjects placing one sample per level is considered high. The fixed integer scale between 1 and 5 was therefore chosen. This forces a

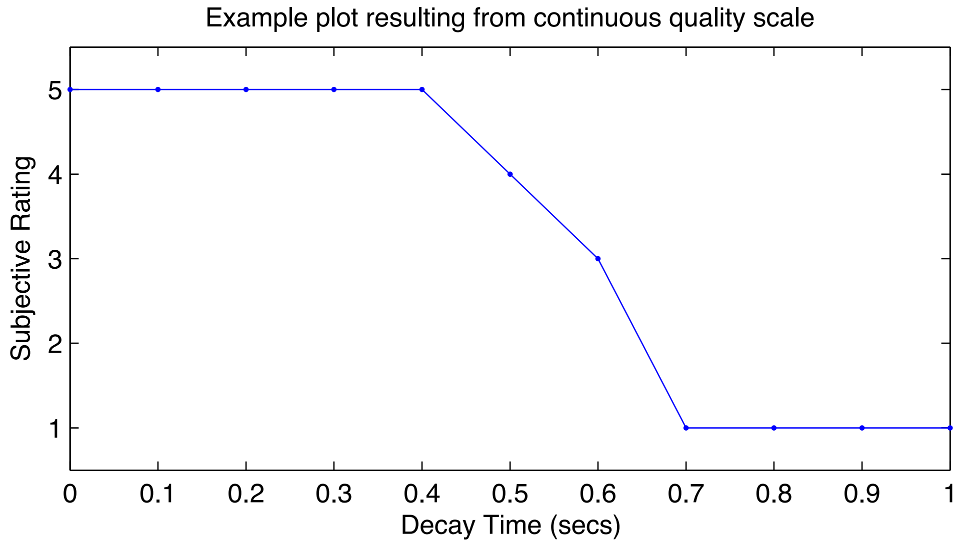


Figure 6.5: Example of a knee point at 0.4 seconds indicating a 'quality threshold' has been reached

number of samples to be rated in the same 'quality category'. The presence of a 'knee point' in the data would thus be indicative of a decay threshold. Figure 6.5 is a hypothetical example of this. Here, the quality is rated as equal for decays up to 0.4 seconds, at which point a drop would occur, suggesting that the decays are now audible and/or producing degrading reproduction artefacts.

Subjects were given full instructions, which included the following text:

> Please rate between 1 and 5, with 5 being the best, and 1 the worst. For every set of 11 samples, **you must score at least one sample 5 and at least one sample 1**.
>
> Please be aware that **although samples may not sound exactly the same, you may decide that they fall in the same quality band in terms of critical listening.**

Furthermore, the subjects were not explicitly instructed that they were listening to differences in decay times, rather, they were instructed to focus on the low frequency region and assess the quality of reproduction as if they were assessing within a critical listening environment. A total of 15 subjects completed the test, in the same session, but prior to, previous absolute decay threshold tests, in order that they were not inadvertently briefed that it was the decay threshold that was under investigation.

Freq. (Hz)	1	2	3	4	5	6	7	8	9	10	11
63	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	1.100	1.200	1.400
125	0.200	0.250	0.300	0.350	0.400	0.450	0.500	0.550	0.600	0.650	0.750
250	0.125	0.150	0.175	0.200	0.225	0.250	0.275	0.300	0.325	0.350	0.400

Table 6.4: Decay times in seconds of each set of eleven samples

6.6.2 Decay Times

The eleven decay times to be modelled were chosen in order to achieve a good spread from a short decay time, close to that at which the mode would have no perceptible effect on audio quality, to one where the mode would be clearly audible. As the sensitivity to these decays improves with increasing frequency, it was necessary to select a different spread of decay times for each of the three test frequencies. The greater the interval between each successive sample, the less accurate the revealed threshold will be, and yet too small a difference will result in a task of too great complexity.

6.6.3 Results

Figure 6.6 shows the mean subjective rating (with standard deviation error bars) for each frequency, decay time and music sample across 15 subjects. It is clear that no obvious knee points, which could be taken as an indication of a perceived quality threshold, are revealed. There is a visible trend, common across all 12 tests, with the expected behaviour of the lowest decay times having the highest perceived quality, and the longest decays the lowest. However, the standard deviation is high. Statistical tests show that there are no significant differences between the different decay times. Many of the subject's comments highlighted the difficultly in making subjective decisions regarding the sample sets. This was particularly so with 11 samples to be rated at any one time. Comments also suggested an increased difficulty in rating the jazz sample which contained the resonant double bass.

It is interesting that the test subjects commented that the samples were similar sounding and therefore difficult to distinguish from each other. This evidence appears to contradict the results from the pilot testing of the same samples using the PEST methodology where a difference could almost always be heard. It is clear that when asked to judge between similar samples, subjects struggle to quantify their perception when posed a question more subjective in nature, and yet, if posed a more clearly defined question, such as identifying whether an absolute difference is audible, they can do so with ease. It is also evident that with eleven samples to rate on a single interface, there is a greater complexity than where only two samples are to be auditioned. It is therefore hoped that the second test, based upon the subject's first impression of a sample, and with only two auralisations to compare at any one time, will reveal more satisfactory results.

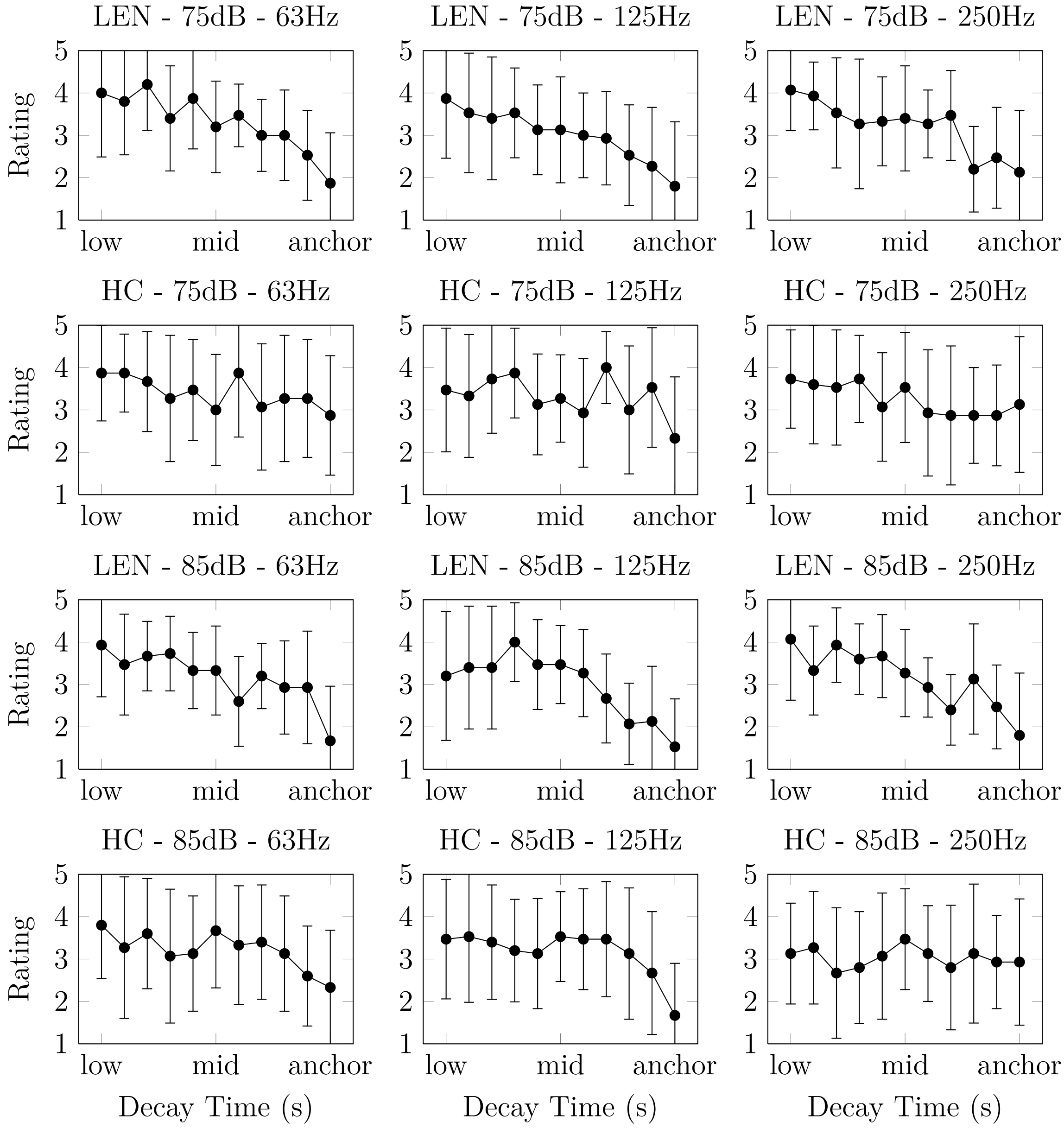


Figure 6.6: Mean Ratings across decay times for each of the cutoff frequencies

6.7 A Forced 'First Impression' Threshold

The second test was to investigate if a threshold could be defined for natural stimuli and realistic room models where a listener's first impression of the decay is considered. It should be noted that the use of such methodology does not imply that within a critical listening scenario, our perception is only determined by initial impressions. Rather, such a method is simply used as a tool to extract more reliable thresholds. Testing was carried out with a modified version of the MATLAB interface used for the artificial stimuli PEST/ABX test.

6.7.1 Method

In keeping with the above test, the same twelve music samples were used - two musical excerpts, cut-off frequencies of 63Hz, 125Hz and 250Hz and replay levels of 75dB and 85dB at the ear.

The PEST rules remained identical to the artificial stimuli test, while the ABX section allowed only a single listen to each of the A, B and X samples. Once the sample had been auditioned, the play button was disabled. In addition to being able to select whether sample X was A or B, an additional ‘Unsure’ button was made available on the interface. Clicking this had the same effect as answering incorrectly - that is, not being able to reliably determine sample X. The benefits of these modifications are considered three-fold. Firstly, it did not allow subjects to constantly refer back and forth between samples, in an attempt to discover very small differences which would allow them to answer correctly. These small differences occur as tonal differences rather than being directly related to decay (not withstanding that the decay change can affect the sample tonally). Secondly, the speed of the test was increased, thereby reducing both fatigue and frustration. Thirdly, the unsure button acts as a ‘get out’ clause which eliminated the feeling that the subject was answering incorrectly. Subjects were encouraged that it was acceptable to press this button if they could hear no clear differences.

A total of ten subjects took this test, at a later date than previous two decay tests. On this occasion they were specifically instructed that the test was searching for a decay threshold, and it was changes in this parameter which should be listened for. Whilst fewer subjects participated, each was specifically invited and had participated in a number of similar tests, assessing the same program material in a variety of situations, over both loudspeakers and headphones.

6.7.2 Results

Thresholds follow a similar trend to those obtained for the artificial stimuli. Figure 6.7 shows the mean and standard deviation for the two samples at each replay level. Again, at the lowest frequencies, the thresholds increase. It would appear that the results are consistent with those of the absolute thresholds in that for frequencies above 100Hz, the results appear to 'level out'. The frequency dependence appears to be similar for both musical samples at both replay levels.

It is interesting to note the thresholds at 250Hz for the Holly Cole (HC) sample. These are particularly low - around 0.1 seconds. The absolute threshold tests were only carried out up to 200Hz, and the thresholds appears to level out above 100Hz. Therefore, even with comparable results between artificial and natural stimuli, we would not expect such low levels for this music sample.

Subject Performance

As with the artificial stimuli analysis, each subject's performance can be evaluated through analysis of PEST runs which failed to converge, the difference in thresholds reported for the first and last repeated tests and also the total time taken. In this experiment, the repeated test was that of the LEN sample at 125Hz cut-off and 85dB playback. Table 6.5 shows these performance measures.

The mean time taken was 32 minutes. This time is a reduction per test when compared with the sine burst tests. There were 13 runs for the natural stimuli as opposed to 10 for the artificial, meaning the average run time was reduced from nearly three minutes to less than two and a half. The PEST failed to converge on only four occasions. Each time this was a result of the maximum number of trials having been reached. It is believed that these performance measures show a similar success in testing as the artificial stimuli, a result which increases confidence in the reported thresholds.

Subject	Max Trials	Reversals	Max Decay	Difference (s)	Time Taken (mins)
1	0	0	0	-0.20	32
2	0	0	0	-0.10	34
3	2	0	0	-0.05	25
4	0	0	0	-0.05	22
5	0	0	0	-0.35	36
6	0	0	0	-0.13	34
7	1	0	0	-0.23	19
8	1	0	0	-0.15	39
9	0	0	0	-0.08	35
10	0	0	0	-0.23	43

Table 6.5: Subject performance for first impression natural stimuli PEST

Frequency (Hz)	63	100	150	200
p value	0.00	0.30	0.03	0.10

Table 6.6: Significance of replay level across four tone frequencies

	Frequency (Hz)		
Sample	63	125	250
LEN p value	0.44	0.73	0.34
HC p value	0.53	0.84	0.80

Table 6.7: Significance of replay level for two music samples across three frequencies

6.8 Comparison of Artificial and Natural Stimuli

A glance at the threshold results for both artificial stimuli (Figure 6.2), and the first impression natural stimuli tests (Figure 6.7) reveal that they are strikingly similar. It should perhaps be expected that using music signals would increase the thresholds, and there is some evidence that this is the case, particularly when using the HC sample. However, even with this sample, particularly at the higher frequencies, it appears that subjects are well able to perceive audible differences as a result of longer modal decays, with thresholds lower than those of both the LEN sample and the sine bursts.

Analysis is now presented which assesses the significance of the results obtained.

6.8.1 Significance of Replay Level

ANOVA has been carried out in order to assess the significance of the replay level. Firstly, Table 6.6 shows the probabilities that the thresholds for the two replay levels are drawn from the same population for each of the test tone frequencies. This shows that it is only at 63Hz that the replay level can be considered significant at the 1% level. At the other three frequencies, it is not significant at that level. It is suggested that at 63Hz, there is a greater difficulty in perceiving the decays at the lowest frequencies, which results in the difference.

Secondly, when analysing the data for the two music stimuli across the frequencies tested, Table 6.7 shows that for both music samples, at all frequencies, there is no significance between the high and low replay levels. Overall, the replay level is therefore considered insignificant and the data for high and low levels are combined for all subsequent analysis.

Frequency (Hz)	63	100-150	200-250
p value	0.00	0.00	0.00

Table 6.8: Significance of stimuli across the three frequency groups

Stimuli	Tones	LEN	HC
p value	0.00	0.00	0.00

Table 6.9: Significance of frequency for each of the three stimuli

6.8.2 Significance of Stimuli

The three types of stimuli, tones and the LEN and HC music samples can also be analysed for significance. As the tested frequencies for the artificial and natural stimuli tests differ slightly, the comparisons made are not an exact like for like match. For example, the tones were replayed at 100Hz and 150Hz, whilst the music samples were tested at 125Hz. Therefore, a linear interpolation of each subject's tone results between 100Hz and 150Hz was taken in order to compare with the 125Hz music samples. Also, the results for music at 250Hz have been compared directly with the tone burst results at 200Hz.

ANOVA results are shown in Table 6.8 and show that the stimulus is a highly significant factor at the 1% level for each of the three frequency groups. However, the ANOVA reveals a significant effect where a single group which can be considered to be drawn from a different sample population. In order to explore the data further, the 'multcompare' method in MATLAB was used. This produces an interactive graph which allows individual group means to be compared to the others. The three multcompare outputs are shown in Figure 6.8, and reveal that in each case, it is the HC sample which differs, while the tones and LEN sample can be considered to be drawn from the same population.

At the two lower frequencies, the HC threshold is higher than the other two stimuli, while at the higher frequencies (200-250Hz), it is lower. Further discussion of these observances follows the analysis of significance of frequency.

6.8.3 Significance of Frequency

Finally, we can consider the effect of frequency. ANOVA is performed for each of the three stimuli and produces three p values. Again, the two replay levels have been combined. Table 6.9 presents the p values, and shows that for each stimulus, there is a significant difference in threshold across frequency at the 1% level.

These data can be explored in a little more depth, again using MATLAB's 'multcompare' method (Figure 6.9). From the three graphs, we can see that for both the

	Threshold (s)		
Frequency group	Highly trained listener	Average listener	All listeners
32Hz	0.59	0.85	1.18
63Hz	0.19	0.44	0.73
100 - 150Hz	0.11	0.24	0.40
200 - 250Hz	0.05	0.15	0.26

Table 6.10: Final decay time thresholds

tones and LEN sample, it is the 63Hz threshold which results in a significance being reported. Frequencies above 100Hz all share a similar threshold. With the HC sample, differences are evident at each of the three frequencies.

6.8.4 Final Thresholds

Considering the significance analysis above, it now possible to produce final thresholds for modal decay. It has been shown that it is appropriate to combine the threshold data for each replay level. It is also shown that, when considering all stimuli, that there is an effect of frequency. Finally, there is the question of stimuli. Whilst it has been shown that the HC sample differs in threshold from the LEN and tones tests, it is argued that it remains valid to combine all stimuli in order to produce usable final thresholds. These can be used by room designers and those seeking to optimise the low frequency decay. For the combined frequency dependent thresholds, the 90th and 10th percentiles have also been calculated. These provide some movement on the thresholds, with the 10th percentile revealing the decays which may be heard by a small number of highly trained listeners, whilst the 90th percentile gives the thresholds above which the majority of listeners would clearly perceive the mode. Table 6.10 presents these final threshold values and they can be seen on Figure 6.10.

6.9 Discussion

With three listening tests conducted in this investigation into the thresholds of modal decay, the implications of the findings are now discussed. Firstly, we may consider the three listening tests in two groups - direct threshold audibility and perceived quality thresholds. The PEST/ABX tests with both sine burst and musical stimuli fall into the first category, with the quality based rating test of the same musical stimuli the second category.

As each tone burst represents a single frequency, these can be seen to reflect a true threshold at that frequency. However, the similarity of these ‘absolute thresholds’ with those obtained using musical stimuli and modelling decays in realistic rooms is particularly interesting. The result leads to the conclusion that the same decay thresholds are applicable regardless of the room or program material.

The analysis reports 10th and 90th percentiles which are useful in providing a ‘weighting’ to the results. Those wishing to eliminate all audible decays should follow the 10th percentile - although it must be observed that this would involve a great amount of effort at frequencies below 100Hz. A safer threshold to target would be the median, whilst decay times in excess of the 90th percentile should almost certainly be avoided.

The success of decay threshold testing gives additional support to the arguments presented in Chapters 4 and 5, that low frequency decay plays a particularly important role in our perception of low frequency reproduction quality.

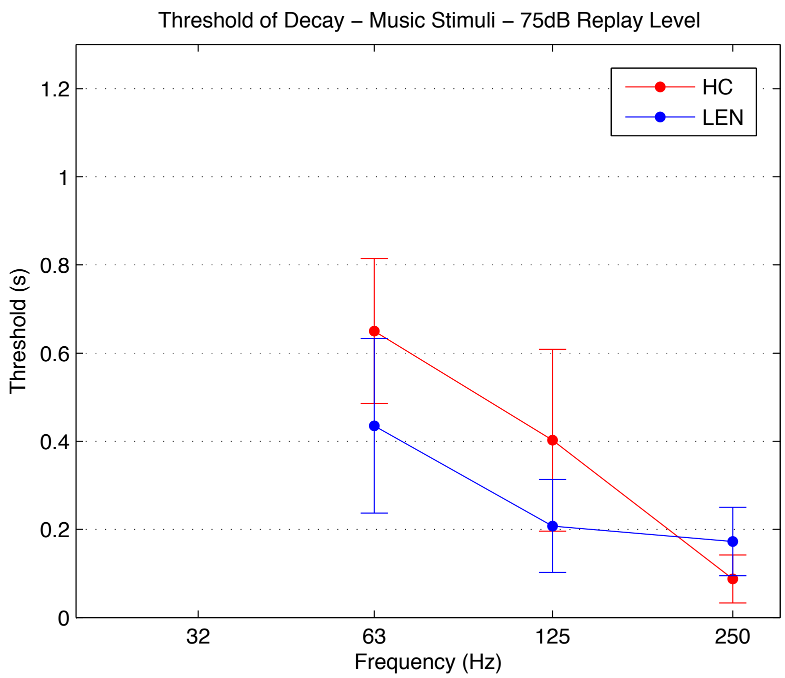
This therefore brings us to the question of perceived audio quality. This chapter introduced a study designed to test this quality in relation to the decay time. Whilst the results gathered did not reveal any clear knee points which would indicate a decay ‘quality threshold’, it is clear that there remains a need for further investigation of the perceived quality of low frequency sound fields. Indeed, it must be remembered that the thresholds of decay observed show the decay time below which we no longer perceive the mode, but do not reveal that this is the decay time producing the highest quality of audio reproduction.

6.10 Summary

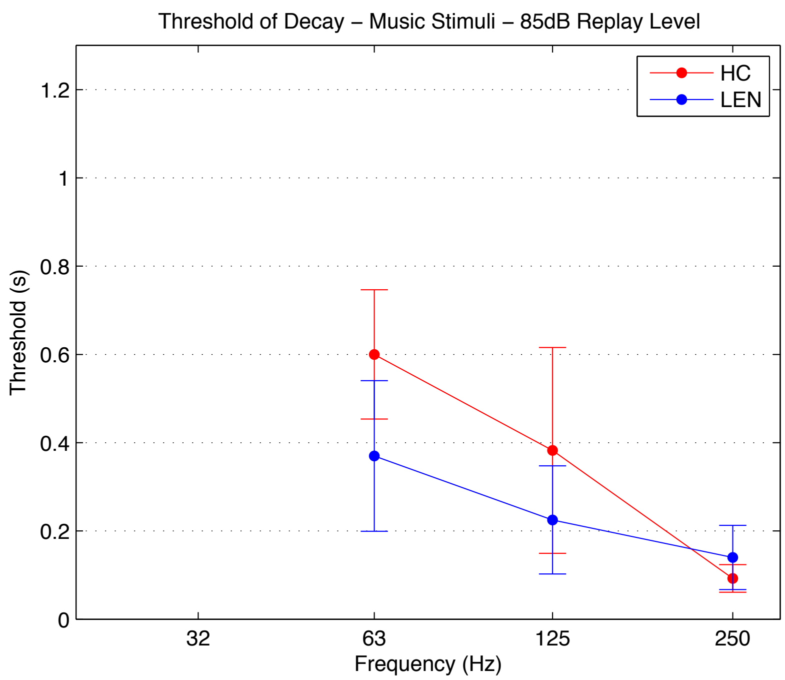
Through a number of listening tests, this chapter has shown that the effects of individual resonances are audible when their decay times exceed defined thresholds. It has been shown that thresholds can be determined for both artificial and natural stimuli, and are independent of replay level. Some dependence on stimuli is observed, although thresholds remain similar, and set of ‘final threshold’ across the low frequency range have been presented which may be considered applicable for any critical listening space. Reduction below the reported thresholds is deemed unnecessary, therefore preventing efforts to reduce decays further than necessary, and incurring a greater cost.

A test was conducted to search for a ‘quality threshold’, using a direct scaling method. This test was considered unsuccessful, and reveals the importance of using a test methodology which presents the listener with a simple task. This confirms the findings of the previous chapter, and this is addressed in Part III.

This concludes Part II of the thesis, which has investigated optimal values and thresholds of modal parameters. The importance of assessing the reproduction quality in terms of preferred room responses, and with appropriate test methodology has been highlighted, and it is to this field of quality assessment that we now turn.

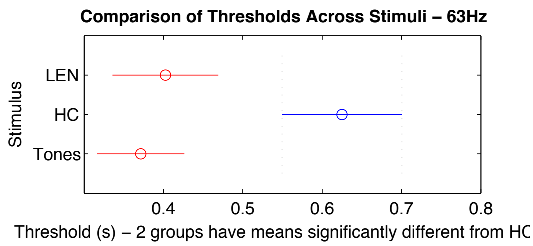


(a)

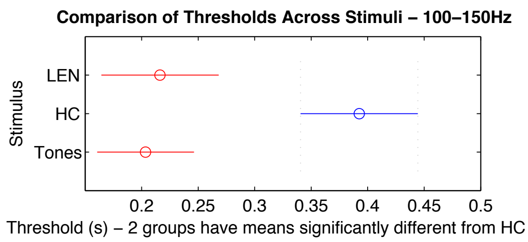


(b)

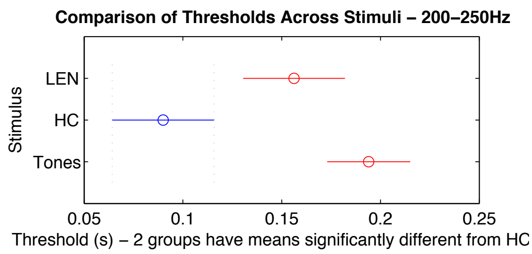
Figure 6.7: Decay time thresholds for the two music samples, a) 75dB and b) 85dB replay levels



(a) 63Hz

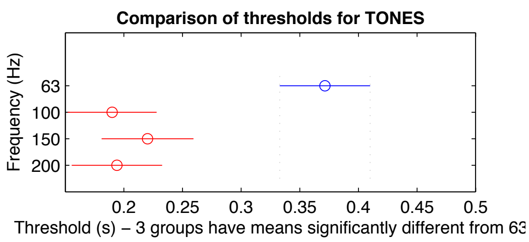


(b) 100-150Hz

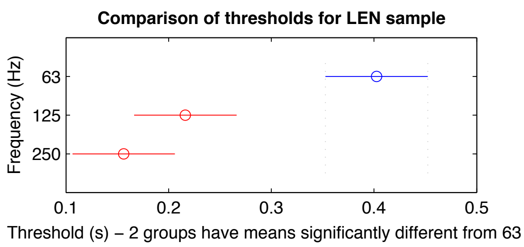


(c) 200-250Hz

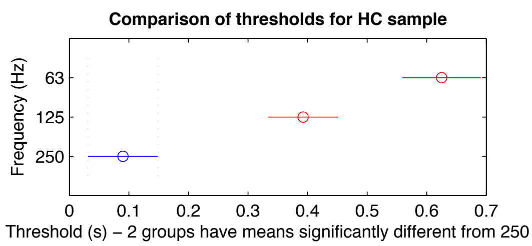
Figure 6.8: Multiple comparisons of thresholds for each stimuli across three frequency groups



(a) Tones



(b) LEN



(c) HC

Figure 6.9: Multiple comparisons of thresholds at each frequency for the three stimuli

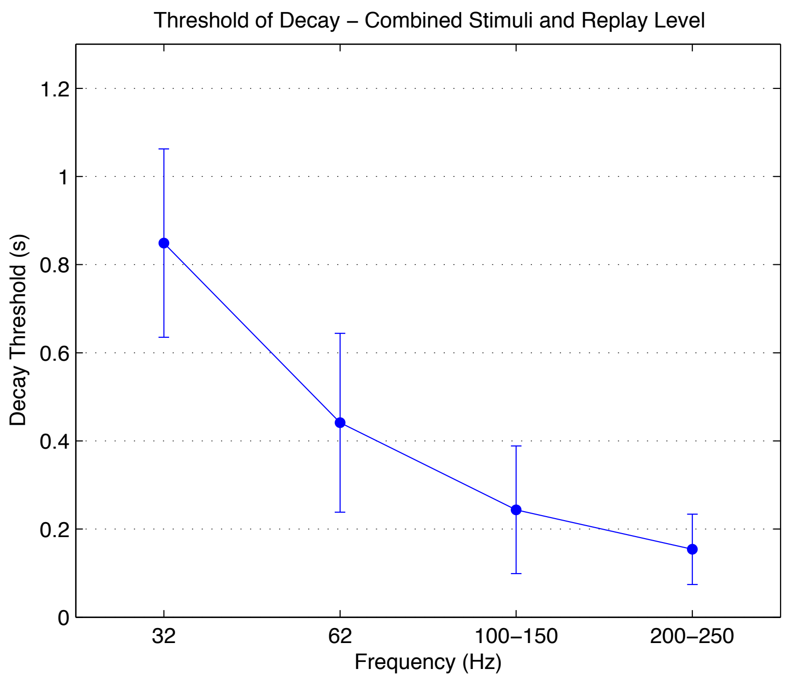


Figure 6.10: Decay thresholds grouped by frequency range across all stimuli and replay levels

Part III

Perceived Quality At Low Frequency

Chapter 7

Exploring Multidimensional Sensation

7.1 Introduction

A number of individual modal parameters have been studied, in search of optimal values and thresholds. The results have highlighted that such optimal values and thresholds are generally only applicable in simplistic scenarios or where assumptions have been made which are invalid within a real room. The importance of a deeper understanding of the perception of individual responses and the associated quality is clear.

There is evidence that low frequency audio quality testing reveals high standard deviations due to subject confusion and evidence of multidimensionality of auditory sensations in the test stimuli (see Chapters 5 and 6 and Wankling and Fazenda (2009)). This chapter therefore applies a technique known as Descriptive Analysis to elicit a set of terms or descriptors which are easier to map to auditory sensation. This has been done in order to discover how each of these sensations impact upon room quality.

The chapter reports on the methodology used to arrive at these terms and concludes with a set of four scales each with an associated description which can be used in the assessment of low frequency audio reproduction.

This is the first time such a term set has been attempted with specific focus on low frequency perception.

7.2 Sensory Relationships

As has been highlighted throughout this thesis, researchers have often attempted to directly map room parameters to reproduction quality in a manner such as Figure 7.1. In truth, it is likely that a more complex relationship exists, more akin to that shown in Figure 7.2. Here, the room parameters continue to influence perception, but their effects are seen through their impact on the unique room response. This room response can be classified through a number of further parameters. The response produces differing amounts of auditory sensations which in turn combine to produce an overall impression of room quality. The exploration of these relationships is the focus of the following chapters.

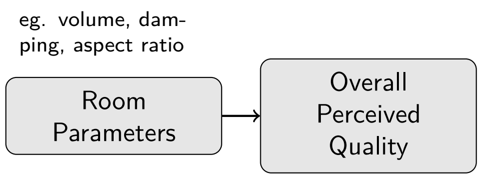


Figure 7.1: Direct relationship between room parameters and perceived quality

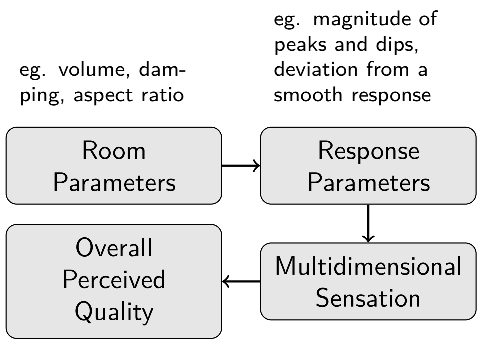


Figure 7.2: A more detailed representation of low frequency sensory relationships

7.2.1 Room and Response Parameters

A *room parameter* is defined as a quantity which is directly attributed to an individual room, such as its physical dimensions, construction materials or damping present. It is the combination of these parameters which produces a unique room response which can be considered in both the time and frequency domains. These parameters remain consistent regardless of aspects such as the listening position within the room. However, as previous chapters have shown, other parameters, or metrics, can be derived from the room response itself which may differ from position to position, or even day to day within the same room. These are therefore referred to as *response parameters*. It is possible to study these independently from the room as they are simply transfer functions of a system. Examples of such parameters may be the average deviation from an ideal frequency response (Cox et al., 2004), the modulation transfer function Fazenda et al. (2006a) or even measures such as the presence of significant peaks and dips.

7.2.2 Multidimensional Sensations

As with the majority of sensory evaluations, audio appears to be multidimensional (Grey, 1977; Bech, 1999). That is, more than one factor contributes to the overall perception. Evidence for multidimensionality, even when considering only the low frequency region, can be seen in previous work (Wankling and Fazenda, 2009). Observations revealed that subjects were confused as how to rate a room auralisation on a single hedonic scale where two distinct characteristics (tight, clear bass and lack of bass) were audible and yet one appears to have a negative impact on sound quality and the other a positive. Bech (1999) has observed the multidimensionality in audio preference and a large amount of work has been conducted into the evaluation of audio using multiple terms to rate stimuli against, rather than a single scale (Bech, 1999; Berg and Rumsey, 2006; Zacharov and Koivuniemi, 2001b; Hatziantoniou et al., 2005; Lorho, 2005; Mattila, 2001). Many of these studies focus on the sensations associated with spatial audio perception. In terms of timbre, perhaps more immediately pertinent to low frequency quality within the context of rooms than spatial attributes, Gabrielsson and Sjogren (1979) have also attempted to use verbal descriptors which can then be used to achieve a greater understanding as to why certain stimuli are preferred. A number of other researchers have also used similar sets of terms, which were determined using a variety of methods (see Section 7.4 for further detail).

7.2.3 A New Term Set

Although the studies referenced above present term sets for the evaluation of audio, they are considered inappropriate for the characterisation of low frequency reproduction for a number of reasons. Firstly, a number relate specifically to spatial perception and therefore include descriptors such as ‘envelopment’, a concept less relevant to low frequency quality. Secondly, many contain descriptors more obviously pertinent to a higher frequency range than that studied in this work, for example, ‘bright’ or ‘airy’. It is also expected that a smaller number of descriptors

(Zacharov and Koivuniemi (2001a) generated twelve for example) will be required for this more focused study. It is interesting to note that in the assessment of low frequency quality of loudspeakers, Harris et al. (2006) asked subjects to respond based upon three low frequency attributes. These were *clarity*, *bass extension and fullness* and *relative level between instruments*. However, no explanation is given as to the origin of these descriptors or the sensation mapping of them.

As a result, this chapter reports on an exercise known as 'elicitation of terms', where new descriptors are defined explicitly for the evaluation of low frequency reproduction quality. Generating new verbal descriptors for the multidimensional audio samples gives us the ability to determine individual sensations, and to what extent they are important in our perception of low frequency reproduction. Having multiple descriptors should also reduce the complexity of the listening task, by making subjective judgements more focused in terms of providing specific auditory sensations which can be more easily quantified. This affords us a greater opportunity to optimise rooms based upon the most important sensations.

In addition to the timbre, the vibro-tactile sensation or 'feel of the bass' has been considered to be an aspect of perceived low frequency quality (Simone et al., 2009). This is however, beyond the scope of this thesis. It is acknowledged that this may influence the perceptual space and therefore remains an interesting area for future research.

7.3 Audio Samples for Elicitation

Although the term set to be developed is envisaged to be of use in a wide range of low frequency assessment scenarios, it is important to retain a focus on room acoustics during the elicitation within this body of work. For this reason, the samples used in both the elicitation of descriptors and the subsequent listening test (Chapter 8) are generated by modifying a number of room acoustic parameters.

The same decomposition model was again used, enabling modification of the modal distribution, the damping and the source and receiver position within the room. The modification of these parameters is seen as an effective way to obtain a wide variety of responses essential for eliciting subjective terms valid in all situations. The terms elicited should cover all likely listening scenarios, not only those of the samples used.

Furthermore, for this elicitation, four musical excerpts were used in order that the descriptors elicited were not related to the characteristics of an individual musical stimuli. Once again, the LEN and HC samples were used. The additional samples were chosen to compliment these and are the FW and FG samples introduced in Chapter 3. The replay level was set at a comfortable listening level, approximately 85dB (calibrated as per Section 6.4.4) at the ear and remained constant across subjects. A total of 23 auralised samples were produced, using combinations of modelled responses and the four music samples.

In order to ensure a complete set of descriptive terms and full magnitude of sensation, the samples were generated to include responses unlikely to be found in real rooms. These cases were produced according to prior experience and findings. For example, a smooth frequency response is known to produce a certain positive sensation and can be simulated by placing the source and receiver within a few centimetres of each other. This would, of course, be unrealistic in a true listening scenario, but is useful to ensure descriptive scales of full magnitude are developed.

A second set of samples, used in rating tests in the following chapter were deliberately kept within the constraints of the parameters usually found in rooms so as to ground those tests in realistic scenarios. This ensures that any relationship found between the modal parameters and descriptors is both meaningful, but also useful in producing guidelines for real rooms.

7.4 Eliciting The Descriptors

There is much literature detailing various methods for determining new sets of verbal descriptors. The techniques find their origin in the sensory evaluation of food products, and have been shown to be successful in the audio field. Bech and Zacharov (2006) summarize these methods succinctly - highlighting that where a direct elicitation is required, the methods fall into two main categories:

1. *consensus vocabulary* – subjects are involved in a panel producing a set terminology (i.e. Descriptive Analysis)

2. *individual vocabulary* – subjects rate samples using their own vocabulary (Repertory Grid Technique (Berg and Rumsey, 2006), Free Choice Profiling (see (Lorho, 2005) for example))

This study uses a modified version of the Descriptive Analysis method (see Bech (1999) for a thorough review of this technique), which has fewer sessions in an attempt to streamline the process. The method consisted of the following three stages:

1. Individual elicitation

2. Group discussion 1 - reduction of terms

3. Group discussion 2 - end points and magnitude agreement

7.4.1 Individual elicitation

During this first stage, subjects were presented with the set of 24 auralised music samples. The subjects were instructed to enter as many descriptive terms as required for each sample. This was done by means of keyboard entry on a graphical user interface in an IEC standard listening room at Salford University. Subjects were encouraged to enter any descriptive term which related to the sample, regardless of the magnitude of sensation (e.g. ‘crisp’, ‘boomy’). They were also instructed to avoid attitudinal terms such as ‘nice’, ‘pleasant’, ‘nasty’ etc.. Subjects were not limited on the number of terms which could be entered for each sample and were free to move back and forth between samples if they desired.

Twelve subjects, each of whom may be considered as an experienced listener with a keen interest in the experimental work, participated in this stage, producing a total of 248 terms. Of these, 74 were unique. Terms which occurred multiple times were either those used by multiple subjects or by a single subject for multiple samples. The most common terms are listed in Table 7.1. Some subjects also quantified terms, for example ‘*very* boomy’. In order to analyse the data, such quantifiers were manually removed, leaving only single word terms.

It is immediately clear that terms commonly associated with small rooms, ‘boomy’, ‘muddy’ and ‘resonant’ occur with a high frequency. It is suggested that these would usually be considered to be negative terms. Additionally, a number of more positive terms, such as ‘tight’, ‘punchy’, ‘dry’ and ‘clean’ also appear regularly. As would be expected, it is evident that a number of differing terms were used to describe the same sensation. For example, a selection of terms used in describing samples 2 and 12 are shown in Table 7.2. These terms are likely to represent similar sensations audible in both samples.

An initial grouping of the terms was performed using a cluster analysis. This process looked at each term, and how it was used across the 24 music samples. The cluster analysis was run using the city block distance method (Hill and Lewicki, 2006) across the 23 most common terms. The output is viewed in the form of a dendrogram (Figure 7.3). The numbered groupings were added subsequently. There appear to be two large groupings of generally positive and negative terms, and at a lower level, six clusters of terms have been identified. Berg and Rumsey (2006) discuss the analysis of such a dendrogram to arrive at a final term set. However, as

Term	Frequency (Hz)
boomy	19
reverberant	15
muddy	14
resonant	14
muffled	13
rattles	10
tight	9
fuzzy	8
punchy	8
wooly	8
distant	8
dry	7
clean	7
strong	7
bassy	6
clear	6
flat	6
undefined	5
precise	4
weak	4
unprecise	4

Table 7.1: Terms used more than three times during the individual elicitation

a further stage of group discussion is employed in this current work, the dendrogram was seen as an initial analysis, useful in guiding the following group discussions.

7.4.2 First Group Discussion

This session aimed to reduce the term set to a smaller number of unique descriptors. All terms were listed on a flip-chart, grouped according to the cluster analysis. Discussion began around those terms which appeared to represent the same sensation. These were collated and a decision made as to which was the most appropriate.

Eleven of the original twelve subjects participated in this discussion, which lasted for two hours. The author of this thesis acted as a panel chair and is not included in these eleven. The discussion took place in a boardroom style meeting room with equipment set up allowing each panel member to listen over headphones to samples being played by the panel chair.

Descriptors were grouped, and after much discussion and debate, the following outcomes were agreed:

There is a:

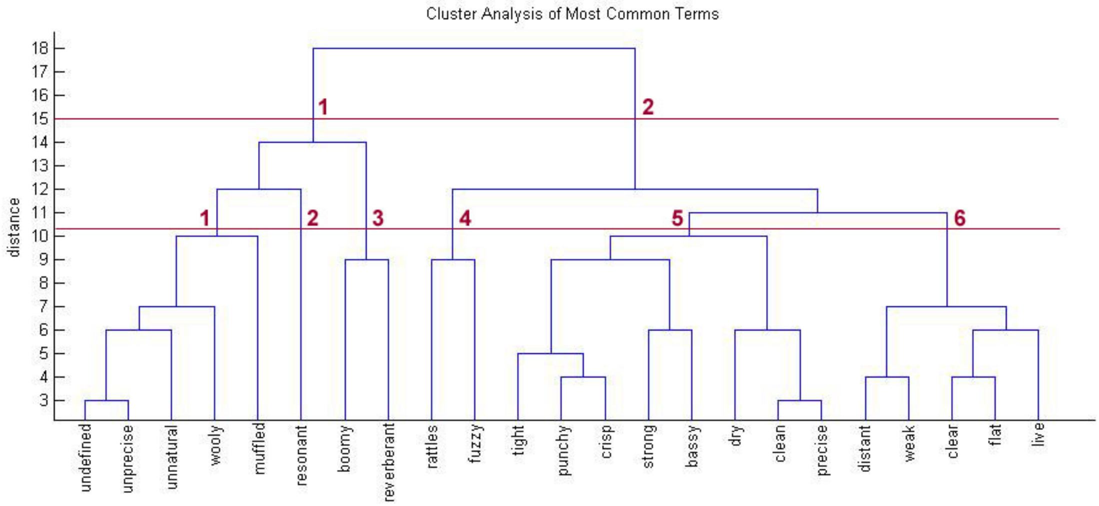


Figure 7.3: A dendrogram showing the cluster analysis from the individual elicitation of terms

Sample 2 Terms	Sample 12 Terms
dry	tight
tight	dry
coherent	crisp
dead	clean
dry	dry
punchy	tight
strong	precise
crisp	flat
tight	clean

Table 7.2: Example of similar terminology used for multiple samples

- single scale ranging between 'muddy' and 'tight'
- unique descriptor 'resonance'
- scale of 'strength' - related specifically to amplitude, or overall bass level
- scale of 'depth' - related to the range of frequency content

At this point the meeting was terminated. It is interesting to note the absence of certain terms, specifically 'reverberant'. This was one of the most frequently used terms in the individual elicitation, but rarely mentioned during the group session. When this was highlighted by the chair, there was almost unanimous agreement that this term should be reserved for describing a higher frequency range, and that the effects which may have been perceived during individual elicitation which prompted the 'reverberant' descriptor, could in fact be adequately described by the scales/terms above. There was also significant discussion regarding the term 'boomy', even though it is absent from the most frequently used terms. It was eventually agreed that this represented a higher level sensation - i.e. a boomy room occurs as a result of the combination of other unique descriptors.

7.4.3 Second Group Discussion

All but one subject from the first discussion were present for a second discussion, where the aim was to finalise the terms. A short time was spent reviewing the previous session, to ensure that the outcomes listed above continued to represent the panel's views. Once this was confirmed, the panel discussed each term, along with *end point descriptors*. For example, the hypothetical term 'bright' may have end points 'very' and 'not at all'. Conversely, it may be decided that the term is itself an end point that is the opposite of another term, i.e. bright/dull.

At this point, the group also decided upon the relative magnitudes of the 24 elicitation samples on each scale. For this process, each subject was given a number of 'voting' cards, numbered 1 to 10. Headphone distribution to all participants was again available, and each sample was replayed a number of times, after which each subject held up a card showing the magnitude of each sensation perceived. Where agreement was not clear, further discussion was held. In a number of cases this process provided further clarification on the meaning of each term. This process was vital to ensure that there was no significant misunderstanding of the terms and the magnitudes exhibited by each sample.

Throughout the process, a short description of each term was refined. These can be given to new listeners who may sit listening tests based upon the descriptors but who were not part of the original panel. Indeed, the listening test in Chapter 8 includes such subjects, and comparison between their results and those of the listening panel reveals the ability of the descriptors to be used in a wider context in future research. Table 7.3 shows the final agreed descriptors.

Descriptor	Scale Values	Definition
Articulation	Muddy	Each sound (or note) has a lack of definition, and could sometimes be described as 'smeared'.
	Tight	Each sound (or note) is distinct, well defined and precise.
Resonance	None High	A resonant sample has some notes which sound louder, ring and last longer.
Strength	Weak Strong	Relates to the loudness of the low frequency when compared to the rest of the frequency range in the sample.
Depth	Shallow	Lacks notes that extend down lower in frequency.
	Deep	Has notes that extend down lower in frequency.

Table 7.3: Final descriptors, end points and definitions

7.5 Validity of the Term Set

With these terms elicited, it is important not only to use them to explore low frequency perception in greater detail, but also to validate the terms themselves. This validation is best achieved through real world testing. This allows us to answer a number of questions: are subjects using the scales in a consistent manner? Do the terms adequately represent independent sensations? (if they do, we should observe little correlation between terms). Do the subjects properly understand and perceive the sensations each term is designed to highlight? Finally, are there differences in the use of the terms between subjects who were part of the original elicitation panel and naive listeners simply presented with the terms and descriptions?

As a new listening test is necessary to answer these questions, it is clear that if properly designed, this test will also be useful in revealing new information about a particular set of room responses. The following chapter therefore reports on a listening test which seeks to validate the term set in addition to investigating three room parameters using the terms.

7.6 Summary

A process known as descriptive analysis has been undertaken in order to determine a number of underlying sensations perceived when listening specifically to low frequency program material. As a result of a number of sessions, four unique terms, or descriptors have been produced, each with resulting end points and brief description. This is the first time such a term set specifically for the characterisation of low frequency reproduction quality has been attempted. It is envisaged that the terms may be used by researchers not only in the field of room acoustic testing, but also in the wider context of low frequency reproduction quality. In order to validate the terms, they must be used in a real listening test scenario, which is the subject of the following chapter.

Chapter 8

Perceived Quality and Room Response

8.1 Introduction

Throughout this thesis, it has been shown that a greater understanding of *perceived quality* in relation to the room response is now needed. Whilst individual parameter thresholds and optimal values have been shown to go some way towards highlighting ways in which we can improve a room's reproduction quality, such individual parameters have been revealed to be limited in their application due to the unpredictability of the response in a real world scenario. Where successful thresholds have been found, such as the threshold of decay, we still lack an understanding of the extent to which a specific decay time affects perceived quality. A simple quality rating of real room responses goes some way towards determining a subjective quality, but it has been observed that there may be confusion between subjects as to which 'elements' of the low frequency reproduction should be considered good or bad (Wankling and Fazenda, 2009). That there is a multidimensional aspect to the perception of low frequency reproduction has been confirmed in the previous chapter through the elicitation of four distinct terms.

With the perceived quality likely to be related to these attributes, it is now important to perform a listening test to both validate the terms produced, and also to utilise them in allowing us to further explore the effect of both room and response parameters. In this chapter, three room parameters are investigated, each of which is considered influential in our overall perception.

Section 8.2 presents the initial room parameters under test, while Sections 8.3 and 8.5 highlight testing and results using a similar methodology and term set within a real room.

The use of the descriptive terms developed are evaluated in Section 8.6, with results also compared to a subsequent test using the descriptors to evaluate a number of optimisation systems in a listening room. Finally, the relationships between the room parameters under test, overall perceived quality and the ratings for each descriptor are explored.

8.2 Room Parameters

In order to relate both the descriptors and room parameters with a perceived quality, two specific tests were employed:

1. Rating of each sample for its perceived quality (quality rating).
2. Rating of each sample for each descriptive term produced (direct attribute rating).

This section discusses the three room parameters, each of which were studied at four levels, creating a total of twelve test samples. It was assumed that changes in these parameters would result in similar perceived qualities and sensory differences regardless of listener position within the room, although it must be acknowledged that this may not always be the case.

8.2.1 Modal Decay

Chapter 6 showed that we can obtain thresholds of decay, but no direct quality rating due to a specific decay time was seen. Decay is therefore a natural choice of room parameter for both the quality and direct attribute rating tests. Although typically, absorption becomes more effective with increased frequency, resulting in a reduced modal decay time, in order to test this parameter directly, the decay time was kept constant across the frequency range of the decomposition model. A 100m^3 room with typical dimensions, source position and receiver position was modelled with four decay times of 0.1, 0.3, 0.5 and 0.7 seconds. As the decay is reduced, the response is typically smoothed as the bandwidth of each mode increases. This smoothing has been associated with an increase in quality (Cox et al., 2004). In the light of the decay time thresholds, these times should represent clear perceptual differences across the entire frequency range under test. A perceptual increase is expected, and it is highly likely that each descriptor will be influenced too.

8.2.2 Room Volume

The second parameter under study was the room volume. An increased volume results in an increased modal density at any given frequency. Perceived differences are expected as there is a striking difference when viewing the frequency responses of rooms with low densities and those with high. One of the key motivations for the elicitation of terms was the difficulty in obtaining frequency response difference for a single scaled rating test (Wankling and Fazenda, 2009). By using the descriptors revealing the multidimensional sensations, it is expected that this difficulty can be overcome. Volumes were modelled with a fixed room aspect ratio, relative source/receiver position and decay time (where decay was now more realistically modelled as frequency dependent). Volumes under test were 20m^3, 100m^3, 250m^3 and 1000m^3. These cover the range from a very small box room, where the lowest modes are widely spaced, to a large studio space where a high modal density exists even at the lowest audible frequencies.

It is acknowledged that this room parameter, volume has previously been used in order to study the modal density, and that from that study, it was difficult to obtain reliable thresholds. However, it was shown that there is always a perceived difference when the volume changes, and the interaction of individual modes produces a unique room response. It is these differences which are sought to be better understood here - particularly if there is a *increase in perceived quality* associated with a particular volume. Section 5.6 asked an initial question on quality based upon changing volume, with some evidence that a greater volume may result in a higher perceived quality. This test seeks to understand how this volume affects the underlying sensations which make up our perception of quality.

8.2.3 Source Position

The positioning of loudspeakers within the room has also been researched previously (see (Pedersen, 2003; Groh, 1974) for example). The room response may be significantly altered through differing positions. For example, a smoothing of the frequency response has been observed when the source and receiver are moved closer together (Wankling and Fazenda, 2009). For this test, the receiver was fixed, in a typical, slightly off centre listening position, whilst the source was moved along the diagonal, from a front tri-corner towards the listener in four equal steps in each of the x, y and z dimensions (see Table 8.1 for exact positions). Even though all of the source positions are not typical of real usage, their responses are realistic and often found in practical situations.

		x (m)	y (m)	z (m)
Receiver		3.40	2.90	1.20
Source (m)	1	0.01	0.01	0.01
	2	1.16	0.89	0.45
	3	2.32	1.77	0.90
	4	3.48	2.66	1.35

Table 8.1: Four values for the testing of source position parameter

Testing for both perceived quality and direct attributes was carried out using samples auralised using this set of twelve virtual rooms. In order to keep the number of trials down and reduce fatigue, the room models were used for auralisation of a single musical excerpt. The 'LEN' sample was considered to be particularly appropriate with its ability to reveal degrading artefacts of low frequency reproduction.

8.3 Methodology

8.3.1 Sample Familiarisation

Before commencing the two tests, subjects were presented with an interface to audition the twelve music samples. They were instructed to familiarise themselves with the samples and listen specifically for differences between them in the low frequency region. After doing so, they could proceed to the quality rating test. In total, 16 subjects were tested, 11 from the elicitation panel (including the panel chair) and five naive subjects.

8.3.2 Perceived Quality Rating

In order to obtain quality ratings, a simple paired comparison method was used. This is an example of indirect scaling where each sample is compared against each of the others, one at a time. By using this method, the complexity of the subject's task is greatly reduced in comparison to direct scaling. The listener need only make a simple judgement as to which sample, A or B, is has a higher reproduction quality. No judgement as to the magnitude of quality difference was required. By using statistical methods, it is possible to place the twelve samples on a ratio level scale (Bech and Zacharov, 2006).

A total of 66 comparisons were made by each subject ($\frac{N^2-N}{2}$, where N is the number of samples). Subjects were instructed to make their judgements for quality according to which sample they believed provided "the best low frequency reproduction for critical listening", the definition for which was given as:

> "Critical listening is the process whereby you listen to the audio program in a way that allows you to evaluate and interpret its characteristics in depth and make decisions regarding any problematic features such as resonances, frequency or level imbalances, lack of definition, etc. An example of a critical listening environment would be in a control room of a recording or mastering studio."

Samples were presented at a loudness level equivalent to that of a 125Hz sine tone played at 85dB SPL at the ears, using the calibration method for music stimuli outlined in Chapter 3, which allows an equivalent loudness level to be determined for a complex audio program (ITU, 2006). Samples were replayed over the same Sennheiser HD650 headphones and testing took place within a semi anechoic room at Salford University. The test administrator enforced a number of breaks, and subjects were instructed to take further breaks at any point if desired.

8.4 Direct Attribute Rating

The direct attribute rating test required subjects to score each of the twelve samples against the four descriptors (attributes) elicited in the previous chapter. They were given a printed sheet explaining the test which included Table 7.3. Each sample could be played as many times as required, and a score was given for each of the four attributes, before continuing to the next sample.

Each scale was from 0 to 100, in increments of 1. These values were reset for each new sample being evaluated. Starting values were the mid point (50) for articulation, strength and depth, and zero for the resonance scale. It was decided that the former three descriptors should begin at the mid point as they could be scaled either positively or negatively. Resonance however, had been created as a scale the with end points 'none' and 'highly' and therefore the default position is most logically set to 'none'.

Presentation orders for all tests were randomised.

8.5 Results

8.5.1 Perceived Quality Rating

Ratings for quality were derived from the paired comparison matrices produced by each subject. Several techniques exist whereby it is not only possible to order the samples based on the comparison matrix, but also to derive an interval scale.

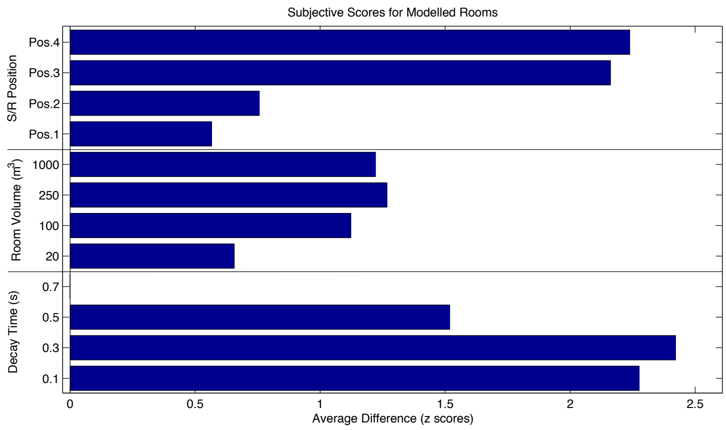


Figure 8.1: Perceived quality scores derived from paired comparison test including all twelve samples

Here, this was done according to the Case V assumption of Thurstone's 'Law of Comparative Judgement' (Thurstone, 1927). As all twelve samples were compared against each other, a full quality scale can be obtained independent of the three room parameters. Figure 8.1 shows the perceived quality results. It should be noted that as a perception scale, there are no specific units on the x axis. The scores relate to the average z score for each sample when compared to each of the others. That is, how consistently each room has been rated better than each of the others.

It is also possible to determine a confidence score from each individual comparison matrix. This is known as the 'coefficient of consistency' and is a normalised ratio of the number of circular triads in the data set compared to the total number possible. A circular triad is evidence of intransitivity. For example:

$$\mathrm{A} > \mathrm{B} > \mathrm{C} > \mathrm{A}$$

is regarded as a circular triad. A more consistent result would be:

$$\mathrm{A} > \mathrm{B} > \mathrm{C} < \mathrm{A}$$

It should be noted that a low score is not necessarily indication of a subject's poor ability, but can often be a result of very small differences between samples (i.e. a complex task). However, four subjects showed a consistency of 0.5 or lower and at this level, the number of circular triads is likely to be the same as if a subject was answering at random. Therefore these subjects were removed before performing the scaling resulting in Figure 8.1. Of these four subjects, two were naive subjects, and two from the descriptive analysis panel.

8.5.2 Direct Attribute Rating

Listener Type

An important comparison can be made between those listeners who were involved in generating the four descriptors, henceforth known as ‘panel’ and those who were regarded as naive listeners, and were only given the information in Table 7.3 and the set of the test samples to audition before performing the test.

Twelve ANOVAs were run, corresponding to each of the three room parameters and each of the four terms. Table 8.2 shows the significance results from these tests (p values). With the exception of three (highlighted) cases, the listener type was not significant at the 5% level (p>0.05). The subsequent analysis therefore combines the data obtained from both panel and naive listeners (16 in total).

		Term (p value)			
		Articulation	Resonance	Strength	Depth
Decay (s)	Attribute Level	0.00	0.02	0.65	0.58
	Listener Type	0.25	0.75	0.47	0.31
Volume (m^3)	Attribute Level	0.00	0.00	0.00	0.00
	Listener Type	0.47	0.76	0.03	0.04
Source/Receiver	Attribute Level	0.37	0.16	0.02	0.02
Position	Listener Type	0.13	0.00	0.06	0.14

Table 8.2: Significance (p values) from ANOVA for the Listener Type and Attribute Level for each room and descriptor.

Room Parameters

Further ANOVAs were run to determine the significance of changes in each room parameter upon the scores given for each descriptor. If the null hypothesis, that scores from across the four levels of a room parameter are from the same population, cannot be rejected, there is evidence that the parameter has no direct influence upon the perception of that sensation. Conversely, a rejection of the null hypothesis suggests that the room parameter *does* contribute to the perceived sensation.

The ANOVA results are shown in Table 8.2 under the ‘Attribute Level’ rows. The individual means are also plotted in Figures 8.2 to 8.4, along with 95% confidence intervals. When considering changes in room volume, all four of the descriptors are affected by the changes, that is, the ANOVA shows a significance at the 5% level (p<0.05). It is however, interesting to note that progressive increase of room volume does not lead to a progressive change in the mean score of any attribute. This suggests that changes to room volume might not be linearly related to perceived quality. This is in agreement with earlier room volume experiments. It shows that the perceived quality, as well as direct sensation, is much more related to the *effects* of the specific volume on the response, rather than the volume directly. When isolating variations in decay time, only the descriptors articulation and resonance are significant (Table 8.2). This result suggests that the perception of both strength and depth are independent of decay time and that is clear also in Figure 8.3. In contrast, the ratings for attributes 'articulation' and 'resonance' follow a defined trend with the former decreasing and the latter increasing for increasingly long decay times. This suggests that decay time has a defined and strong effect on perceived quality. It is argued here that the 'distortion' mentioned in previous chapters is highly related to the sensation described by the panel as articulation. The results observed for source position reveal that 'strength' and 'depth' are significant terms whereas 'articulation' and 'resonance' are not. This might be explained by the fact that changing the source position merely affects the way certain room modes are excited, by strong coupling or otherwise, leading to more or less low frequency content and variations in its amplitude. Other aspects of the response related to the time response, such as decay, are not strongly affected, thus not provoking a change in the perceived attributes of articulation and resonance.

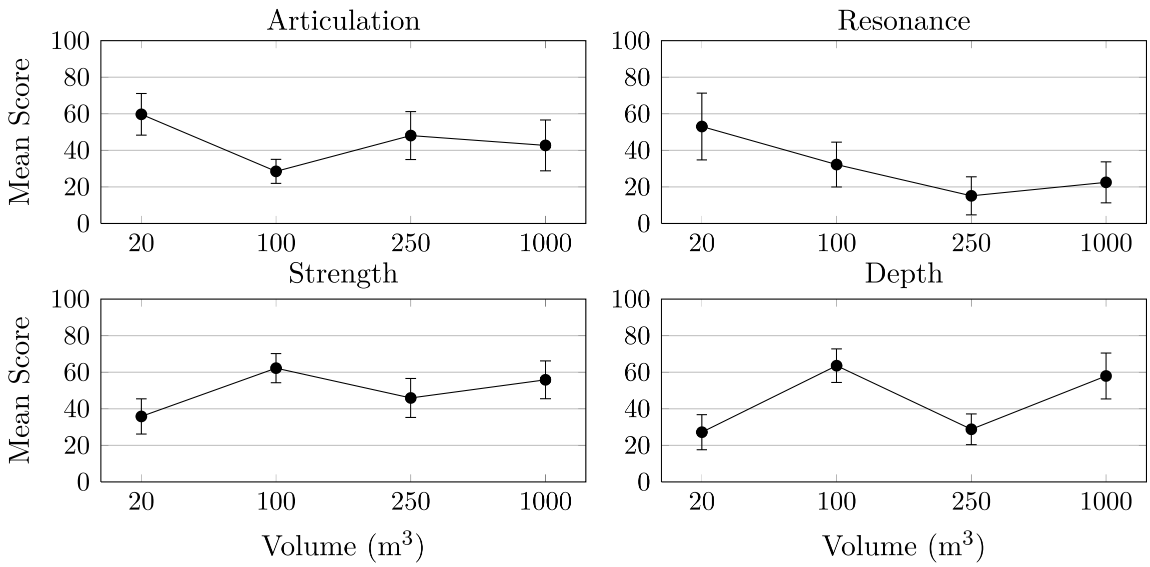


Figure 8.2: Change in volume - mean scores and 95% confidence intervals

Reliability

As stated, one aim of this work was to increase the confidence in subjective testing methods. The standard deviation across subjects is used here as a measure of the success of this testing method. Over the 48 ratings that were performed by each

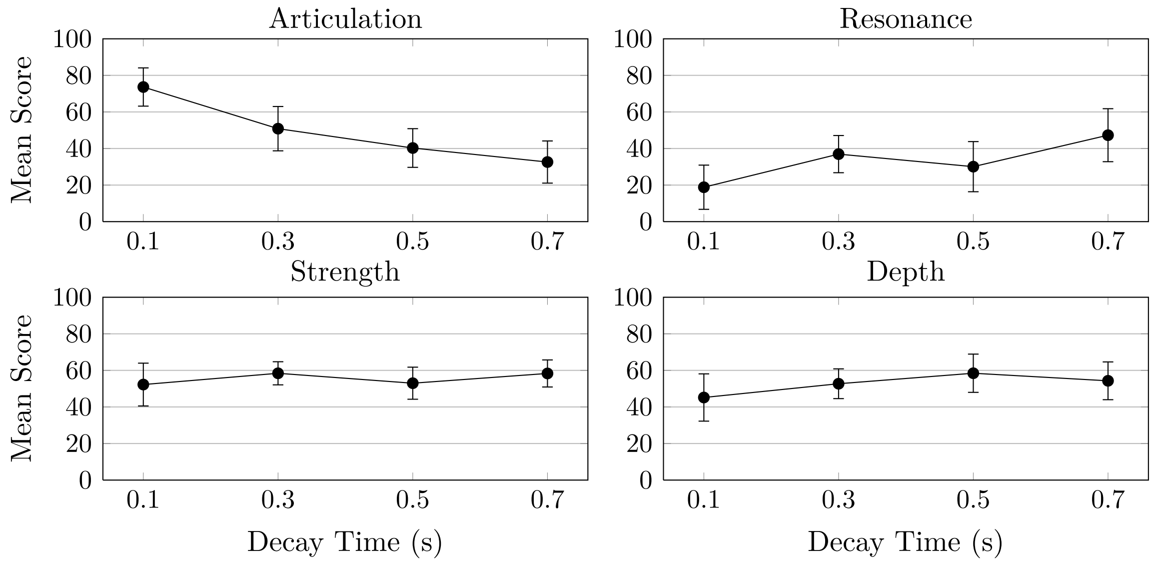


Figure 8.3: Change in decay time - mean scores and 95% confidence intervals

subject (12 samples, four descriptors), the mean standard deviation was 21.7.

Table 8.3 shows the mean standard deviations for each of the four descriptors.

	Articulation	Resonance	Strength	Depth
Std	21.1	27.3	17.7	20.6

Table 8.3: Mean standard deviations for direct ratings of the four descriptors

The implications of these reliability results are discussed further in Section 8.6.2.

Correlation

Correlation analysis has been conducted for both the direct attribute ratings and overall perceived quality. Table 8.4 shows both the correlation coefficients and the associated significance value for each of the four descriptors and perceived quality. The descriptive analysis technique aims to achieve orthogonality between terms. Therefore, ideally, if they are indeed independent of each other, there should be little correlation between pairs of attributes. Spearman's correlation coefficients were calculated, taking into account the full data set, i.e. every subject's responses whenever a comparison was made between terms, regardless of room parameter. It may be observed that the majority of the coefficients are small. However, notable results are a positive 0.59 correlation between the terms strength and depth, and a negative -0.32 correlation between articulation and resonance. Whilst these scores would not usually be considered particularly high, it would seem intuitive to explore

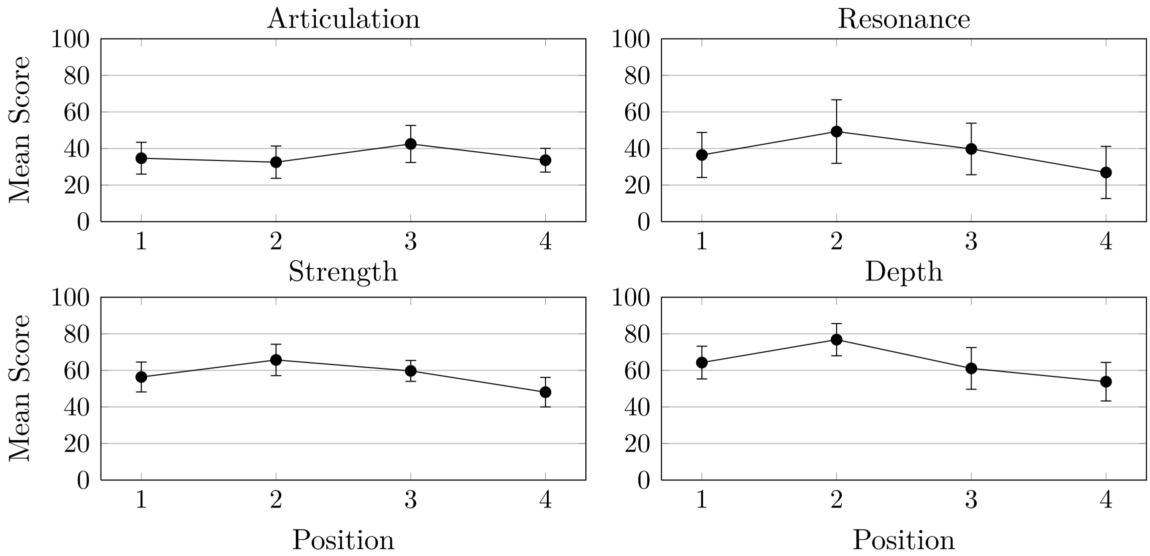


Figure 8.4: Change in source position - mean scores and 95% confidence intervals

		Term				
		Articulation	Resonance	Strength	Depth	Quality
Articulation	rho		-0.32	-0.10	-0.14	0.20
	p val		0.00	0.19	0.05	0.01
Resonance	rho	-0.32		0.12	0.10	-0.20
	p val	0.00		0.09	0.17	0.01
Strength	rho	0.10	0.12		0.59	-0.02
	p val	0.19	0.09		0.00	0.78
Depth	rho	-0.14	0.10	0.59		-0.03
	p val	0.05	0.17	0.00		0.64
Quality	rho	0.20	-0.20	-0.02	-0.03	
	p val	0.01	0.01	0.78	0.64	

Table 8.4: Correlation coefficients and associated significance values for each term and overall perceived quality

the possibility that there may be some linkage between terms, and will be discussed further in Section 8.6.1.

8.5.3 Principal Components

Anecdotal evidence from test subjects concurred with the correlation analysis, with a number commenting that making independent judgements for strength and depth was difficult. The data were therefore subjected to a principal component analysis (PCA) to determine whether the term set can be reduced, and to better visualize the data. Figures 8.5 and 8.6 show projections on the first three principal components, which account for 91% of the variance. The loadings from each term are shown, along with the mean room rating for each of the twelve samples (selected rooms are

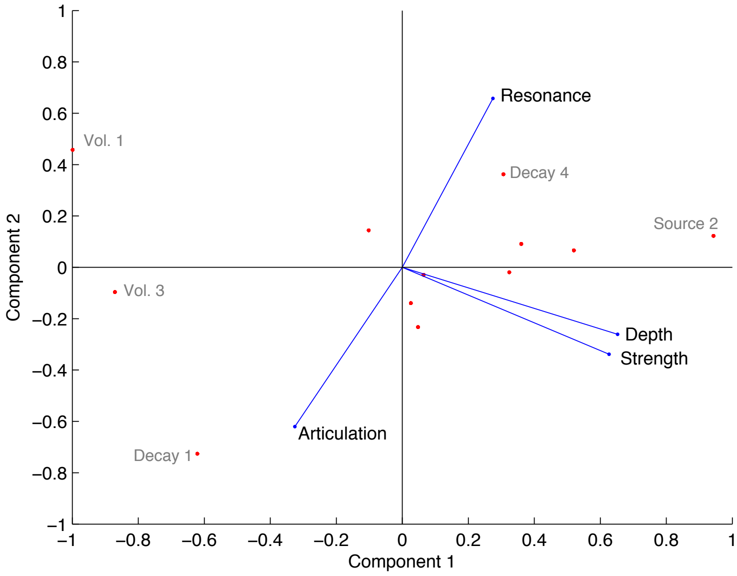


Figure 8.5: Principal component analysis showing the first two components, projections of each term and the mean subjective score for each room (five highlighted)

highlighted). The implications of these results are now discussed.

8.6 Validity of Low Frequency Descriptors

The results of this test are now used to asses the validity of the four low frequency descriptors presented in Chapter 7.

8.6.1 Independence of Scales

Any set of descriptive terms must only contain independent sensations. Where there is duplication, the result is inefficient testing. It is noted that, from both the raw analysis of the descriptors used during testing and anecdotal evidence given during feedback, there appears to be some discrepancy with the original panel discussions. There would appear to be some positive correlation between the terms strength and depth, and some negative correlation between articulation and resonance. Given further consideration, both these results are understandable.

1) Strength and depth may both be considered as ‘quantities which relate to low frequency energy’, whether this be in amplitude or low frequency extension. During the original panel discussions, it was suggested, and subsequently agreed, that a sample *could* be classed as strong but not deep, but also deep but not strong. This was not shown in the results. One explanation for this may be the use of samples

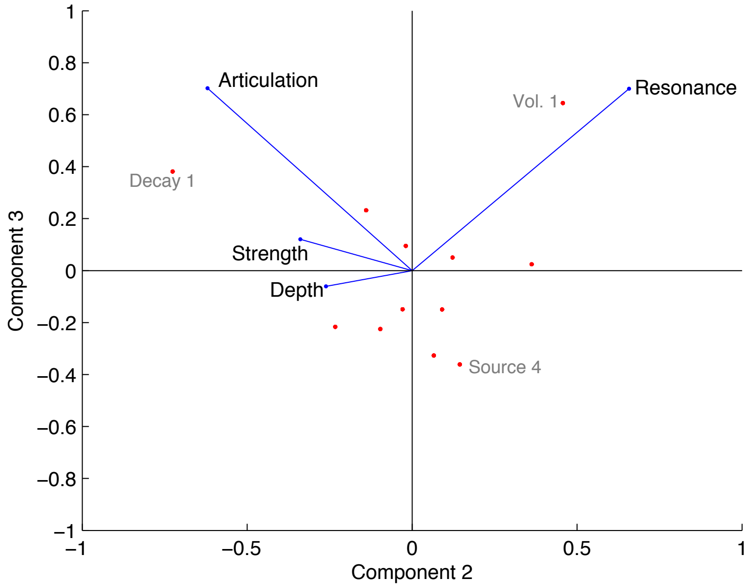


Figure 8.6: Principal component analysis showing components two and three, projections of each term and the mean subjective score for a each room (three highlighted)

with exaggerated features during the elicitation process. When rating samples generated using parameters much more likely to be found in real rooms, the descriptors were somewhat synonymous with each other. This effect of exaggerated features is also noticeable when viewing Figures 8.2 to 8.4 - a relatively small percentage of each scale was used. In other words, very rarely were the samples considered to have evoked a sensation of full magnitude. There is of course, some question as to what should be classed as ‘full magnitude’ for each scale. Whilst the second panel discussion was designed to determine this, in practice, particularly with a new sample set, this remains highly subjective. The case for using a larger set of audio samples here is interesting. Providing a large sample set with many different musical excerpts and virtual rooms may assist in training the listener, although it may then become difficult to determine which samples/rooms should be included in the test. As these descriptors, in this test at least, refer specifically to room acoustics, the magnitudes of maximum and minimum could have been defined at the elicitation stage as corresponding to the levels of each parameter *likely* to occur in rooms. However, this somewhat precludes the exploratory nature of the study and limits the scales in terms of their wider use in low frequency subjective testing.

2) The negative correlation between the terms articulation and resonance is also logical. It could be argued that a resonant artefact inherently reduces the articulation, ‘muddying’ the overall sound. However, it is suggested that the correlation is

not particularly strong due to the fact that there are clearly cases where an individual resonance may be heard in an otherwise tight sample, as is the case with the room of volume 20m^3, scoring highest on both articulation *and* resonance. These results appear to suggest that further training and explanation as to the exact meaning of the descriptors would reduce the correlation further. It is therefore proposed that they should still be seen as valid independent scales.

The PCA further clarifies this conclusion. Figures 8.5 and 8.6 reveal the underlying dimensions within the dataset. It should be noted that some care must be taken when interpreting these axes, as the components show the underlying data structure, rather than specific sensory scales. From the first two components, accounting for 73% of the variation, we can see that the first appears to reveal a single scale closely related to both strength and depth. In addition to these two terms sharing a similar space, it is noted that the highest score on dimension one was that of the source in position two (marked Source 2), and the lowest scores were those of the first and third room volumes (Vol.1 and Vol.3). Figures 8.2 to 8.4 suggest that these were indeed the samples rated the strongest/deepest and weakest/shallowest respectively. Furthermore, strength and depth project similarly onto dimensions two and three. It is therefore suggested that the two descriptors may be combind into a single sensory attribute and henceforth be described as 'bass energy'.

The second dimension would appear to relate to scales of articulation and resonance. One could read that these two terms are negatively correlated. The sample with the shortest decay (Decay 1) was rated with high articulation, and the samples with both the longest decay time (Decay 4) and smallest room volume (Vol. 1) were given subjective ratings of highly resonant. These can be seen to correspond fairly well with the second dimension. However, we must not be too hasty in combining these two descriptors into a single perceptual dimension. On the third PCA dimension, the projections of articulation and resonance are of a similar magnitude. It appears that, as previously discussed, both resonance and articulation should be considered separate sensations, both having an important place in the terminology defining low frequency audio.

8.6.2 Consistency of Scale Use

The standard deviation observed in the direct attribute ratings is seen as a measure of the consistency in using the scales across all subjects. In previous testing, high standard deviations have been observed, particularly when asking for a quality rating (6.5, (Wankling and Fazenda, 2009)). By having the four descriptors providing rating scales more obviously related to specific objective quantities, the standard deviation is expected to be reduced. The standard deviations for the direct attribute ratings here are shown in Table 8.3, showing similar ranges for each of the four direct attributes (17.7 - 27.3), with a mean of 24.4. Although sensation magnitudes were generally agreed upon in the second discussion session, they appear not to have translated to a small standard deviation. One explanation may be related to the fact that with the more realistic room models, the range of perceived sensation was reduced. Under these conditions, the complexity of the task increased, and the use of the scales may have differed between subjects. For example, it is likely that some subjects 'compensated' for the smaller differences by exaggerating their scaling.

It has been reported that training of subjects is an important process (Bech and Zacharov, 2006), and it is possible that if further training had taken place, with more detailed examples and explanation of each sensation, the observed variation could have been reduced. Furthermore, the use of high and low anchors within the test itself may prove beneficial in future testing. However, with the success of the paired comparison quality test methodology, it appears that ultimately, decreasing the complexity of any subjective task is the most effective solution. Therefore, as there are just four descriptors to be rated, it is suggested that the paired comparison method could be used for each of these. Making a forced choice as to the most resonant of two samples, or deciding which has the highest articulation may further increase reliability.

8.7 Discussion of Tested Room Parameters

We may now focus our discussion upon the analysis of the terms and overall perceived quality with relationship to the room parameters under test. Firstly, we consider the room volume parameter. In an attempt to continue the discussion of findings in Chapter 5, it is interesting to consider the results with reference to Schroeder Frequency - the transition between a modal sound-field and one which may be considered diffuse, occurring where the statistical modal density reaches that of three modes per bandwidth (Schroeder, 1987). Traditional Schroeder Frequency theory interpretation argues that higher density and thus larger volumes should result in a reproduction free of modal artefacts. However, these results (see Figure 8.1) indicate that there are no significant differences in the overall perceived quality across the four volumes, in agreement with the results of Chapter 5. Whilst the smallest room of $20m^3$ may appear to score lower at first glance, further investigation shows that its score is less than one z score from the larger volumes, which are all rated similarly. We must therefore look further into the individual attributes associated with a change in volume. Here, it is evident that there are no real linear trends for any of the four attributes, perhaps with the exception of resonance. We must ask why this is.

The answer lies in the complex interaction of modes sharing the same bandwidth. In any room where modal bandwidths overlap substantially, there is evidence of dips in the magnitude response caused by destructive interference. These destructive interference dips prevent higher densities from smoothing the magnitude frequency response as would otherwise be expected from simple, in phase, modal interaction. This causes audible degradation which appears to be particularly associated with the 'articulation' descriptor. No linear relationship exists, as the interference patterns appear at different frequencies and with different bandwidths as volume increases. Hence, no clear improvement in reproduction quality is necessarily afforded when increasing the volume of the room, adding weight to the argument that the Schroeder Frequency should not be used as a subjective indicator.

However, it does seem feasible that a high level for the resonance attribute could be expected in rooms with small volumes due to the lack of interaction between modes. In this case, individual modes are much more likely to be heard in isolation and perceived as resonance. Even though a smaller room scores higher articulation, it also scores higher resonance. As this room appears to have a slightly lower perceived quality (Figure 8.1), it is suggested that high resonance is particularly detrimental to the overall perception of reproduction quality.

Room volume as a design parameter, however, must not be considered in isolation from other parameters such as modal decay and source/receiver position, which also contribute to the perception of the room. Indeed, it is important to remember here that the Schroeder Frequency is dependent on both volume *and* decay - the tested room parameter now discussed. As the decay time is reduced, the Schroeder Frequency lowers, implying an improved perception at lower frequencies for a room of the same volume. This testing has shown that this is indeed the case. A reduction of decay time clearly shows an increase in perceived articulation and a decrease in perceived resonance (Figure 8.3). In terms of a relationship with the descriptors developed, one explanation for this is that the response becomes less sensitive to particular modal interaction due to higher damping, thus leading to a perceptual improvement through increased articulation.

From Figure 8.1, it is clear that reducing the decay time leads to a clear perceptual improvement. The 0.7 second decay time condition produced the worst reproduction overall (including the various levels for other room parameters under test). Rooms with 0.3 and 0.1 second decays appear to have similar quality ratings.

One explanation is that a threshold below which we no longer perceive a difference (where decay is around 0.1 to 0.3 seconds) has been reached. This is in line with results from Chapter 6. We propose that it is this threshold rather than the Schroeder Frequency which should be the dominant factor for room design. A reduction in decay affords an improvement not because it lowers the transition frequency, but rather because a) any ringing of modes is shorter and therefore less noticeable, and b) modal interference is reduced, leading to a smoother magnitude frequency response and an improved perceived quality through achieving a high articulation.

Finally, we consider the third parameter - source and receiver position. Using the relationship between these positions within a room, it is possible to achieve a smoother frequency response - a point which has surprisingly, thus far remained uncovered by most aspect ratio optimization literature. Figure 8.1 shows a higher perceived quality in rooms where the source is positioned closer to the receiver. Note that even though these rooms had a high decay time at the lowest frequencies, their rating is comparable to those where decay was 0.1 and 0.3 seconds. However, these scores are only achieved where the source is placed unrealistically close to the receiver. Realistic positions, such as with the source in a tri-corner, are shown to produce some of the lowest scores. It would be expected that the source/receiver position impacts the articulation rating, as moving these affects the coupling of loudspeaker/listener to each mode, modifying the interactions and therefore the relationships between peaks and dips in the response. However, the statistical significance obtained through the ANOVA (Table 8.2) shows this not to be the case. Rather, results suggest a significant influence on bass energy. This would be expected as the modal coupling decreases as the source moves further from the corner. This result therefore confirms the conventional wisdom, that in order to alter the bass energy, adjusting the source position would be the most appropriate course of action, but it is encouraging to note that this should not affect the articulation - a perceptual aspect shown to be closely related to the overall perceived quality.

8.8 Summary

This chapter has presented a listening test which was two-fold in purpose - to test the validity of the descriptors developed and to use them to further explore possible relationships with room parameters which may have been overseen by previous testing methods.

Through statistical methods and anecdotal evidence, it has been suggested that the descriptors ‘strength’ and ‘depth’ be amalgamated into a single term ‘bass energy'. Whilst there is some suggestion that articulation and resonance could be negatively correlated, the discussion argued that these do represent independent sensations and should remain as unique scales. The use of these terms across subjects reveals a higher than expected standard deviation. However, it is argued that this does not suggest an unreliability of the scales, but rather, methods to improve this consistency have been suggested for use in further studies. A greater emphasis on training is suggested, alongside the reduction of task complexity when rating against these scales.

The listening test produced two sets of results - an overall perceived quality for the twelve modelled rooms, and the ratings for each against each descriptor. Three commonly used room parameters were investigated, with four levels of each. The results indicate:

- the primary importance of reducing the decay time within a room. This leads to an increased articulation which is considered to be the related to perceived quality. Strength and depth are not directly influenced by decay time.
- the positioning of source and receiver appear next in the relative importance of design parameters. In some cases, perceptual improvement is at the same level as the effect of reduced decay, although in this test, these cases correspond to unrealistic source positions. A clear reduction in perceived quality was shown with the loudspeakers close to a tri-corner however, suggesting the increased excitation of modes is detrimental, as is an excessive bass energy. Positioning allows balancing of the bass energy without influencing the articulation or resonance.
- a small room volume was shown to be associated with increased perceived resonance, which leads to a reduction in perceived quality. Differing volume rooms show a similar quality, casting doubt on a common interpretation of the Schroeder Frequency - that increasing density necessarily reduces the perception of modal artefacts.

Chapter 9

Perceived Quality and Response Parameters

9.1 Introduction

Throughout this thesis, it has been shown that any difference in room response will be perceptible, and yet perception of difference does not in itself reveal anything of the responses' quality.

The previous chapter therefore explored listener perception when assessing low frequency room responses. In particular, the paired comparison method was found to be successful in assessing quality. This indirect scaling method reduced the complexity of the task when compared to hedonic rating scales requiring subjects to use sliders or point scales to rate quality directly.

Figure 7.2 showed that the room parameters such as those studied will affect the *response parameters*. The findings of this thesis point to these response parameters as ultimately responsible for the perceived quality. To conclude the thesis, we may now use the same paired comparison method to assess the reproduction quality with direct reference to a set of response parameters.

9.2 A Reversal in Approach

As has been highlighted, many attempts to determine the best sounding room responses take what is here defined as a 'parameter to perception' approach. Such an approach determines some metric which can be calculated from the parameters of the room (the modal density for example). A number of samples are then produced and tested where the independent variables are one or more parameters, and the dependent variable the room quality, or a threshold/optimal value. The attempt here is to determine to what extent the metric affects perception.

Both previous research, and the experimental work of this thesis show that these methods are often frustratingly inadequate. This is particularly so where one metric appears to have no linear relationship with the desired result, as is the case with modal density (Chapter 5). Worse still, a result is observed, and yet the metric is either a) based upon an assumption which is invalid in real rooms, b) the best scenario observed is infeasible within a room, or c) the response of a real room doesn't exactly match that expected. In each of these cases, the optimisation becomes at best ineffective, or at worst, results in an even greater degradation.

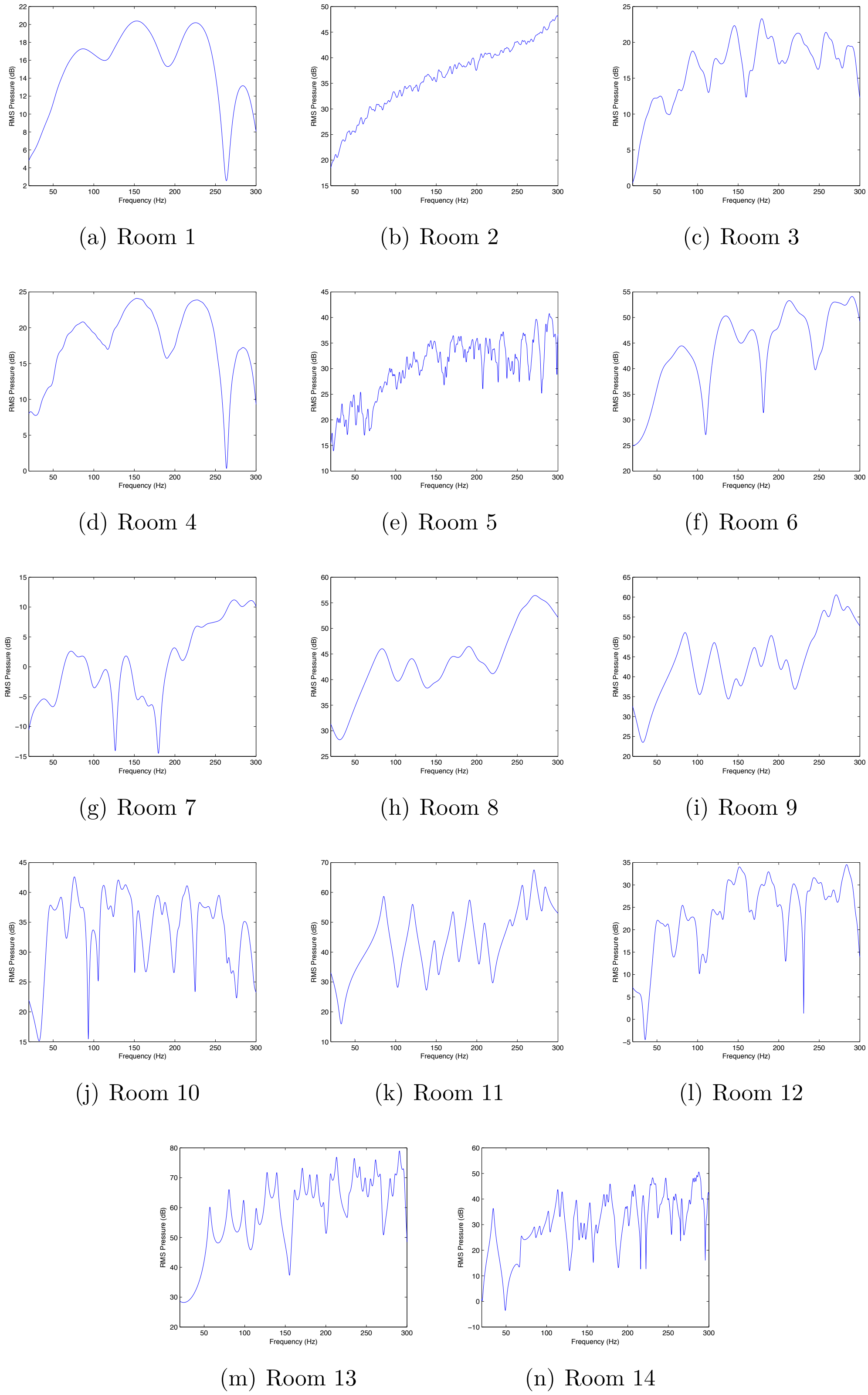
The above observation leads to the suggestion of taking the opposite approach - that of 'perception to parameter'. This takes the room response as the starting point, and determines the perceived audio quality. Each room response can be seen as arbitrary, and it is the response *features*, rather than the room parameters which are considered as potentially influential. It is only where a significant quality increase related to a particular response parameter is found that an attempt can be made to uncover what optimisation techniques can be applied to achieve the highest quality according to that parameter.

This chapter takes this new approach, considering a number of room responses, and determining the perceived quality before analysing these in terms of a number of response parameters.

9.3 Test Room Responses

The previous chapter reported on the evaluation of twelve room responses. These were generated based upon successive levels of three room parameters. In order to distinguish the response parameters under test here, a further set of fourteen rooms have been generated. The parameters were modified in such a way that a range of responses were produced. This was achieved by selecting model parameters based upon prior knowledge gained through these experiments, and ensured a set of samples which, whilst realistic in nature, were not based directly upon a specific room parameter.

The fourteen frequency responses are shown in Figure 9.1.



(a) Room 1 (b) Room 2 (c) Room 3

(d) Room 4 (e) Room 5 (f) Room 6

(g) Room 7 (h) Room 8 (i) Room 9

(j) Room 10 (k) Room 11 (l) Room 12

(m) Room 13 (n) Room 14

Figure 9.1: Frequency responses of the fourteen rooms

9.4 Listening Test

A test for perceived quality was run using the same paired comparison method outlined in the previous chapter. As before, the results may be analysed to produce a quality scale. The scores can be correlated to any metric calculated, thus revealing if the metric is a useful predictor of subjective quality.

The 'LEN' music sample from previous tests was used as a single stimulus, having been shown to be particularly useful in highlighting low frequency artefacts. It is acknowledged that, while a number of researchers have shown that music sample is not a significant factor in subjective tests of this nature (Chapter 6; Antsalo et al. (2003); Wankling and Fazenda (2009); Fazenda et al. (2005)), there remains the possibility that certain *response parameters* may be perceived differently dependent on source material.

A total of 19 subjects were tested, with each of the fourteen samples compared against the others once. Therefore a total of 91 comparisons were made by each subject. With a large number of comparisons, the risk of listener fatigue was increased, and so regular breaks were enforced. Playback order was randomised between subjects.

9.5 Perceived Quality Results

Figure 9.2 shows the results for the overall perceived quality rating of the fourteen samples. As in the previous chapter, the scaling is again derived using Thurstone's analysis based upon the Case V assumption (Thurstone, 1927). It is noted that higher z scores are observed in this quality test, compared to that of the previous chapter, suggesting a greater differentiation between the samples was perceived.

9.6 Response Parameters

It is possible to derive an almost limitless number of response parameters from an individual room response. However, it is sensible to determine those which are most likely to be perceptually relevant and those which have been suggested in the literature. Throughout this thesis, a number of observations have been made as to the factors most likely to have a significant effect on the overall perceived quality. A selection of response parameters are now described, and metrics derived from them. Once introduced, these metrics are run against each of the fourteen responses to produce an associated score for each room.

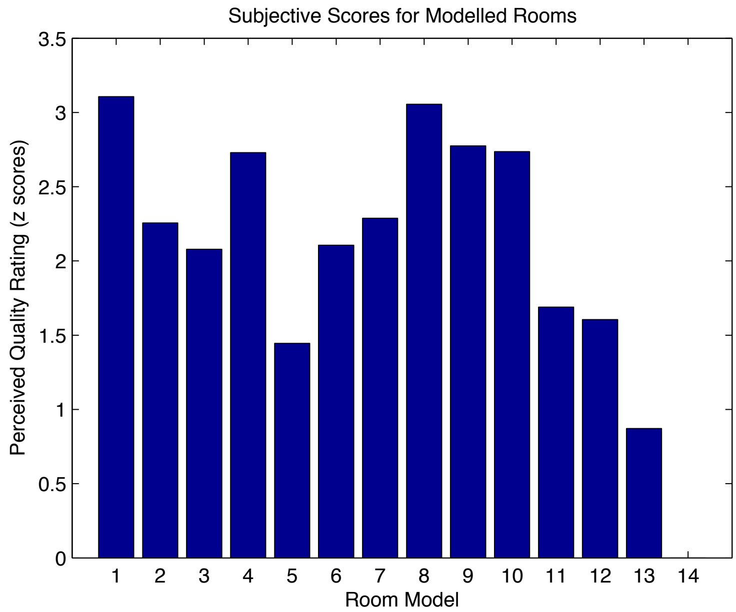


Figure 9.2: Quality scores for the fourteen room responses

The metrics chosen have been broken down into several classes - decay rate, signal reproduction, frequency deviation and frequency artefacts.

9.6.1 Decay Rate

Results have shown the undesirable characteristics of excessive decay times. Indeed, Chapter 4 determined the optimal spacing of two resonances in order to reduce this decay time and there was a increase in perceived quality for rooms with shorter decays in Chapter 7.

Whilst it is common to consider each resonant mode as having its own individual decay time (Howard and Angus, 2001), wherever there are multiple modes in a small frequency region, it is likely that they will share a *similar* decay rate. It is therefore suggested that considering the *overall* decay time at low frequency is a valid metric. This is particularly apparent in cases of high modal density, where the low frequency sound-field may well be considered statistical according to Schroeder's equation (3.31). In such cases the general decay time across the low frequency region may be perceived as a single entity.

In order to calculate a metric based on the low frequency response, the Schroeder backward integration method is again employed (see Section 6.5.2). The energy decay curve is analysed to determine the points at which the energy first drops below -5dB and -35dB. The gradient is then calculated from these points and extrapolated to determine the time taken for a drop of 60dB, analogous to the commonly used acoustical measure of RT_{60}. For this 'overall low frequency decay time', the frequency range for this calculation was between 20 and 250Hz.

In addition to the overall decay across the low frequency range, decay times were calculated for each third octave band. Chapter 4 showed that decays are perceived differently across the frequency range and Chapter 6 revealed the decay time thresholds which are frequency dependent. By analysing the measured decay time at each band, it is possible to investigate whether those at specific frequencies have a higher correlation to perceived quality.

9.6.2 Signal Reproduction

The second class of metrics is referred to as 'signal reproduction'. These are measures of the ability of the system (in this case the model room) to accurately reproduce the input signal. For this study, the Modulation Transfer Function (MTF) is considered.

Initial studies have shown that the MTF is a promising predictor of audio reproduction quality (Houtgast and Steeneken, 1985; Harris et al., 2006; Fazenda et al., 2006a). The function measures a system's ability to preserve amplitude modulations of a signal over a set frequency range. The modulation frequencies are defined as representative of audio signals and in particular those found in speech (this technique is applied to define the Speech Transmission Index). Research has been conducted into a set of modulation frequencies more appropriate for music reproduction (Harris and Holland, 2008) and as such, the modulation frequencies used in this metric were 0.8, 1, 2, 4, 6 and 8Hz. The algorithm used to calculate the MTF is defined by Schroeder (1978):

$$m(F) \approx \frac{\sum_{0}^{N} h_f^2(n) e^{-j2\frac{F_n}{Fs}}}{\sum_{0}^{N} h_f^2(n)} \tag{9.1}$$

where F is the modulation frequency, $h_f(n)$ is the discrete impulse response of the system band-passed with centre frequency f, and f_s the sampling frequency of the impulse.

The function therefore takes the modelled room impulse response and calculates the result for each modulation frequency. These may then be averaged and the result lies between 0 and 1 for each frequency band, where 1 represents the system preserving an an exact copy of the input signal. A further average is taken across each frequency band, which are the third octave bands between 31 and 200Hz, producing the final score.

9.6.3 Deviation from Ideal Response

A number of researchers have proposed metrics based upon the deviation of the frequency response from an 'ideal' case. For example, Cox et al. (2004) base a room optimisation algorithm upon the deviation from a straight line through a modelled response.

Another similar metric has been referred to as a 'figure of demerit' (Vanderkooy, 2007), which was used in the assessment of active loudspeaker arrangement simulations. It considers the frequency range of 20-150Hz, tapered to give precedence to frequencies in the 40-100Hz range. The final score is the deviation of the room response from a one octave smoothed version of itself. Vanderkooy goes on to state that this metric may not have 'a strong psychoacoustic basis', but that it may be used nonetheless in the 'absence of relevant research'. It is hoped that this thesis helps to alleviate this absence.

Within this class of metrics, in addition to the examples shown from the literature, it is suggested that the deviation from a simple 3rd order polynomial curve fit of the response may offer correlation to quality. The rationale behind this is that such a curve should more closely follow the general characteristics of the response. The 'ideal' target remains smooth however, and therefore, any response closely aligned to this is likely to have a short decay time. A similar deviation from a simple 'flat' response case, is also included.

Finally, the deviation is calculated where the ideal response is 'tilted' in favour of the lower frequencies. Wankling and Fazenda (2009) suggest that a lack of bass may be perceived as a negative attribute of the room by certain subjects depending on their personal preference, even where the response is otherwise articulate. In larger listening spaces, such as concert halls, Beranek (2003) suggests that a rise in the low frequency energy (bass ratio) is necessary, and that a greater amount of low frequency energy is required to 'support' the music. Furthermore, Soulodre and Bradley (1995) observe that the perception of low frequencies is best correlated to the strength of those frequencies. A corresponding metric is therefore determined by defining a weighted line of best fit with a 10dB increase from 200Hz down to 20Hz.

It should be remembered that while the deviation metrics discussed are intuitively satisfying, they are based upon an *average deviation* across the frequency range. There is no guarantee that an individual resonance might cause significant audible degradation, and yet as part of the overall response, not decrease the average deviation significantly. It is possible that it is not the 'average' smoothness of the response, but the individual elements of it which are perceived. Figure 9.3 shows

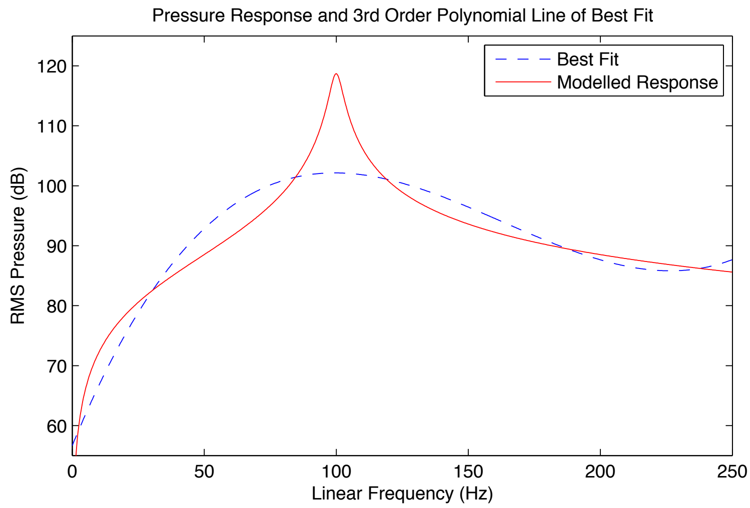


Figure 9.3: A response with a single mode at 100Hz and its associated 3rd order polynomial best fit curve

an example case of a single resonance which is highly likely to cause degradation in the room, and yet scores reasonably highly (0.71) on the polynomial curve metric. Such individual elements are therefore the basis for the final class of metrics.

9.6.4 Frequency Artefacts

The final class of metrics considered are known as 'frequency artefacts'. These are defined as isolated characteristics of the frequency response which may be significant in the perceived quality of reproduction. These relate to the general shape of the response which may be observed through a visual inspection. For example, a number of significant peaks may be clear, and acousticians will often flag such frequency imperfections for optimisation. By including such metrics in this analysis, it is possible to determine the validity of such a visual approach for calculation of expected perceived quality.

Obvious artefacts such as peaks, troughs and dips of the low frequency room response have been the focus of much research (Bucklein, 1981; Olive et al., 1997; Toole and Olive, 1988). Olive et al. (1997) conclude that resonant peaks are more degrading than comparable dips. Where modal equalisation has been attempted, it has often been the significant peaks which are selected for treatment by the equalisers (Makivirta et al., 2003). Therefore, a 'number of peaks' metric is considered. For this study, complex mode parameter detection such as in (Makivirta et al., 2003) has not been undertaken. Rather, the simple metric is determined by taking each frequency response and calculating the number of peaks present in the 20-200Hz

Class	Description	Short Name
Decay	RT_{60} across all frequencies	DEC_ALL
	RT_{60} at 32Hz	DEC_32
	RT_{60} at 63Hz	DEC_63
	RT_{60} at 80Hz	DEC_80
	RT_{60} at 100Hz	DEC_100
	RT_{60} at 125Hz	DEC_125
	RT_{60} at 160Hz	DEC_160
	RT_{60} at 200Hz	DEC_200
Signal Reproduction	Modulation Transfer Function	SR_MTF
Deviation	Vanderkooy Figure of Demerit	DEV_VAN
	3rd Order Polynomial Best Fit	DEV_SMOOTH
	Flat response	DEV_FLAT
	Bass Tilted response	DEV_TILT
Frequency Artefacts	Significant peaks	ART_PEAKS
	Significant dips (>6dB)	ART_DIPS6
	Significant dips (>12dB)	ART_DIPS12

Table 9.1: Table showing each metric under test and its associated class

range.

In addition to peaks, a metric is proposed based upon the number of significant dips. This is calculated by determining the number of points in the response which are greater than 5dB lower than the previous peak. A second significant dips metric follows the same method, but with significance attributed to a 10dB drop.

Table 9.1 summarises each metric under study along with a short code which is used when presenting the data.

9.7 Analysis

9.7.1 Correlation

Correlation analysis was performed on the sixteen room metrics and the preference scores across the fourteen room responses. The results can be seen in Table 9.2, which shows the Spearman correlation coefficients for the quality and each metric. Where a score is above 0.7, a link in perception is considered between the perceived quality score and the metric score. There are five such correlations - overall decay, decay at 160Hz and 200Hz, the number of peaks, and also the deviation from a tilted response. With the exception of Signal Reproduction, each class of metrics is represented here.

Furthermore, correlation with perceived quality for each of these metrics makes

Metric	Correlation with perceived quality score
DEC_ALL	0.74
DEC_32	0.19
DEC_63	0.56
DEC_80	0.56
DEC_100	0.53
DEC_125	0.68
DEC_160	0.78
DEC_200	0.80
SR_MTF	0.60
ART_PEAKS	0.74
ART_DIPS6	0.59
ART_DIPS12	0.51
DEV_SMOOTH	0.44
DEV_FLAT	0.70
DEV_TILT	0.83
DEV_DEMERIT	0.65

Table 9.2: The correlation of the 16 metrics and perceived quality

good sense, and it is suggested that these can be related to the three auditory sensations (articulation, resonance and bass energy) defined in Chapters 7 and 8. In the previous chapter, decay was shown to relate to articulation, with a high articulation resulting in increased quality. The overall decay takes into account the higher frequencies, and of the individual frequency band decays, it is the two highest which show a correlation. We have seen in Chapter 6 that we are more sensitive to decays above 100Hz, and so higher correlation with perceived quality of decays at higher frequencies is logical.

It is suggested that the number of peaks may relate to the perception of resonance. Low scores of this metric (high numbers of peaks) suggest the presence of distinct resonant modes which are audible and degrading. Finally, the higher quality of rooms scoring highly on the tilted response metric appear to relate to the perception of 'bass energy'.

It is acknowledged that correlations of 0.7 - 0.8 do not provide conclusive evidence of a direct relationship, and the above suggestions are seen as indicative only. It is usually expected that such audio quality tests will have much larger number of participants, and this is seen as critical for further work studying the perceived quality. The results are nonetheless of interest, and are discussed further in the light of the overall conclusions of thesis in Chapter 10.

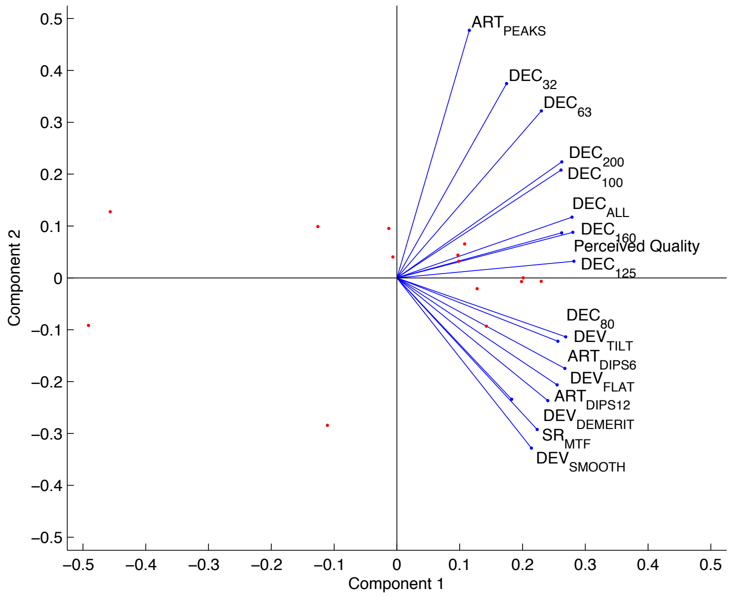


Figure 9.4: Principle Component Analysis for Sixteen Metrics and Perceived Quality

9.7.2 Principle Component Analysis

It is noted that many of the metrics studied are related. For example, a low decay time is likely to also result in fewer peaks. In order to explore these relationships, a principal component analysis was performed on the the sixteen figures of merit, along with the perceived quality scores. The first two dimensions are plotted on Figure 9.4, and account for 71% and 15% of the variance respectively. By analysing the loadings of each metric on the PCA, the first dimension appears to relate to the decay time. It is unclear as to the effect explained by the second component, although as it explains 15% of the variance, must be considered an improtant factor.

In addition to the scores for the fourteen rooms, Figure 9.5 shows the PCA for these twelve responses tested in the previous chapter. It is noted that the first two dimensions of the PCA in this case only account for 58% of the variance. However, the first dimension once again appears to reveal perceived decay. The circled rooms on this plot show this clearly, as the scores across the four rooms where the parameter under test was decay increase linearly along the first dimension. However, when attempting the correlation analysis with these twelve rooms, the highest coefficient observed was 0.7, and coefficients as low as 0.2 are reported for the decay at 160Hz metric. It is believed that this is as a result of the twelve samples being much more similar that the fourteen tested in this chapter. This would explain the lack of correlation with the perceived quality scores for the metrics considered here.

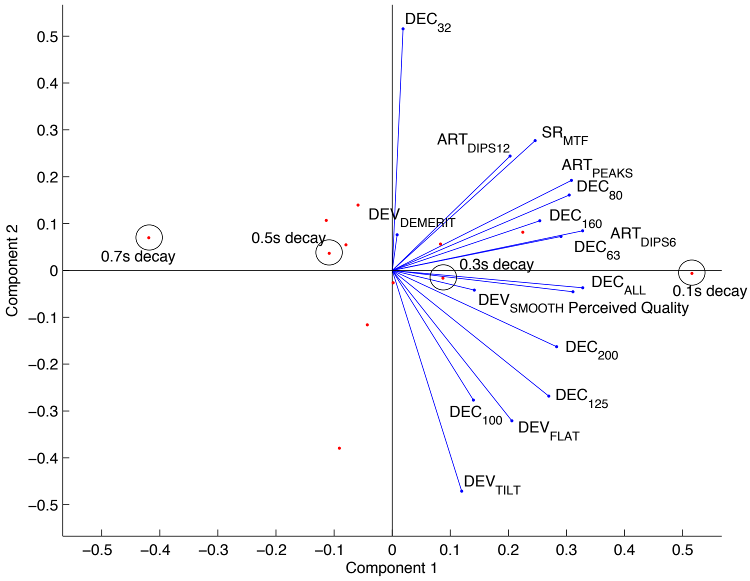


Figure 9.5: Principle Component Analysis for Sixteen Metrics Calculated from the Twelve Room Responses Of Chapter 8

9.8 Discussion

The analysis of sixteen response parameters reveals the importance of the low frequency decay in the perceived quality of reproduction. It is clear from the principle component analysis for the first listening test that each metric loads positively upon the first component, with the overall decay time loading particularly strongly. Furthermore, we may observe that perceived quality also loads on this component. There are a number of metrics which also load upon the second component, some positively (e.g. the number of significant peaks) and some negatively (e.g. the deviation from a smooth response). It is unclear as to whether this dimension reveals a specific percept. However, it is suggested that, with much of this thesis revealing that subjects often perceive differences between room responses independent of quality, the existence of this second dimension seems sensible. That two dimensions are present adds weight to the argument that the perception is multidimensional, and yet the perceived quality is most strongly associated with the decay times of the low frequencies.

As has been shown, a number of metrics have been derived in the literature. Some of these are more complex than others. For example, the Modulation Transfer Function which uses an algorithm to look at the preservation of modulation by the system. Furthermore, the Vanderkooy figure of demerit considers the pressure response, weighting certain frequencies and the relative importance of peaks and dips. Others such as the deviation from a flat response are much less complex. However, each metric tested appears to gain its potential subjective relevance through its relationship with the decay time. It is clear that a response with a low decay time will deviate little from most interpretations of an 'ideal' response. It will also be more likely to reproduce modulations faithfully. It is therefore argued that each of these metrics, whilst maybe visually pleasing and intuitive, offer little advantage over a simple decay time calculation.

9.9 Summary

This chapter has taken the opposite approach to previous work which has often focussed on determining thresholds or the perceived quality of low frequency reproduction based on a particular parameter. Here a listening test was performed with fourteen room responses unrelated to a specific room parameter. After the testing phase, the responses were analysed using a number of response parameters, or metrics, and the correlation between these metrics and the perceived quality scores analysed. Whilst the correlations are informative, in order to draw firm conclusions, it is necessary to extend this methodology to include a larger number of participants.

In line with the results of the previous chapter, it has been confirmed that metrics relating directly to the length of the modal decay are the most strongly associated with perceived quality of audio reproduction. It has been shown that the calculation of overall decay, rather than the decay of individual modes, or frequency bands is acceptable. The analysis has shown that a number of other metrics which have been used in the literature may also help to reveal audio quality, although these are themselves often related to the modal decay time.

The following chapter considers the results reported here, and presents a discussion of the findings of the experimental work conducted throughout this thesis.

Chapter 10

Discussion and Conclusions

10.1 Introduction

This thesis has presented an investigation into the assessment of low frequency reproduction quality in critical listening spaces. In such spaces, it has been shown that due to the complex interactions of resonances, known as modes, the low frequency region often displays less than ideal reproduction. Many researchers have attempted to produce optimisation methods in order to control the low frequency sound-field. Chapter 2 reviewed these techniques. It was highlighted that many objective metrics exist by which rooms and optimisation attempts have been measured. Whilst some work has been carried out in order to understand our perception in more subjective terms, the review identified the need for a greater understanding. This thesis has therefore presented a number of listening tests in order to achieve this.

This chapter considers the results of these experiments in the context of the research area. A number of key points are discussed and conclusions drawn.

10.2 Methodology

Although a number of test methods were employed, there was commonality in that all experiments utilised virtual room auralisation with playback through headphones. Where necessary, individual resonances were generated, although most often, rooms were auralised to simulate the low frequency with respect to the variable under test. Such an approach has been demonstrated previously (Fazenda, 2004) and was considered successful again. Where results were shown to be less satisfactory, this was due to either the complexity of the task, or the specific test method, rather than the performance of the model/auralisations. It is true that more accurate room models exist, but each experiment in this thesis was designed on the basis of the *general* room response, rather than an exact representation.

A novel approach taken in this thesis was the combination of PEST and ABX. Previous research asked subjects if a difference is audible between two samples (Avis et al., 2007; Goldberg, 2005). The response informs the PEST rules for the next independent variable level. This relies on the subject's answer aligning with that of their true perception. This may not always be the case, such as if they feel they are performing 'better' if they answer a certain way. The addition of the ABX stage in the experiments reported on in the previous chapters requires subjects to 'prove' their answers relate to their perception. Whilst this may increase the time taken for each trial and therefore have a negative impact on fatigue, it is argued that this is offset by the greater accuracy, particularly as an 'incorrect' answer is immediately registered.

A number of methods for testing of perceived quality were also implemented. The paired comparison method was used for the first time with specific reference to low frequency reproduction. Throughout this thesis, it has been shown that the correct methodology is essential in order to obtain reliable quality results. The complexity of the task must be kept minimal, and in this respect, it is concluded that the paired comparison method is advantageous over hedonic direct rating scales.

The following sections now discuss and summarise the key research findings with reference to the aims set out in Chapter 1.

10.3 Perception of the Modal Distribution

Chapters 4 and 5 considered the perception of the modal distribution. Both the modal spacing and the modal density have been shown to be widely quoted in the literature, and both have been used in the derivation of objective metrics. Therefore, although both relate to the concept of the modal distribution, each was investigated separately.

In terms of modal spacing, each of the early room aspect ratio studies were based upon the assumption of some ideal spacing or alignment of modes. Different researchers chose different 'ideal' criteria and, once defined, rooms of differing dimensions were assessed. The investigation in this thesis takes the opposite, novel approach, searching for the spacing itself which listeners find optimal, and then assessing this with respect to room design.

The results from the simplified scenario under test in Chapter 4, of the effect of the spacing between two resonances, highlights the importance of considering the interaction of those two resonances. Where the 'optimal' spacing is defined as that point where the overall decay is reduced and where audible effects such as first order beats are not perceptible, the spacing is equal to one third of the modal bandwidth (with the assumption that the bandwidth of each resonance is equal). As the optimum is related to bandwidth, resonances of different Qs require a different spacing, as do resonances at different frequencies. Typically, bandwidths are small - 5.5 Hz for a resonance at 100Hz in a room with a decay time of 0.4 seconds. This leads to an optimal spacing of just 1.83Hz for this example, which would be problematic to achieve in practice. This leads to the conclusion that to achieve such an optimal spacing, a room with two almost identical dimensions is required. Interestingly, such a suggestion opposes the conventional wisdom that the dimensions should be designed in such as way as to *avoid* similar modal frequencies. Further practical complications arise when considering construction tolerances, exact and uniform absorption and also the introduction of the third dimension.

Indeed, the chapter continued to extrapolate results to the spacing of three modes, where, if the same perceptual model is used, a smaller spacing of 23% of the bandwidth is required. A virtual Q concept was introduced, and the effect of interactions in anti-phase shown. However, the investigation continually highlights the real problem with the concept of an optimal spacing - the assumptions made which are not valid within a real listening space. Even with a positive optimal spacing between two resonances defined, it is concluded that this too is based upon the assumption that all modes add in-phase, and with maximum magnitude. This takes no account of source or receiver position within the room. It was shown that an out of phase mode at the same 'optimal' frequency produces a very different response, which, according to the conclusions of the two spaced resonance test, will be both audible and degrading.

In summary, it is concluded that an optimal spacing is not feasible within realistic scenarios and it is particularly clear that distribution metrics resulting in aspect ratios are not perceptually beneficial.

Personal experience through research, discussion and attendance at acoustic conferences suggests that the modal density is more widely considered to have a subjective relevance than spacing. A number of texts imply a 'flattening' of response as the density increases, and there is the much quoted Schroeder Frequency which again, although not explicitly stated, is used to imply the problems associated with modal activity in a room are alleviated above the crossover frequency.

Chapter 5 examined this density, and a number of listening tests have been reported in search of an optimal density. To begin, in line with the optimal spacing, shape functions were omitted from the decomposition model, simulating the assumptions of many metrics which consider only the dimensions of the room. As frequency increases, density also increases, and with a greater number of modes, a pressure increase with frequency is observed, along with a clear 'smoothing' of the response. When compared to a reference case with a high enough density for the response to be considered completely smooth, at three frequencies, 63, 125 and 250Hz, an optimal density was observed using the PEST/ABX methodology. At 63Hz, this was a density of four modes per bandwidth. Comparison can be drawn with Schroeder's assessment that the transition between modal and statistical sound-field characterisation occurs where there are at least three modes per bandwidth. Therefore, at 63Hz, the statistical region could be considered analogous to a 'perceived smooth response'. However, at 125Hz and 250Hz, the required density in order to perceive this smooth response increased significantly. Therefore, even where no account of modal coupling is considered, it is concluded that the Schroeder Frequency is unable to predict the modal density above which we perceive reproduction as we would were there a smooth frequency response.

However, as with the spacing, the assumptions made by omitting shape functions produce unrealistic listening scenarios. A similar test was attempted with including the shape functions. Here it was observed that no convergence can be obtained using the PEST/ABX method. The test confirmed the perception of differences regardless of density. Rooms with a typically low density ($100m^3$ - $500m^3$) were compared to rooms with a much larger density ($1000m^3$ - $5000m^3$). It was shown that at both low and high density, differences in samples convolved with the room models could be distinguished.

In terms of the narrative of this thesis, the investigation into an optimal density brings clarity in two particular aspects. Firstly, it is clear that any metric which is derived from room parameters but does not account for the source and receiver position, the volume or the damping present is unable to predict the reproduction quality of that room. Secondly, whenever a room response differs, listeners are able to detect differences. This is shown to be the case regardless of density. Stimuli within the room will be perceived as different unless the responses are very similar - in the example tested, within 10% of the room volume. This does not however give an indication as to the perceived *quality* of reproduction. It is stressed that two rooms may *sound* different, but be judged to have the *same ability* to reproduce low frequency audio.

In conclusion, optimal values for the distribution may only be observed if assumptions are made which are invalid in realistic scenarios. Furthermore, when these assumptions are not made, and perceived quality assessed, the modal density is not related to a linear scale of quality.

10.4 Perception of Modal Decay

Chapter 6 reported the search for thresholds of modal decay. The chapter presented results for both an absolute threshold and also the thresholds in the presence of more complex musical signals. Of particular interest is the similarity of these two sets of results.

In line with the previous studies, absolute thresholds were found to be around 0.2 seconds above 100Hz, with a sharp increase as frequency drops below this point. The thresholds where musical stimuli and more realistic room models were employed reveal similar results - around 0.2 seconds above 100Hz, increasing below this. Testing such thresholds proved challenging. As has already been discussed, any differences in room response can usually be heard unless those differences are particularly small. This is especially the case with the PEST/ABX methodology which allows the listener to audition the samples as many times as they like, listening for very small changes. These changes are not considered to be as a direct result of a perceived difference in the decay time, but rather, slight timbral differences as a result of the differing interaction between the modes as the damping changed. However, thresholds could be obtained using the 'first impression' method. Such a method is a novel way of determining thresholds of perceptual parameters in the context of musical signals and realistic modelled systems, and is highlighted as a useful method for future testing. Thresholds for both artificial and natural stimuli were shown to be independent of replay level, and after statistical analysis, it was considered valid to present frequency dependent thresholds which can be used regardless of stimuli.

Although these thresholds themselves are seen as useful to room designers / optimisers in offering targets below which it is not necessary to reduce decay further, they should not be considered as the levels below which there is no increase in *perceived quality*.

In a first attempt to define a threshold of perceived quality, the chapter also reported a listening test asking subjects to rate an audition sample in terms of its reproduction quality in a critical listening space. It was hoped that if decay is an indicator of quality, a knee point would be found above which the quality is perceived to deteriorate. This knee point would reveal the threshold of perceived quality for decay. The results however, revealed no such points. The key point here is the effectiveness of the test method. At each frequency, music sample and listening level, 11 auralisations with varying decay times were presented. Listeners found it difficult to rate each of these accurately. With so many samples, the task becomes particularly complex. Again, a clear conclusion is the necessity to improve the methodology for testing perceived quality at low frequency.

To summarise, defining a threshold for the modal decay parameter is an important study, where results are not only applicable to critical listening space design, but a number of other professional audio applications such as transducer manufacture. This thesis has presented an in depth investigation into modal decay thresholds, with results revealing similar thresholds for both artificial and natural stimuli, providing a baseline which can be used by room designers and optimisers, although they should not be confused with threshold of reproduction quality.

10.5 Perceived Quality at Low Frequency

A number of initial studies into perceived quality of audio reproduction in the modal sound-field were reported in Chapter 5 - the quality of samples with different modal densities, and Chapter 6 - the quality as a response to increasing absorption.

Each of these asked the subject the question - which virtual room is responsible for the highest quality of audio reproduction? Both tests reveal the necessity for a more robust methodology for testing perceived quality at low frequency. Furthermore, it has become clear that low frequency perception has a multidimensional aspect and, similarly to areas such as spatial perception, low frequency perception should be considered in these terms.

In an attempt to better understand the overall perception in the low frequency region, Chapter 7 introduced the concept of assessment of multidimensional perception. Four terms, or descriptors, were elicited by a trained listening panel. It was agreed that the terms represent independent perceptual sensations. The terms articulation, resonance, strength and depth were agreed. The Descriptive Analysis method was used to determine these, and whilst this has been shown to be successful in a number of other areas of audio perception, is a novel contribution of this thesis in the specific context of low frequency listening. Chapter 8 sought to validate these terms and utilise them in order to determine the effect of each on the overall perceived quality of three ‘room parameters’ - decay, volume and source/receiver position. Whilst these were studied as individual parameters in previous chapters, Chapter 8 considered them with specific relation to perceived quality and perceived sensations. Furthermore, they can now be studied a) with the descriptive terms and b) a refined test methodology for obtaining overall perceived quality.

Chapter 8 itself presented a detailed discussion of the results. In particular, it is suggested that the four terms should be reduced to three, with articulation and resonance still defined as independent, but strength and depth being combined into the single term 'bass energy'. Furthermore, when we consider the relationship of these sensations to audio quality, we see on the principal component analysis, a stronger relationship between the perceived articulation and perceived quality. Throughout testing, what is less clear is the relationship between the amount of low frequency energy and perceived quality. Running a similar perceived quality test and considering the response parameters as opposed to the room parameters in Chapter 9, there is evidence to suggest a higher perceived quality associated with a higher bass energy. It is therefore suggested that there are two separate responses which a listener has to the reproduction of audio within a room:

1. A *critical* response - specifically, the response to quality in terms of the rooms articulation - the ability to reproduce audio which could be described, according to some of the original list of descriptors elicited as tight, clear, punchy etc. The amount of resonance would also impact this critical response, with excessive resonance considered degrading.

2. An *affective* response - encompassing the response to the low frequency reproduction which is not explained simply through a critical evaluation, but encompasses the listeners personal tastes and preferences. Considering the results of the previous chapters, it is suggested that this affective response is particularly related to the low frequency energy. Both the level of bass and the extension to the lower frequencies may be considered to either increase or decrease the reproduction quality dependent on the individual listener. It is argued that this personal preference would not affect the 'critical' response (with the possible exception of a clear excessive bass level at a particular frequency resulting from an isolated resonance).

Whilst observations throughout this research have led to this conclusion, it should be properly tested and forms part of the suggested further work below. However, a similar model has been suggested by Fog and Pederson (1999), where overall perception is considered a result of two 'filters', the first of which receives the 'physical stimulus' and whose output is the 'perceived stimulus'. This would be analogous to the critical response. A second filter known then accounts for expectations, interests, emotions etc. and equates to the above suggestion of an affective response.

Such a dual response makes a precise model for the quality of low frequency reproduction difficult. However, Chapter 8 showed that the different auditory sensations can be related to different room parameters, and it is suggested that this offers the designer the best chance of achieving a 'good' critical listening space. A reduction in the low frequency decay time clearly affords a higher level of articulation, and reduces the perception of resonance. Avoiding a very small listening space is beneficial in reducing the occurrence of single discrete, degrading resonances. The position of the subwoofer was shown to affect the levels of strength and depth (bass energy) independently of articulation and resonance. Therefore, a reduction of decay in a large enough space, should allow individual listeners to place a subwoofer according to their preference, without affecting the perceived quality as attributed by the critical response.

10.6 Metrics for Low Frequency Assessment

Finally, Chapter 9 took the paired comparison method for obtaining perceived quality and mapped this to a series of 'response parameters'. The parameters chosen each attempt to relate aspects of the room response to a score. Results again confirm the perceived quality to be correlated to the overall decay time. There was no significant correlation increase for decay at any individual band. A PCA revealed two principal dimensions for the assessment of samples. The first is strongly weighted to suggest it measures perception of decay. This is most likely analogous to the 'critical response' proposed. The literature reveals that it may be 'dangerous' to associate decay times based on the standard measure of RT_{60} to a modal system, as each mode will have its own decay time (Howard and Angus, 2001). However, results here suggest that this overall time is indeed a satisfactory measure.

It is difficult with the data gathered to properly label the second component. However, as the individual metrics load differently for the two sets of samples, response parameters and room parameters, it seems reasonable to suggest that components other than the one relating directly to decay may relate to the 'affective response', and we might relate this to bass content using metrics such as the deviation from a ideal titled response.

The advantage of this chapter's approach to perceived quality mapping is the elimination of any specific room parameter, meaning that any correlation found between response metric and perceived quality could be used to evaluate an existing room with a simple measurement of its response.

10.7 Further Work

10.7.1 A Large Scale Test into Perceived Quality

Chapter 8 reported on the use of multidimensional scaling and perceived quality mapping techniques. Further work is required extend this mapping of low frequency descriptors and overall quality. The number of subjects tested in these experiments was considered sufficient for the direct attribute ratings. However, a number of authors suggest that greater numbers are required when considering the overall perceived quality, particularly if this is to be representative of a common user group (Martin and Bech, 2005; Bech and Zacharov, 2006).

It is expected that a large scale test of perceived quality will produce more accurate data for further quality mapping. A further carefully chosen set of room responses, incorporating changes in both room and response parameters would be used. Such an experiment may also reveal a regression model for the overall perceived quality in relation to the low frequency descriptors, which was not possible in the work reported in this thesis.

10.7.2 Critical and Affective Responses

This chapter has suggested that perception at low frequency could comprise separate critical and affective responses. In order to properly assert this, a further listening test is envisaged which would consider a number of different listening rooms, and pose questions to the subjects such as rating in terms of the room's articulation, or ability to reproduce the audio content accurately, but also ask for a personal response. Room models could be modified in such a way as to expect the articulation to remain, for example, keeping the room parameters constant, but apply filtering to remove or add low frequency content.

If the suggestion of two separate responses is correct, the critical response, i.e. ratings given for articulation would remain constant, where as the affective response will vary to a greater extent dependent on the individual subject's preference. With such an experiment, both sets of data could also be analysed relation to the response parameters, which would aid the continuing refinement of a model for low frequency perception.

10.7.3 Evaluation of Real Room Responses

Within the suggested further testing, real room responses could be incorporated, either as a separate test, or part of the wider large scale perceived quality testing.

A series of real listening spaces would be measured, and their impulses convolved in the same way as the virtual room models. This way, patterns revealed by the quality mapping can be proved to be valid in real listening scenarios.

As part of this real room testing, the physical effect of the low frequency energy, the so called ‘feel of the bass’ as vibrations are received through the body may also be evaluated. Although this sensation was considered out of scope for this thesis, the conclusion that an affective response may constitute a part of the overall perceived quality indicates that the physical effect of the low frequency may be of high importance. It would also be interesting to determine whether this sensation has a direct effect the critical ratings.

Appendix A

The following paper has been published in the May 2012 edition of the Journal of the Audio Engineering Society. The paper contains work presented in Chapters 7 and 8 of this thesis. Please note that this was published using the author's pre-marital name, Matthew Wankling.

References

Allen, J. B. and Berkley, D. A. (1979). Image Method for Efficiently Simulating Small-Room Acoustics. *J. Acoust. Soc. Am*, 65(4):943 – 950.

Allison, R. F. (1976). The Sound Field in Home Listening Rooms, ii. *J. Audio. Eng. Soc*, 24(1).

Antsalo, P., Karjalainen, M., Makivirta, A., and Valimaki, V. (2003). "Comparison of Modal Equaliser Design Methods". In *Proc. of the 114th AES Convention*, Amsterdam.

Avis, M. (2000). *"The Active Control of Low Frequency Room Modes"*. Doctoral thesis, University of Salford.

Avis, M. R. (2001). IIR Bi-Quad Controllers for Low Frequency Acoustic Resonance. In *Proc. of the 111th AES Convention*.

Avis, M. R., Fazenda, B. M., and Davies, W. J. (2007). "Thresholds of Detection for Changes to the Q-factor of Low Frequency Modes in Listening Environments". *J. Audio Eng. Soc*, 55(7-8):611–622.

Ballagh, K. O. (1983). "Optimum Loudspeaker Placement Near Reflecting Planes". *J. Audio Eng. Soc*, 31(12):931–935.

Baskind, A. and Polack, J. (2000). Sound Power Radiated by Sources in Diffuse Field. In *Proc. of the 108th AES Convention*, Paris.

Bech, S. (1999). "Methods for Subjective Evaluation of Spatial Characteristics of Sound". In *Proc.of the AES 16th International Conf.*, pages 487–504, Rovaniemi.

Bech, S. (2002). "Requirements for Low-Frequency Sound Reproduction, Part i: The Audibility of Changes in Passband Amplitude Ripple and Lower System Cutoff Frequency and Slope". *J. Audio Eng. Soc*, 50(7-8):564–580.

Bech, S. and Zacharov, N. (2006). *The Perceptual Audio Evaluation: Theory, Method and Application*. WileyBlackwell.

Beranek, L. (2003). "Subjective Rank Orderings and Acoustical Measurements for Fifty-Eight Concert Halls". *Acta Acustica*, 89:494–508.

Berg, J. and Rumsey, F. (2006). "Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique". *J. Audio Eng. Soc*, 54(5):365.

Bhatt, N. B. (1939). Effect of an Absorbing Wall on the Decay of Normal Frequencies. *J. Acoust. Soc. Am.*, 11:160.

Blaszak, M. A. (2007). Acoustic Design of Small Rectangular Rooms: Normal Frequency Statistics. *Applied Acoustics*.

Bolt, R. H. (1939a). Frequency Distribution of Eigentones in a Three-Dimensional Continuum. *J. Acoust. Soc. Am.*, 10(3):228–234.

Bolt, R. H. (1939b). Normal Modes of Vibration in Room Acoustics: Angular Distribution Theory. *J. Acoust. Soc. Am.*, 11(1):74–79.

Bolt, R. H. (1946). Note on Normal Frequency Statistics for Rectangular Rooms. *J. Acoust. Soc. Am.*, 18:130.

Bonello, O. J. (1981). "A New Criterion for the Distribution of Normal Room Modes". *J. Audio Eng. Soc*, 19:597–606.

Botteldooren, D. (1995). Finite-Difference Time-Domain Simulation of Low-Frequency Room Acoustic Problems. *J. Acoust. Soc. Am.*, 98:3302.

Bucklein, R. (1981). "The Audibility of Frequency Response Irregularities". *J. Audio Eng. Soc*, 29(3):126–131.

Celestinos, A. and Nielsen, S. (2008a). "Controlled Acoustic Bass System (CABS) A Method to Achieve Uniform Sound Field Distribution at Low Frequencies in Rectangular Rooms". *J. Audio Eng. Soc*, 56(11):915–931.

Celestinos, A. and Nielsen, S. (2008b). Low-Frequency Loudspeaker-Room Simulation Using Finite Differences in the Time Domain - Part 1: Analysis. *J. Audio Eng. Soc*, 56(10):772–786.

Ciskowski, R. D. and Brebbia, C. A. (1991). *Boundary Element Methods in Acoustics*. Springer.

Conrad, R. (1964). Acoustic Confusions in Immediate Memory. *British Journal of Psychology*, 55:79–80.

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

Cox, T. J., D'Antonio, P., and Avis, M. R. (2004). "Room Sizing and Optimization at Low Frequencies". *J. Audio Eng. Soc*, 52(6):640–651.

Crowder, R. and Morton, J. (1969). Precatorgorical Acoustic Storage. *Perception and Psychophysics*, 5:367.

Darlington, P. and Avis, M. R. (1996). Time/Frequency Response of a Room with Active Acoustic Absorption. In *Proc. of the 100th AES Convention*, New York.

Davis, D. and Patronis, E. (2006). *Sound System Engineering, Third Edition*. Focal Press, 3 edition.

Elliott, S. J. and Nelson, P. A. (1989). "Multiple-point Equalization in a Room using Adaptive Digital Filters". *J. Audio Eng. Soc*, 37(11):899–907.

Elliott, S. J. and Nelson, P. A. (1995). *"Active Control of Sound"*. Academic Press, Cambridge.

Everest, F. (2001). *The Master Handbook of Acoustics*. McGraw-Hill, New York ;London, 4th ed. edition.

Farnsworth, K. D., Nelson, P. A., and Elliott, S. J. (1985). Equalization of Room Acoustic Responses over Spatially Distributed Regions. In *Proc. Inst. Acoust.*

Fazenda, B., Avis, M. R., and Davies, W. J. (2005). "Perception of Modal Distribution Metrics in Critical Listening Spaces-Dependence on Room Aspect Ratios". *J. Audio Eng. Soc*, 53(12):1128–1141.

Fazenda, B., Holland, K., and Newell, P. (2006a). Modulation Transfer Function as a Measure of Room Low Frequency Performance. *Proc. Inst. Acoust. (UK)*, 28(8).

Fazenda, B. M. (2004). *The Perception of Room Modes*. Doctoral thesis, University of Salford.

Fazenda, B. M., Avis, M., and Davies, W. J. (2006b). "Comments on 'Perception of Modal Distribution Metrics in Critical Listening Spaces - Dependence on Room Aspect Ratios' ". *J. Audio Eng. Soc*, 54(5):412–413.

Fazenda, B. M. and Davies, W. J. (2004). “The Views of Recording Studio Control Room Users”. *Proc. Inst. Acoust.*, 23(8).

Fletcher, H. and Munson, W. A. (1933). Loudness, Its Definition, Measurement and Calculation. *J. Acoust. Soc. Am.*, 5(2):82–108.

Fog, C. L. and Pederson, T. H. (1999). Tools For Product Optimisation. *Human Centered Processes.*

Fryer, P. (1977). Loudspeaker Distortions, Can We Hear Them? *Hi-Fi News Rec. Rev*, 22:51–56.

Fryer, P. A. (1975). Inter-Modulation Distortion Listening Tests. In *Proc. of the 50th AES Conference*, volume 23, page 402.

Gabrielsson, A. (1990). Perceived Sound Quality of Reproductions With Different Frequency Responses and Sound Levels. *J. Acoust. Soc. Am.*, 88(3):1359.

Gabrielsson, A. (1991). Loudspeaker Frequency Response and Perceived Sound Quality. *J. Acoust. Soc. Am.*, 90(2):707.

Gabrielsson, A. and Sjogren, H. (1979). “Perceived Sound Quality of Sound-Reproducing Systems”. *J. Acoust. Soc. Am.*, 65(4):1019–1033.

Geddes, E. R. and Porter, J. C. (1988). Finite Element Approximation for Low-Frequency Sound in a Room With Absorption. *J. Acoust. Soc. Am.*, 83:1431.

Goldberg (2005). “A Listening Test System For Measuring The Threshold of Audibility of Temporal Decays”. *Proc. Inst. Acoust.*, 27(5).

Goldberg, A. (2006). Measuring the Threshold of Audibility of Temporal Decays. In *Proc. of the 120th AES Convention.*

Goldberg, A. (2009). Windowed Sine Bursts: In Search of Optimal Test Signals for Detecting the Threshold of Audibility of Temporal Decays. In *Proc. of the 126th AES Convention*, Munich.

Grey, J. M. (1977). “Multidimensional Perceptual Scaling of Musical Timbres”. *J. Acoust. Soc. Am.*, 61(5):1270–1277.

Groh, A. R. (1974). “High-Fidelity Sound System Equalization by Analysis of Standing Waves”. *J. Audio Eng. Soc*, 22(10):795–799.

Harris, L., Holland, K., and Newell, P. (2006). Subjective Assessment of the Modulation Transfer Function as a Means for Quantifying Low-Fequency Sound Quality. *Proc. Inst. Acoust.*, 28(8).

Harris, L. K. R. and Holland, K. (2008). Evaluating Loudspeaker Reproduction Quality at Low Frequencies: Optimisation of a Music-Focused Modulation Transfer Function Technique. In *24th Inst. of Acoustics Reproduced Sound Conference*.

Hatziantoniou, P., Mourjopoulos, J., and Worley, J. (2005). "Subjective Assessments of Real-Time Room Dereverberation and Loudspeaker Equalization". In *Proc. of the 118th AES Convention*.

Herzog, P., Soto-Nicolas, A., and Guery, F. (1995). Passive and Active Control of the Low-frequency Modes in a Small Room. In *Proc. of the 95th AES Convention*.

Hill, A. J. and Hawksford, M. O. (2010). Wide-Area Psychoacoustic Correction for Problematic Room Modes Using Non-Linear Bass Synthesis. In *Proc. of the 129th AES Convention*.

Hill, T. and Lewicki, P. (2006). *Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining*. StatSoft, Inc.

Holland, K. R., Newell, P. R., Castro, S. V., and Fazenda, B. (2006). Excess Phase Effects and Modulation Transfer Function Degradation in Relation to Loudspeakers and Rooms Intended for the Quality Control Monitoring of Music. *Proc. Inst. Acoust.*, 27(8).

Holland, K. R., Newell, P. R., and Mapp, P. (2004). Modulation Depth as a Measure of Loudspeaker Low-Frequency Performance. *Proc. Inst. Acoust.*

Houtgast, T. and Steeneken, H. J. M. (1985). A Review of the MTF Concept in Room Acoustics and its Use for Estimating Speech Intelligibility in Auditoria. *J. Acoust. Soc. Am.*, 77:1069–1077.

Howard and Angus (2001). *Acoustics and Psychoacoustics*. Focal Press, Oxford ;;Boston, 2nd ed. edition.

Huang, Y., Di, G., Zhu, Y., Hong, Y., and Zhang, B. (2008). Pair-Wise Comparison Experiment on Subjective Annoyance Rating of Noise Samples With Different Frequency Spectrums But Same A-Weighted Level. *Applied Acoustics*, 69(12):1205–1211.

Ingard, K. U. (1994). *Notes on Sound Absorption Technology.* Noise Control Foundation, 1st edition.

ITU (1994). "ITU-BS.1116 - Method for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems". Technical report, ITU, Geneva, Switzerland.

ITU (2006). "ITU-BS.1770 Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level". Technical report, ITU.

Karjalainen, M., Antsalo, P., Makivirta, A., and Valimaki, V. (2004). "Perception of Temporal Decay of Low-frequency Room Modes". In *Proc. of the 116th AES Convention*, Berlin.

Karjalainen, M., Esquef, P. A., Antsalo, P., Makivirta, A., and Valimaki, V. (2002). Frequency-Zooming ARMA Modeling of Resonant and Reverberant Systems. *J. Audio Eng. Soc*, 50(12):1012–1029.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (2000). *Fundamentals of Acoustics.* John Wiley & Sons, 4th edition edition.

Kuttruff, H. (1991). *Room Acoustics.* Elsevier Applied Science, London and New York, 3rd edition.

Kuttruff, H. (2007). *Acoustics: An Introduction.* Routledge.

Lorho, G. (2005). "Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction". In *Proc. of the 119th AES Conference*, New York.

Louden, M. M. (1971). "Dimension Ratios of Rectangular Rooms with Good Distribution of Eigentones". *Acustica*, 24(5):101–04.

Maa, D. Y. (1939). The Distribution of Eigentones in a Rectangular Chamber at Lower Frequency Ranges. *J. Acoust. Soc. Am.*, 10:235–238.

Makivirta, A., Antsalo, P., Karjalainen, M., and Valimaki, V. (2003). "Modal Equalization of Loudspeaker-Room Responses at Low Frequencies". *J. Audio Eng. Soc*, 51(5):324–343.

Martin, G. and Bech, S. (2005). Attribute Identification and Quantification in Automotive Audio-Part 1: Introduction to the Descriptive Analysis Technique. In *Proc. of the 1218th AES Convention*, volume 6360.

Mattila, V. V. (2001). “Descriptive Analysis of Speech Quality in Mobile Communications: Descriptive Language Development and External Preference Mapping”. In *Proc. of the 113th AES Conference*.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*. Academic Press Inc, 5th edition.

Moore, B. C. J. and Tan, C. (2003). Perceived naturalness of spectrally distorted speech and music. *J. Acoust. Soc. Am.*, 114(1):408.

Morse, P. M. C. (1936). *Vibration and Sound*. McGraw-Hill New York, st edition.

Morse, P. M. C. (1948). *Vibration and Sound*. McGraw-Hill New York, 2nd edition.

Mourjopoulos, J. (1985). On the Variation and Invertibility of Room Impulse Response Functions. *Journal of Sound and Vibration*, 102(2):217–228.

Mourjopoulos, J. and Paraskevas, M. A. (1991). Pole and zero Modeling of Room Transfer Functions. *Journal of Sound and Vibration*, 146(2):281–302.

Nelisse, H. and Nicolas, J. (1997). Characterization of a Diffuse Field in a Reverberant Room. *J. Acoust. Soc. Am.*, 101:3517.

Newell, P. (2007). *Recording Studio Design*. Focal Press.

Olive, S. E., Schuck, P. L., Ryan, J. G., Sally, S. L., and Bonneville, M. E. (1997). “The Detection Thresholds of Resonances at Low Frequencies”. *J. Audio Eng. Soc*, 45(3):116–127.

Olive, S. E., Schuck, P. L., Sally, S. L., and Bonneville, M. E. (1994). The Effects of Loudspeaker Placement on Listener Preference Ratings. *J. Audio Eng. Soc*, 42(9):651–669.

Olson, H. and May, E. (1953). Electronic sound absorber. *J. Acoust. Soc. Am.*, 25(4):829.

Osipov, A., Mees, P., and Vermeir, G. (1997). Low-frequency airborne sound transmission through single partitions in buildings. *Applied Acoustics*, 52(3-4):273–288.

Pedersen, J. A. (2003). “Adjusting a Loudspeaker to its Acoustic Environment: The ABC System”. In *Proc. of the 115th AES Convention*.

Pierce, A. (1989). *Acoustics: An Introduction to Its Physical Principles and Applications*. Acoustical Society of America.

Rasch, R. and Plomp, R. (1999). *The Perception of Musical Tones*. Academic Press Elsevier, 2nd edition.

Robinson, D. W. and Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5):166–181.

Rumsey, F. (2004). New Horizons in Listening Test Design. *J. Audio Eng. Soc*, 52(1-2):65–73.

Santillan, A. O. (2001). “Spatially Extended Sound Equalization in Rectangular Rooms”. *J. Acoust. Soc. Am.*, 110:1989.

Schroeder, M. R. (1965). New Method of Measuring Reverberation Time. *J. Acoust. Soc. Am.*, 37:409.

Schroeder, M. R. (1978). Modulation Transfer Functions: Definition and Measurement. *IEEE Trans. Acoust. Speech*, 99(5):3240–3241.

Schroeder, M. R. (1987). “Statistical Parameters of the Frequency Response Curves of Large Rooms”. *J. Audio Eng. Soc*, 35(5):299–305.

Schroeder, M. R. (1996). The “Schroeder Frequency” Revisited. *J. Acoust. Soc. Am.*, 99(5):3240–3241.

Simone, G., Olive, S. E., and Welti, T. (2009). “The Effect of Whole-Body Vibration on Preferred Bass Equalization in Automotive Audio Systems”. In *Proc. of the 127th AES Convention*, Berlin.

Skålevik, M. (2011). “Schroeder Frequency Revisited ”. In *Proc. of Forum Acousticum*, Aalborg.

Soulodre, G. A. and Bradley, J. S. (1995). Subjective Evaluation of New Room Acoustic Measures. *J. Acoust. Soc. Am.*, 98(1):294–301.

Stefanakis, N., Sarris, J., and Cambourakis, G. (2008a). Source Placement for Equalization in Small Enclosures. *J. Audio Eng. Soc*, 56(5):357.

Stefanakis, N., Sarris, J., Cambourakis, G., and Jacobsen, F. (2008b). Power Output Regularization in Global Sound Equalization. *J. Acoust. Soc. Am.*, 123(1):33–36.

Stone, H. and Sidel, J. L. (2004). *Sensory Evaluation Practices*. Academic Press.

Taylor, M. M. and Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *J. Acoust. Soc. Am.*, 41(4A):782–787.

Taylor, M. M., Forbes, S. M., and Creelman, C. D. (1983). PEST Reduces Bias in Forced Choice Psychophysics. *J. Acoust. Soc. Am.*, 74:1367–1374.

Thurstone, L. L. (1927). "A Law of Comparative Judgment". *Psychological review*, 34(4):273–286.

Toole, F. (2008). *Sound Reproduction: Loudspeakers and Rooms*. Focal Press.

Toole, F. E. (1986). Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1. *J. Audio Eng. Soc*, 34(4):227–235.

Toole, F. E. (2006). Loudspeakers and Rooms for Sound Reproduction - A Scientific Review. *J. Audio Eng. Soc*, 54(6):451–476.

Toole, F. E. and Olive, S. E. (1988). The Modification of Timbre by Resonances: Perception and Measurement. *J. Audio Eng. Soc*, 36(3):122–141.

Vanderkooy, J. (2007). Multi-Source Room Equalization: Reducing Room Resonances. In *Proc. of the 123rd AES Convention*, New York.

Wald, A. (1947). *Sequential Analysis*. John Wiley & Sons Inc.

Walker, R. (1996). "Optimum Dimension Ratios for Small Rooms". In *Proc. of the 100th AES Convention*.

Wankling, M. R. and Fazenda, B. M. (2009). "Subjective Validity of Figures of Merit for Room Aspect Ratio Design". In *Proc. of the 126th AES Convention*, Munich.

Weisser, A. and Rindel, H. (2006). Evaluation of Sound Quality, Boominess, and Boxiness in Small Rooms. *J. Audio Eng. Soc*, 54(6):495–511.

Welti, T. (2009). Investigation of Bonello Criteria for Use in Small Room Acoustics. In *Proc. of the 127th AES Convention*, New York.

Welti, T. and Devantier, A. (2006). "Low-Frequency Optimization Using Multiple Subwoofers". *J. Audio Eng. Soc*, 54(5):347.

Zacharov, N., Bech, S., and Meares, D. (1998). The use of subwoofers in the context of surround sound program reproduction. *J. Audio. Eng. Soc*, 46(4):276–287.

Zacharov, N. and Koivuniemi, K. (2001a). "Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training". In *Proc. of the 111th AES Convention*.

Zacharov, N. and Koivuniemi, K. (2001b). "Unravelling the Perception of Spatial Sound Reproduction: Techniques and Experimental Design". In *Proc. of the AES 19th Conf*, volume 19.

Zha, X., Fuchs, H. V., Nocke, C., and Han, X. (1999). Measurement of an Effective Absorption Coefficient Below 100Hz. *Acoustics Bulletin*, 24:5–10.

Zielinski, S., Rumsey, F., and Bech, S. (2008). On Some Biases Encountered in Modern Audio Quality Listening Tests. *J. Audio Eng. Soc*, 56(6):427–451.