Semantic Design Methodologies for Acoustic Modelling Interfaces

Philip Oke

MSc (by research)

University of York

Electronic Engineering

December 2021

Abstract

While there have been developments in semantic interfaces for digital audio effects in order to allow non experts in audio engineering to understand and apply audio transforms within an audio production workflow, there is relatively less research available in applying these principles towards interfaces that allow non experts to treat live environments to induce desired perceptual factors of sound played within them. The aim of this research was to propose methods to derive associations between semantic descriptors of audio that would be used within a hypothetical semantic interface, and transformations of an acoustic environment. These methods were then investigated for their effectiveness via deriving associations for two selected perceptual factors, 'brightness' and 'closeness', in relation to two room modification variables, 'source/receiver distance' and 'absorption coefficient of wall surfaces'. The hypothetical interface was conceptualised as a three schema arrangement; with the perceptual terms in the external layer, acoustic measurement in the conceptual layer, and room modification variables in the internal layer. Experiments with auralisation models were performed to derive associations between perceptual terms and acoustic measurements, while a listening test was undertaken to derive associations between acoustic measurements and room modification variables. The results from these two experiments were cross referenced with each other to derive associations between perceptual terms and room modification variables. It was concluded that absorption coefficient was the primary factor influencing perceptual brightness, while source/receiver distance was the primary factor influencing perceptual closeness. It is recommended that these methods be further investigated for more complex associations to be made for a greater variety of perceptual terms to move towards the development of a semantic interface that can be used with acoustic models. Future applications of these methodologies could integrate machine learning methods to allow for more complex systems and a greater range of applications.

Contents

List of Figures

List of Tables

Acknowledgements

My endless gratitude towards my supervisors Damian Murphy and Frank Stevens for their tireless support throughout what has been an unprecedented time in history to do research work. Your knowledge and enthusiasm has helped keep me going throughout this whole period, and it's genuinely been a more positive experience as a whole working with you both.

To everyone else at Audiolab, whose warm and insightful presence helped me feel like a part of the campus while I was working remote, and who took time to help distribute my listening test among many other things. I'm super glad I got to meet you all at least once in person.

To the folks at the Maths Skills Centre, who gave me useful help and advice in developing a statistical analysis framework for this research project, so many aspects of this project would have been an endless struggle without their help. I cannot commend them enough.

To Gus and the gang for always being there, even when it seemed like the world was falling apart. I miss those old Sheffield days like the world.

And to Mum and Dad, for putting up with everything.

Declaration of Authorship

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Chapter 1

Introduction

1.1 Background

Acoustics in a broad sense can be defined as dealing with sound both as a physical phenomenon and in terms of how it is perceived. The perceived quality of a sound, its timbre, is the foundational element of all other aspects of a sound, allowing listeners to distinguish between different types of musical instrument, phonemes in speech, and emotional responses to different genres of music.

One of the most common practical applications of acoustics is in architectural and environmental design. An example of this can be seen in the area of noise control, which involves dealing with a common problem in cities and other urban environments by using the core principles of the propagation and absorption of sound to inform decisions around architectural design and urban planning [1]. This area of acoustic engineering work draws from fundamental and objectively measurable properties of sound, such as pressure, energy density, and spectra [2]. With noise control, there exists a quantifiable measure of desirably perceived sound, the lower the sound level is from a source, the better. However, for more complex applications of acoustics this is not the case.

While many elements around how sound is perceived are derived from its innate physical properties; many elements are instead dictated by properties that are subjective, and that can vary to due a large range of factors. Objective elements of timbre are often described in relation to subjective observations, and a contemporary understanding of timbre draws

from the field of psychoacoustics [3]. Subjective perceptual elements of sound dictate many elements of practical acoustic design, a common example is through the design of concert halls, and how approaches have changed over the years. Halls are designed to facilitate the music performed in them, and induce desirable qualities in the performed sound in terms of how a listener hears it. This varies from space to space, and is dependent on many other factors like the genres of music the space will accommodate, or the type of instruments played within the space.

Audio engineering informs architectural design in order to induce desirable characteristics in these spaces while mitigating undesirable characteristics. In concert halls and arenas, one would want performances to be clearly heard by audience members throughout the spaces, as well as performers being clearly being able to hear themselves and other performers. In a voice booth one would want an environment with minimal noise and reflections in order to get dry and clean vocal takes. But beyond these assumptions, people desire many differing properties in these acoustic environments for a variety of different reasons.

Acoustic modelling software can help engineers design these spaces by allowing them to test different geometries and room configurations and observing how this changes properties of the resultantly generated sound, all without having to work in and modify a real space. In the field of acoustic modelling, software is often designed with experts in mind, the experts themselves being able to translate the intent of an end user into modifying elements of a an acoustic space and using the software as a tool to help with that.

In contrast, in field of music recording and editing, Digital Audio Workstations (DAWs) have developed a more user friendly approach to the user experience as they have iteratively evolved over the years. These programs will often contain resources and UI elements within the software itself to accommodate non experts, specifically for the primary purpose of a DAW; to facilitate editing, mixing, and mastering recorded or digitally produced music. Yet the majority of DAW software still operates through traditional interfaces, describing sound and signal processing through conventional means. Take for instance the default reverb effect in Ableton Live 9, shown in figure 1.1, which provides descriptions for each parameter as they are hovered over.

In recent times, as a result of the emergence of VR applications and a maturation of rendering techniques, there has been a renewed interest in modelling 3D audio environments,

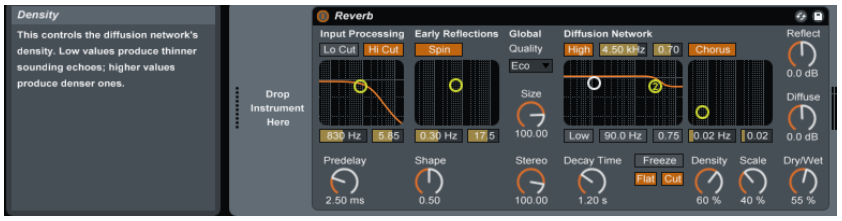

FIGURE 1.1: The default Ableton Live 9 reverb effect.

specifically in auralisation, the active rendering of a space that creates a 3D spatial audio impression for a user[4]. As the creative potential of 3D audio soundscapes becomes more apparent, more non-experts are becoming perceptive to how the acoustic properties of a virtual (or indeed real) environment can lead to emotional responses in listeners. However, in the current audio engineering landscape there are still roadblocks to further developments arpund these trends; and while there have been developments within the music industry around creating intuitive interfaces for DAWs to be used by non-expert musicians, there is much less research around equivalent solutions for acoustic modelling software.

A problem arises in making these types of software approachable in that how sound is conventionally and colloquially described is via common subjective terms like ‘bright‘ or ‘full‘, rather than technical descriptors. So when acoustic engineering work is contextualised in terms of an end user, and when a space is modified to their preference, there needs to be a way to translate the former into the latter. A way of assigning a meaning to an non-expert’s preference that can be applied in the context of a set of modifications within a space. Typically this is done via the consultation of an expert, but there is value in making this process more intuitive and giving users more agency by allowing them to tinker with the elements of the space themselves.

Subjective descriptors used when describing audio, like warmth of closeness, or harshness and so on, are often categorised as semantic descriptors [5]. Semantics as a field is the study of how meanings are assigned to concepts, and how concepts relate to other concepts. In regards to the context of this project, and important aspect of semantic research is that it allows the quantification of subjective data, data based on user preferences and emotive responses, which is the primary framework in which people discuss music and audio in a non-technical environment.

In order to facilitate effective interfaces for non-experts, there needs to be a process where the subjective desires of a user can be interpreted and acted on in terms of quantifiable

properties in a way that is intuative on the user end and effective on the application end. This can be abstracted as a simple black box model, shown in the figure below:

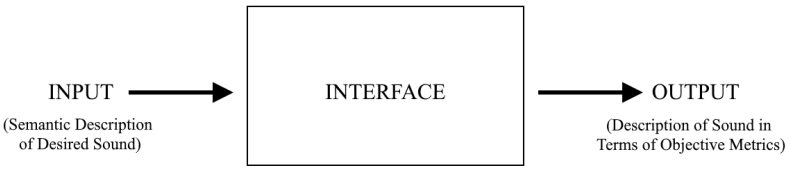

FIGURE 1.2: A simple abstraction of the hypothetical semantic interface in terms of a black box model.

There are complications that need to be resolved before this is possible. There needs to be significant research in order to create well defined meanings of semantic terms that are both commonly understood and useful in the context of this work. There also needs to be work in understanding how subjective response to sound within an acoustic space varies depending on the type of sound source; a melodic source may be described differently to a percussive source, or speech.

This research aims to develop processes that allow mathematical associations between the black box inputs (semantic terms) and outputs (practical modifications of the acoustic environment) of a hypothetical semantic interface. Hypothetically, this interface will allow a user to adjust the strength of a subjective parameter (brightness, as an example) and the acoustic model will modify elements within the space to facilitate that adjustment, so something playing within that space sounds more or less bright. This research work has the potential to be used to inform more automated and intelligent auralisation models in order to fit the subjective needs of end users; which in itself can inform the intelligent optimisation of acoustic environments for the same purpose.

While this specific project focuses on a set number of defined terms, future research could potentially apply these principles for a greater array of subjective descriptors of sound; or even through a sound reference being used as a point of comparison for the desired qualities of a room, with subjective timbral elements being disseminated through analysis and then applied to a room model in a similar fashion to this work.

1.2 Statement of Hypothesis

In referring to the wider topics that inform this work, the core hypothesis behind this investigation can be summarised as:

"Principles derived from semantic audio can be used to develop intuitive interfaces for auralisation models in order to fit the subjective needs of end users; which in itself can inform the intelligent optimisation of acoustic environments for the same purpose"

1.3 Thesis Structure

This section will give an overview of the structure of this thesis, outlining what chapters there are in the rest of this document, and detailing the content within each chapter.

Chapter 2, 'Literature Review' This chapter presents the current state of research and knowledge relating to the key areas of research being drawn from for this project. With this work aiming to investigate how the fields of auralisation and semantic audio relate to each other, this section will focus on key theoretical principles and notable breakthroughs in contemporary research for both of these areas.

Chapter 3, 'Project Design' This chapter contains a discussion about the steps taken to develop an appropriately bounded deliverable to work towards that is a worthy investigation of the key research questions relating to this project. In this section there is discussion about the perceptual factors that were selected for this work, how they were selected, and the importance of an explicit process of selection. There is also discussion of the iterative moves toward what the 'semantic interface' was defined as in this specific project. The aim is to provide context for the project as a whole and help inform further research around this field

Chapter 4, 'Auralisation Model Experiments' This chapter is a presentation of the series of experiments with room auralisations in ODEON software that aimed to develop associations between modifying variables in an acoustic space and the resultant changes in objective acoustic measurements of the space. This section is broken down into a series of individual experiments, each with data presented and a brief discussion. These experiments

were based around the same core approach, but the step by step methodology iterated over time as the complexities around the modelling software and the IR analysis tools became known.

Chapter 5 'Perceptual Acoustic Testing - Listening Test' This chapter focuses on an investigation of the associative relationships between objective acoustic measurement parameters and the relative strength of perceptual elements of sound, and how a listening test was designed in order to gather subjective perceptual data related to this task. In this section there is a discussion around the various existing methodologies for perceptual testing both in and out of the field of audio, and what elements from those are appropriate for the data needed for this specific task. In addition there is discussion around the finalised listening test methodology, and a presentation of and discussion around the results from this work.

Chapter 6, 'Conclusions and Further Work' This chapter draws from they key findings in the auralisation and listening test experiments to derive proportional representations of the associations between the defined subjective variables in this work and the defined room variables in this work. This chapter also outlines the key takeaways from the project as a whole, outlining the effectiveness of the methodologies used to develop the associations between semantic terms and room transformations, and how these techniques could be developed and expanded on in the future. The validity of the initial hypothesis is discussed in relation to the final results of the project There is also discussion around the wider applications of this work and how the broader concept can be potentially implemented in use cases of larger scope than described in this thesis.

Chapter 2

Literature Review

This chapter outlines the theoretical concepts that will be drawn from for this research work; auralisation, the assessment of acoustic spaces, audio applications of research related to perceptual and subjective data, and foundational aspects of perceptual sensory testing. A foundational theoretical background is provided for each area of research, followed by discussions around historical and contempary research in the field, with relevent and novel papers and commercial products being highlighted. This literature review covers the folowing:

Section 2.1 - Auralisation and Modelling Reverb in 3D Space. This section outlines the principles behind three dimensional modelling of sound, and the emergence of auralisation as a specific application of these principles. This section goes on to outline the early history of auralisation as a field and recent developments in rendering techniques, with descriptions of some applications of auralisation techniques in commercial software.

Section 2.2 - Reverberation Assessment of Acoustic Spaces This section firstly describes fundimental elements of acoustics that are key to how the acosutic elements of a space are commonly quantified and assessed by experts, with a discussion of where these frameworks for acoustic assessments come from and wider discussions in the field of how reliable these frameworks are. There are then descriptions of a variety fo case studies around the assessment of reverberation in acoustic spaces for a wide range of contexts, all of which draw from the frameworks outline sin the beginning of this section.

2.3 - Semantic Audio and Music Information Retrieval This section firstly describes the concept of semantics, and it's common application within the fields of information and data science. This section then outlines how these concepts have been applied in audio contexts, especially in the emerging field of 'Music Information Retrieval', with discussions around various case studies.

2.4 - Subjective and Objective Perceptual Sensory Testing This section outlines and discusses approaches various existing methodologies and approaches in assessing perceptual sensory factors, both in and outside the field of audio; with perceptual sensory factors being objective and subjective characerstics of stimuli that is processed via the human senses such as hearing or taste. This section discusses literature where sensory testing methods outside of an audio context are drawn from to inform the design of subjective audio assessment methods; as well as discussion around MUSHRA, a listening test methodology designed to parse obejctive qualitiative elements from various lossy encoding processes.

2.1 Auralisation and Modelling Reverb in 3D Space

2.1.1 Principles of Auralisation

Fundamentally, Auralisation is a form of acoustic modelling, representing an acoustic space in a virtual three dimensional environment. Auralisation was initially conceptualised as the audio equivalent of visualisation, which refers to rendering a visual representation of a real space virtually [6]. Whereas one can visualise a three dimensional space in the form of a 3D model, one can 'auralise' how a three dimensional space via it's 'soundfield' or 'soundscape', representing the acoustic qualities of a space and how that effects sound within it.

The basic principle of auralisation is based on the concept of an impulse response, impulse responses are generated by recording a brief excitation signal (the impulse) within a live space. The mathematical principle behind impulse responses is that when a impulse is inputted into a dynamic system then the resultant output contains information about the system, in this case the audio output of the recording of an impulse within a space contains acoustic information about that space. An example of an impulse response of a space in

the time domain is shown in the figure below, where one can observe the intial excitation and immediate decay:

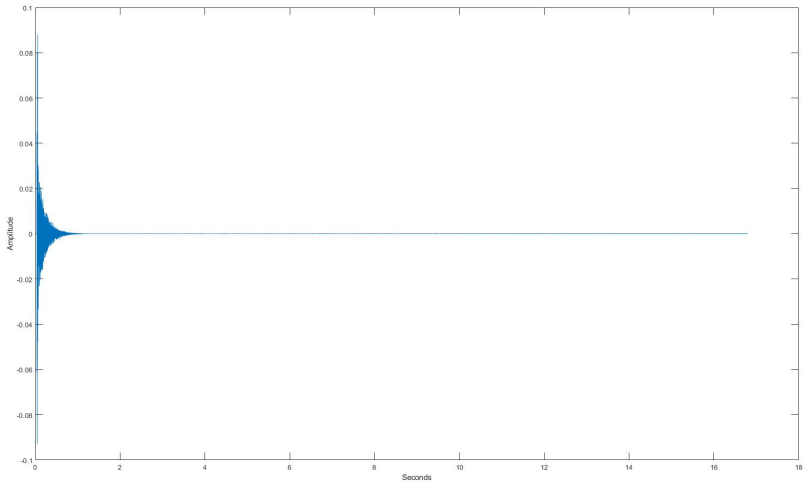

FIGURE 2.1: An Impulse Response in the Time Domain

The most basic abstraction of an auralisation system is a convolution of the impulse response of a space and a sound source. Convolution is a mathematical function that derives the integral of the end product of an input function and a second input function that has been reversed and shifted. Convolutions are used to show how one function is modified by another, in this case how the acoustic properties of a space (contained in it's impulse response) effect a sound source. This convolution can be written as shown below, where r(t) is the output signal, h(t) the impulse response and s(t) is the input signal in the time domain:

$$r(t) = h(t) * s(t) \tag{2.1}$$

This equation outlines the basic principle of auralisation, that it allows us hear how sound sources will sound in different acoustic environments via modelling techniques. The impulse response of a particular space is it's room impulse response (RIR). RIRs are measured by recording an exciation source, such as a handclap or a baloon pop, with an appropiate microphone setup, and example of this is shown in figure 2.2.

Practically, auralisation systems expand far beyond this definition; modelling how sound waves propagate through the air and reflect off surfaces. This is achieved via modelling the

FIGURE 2.2: A Practical Example of Recording an Impulse Response of the Inncocent Railway Tunnel [7]

paths of the reflected waves themselves, usually via either a ray tracing or wave tracing method; in addition to modelling the materials of the surfaces themselves and how much of a reflecting wave is absorbed, with a materials reflective properties often defined by their absorption coefficient. In an acoustic context, the absorption coefficient is effectively how much sound is reflected by a material vs how much sound is absorbed, acting as a scalar simplification of surface impedence (which consists of the magnitude and phase of the impedence). In actuality absorption coefficient is only one component of a materials overall attenuation coefficient, the other being its scattering coefficient. However, in this context absorption coefficient is used to quantify how materials differ from other materials, since while the scattering coefficient can be associated with the shape and patterning along the surface of the space, absorption is an innate property of the material itself.

In addition to this, auralisation modelling allows for the spatial representation of sound in relation to the user. Simple stereo arrangements are not sufficient for full three dimensional spatial audio, so binaural techniques are used. Binaural audio differs from standard audio in the fact that it renders audio in relation to the natural ear spacing of a user and the associated interaural time differences (ITDs) and interaural level differences (ILDs) related

to where a user is positioned within a virtual environment. Interaural time difference is the delay between sound reaching one ear and the other ear, while interaural level differences are the difference in amplitude and frequency between the sound reaching one ear vs the other. Both of these phenomena factor into how the human brain processes audio to create a three dimensional spatial impression, especially at close distances [8].

Practically, rendering binaural audio involves the used of a special series of impulse responses called head related transfer functions (HRTFs), which represent how the human ears perceive sound at a particular position around the head and are generated via recordings of dummy heads. convolving HRTFs with sound sources allows a user to feel full three dimensional spatial impression of a space via headphones or a suitable speaker arrangement, and pairing this with the convolution of the RIR of a space allows spatial impression of audio for a defined space and the users position within it.

2.1.2 Development of Auralisation Methodologies

Within the field of spatial audio modelling, auralisation is a newer and more novel approach to 3D rendering. Early work in spatial audio rendering was undertaken as a means to localise sound sources such as vocals and instruments within a conventional listening environment. The foundational principles of this spatial audio rendering work can be attributed to work in the early 1980s, a notable computer music journal paper from F. Richard Moore [9] lays out a conceptual model for spatial processing work to inform implementation in various sound synthesis programs. It outlines a geometrical acoustic environment, loudspeakers (sound sources) as objects within this environment; and then the simulation of sound propagation through the radiation of sound sources, early echo response, and the wider global response of the acoustic environment.

A key aspect of auralisation methods is how they model wave propagation. Wave field synthesis utilises an array of loudspeakers in order to generate an artificial wavefront emulating the simulated environment. It operates on the Huygens-Fresnel theorem that a waveform is the summation of a series of wavelets and as such a combination of these wavelets will reproduce the waveform; the loudspeaker array is designed around this principle, where each sound source models a spherical wavelet to produce a waveform emulating the source. Part of the appeal for wave field synthesis is that it is invariant to

where a listener is located in comparison to traditional stereo arrangements [10]. Spatial modelling of this nature was thought to be the next step in producing audio mixes that conveyed the nature of a recorded in environment in higher fidelity, but notably seemed to mainly apply this within traditional mixing contexts [10].

A greater understanding of spatial hearing, developments in 3D audio technologies [11] and the emergence of new media such as virtual reality all led to increasing focus in the modelling of 3D audio environments in the 1990s. Through this trend, ray tracing methods became much more common in spatial audio research. Contrasting with wave field synthesis, ray based approaches, mainly ray tracing, calculate the paths and reflections of individual rays. The origins of many ray tracing methods can be traced back to a paper from A. Krokstrad et al outlining the Ray Tracing Technique (RTT) to derive acoustic room response [12], where a wave is broken down into series of rays emanating from a single point and with a uniformly distributed three dimensional directional vectors in order to simulate a scattering effect; as the path of these rays are computed, reflections are calculated for every time a ray hits a surface. Another ray based approach is the image source technique, based on the idea that the specular reflections of a source can be generated my mirroring the original signal. However, there are problems with both of these methods as well; an RTT approach leads to less accuracy than an image source approach, while in terms of the number of calculations required the image source method scales poorly to more complex geometric shapes [13].

There are bottlenecks to the effectiveness of wave based rendering techniques in more complex modelling applications. For wave field synthesis, if a loudspeaker array has a non-uniform distribution then spatial aliasing occurs which has the potential of generating artefacts. Wave based techniques also scale poorly to more complex models, finite element method (FEM) and boundary element method (BEM) techniques require a much larger amount of elements for higher frequency applications [6]. Auralisation as a field has a fairly robust definition, but still broad enough to foster differing approaches to research. So while there have been cases where research has approached wave field based spatial modelling under the banner of 'auralisation' [14], in most cases wave field approaches have become less prominent in comparison to ray tracing based models.

One example of an early auralisation system architecture was the Digital Interactive Virtual Acoustics (DIVA) system, created in 1994 by the Helsinki University of Technology.

DIVA segments the overall task of auralisation into three phases; modelling the source either through defining an audio signal as an object or full synthesis (in addition to source direction), environment modelling or the modelling of acoustic spaces, and listener modelling which aims to produce a binaural environment via head related transfer functions (HRTFs). For each stage of auralisation the subject is defined, modelled and reproduced [13], an example of this is shown in the figure below.

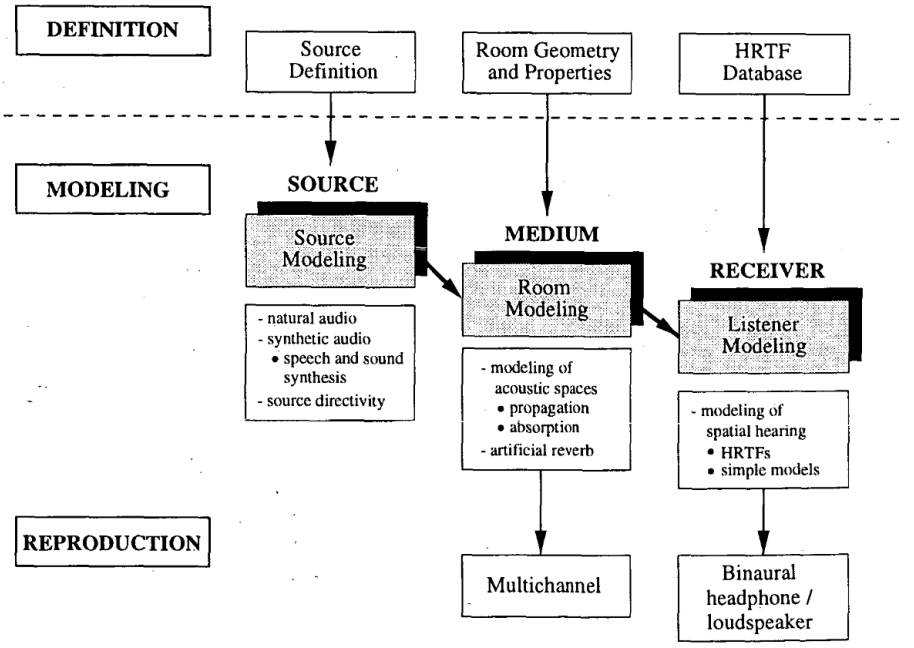

FIGURE 2.3: The three stage auralisation process outlined by DIVA [13]

While a typical approach to render an acoustic environment is to perform a convulsion of an audio signal with a rooms impulse response, the designers of DIVA found this to be limiting. A single impulse response does not contain real-time positional information, in order for the model to run in real-time the impulse response needed to be constantly recalculated, which proved taxing for the computers available at the time. The solution to this was to segment the overall reverb profile into real-time and non-real-time elements; with the direct source signal and early reflections being rendered in real-time and late reflections being artificially generated [15].

This case study is useful in providing an overview of how auralisation systems are typically laid out and the underlying thought processes behind the derivation of auralisation models. The overall aim is to model how a source is perceived by a listener as it travels through a

certain environment. It's no surprise that the methodology behind the design of the DIVA system is heavily cited in research on a wider variety of spatial audio applications.

2.1.3 Commercial Products and Emerging Research

Auralisation methodologies incorporating ray tracing techniques have existed in commercialised products for decades. ODEON is a commercial program that acts as a multipurpose acoustic simulation tool through versatile 3D auralisation modelling. Geometric information can be fed into the software through 3D drawings imported from CAD software, sources and receivers can be placed as objects in the environment via coordinates; and simulations can be run to measure single point response, multi point response, and grid response, in addition to more complex auralisation simulations. The origins of the software can be traced back to research from the Technical University of Denmark in 1991 [16], where the underlying methodology displayed similarities to the DIVA system, utilising a hybrid method derived from the ray tracing method and image source method. The paths of traced rays were calculated as they reflected off surfaces but only rays which contributed to the directional perception of the listener were accounted for in further reflection calculations, this allows for a trade-off between accuracy and computational cost[17].

Acoustic simulation software like ODEON has been implemented significantly in the field of audio engineering; in heritage acoustics, in the architectural design of buildings and rooms and informing work on music recording and mixing. The versatility of these programs is weighed with the fact that these are tools aimed at and marketed to audio engineering professionals with an understanding of the scientific theory behind sound propagation.

The continuing increase in available computational power has led to newer approaches towards faster and more accurate auralisation techniques being investigated. There have been investigations into the usage of neural networks to increase the speed of auralisation models [18], as well as revisiting wave based techniques to model complex geometry [19]. The nature of this research work is iterative, attempting to find novel variants of already existing and well documented techniques. Other research work focuses on case studies for large scale auralisation models, from heritage acoustics to the renovation of concert halls [20].

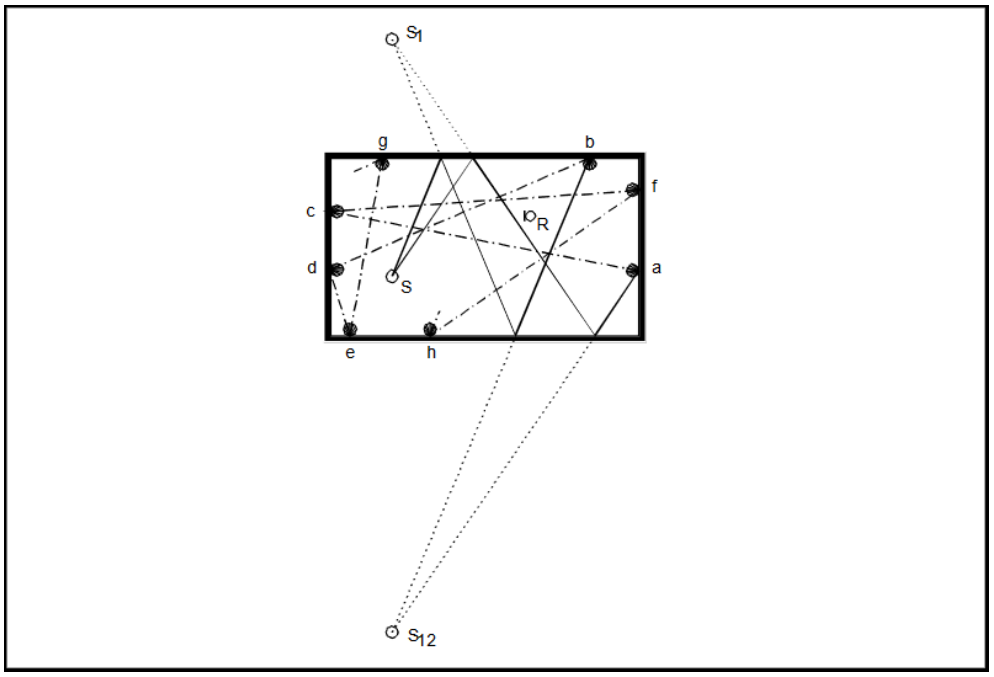

FIGURE 2.4: The 'hybrid model' used in ODEON, early reflections from ray tracing are considered as sources which emit their own waves [16]

2.2 Reverberation Assessment of Acoustic Spaces

2.2.1 Principles of Room Acoustics

This research work focuses on a series of environmental acoustic properties such as early decay time (EDT), T30 reverberation time, and clarity (C80); all defined in ISO 3382-1 as key parameters in defining the distinct acoustic qualities of a particular space. These factors act as descriptors for fundamental acoustic properties relating to the treatment of spaces.

The innate acoustic qualities of a room can be codified in terms of its room impulse response (RIR), and so in practice the work in measuring the aforementioned acoustic parameters will focus on generating a RIR within a real space. Practically, the impulse of an impulse response is approximated via a short excitation from a sound source, such as a hand clap or a pistol shot. A trending line of the decay following the initial excitation of an impulse response can be derived via an inverse integral of the squared impulse response within the time domain, this is known as the energy decay curve (also known as the schroder curve) and is used to derive values of reverberation time.

EDT and T30 are two separate measurements of reverberation time via the energy decay curve, with EDT being the time period of decay between 0 and -10dB and T30 being a measurement of the decay between -5 and -35dB. These measurements indicate different acoustic qualities of an impulse response, EDT is often described as being a measure of perceived reverberance, while T30 acts as a measure for the physical properties of a space.

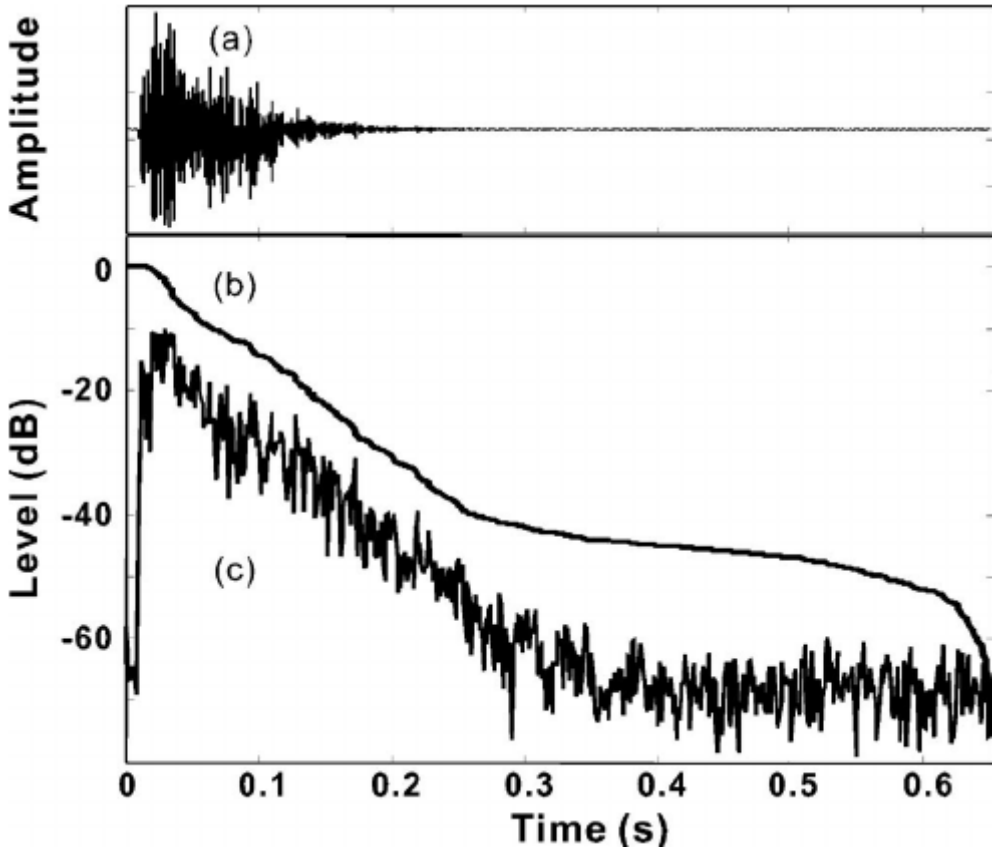

FIGURE 2.5: An example of the time domain of a impulse response (a) and it's associated energy decay curve (b) as well as it's energy time curve (c) [21]

Clarity is defined in acoustic terms as the measurement of the ratio between early and late arriving energy. The energy of an impulse response can be calculated as the integral of the energy decay curve, with the early time limit t_e being the cut-off point for early arriving energy and the start point for late arriving energy. It can be described as shown, with t being time and p being the measurement of acoustic pressure:

$$\frac{\int_0^{t_e} p^2\,dt}{\int_{t_e}^{\infty} p^2\,dt} \tag{2.2}$$

C80 is a calculation of clarity where t_e is 80ms, it is often defined as a good measure for musical clarity rather than speech clarity, which is often quantified in terms of C50 (where t_e is instead 50ms).

Clarity is both an acoustic measurement of early arriving vs late arriving energy, one which hypothetically would be effected by distance, and is also a term intuitively associated with how loud or close something is relative to a listener.

This variable is an abstraction of the practical surface treatment work done within acoustic spaces, which focus on mitigating or extending reflections of sound sources off those surfaces, these reflections defining the reverb of a space. Therefore it can be deduced that the perceived reverb of a space can be associated with absorption coefficient, this perceived reverb is often linked with the T30 measurement.

2.2.2 The ISO 33382-1 Standard

The official standard for acoustic space measurement, ISO 3382-1, is often referenced as the base guideline for approaches to this topic [22]. The standard lays out that the perception of an acoustic spaces can be codified as a collection of key parameters representing four types of measure; decay times, sound strength, clarity, and spatial impression.

- The decay time is represented by T60, and the early decay time (EDT) both of which are derived from the time decay curve of an acoustic measurement. They are the most frequently cited parameters in this field of research in terms of deriving the characteristics of an acoustic space. EDT and T60 typically correlate with each other.
- Sound strength is represented by its own measurement (G), and is defined as the difference between sound pressure at the measured position and at a 10m distance away in free space [23]. Research has indicated the usefulness of having defined Gearly and Glate measurements [24].
- Clarity measures are represented by C50 and C80 clarity values, which themselves are derived from the definition (D) measurement, as well as the centre time (TS). Definition is the energy ratio of early reflections to total energy in the first 50ms of

Type of Measure	Measures	Notes
Decay times	T60, reverberation time	Physically important
	EDT, Early decay time	Subjectively important
Sound strength	G, Strength	Hall effect on sound levels
Clarity measures	D50, Definition	Clarity of speech
	C50, Clarity	
	C80, Clarity	Clarity of music
	Ts, Centre time	
Spatial impression	Lfearly, Early lateral energy fraction	Apparent source width
	IACCearly, Early inter-aural cross correlation	
	GLL, Late lateral sound level	Listner envelopment

TABLE 2.1: An overview of key measurable parameters according to standard ISO 3382-1 [22]

sound transmission. Clarity parameters are logarithmic expressions of the energy ratio between early reflections and late reflections, and spatial time is defined as the centre of gravity of an impulse response.

- Spatial impression is represented by early lateral energy fraction (LFearly), early interaural cross correlation (IACCearly) and the late lateral sound level (GLL). Research has indicated a listeners perceived envelopment is formed of a envelopment component represented by the lateral sound level being in addition to a spatial width element represented by the early lateral energy fraction [25].

A summary of these parameters and their groupings is shown in in table 2.1.

The standard is the encapsulation of decades and even centuries of room acoustics research, preliminary work can be traced back to the 1970s, where researchers investigated how these parameters can provide a quantitative profile of acoustic spaces [26].

While ISO 3382-1 has been a very prominent and widely referenced standard within the field of room acoustics, there has also been lots of discussion about potential modifications to the standard, with a notable focus on the viability of measuring its stated parameters and how measurements can lose coherence in various environments. There are concerns about the variance in the validity of measurements depending on frequencies [22], as well as suggestions for noise compensation techniques to be implemented in the standard [27]. International standards encapsulate the wider research work around their constituent topics,

so reading these discussions can help gleam insight to the wider questions surrounding acoustic measurement; while research has mostly coalesced into considering a few key parameters for investigation, measuring reverberation is a complex and multifaceted process, and there are deviations regarding the appropriate scope and focus of these core measurements as well as what parameters should be measured.

2.2.3 Subjective Assessments of Acoustic Environments

A 2014 article from Taplo Lokki published in Physics Today outlines an investigation into the sensory evaluation of concert halls through listener feedback; drawing from analogies to wine tasting [28], the article proposes a methodology of the subject analysis of acoustics based on consensus vocabulary profiling, hierarchical clustering, and multi factor analysis. The investigation focused on the assessment of 20 concert halls from 20 listeners through the virtualisation of these halls via a virtual orchestra of 34 loudspeakers [29]. Consensus vocabulary profiling is where assessors define their own terms for how to describe their perception of stimuli, in this case the characteristics of sound within a room; this method is ubiquitous in research relating to food and drink taste testing [30]. These terms were then placed into common groups via hierarchical clustering; in this case 102 attributes were grouped under seven categories: definition, clarity, reverberance, loudness, envelopment, width of sound (bassiness), and proximity. The groupings of these 102 attributes is shown below in figure 2.6 (note the presence of an 'ungrouped' cluster). From this, data could be then be evaluated and interpreted in many novel ways.

Group	Individual attributes (translated to English)	N
Reverberance_1 (size of the space)	reverberance (X41), reverberant (X77), reverb (X34), sonority (X103), amount of reverb (X94), drr (X60), size of the space (X105)	7
Reverberance_2 (envelopment)	reverberance (X26, X3, X67, X86), reverberation (X50), broadness (X55), reverb (X106), envelopment (X61), width (X46), emphasis on bass (X5)	10
Width of Sound (bass)	width of sound (X39), wide (X13), wideness (X95, X80), width (X92), sense of space (X10), 3-dimensional (X20), focused sound (X107), envelopment (X83), naturalness (x7), bass (X109), balance between warm and cold (X71), amount of bass (X99)	13
Loudness	loudness (X37, X2, X43, X96, X69), full-flavored (X8, X85), dynamics (X57), volume (X47), approach of sound (X91)	10
Distance	distance (X82, X24, X28, X48, X44, X88, X100, X108, X97), distant (X76), closeness (X18, X65)	12
Ungrouped	spread of sound (X17), breadth (X74), neutral (X78), brightness (X66), liveness (X64), muddy (X98), stand out (X9), intimacy (X90), eq (X62), sharpness (X104), width of sound (X23)	11
Balance	balance (X31), directed (X52), symmetry (X11), brightness (X38, X36), balanced (X6, X111), clearness (X16)	8
Openness	soulless (X15), naturalness (X14, X73), openness (X84), depth (X70), clearness (X30, X75, X89), pronounced (X79), presence (X81), definition (X87), discrimination (X40), distance of source (X32), intensity (X72), closeness (X4, X54)	16
Definition (separability, clarity)	definition (X27, X35, X102, X53), distinctness (X59), clarity (X58), localizability (X63, X101), treble (X110), transparency (X22), tone color (X56, X33), precise (X12), softness (X42), texture (X19)	15

FIGURE 2.6: Cluster Groups for 102 defined attributes for acoustic spaces. [29]

The paper shows the comparison of three of the demonstrated halls via a hexagonal sensory profile, with an attribute cluster at each point; within each hall various different measured points in the environment were mapped onto these profiles, in some halls there was extreme variance in sensory profiles depending on where one was listening, in others there was little to no difference. Through multiple factor analysis it was also found that the attributes in the proximity and loudness clusters accounted for more than 50 percent of variance regarding the evaluation of the sampled environments, in other words the most important factors behind the subjective 'enjoyment' of an acoustic space. A notable takeaway from this is that the loudspeaker array was recorded at a set distance of 12m for every virtualised hall, and yet the perceived distance varied for each hall, a separate investigation [31] from the same author found that this observation may be due to reflections from various surfaces, indicating that different acoustic sensations could be induced within the same environment with changes to how the environment reflected sound.

The methodology of the investigation itself and the clustering and multi-dimensional analysis techniques used in analysing the results evoke semantic audio research, providing credence to the fact that outside of heavily publicised research initiatives there has been a long line of research with audio engineering that has incorporated and focused on semantic elements; especially in subjective perception of acoustic spaces.

Another investigation into the subjective perception of acoustic spaces focussed on virtual acoustics, the usage of speaker arrays to simulate an acoustic environment in a live space [32], not just as a means for testing but as a system to aid performers. What's notable is that the scope of these investigations focused on listening related to performance, how easy it was for performers to play instruments amongst an ensemble in these virtual environments and the 'naturalness' of the virtual environment; this work is in contrast with many other similar papers which focus on acoustic quality from an audience perspective. The experiment focused on evaluating three pre-defined acoustic conditions; with independent parameters T30, C80, stage support, IACC and LF being used. Participants were asked to answer questions pertaining to subjective quality. For analysis purposes the questions were grouped into three loose categories; 'spatial impression', 'stage support and clarity', and 'tonal balance'. Multiple regression analysis demonstrated that spatial impression was the most pivotal factor behind the qualitative assessment of the system, variances in the other two categories made little to no difference in participants overall satisfaction.

Conclusions were drawn that from a performers perspective, 'naturalness', 'ease of hearing each other' and 'height sensation' were the most important factors for a satisfactory live acoustic environment. There is some overlap with the methodology behind this experiment and related ones, with common qualities such as 'clarity', 'envelopment', and the level of reverberation being identified.

2.3 Semantic Audio and Music Information Retrieval

2.3.1 Principles of Semantic Data

To understand what semantic audio means within the context of this work one must first understand the wider field of semantic data. Within the field of software engineering, a data model is an organised, quantitative representation of a various assemblage of related data, a means of abstraction to give data a defined structure that allows for more efficient analysis [33]. Computational semantics is a field of research within computer science that aims to provide representations and associated meaning of human language expressions in a machine readable form in order to facilitate automated processes [34]. A semantic network is a formal representation of how various objects, abstract concepts, and other forms of data are associated with each other through a uniform knowledge representation. In a broad sense, semantic modelling can be seen as the formulation of a common vocabulary for complex, interconnected processes.

A semantic network is a graphical network formed of nodes and connecting vertices, with nodes representing a concept and connections representing semantic relations [35]. The collective semantic relations for a given object define a concept's ontology; within information science the ontology of a concept is its wider relevant purpose within a certain frame of context. A greater amount associations between fields leads to more well defined ontologies. The most common semantic association is a semantic triple [36], where a subject and an object are linked via a predicate. An example would be 'A dog is a type of animal' or 'sodium is an element', and a network of these triples defines objects in multiple dimensions and allows the usage of this data in more complex processes, often with a focus on machine learning applications.

The broad field of semantics has wider applications beyond information science. In studies relating to various elements of human perception and subjective interpretation, subjective descriptors are often assigned a meaning in relation to the stimuli being observed. For example, taste and food perception is often described as highly variant and subjective; dependent on a range of subjective sensory, psychological, and cultural factors [37]. This is a wider application of the key principle of relating concepts (processed stimuli) with other concepts (subjective descriptors) as a representation of data, and as such these subjective descriptors are often described as 'semantic' terms. Audio is a form of stimuli with quantifiable properties inherent to it that nonetheless induces subjective responses through the perception of it by a user. As such, there is the potential for principles inm computational semantics to have useful applications in the field of audio.

2.3.2 Semantic Approaches to Digital Reverb Effect Design

A common element of both the subjective analysis of acoustic spaces and music information retrieval for digital audio effects is that they both involve the assessment of reverb; quantifying and emulating a natural reverb profile for the former and transforming a recorded or synthesised signal in order to achieve a desired effect in the latter. For the purposes of this project, an investigative focus on digital reverb effects in comparison with equalisation, distortion and compression effects is most valuable in providing insight for further work.

A notable aspect of research in this field is while there has been work undertaken into the automation of fading, EQ, and panning effects through machine learning related strategies; there has been relatively little work towards automating reverb effects. Researchers have suggested that this is due to the fact that while most digital effects are based on common principles, the architectures of various reverb effects differ significant in terms of design and are based off a much wider common pool of modifiable parameters, making applying work done with one type of reverb effect harder to apply to another one [38].

Ircam's Spatialisateur (SPAT) software is a primary example of commercially available digital reverb effect that has built in 'semantic' functionality [39]. SPAT was initially designed as a MAX/MSP plugin in the early 1990's and eventually evolved into a framework that accommodates a broad range of use cases; with the modularity of its core software

meaning that it can be used in live, mixing, and modelling contexts. SPAT can be described as an algorithmic reverberation engine with a modifiable user interface on top of it, notably these processing and interfacing layers are stored in separate libraries, and operate independently from MAX/MSP, enabling interfacing with external software such as VSTs or MATLAB and increasing the versatility of the effect.

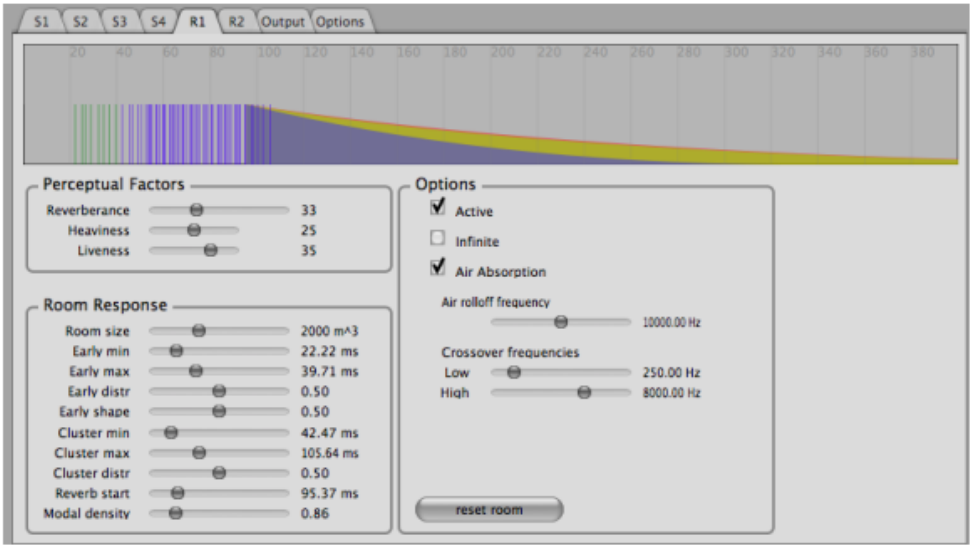

FIGURE 2.7: The reverberation module for Ircam SPAT [39]

While the full specification of the reverberation engine is both unlikely to be available for public reading and out of the scope of this literature review, it is useful to envision a simplified of this engine as the summation of four distinct phases of reverberated sound. The original signal, early reflections, late reflections, and the long reverberation tail. This is congruent with the broad terms of how reverb measurement and modelling are described. These phases are each generated through a combination of delay chains, feedback loops and filters, fundamental elements of reverb effect design. From here various methods, such as higher order ambisonics and binaural synthesis, are used to pan the reverberated signal in order to achieve a 3D sound field.

SPAT is unique as a reverb effect due to the fact that its primary interface elements were designed with perceptual features derived from psychoacoustic research into the subjective experiences of acoustic spaces. These features are combination of source characteristics, room characteristics, and characteristics attributed to the interfacing between the source and room, shown in table 1. Each characteristic has a prescribed parameter attached to it which can be modified not unlike on a typical VST GUI.

SPAT is informed by principles related to both auralisation and the subjective perception of acoustic spaces, and as such stands out as a digital reverb effect that aims to convey the characteristics of an acoustic environment in the way that the listener perceives it . In this sense it forms the bridge between work done in live spaces and digital recording spaces, but nonetheless is designed to act as a standard audio effect.

There has also been work done in creating a methodology to derive perceptual descriptors for reverb effects that maps toward broader reverberation terminology rather than specific parameters in order for it to be potentially applicable to any reverb effect [40]. This example, in comparison to SPAT defines descriptors in broader terms (bright, clear, boomy, bathroom-like, and church-like) which themselves are mapped onto a complex combination of multiple parameters, such as echo density, clarity, central time, and spectral centroid. The experiment outlined in the associated paper involved surveying participants who ranked a training set of 60 audio clips modified by an impulse function in terms of how much they captured the outlined perceptual descriptors; then a truncated version of the training set was modified by a low level parameter, one at a time, with users assessing results in the same way. These results were mapped onto control sliders for each perceptual descriptor, with these sliders then being assessed for accuracy by participants.

2.3.3 Music Information Retrieval

In recent decades there has been an renewed interest in how semantic data based technologies and principles can be applied in the music industry, specifically within the field of music information retrieval (MIR). MIR draws from and related to, but is ultimately distinct from audio engineering and acoustics. One can define MIR as data science towards the formation of knowledge about how music is performed, recorded, heard, and classified in a social and psychological context [41]. The International Society for Music Information Retrieval (ISMIR) is the most prominent source of music information retrieval research within academia; labelling themselves 'the world's leading research forum on processing, searching, organising and accessing music-related data'. The nature of their work, especially regarding data processing for front facing digital music storefronts and consumer profiling, has led to heightened interest from major players in the wider music industry. Sponsors for their 2020 conference included Spotify, Sony, Adobe, Google, and Dolby, amongst other parties [42].

In evaluating the proceedings of recent ISMIR conferences there are some distinct and continuing trends in research work. Research related to AI generated music, both in machine learning methodologies of the generation of music itself and in parsing out perceivable features of music, is a persistent theme. On the other end of the spectrum there has also been significant research in the ordered classification of music for marketing and playlisting purposes. One paper from the 2019 ISMIR proceedings [43] presents investigations into making machine generated audio content; using multiplicative long term short memory neural networks with adjustable weights of various emotive modifiers, such as a scale of valence from negative to positive, music can be generated that fits a certain intentional sentiment. Another paper outlined a methodology using neural networks to map perceptual mid-level features (dissonance, articulation, rhythmic complexity, etc.) to emotional responses (fear, happiness, anger etc.) [44]; a pairwise correlation matrix is shown in figure 2.8.

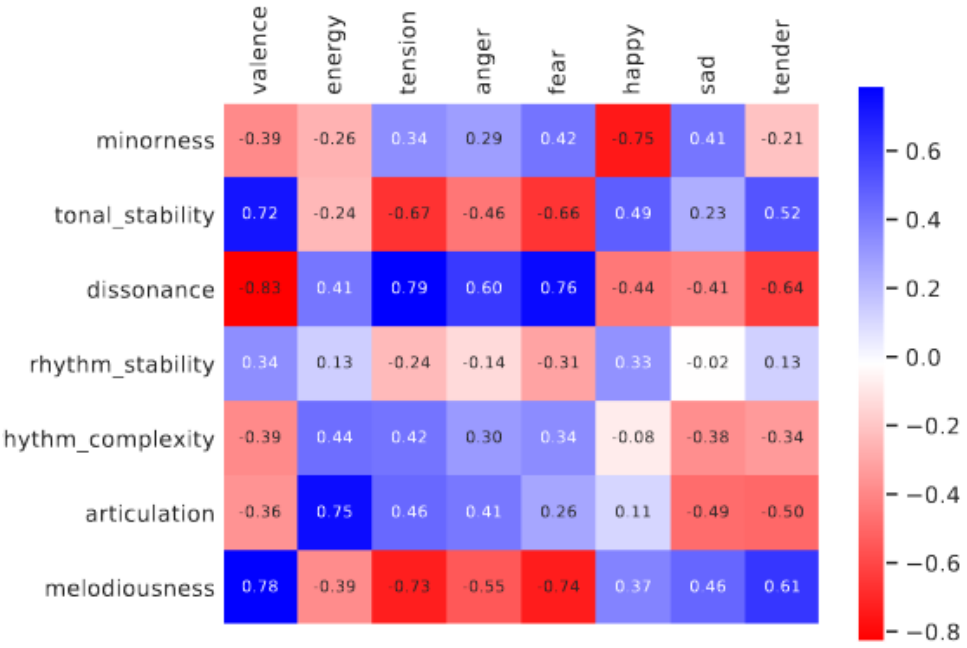

FIGURE 2.8: Associative matrix between mid-level music features and high level emotive responses [44]

These two investigations are indicative of the typical research work output from ISMIR; a common focus on the application of machine learning techniques, in this case deep learning neural networks, in order to derive associations between high level perceptive terminology and mid-level music features within a certain context. Within ISMIR there is a focus on MIR related to musical performance and the playback of recorded music, and while there is

a variety of work that explores a diverse range of fields within this scope there is very little output from outside that scope. For instance, there is very little work on MIR relating to acoustic environments, and while there is a significant amount of work dedicated to automated music generation there is less work than expected on how machine learning methods can inform music recording and mixing, with little work on low level audio effects.

2.3.4 Semantic Audio - The FAST Project

FAST was a five year research project for the EPSRC from Queen Mary University of London launched in 2014. it's overarching aim was the investigation of practical applications of semantic modelling principles within the music industry, specifically in regards to the pipeline of music production from performer to producer to listener [45]. The project aimed to use techniques analogous to the semantic web and related technologies. The impetus behind the project was that the music industry has historically been late to developments in digital technology, and that an investment into the development of semantic knowledge and increased automation will provide benefits to both producers and consumers.

A key element of FAST research work is in the development of various audio ontologies. These ontologies were focused on various elements of music production and consumption, like 'musical performance' or 'digital effects'. Ontologies have been developed for audio features, the properties of musical instruments, the studio workflow, and the properties of recorded music. These ontologies interface with each other to form a wider knowledge representation of musical data, the overall high level semantic model of these related ontologies is known as a whole as the music ontology [46], shown in figure 2.9:

In these ontologies, various concepts are modelled as classes and subjective phenomena are modelled as properties that classes are associated with. For example, in the audio features ontology; there is a class of *PersonSpeaking* and a property of *EmotionalIntensity*, meaning that a certain quantitative value of emotional intensity can be ascribed to a certain segment of a person speaking. The class of *PersonSpeaking* is itself a sub class of *SpeechSegment* which is a sub class of *StructuralSegment*, and so on. Ontologies factor into other ontologies as well, the *structure* class in the audio feature ontology is a sub class of the event *class* which exists in its own 'event' ontology.

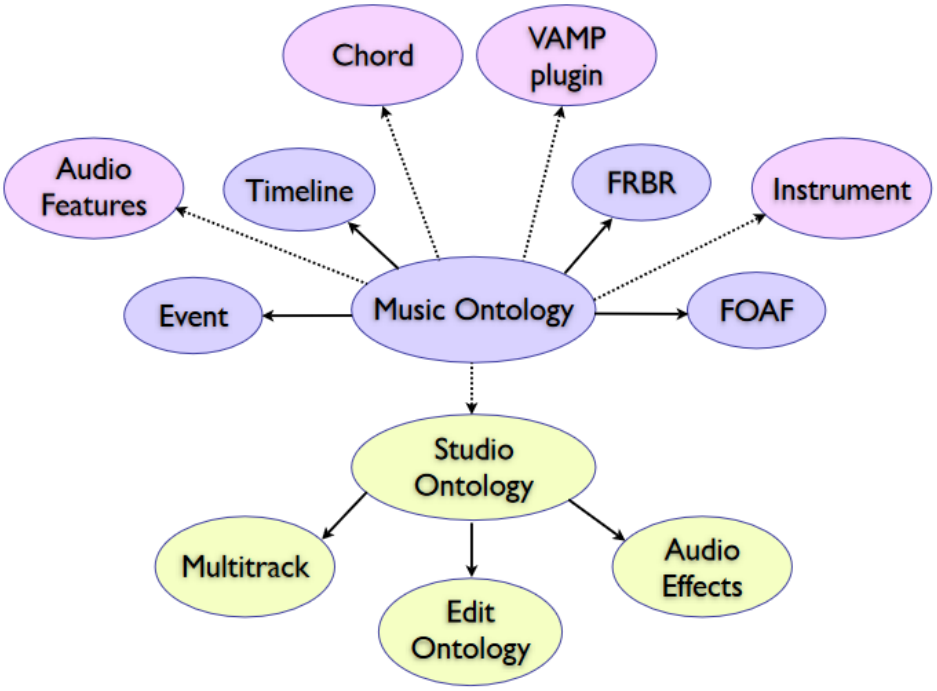

FIGURE 2.9: Semantic Web outlining the relations between various ontologies created by the FAST project [46]

The majority of published papers and tools emerging from this project are based off these ontologies. The project generated 120 published research papers, and most papers involve the implementation of these ontologies or related semantic web technologies in conventional MIR topics. In comparison to research found in ISMIR conferences, there is a greater amount of research output relating to low level audio transformations, specifically towards semantic interpretations of transformations via audio effects used in digital audio workstations. Within the FAST project, low level transformations are defined as fitting into one of four types; compression, distortion, equalisation, and reverberation. The Audio Effect Ontology (AUFX-O) is a representation of a typical production workflow in regards to standard audio effects used in music production[47]. A comparison between the layers of abstraction for AUFX-O and another semantic model (Functional Requirements for Bibliographic Records) is shown in figure 7, AUFX-O focuses on associations between user inputs on the implementation level and low level transformations on the device level.

A related project is the Semantic Audio Feature Extraction (SAFE) project, which functions as the implementation of AUFX-O within a DAW environment in order to parse semantic descriptors of timbral changes induced by audio effects to inform development of digital

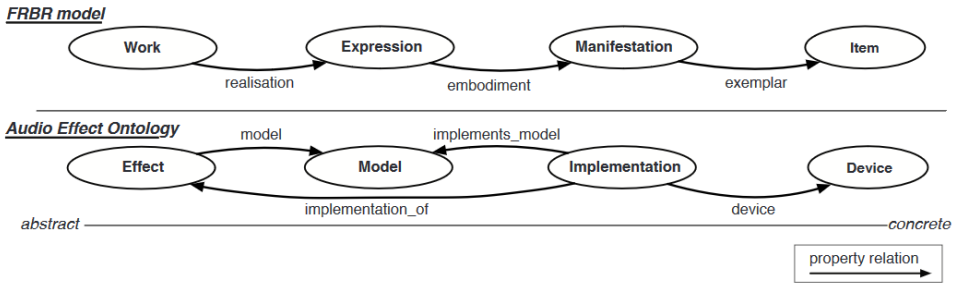

FIGURE 2.10: A comparison between abstraction layers in AUFX-O and FRBR [47]

audio effects with semantic interfaces [48]. SAFE currently consists of bespoke distortion, EQ, compression and reverb VST plug ins. With these effects users can enter timbral descriptors for the audio they are working with, which gets uploaded to global database, from this database presets are derived for a bank of descriptive terms which can then be loaded by the user; as an iterative process they more data they collect the more fleshed out these plugins will be, but a key thing to note is that these plugins are static, which is to say that there is no way to change the intensity of perceptual factor, or work with two factors at the same time. These plugins seemingly associate all of their modifiable parameters to a single term, effectively acting as a more user friendly approach to presents in conventional digital effects, while this helps in bridging the knowledge gap between experts and non-experts, in their current state these plugins do not contain significant elements of automation.

2.4 Subjective and Objective Perceptual Sensory Testing

2.4.1 Subjective Testing Approaches - Food Testing Methods

Existing literature around the subjective analysis of acoustic reverb often draws from work done in adjacent sensory analysis fields to develop methodologies and frameworks for analysing subjective perceptual data. Taste testing for food and drink is commonly referenced as a useful analogue to draw from, since the emotive descriptors used within the field share similarities to how audio is perceived by listeners [49]. There have been techniques developed around the assessment of taste via perceptual testing, where a series of stimuli are evaluated in relation to the strength of defined perceptual descriptors, the 'Sensory evaluation of food: principles and practices' book by Harry T. Lawless and

Hildegarde Heymann [50] has been heavily referenced in literature the development of listening tests and other assessments of audio. An investigation into biases in modern audio quality testing drew from similar work in the food industry outlining biases in food quality assessment procedures [51], for example.

Wine tasting perceptual test methodologies prioritise participants being able to sample all of the stimuli at the same time in the same environment [49], in this case being able to taste all the wines in the same environment and whenever they please during the test. Acoustic modelling is a crucial aspect in taking these principles and applying them in the assessment of live spaces. Speaker arrays have been used to create elaborate virtual orchestration environments for perceptual testing [52]; in lieu of physically travelling between live environments, untenable for the practical implementation of a listening test, these speaker arrays can provide the characteristics of sound played in various live spaces through acoustic modelling instead.

2.4.2 MUSHRA - A Standardised Testing Approach

There have been many standardised methods for the assessments of perceived sound in various contexts. One area in which perceptual assessment have become crucial is the evaluation of lossy audio codecs, which are used to quantify the effectiveness of audio compression techniques in retaining sound quality [53]. One methodology that is prominent within this space is the 'Multiple Stimuli with Hidden Reference and Anchor' or MUSHRA test. The MUSHRA test is a double blind listening test, meaning neither the participant or the tester are aware of the specific set of stimuli being assessed at a given time, this is normally achieved through some level of randomisation.

In a MUSHRA test, a participant evaluates a series of stimuli in comparison to a reference, scoring them from 0 to 100 for the quality of the sound, along an ordinal scale of categories (bad, poor, fair, good, and excellent). An example of a MUSHRA test question is shown in figure 2.11:

A core aspect of MUSHRA is the concept of an 'anchor', in the context of MUSHRA an anchor is a duplicate of the original source that the other samples are being compared with that is placed along with said samples in order to evaluate a participants capacity to discern artefacts within the tested samples. Ideally the result for an anchor within a question

FIGURE 2.11: An example of a MUSHRA question displayed on a computer panel [54]

should be zero, as there is no difference between it and the reference. MUSHRA tests use anchors due to the low level of variance in impairments between samples, where evaluating the anchor can inform how valid the other results are. In a MUSHRA test, the anchor is hidden, meaning the existence of a duplicate source is not stated to participants. The idea behind using an anchor in these tests is to evaluate the effectiveness of a participants results, if the anchor results are beyond a certain difference threshold then all of the participants answers are discounted.

MUSHRA is a standardised methodology, with specific recommendation relating to the context of assessing impairments in lossy codecs [54]. MUSHRA relates to objective measurements of sound, and the observation of those objective metrics by experts. It is specifically designed around being valuable with relatively fewer participants than other methodologies due to the fact that it is designed around expert participants. Standards state that no more than 20 participants are required for effective conclusions to be drawn from results, the appropriate amount of participants van be determined via the expected variance and resolution of the listening test's scoring system.

Although MUSHRA has been a prominent methodology for its specific context; an API based off the method, webMUSHRA, has emerged as an approach to designing online based listening tests. It was within this context that the MUSHRA methodology was investigated for this work.

Chapter 3

Project Design

The aim of the project design is to provide a proof of concept demonstration within a restrained scope of the methodologies that can be implemented to assess the semantic relationships between elements of room acoustics and subjective descriptors used for sound. The project design consists of a series of tests that will hypothetically generate quantitative data that can be fed into the hypothetical semantic interfaces that this project as a whole is investigating. Said constraints will be related to the amount of room acoustic elements and perceptual terms that will be assessed in this project work and the complexity of the testing methodologies implemented, as elaborated in this chapter the scale of these aspects are reduced due to time limitations.

This section outlines the early stages of the project, which involved setting an appropriate scale and scope for this work, and outlining the specifics of the rest of the project. This section tackles various elements of this work in detail, with discussions about how they were decided upon and how these elements can be potentially expanded upon in future research to develop more full realised and complex semantic interfaces. This chapter contains sections covering the following:

Section 3.1, Deriving a Conceptual Representation of the Proposed Model. This section describes the development of a conceptual representation of hypothetical semantic interface functionality; this abstraction of the flow of data within said interface formed the basis for the development of the rest of the project.

Section 3.2, Selecting Subjective Perceptual Factors. This section outlines the work involved in selecting the descriptive terms used in this project, which involved work towards a wider understanding of what descriptive terms are appropriate for this line of work, and how increasing the amount of terms increases the complexity of a semantic interface. This wider work is discussed in the context of it's usefulness for further work in this field.

Section 3.3, Selecting Physical Properties of Acoustic Environments. This section discusses the selection process for what modifiable elements of an acoustic space were to be focused on in this project. This involved developing an understanding around how people practically treat acoustic environments within the context of the potential use cases of this work.

3.1 Deriving a Conceptual Representation of the Proposed Model

The wider context behind this work was moving towards the development of a proof-of-concept semantic interface designed to be used with ODEON software; the foundations of which would be a mathematical model that took user inputs and provided outputs in the form of modifications of the acoustic space. The idea was that the mathematical model would provide a basis for an interface where a user would be able to modify a series of independent scalar perceptive values in order to produce desired characteristics for a set sound source within in the defined space via the physical modification of said space. Practically this model would be implemented by taking outputs generated from the interface and inputting them manually into the software itself.

Three schema design is used to build information management systems and is a key element of many formative semantic data frameworks, including the widely referenced Semantic Data Model (SDM) [55]. Three-schema concept architecture consists of an external, conceptual, and internal layer. The internal layer consists of how data is stored within the database, and the external model is how an end user perceives that data, the conceptual layer is the interface between the internal and external layers [56]. Examples of three schema conceptual arrangements are shown in figure 3.1; which map out the concepts

of auralisation, semantic data, and music information retrieval outlined in the previous section in the form of a three schema layout:

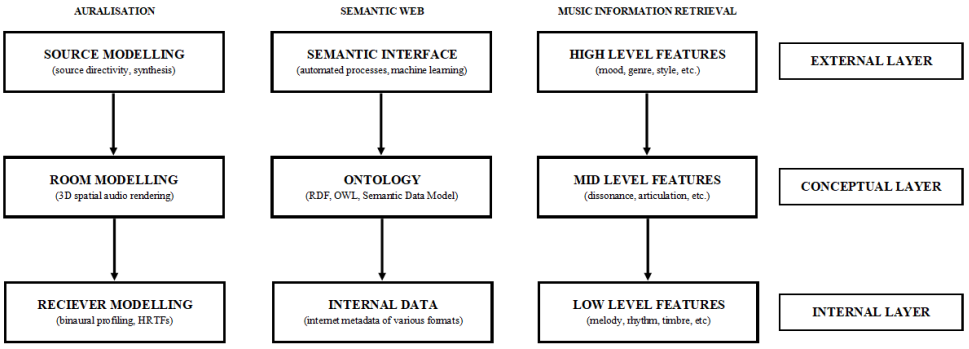

FIGURE 3.1: Typical three schema conceptual arrangements [56]

Drawing from these principles, the proposed mathematical model can be abstracted as facilitating flow of data from a user input to outputs into auralisation software via a three schema concept architecture, shown in figure 3.2:

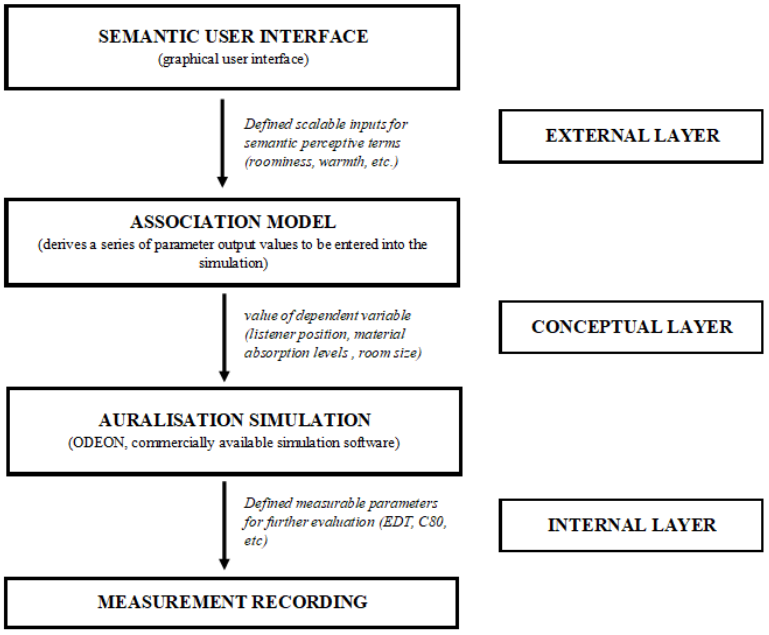

FIGURE 3.2: A three Schema Conceptual Representation of the proposed model

This research work focused on the association model, a means of defining terms from the external layer (perceptual terms within the semantic interface) in the form of variables within the conceptual layer (variable physical properties of an acoustic environment). In order for this to happen, the variables within the external layer and conceptual layer needed

to be decided upon. The aim was to describe perceptual terms in the form of a series of physical properties. The hypothetical ideal of this type of association can be described in the form a generic equation displayed below, where z is a perceptual term, x and y are physical elements of the space, and a and b are weighting factors:

$$z \propto ax + by \tag{3.1}$$

This generic equation outlines that the resultant associative equations derived in this work are designed to outline the type and scale of proportionality for each variable rather than a numeric relationship between the perceptual term and every variable. In other words, this work aims to outline whether a observable proportionality exists at all, whether it is positive or negative, and whether it is direct or inverse for associations between each independent variable (perceptual factor) and dependent variable (room variables). This work aims to outline the relative strength of factor a compared to b, and vice versa, in inducing a desired change in z, effectively stating the significance of each the room modification variables in comparison to each other in determiing changes for each perceptual factor.

In order to derive how perceptual factors relate to physical properties of an acoustic environment, this work aimed to cross reference both perceptual and physical factors in terms of objective acoustic measurements. Where each perceptual term could be defined as being proportional to series of measurement parameters, and physical properties would be defined in terms of the same parameters. Through this common reference the desired associative links between perceptual terms and physical properties could be derived.

This work would take the form of two investigations. Firstly, the analysis of physical properties and resultant changes in measurements via experiments with auralisation software, outlined in chapter 4. Secondly, the investigation of how the strength of a perceptual term changes as measurement parameter changed, which took the form of a listening test outlined in chapter 5. To select measurement parameters to be used in this project, insight was drawn from the ISO3382-1 standard, which outlines a series of parameters and what aspects of perceived reverb they relate to. From this list, three parameters were decided upon, shown in table 3.1:

Parameter	Room Property
Early Decay Time (EDT)	Perceptive Reverberance
Reverberation Time (T30)	Physical Reverberance
Musical Clarity (C80)	Balance Between Clarity and Reverb

TABLE 3.1: Selected reverb parameters for model development, drawn from ISO3382-1 [22]

EDT and C80 were deemed significant due to the fact that they are measurements that define perceived qualities of sound according to the IS3382-1 standard. Reverberence and clarity are frequently reference as key elements of optimising acoustic spaces in the context of recording and performance of music, as outlined in section 2.2. Since EDT is extremely dependent on position, T30 was later also incorporated into experiments. In comparison to EDT, T30 is described as being an objective parameter for reverberence, and the assessment of both decay time parameters allows for a wider perspective to be gleamed from analysing the results of experiments. The other key parameters outlined in the standard were deemed to be out of scope for this work.

3.2 Selecting Subjective Perceptual Factors

A primary query that this research work aims to investigate is the question of what perceptual terms are appropriate to use within a semantic audio interface for acoustically modelled spaces. In examining this aspect of the work there are a series of questions that need to be addressed regarding how useful a perceptual term would be for an end user operating an interface:

Firstly, perceptual terms need to be *valid*. In order to understand this criteria, there needed to be a clear understanding of a discrete number of distinct elements that a user would want to modify within a space. In the context of this work, which involves using modelling to emulate real spaces, these factors are elements of the natural reverberation of a space. A framework exists, derived from a large range of sources of research into the perception of room acoustics, that frames the concept of acoustics within concert halls as a series of descriptors grouped in a number of categories [57]. These groupings, such as 'timbre', 'reverberence', 'intamacy', and 'spatial impression', provide broad examples

of the key distinct elements of a space that a end user would want to modify. A visual represenation of this work, 'The Wheel of Concert Hall Acoustics', is shown in figure 3.3:

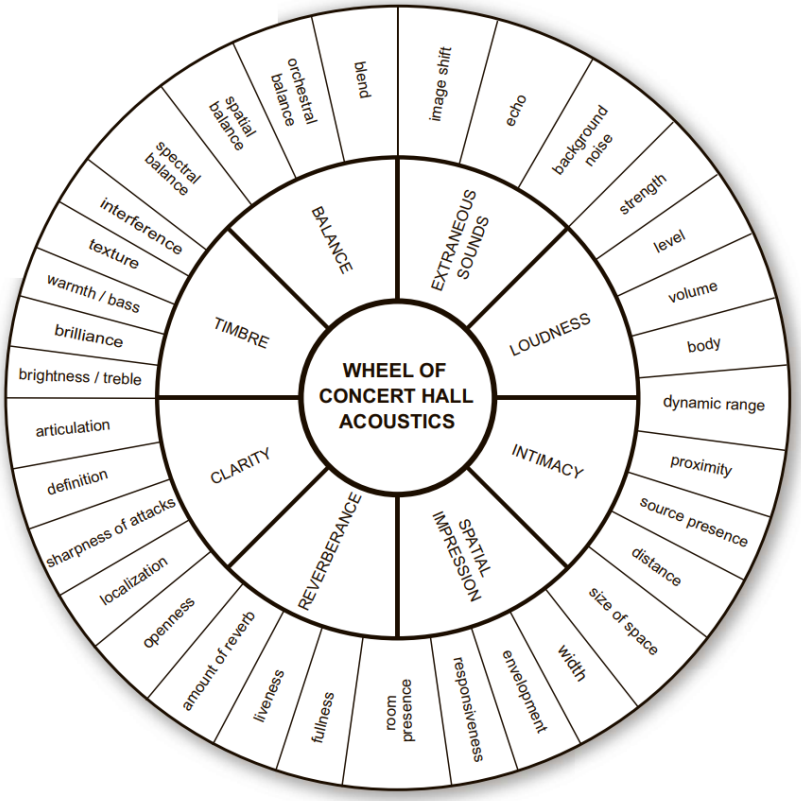

FIGURE 3.3: 'The Wheel of Concert Hall Acoustics', a visual represenation of a framework derived from perceptual acoustics studies [57]

There needed to be information about what perceptual terms are *relevant* to the transformation of sound within an acoustic environment. When referring to perceptual descriptors of audio, one is evoking a large pool of words and phrases in a wide variety of contexts. A variable in a hypothetical semantic interface should be something that relates to the natural reverberation of a space. An example of why this is important can be seen in a investigation into a crowdsourced database of reverb descriptors for a semantically driven digital reverb interface, which lead to a large variety of terms added [58]. Some terms like 'underwater' were very common with participants, but actually attempting to associate this word with elements of reverb within a live space may prove problematic since in comparison to digital reverb effects, the hypothetical semantic interfaces in this project are for accurate renderings of three dimensional environments and not user defined soundscapes within conventional listening environments (headphones, speakers, etc). A word bank for a proof

of concept demonstration of the reverb effect is shown in figure 3.4, with various emotive descriptors which would be out of scope for use in this work.

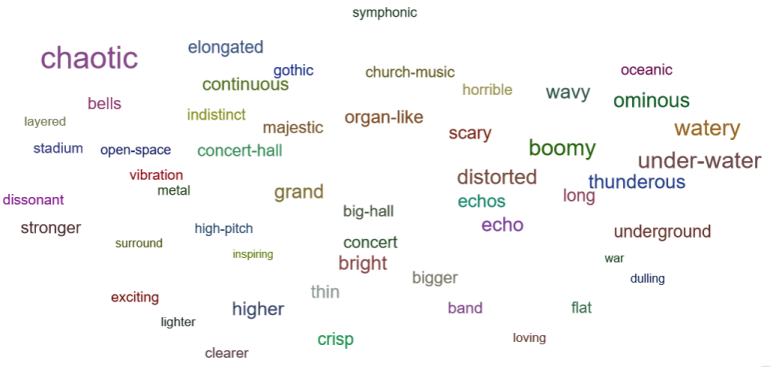

FIGURE 3.4: A Word Bank for the Demonstrative 'Audealize' Project [58]

Perceptive terms need to be *scalable*, which is to say that a term can be described as a continuous variable rather than a discrete variable. The scalar nature of a term must also be somewhat intuitive. As an example, brightness is one of the most common terms used in descriptions of musical timbre and is useful in the fact that the idea of low and high brightness is intrinsically understood as an auditory analogue to visual luminosity [59], where an the brightness of an image is percieved as on a scale between full bright and completely dark, rather than a binary of being bright or dark.

Whichever terms that were used would also have to have a high amount of *objectivity*, in this case meaning a common definition from person to person. Terms that are common in studies of timbre and the perception of audio may prove problematic within the context of this research work, as these terms may have a large amount of variance in how they are understood and interpreted. Studies into the proportion of participants who agree on definitions of common timbral descriptive terms have been conducted [60].

There was also a need in these initial investigations to make sure that terms were *unique* and distinct enough from one another so as not to describe overlapping elements of sound in terms of both the reverberation of a live space and how sound is perceived by a user. Many investigations into the subjective perception of timbre use multidimensional scaling as a form of analysis in order to codify the multitude of factors that make up the timbre of as a smaller set of dimensions. Each of these dimensions is representative of a group of scalar descriptive terms.There may be small differences in the common perception of the terms 'brightness' and 'warmth' in the context of sound, but in these investigations

these terms are often grouped within the same multidimensional factors and show high correlation with one another[5]. In this work for example, focusing on both brightness and warmth could prove problematic in terms of making terms clear for the end user.

The first perceptual term selected for this project was brightness. Brightness is one of the most commonly referenced descriptors for both musical timbre and acoustic reverb studies. It is often used interchangeably, and placed in the same perceptual dimension as warmth both brightness and warmth have existing analogues in visual luminance and temperature respectively which lead to strong a scalable definitions within an acoustic context. Brightness was seen as a more pertinent choice for this work as many theoretical elements of both auralisation and room acoustic assessment research draw heavily from visualisation and perceptual studies relating to the visual image. Within the context of timbre, brightness is often linked with high frequency flux (rate of power change) and spectral centroid [61], but there have also been studies suggesting a more multifaceted explanation of perceptual brightness. In the context of reverb, contemporary research does not outline a consensus on what elements of a reverberant space 'brightness' would be associated with, so further investigations into how brightness is link with a series of different physical properties of a space may prove useful.

The second perceptual term selected for this project was closeness. As a descriptor of audio, closeness if often interchanged with intimacy, in this study the former was used as a term instead of the latter due to the emotional aspect of intimacy as term having the potential to lead to a lack of clarity about what aspect of the sound within a space was being altered. Closeness as a term is often intuitively linked to the distance between a source and a listener, but studies into perceived sound within varying acoustic spaces outline that while source/receiver distance is a significant factor in perceptual closeness, it is not the only factor. Initial time delay gap (ITDG) sound reflections, which contain auditory spatial information about a space, can in circumstances convey to a listener that a room is smaller than it physically is (albeit not larger than it physically is). [62]. This study aims to investigate these factors as a case study for how methodologies such as the ones outlined in this paper can provide robust and reliable definitions of subjective audio descriptors like closeness in the form of a series of transformations within a space, rather than an association with a single element within said space. Meaning more effective

interfaces that can operate on a greater level of complexity and allow for more control by the end user.

Not only is closeness describing distinctly different acoustic phenomena than brightness, it is also less straightforward in existing literature what aspects of either timbre or reverb can be related to the term itself, in comparison to brightness which has large amounts of research outlining its relation to timbre. In investigating both of these terms the aim is to outline how these methodologies can derive definitions for both unambiguous and ambiguous terminology, and the changes in approach needed for each.

The work involved in deciding on these perceptual terms involved lengthy investigations into historical and contemporary research to evaluate the terms in relation of the factors listed above, this can be shown in the form of the table shown below:

Selection Criteria	**Brightness**	**Closeness**
Similar Terms	Warmth, Luminance, Colour [57][59]	Breadth, Intimacy, Naturalness, Presence. [57]
Validity	Common usage in reverberation and timbre description [59]	Usage in descriptions of reverb, intuitively linked with 'loudness' which is used in descriptions of timbre [63]
Relevance	'Bright' sound often desired in concert hall environments [64]	Close sounding audio ideal for immersive audio experiences [65]
Scalability	Highly scalable, established analogue to widely understood concept of visual luminosity [59]	Highly scalable, established analogue to widely understood concept of near/far distances [66]
Objectivity	Studies have established strong consensus on common definition of brightness in relation to audio [63]	Studies have established strong consensus on common definition of closeness/intimacy in relation to audio [63]
Uniqueness	Studies show brightness conceptually representative of a distinct class of timbral elements, primarily based on high frequency presence and spectral flux [5]	Concert hall studies indicate proximity as a well-defined unique category of reverb perception with little overlap with other elements. [57]

TABLE 3.2: Brightness and Closeness assessed using selection criteria

There are drawbacks to this approach of perceptual term selection that should be noted. Existing literature around both subjective assessments of room acoustics and commonly

used semantic descriptors for sound do not coalesce toward a singular methodology or set of methodologies, leading to wide variety of conclusions towards what perceptual terms are most appropriate to use according to the selection criteria outlined in this approach. As a result of this, different researchers may draw from the same literature as referenced and discussed in this section and conclude that different perceptual factors more fit the validity selection criteria. In more ideal conditions, a wider array of perceptual terms would be selected to mitigate this.

In addition to this, the perceptual terms that are selected lead to constraints for what spaces can be assessed. Since relevance is a key selection criteria, spaces where the selected perceptual factors are not relevant will not be of use for this experimental work, and different room acoustic properties are relevant in different spaces. Future work around this topic should take this into consideration, with the context for a hypothetical semantic interface driving the decisions made about what perceptual factors are 'relevant'.

3.3 Selecting Physical Properties of Acoustic Environments

This project involves studying two elements of an acoustic space which are investigated in detail; the distance between a source and receiver in the space, and the absorption coefficient of wall materials within the space. These two factors are representative of different elements of how people can practically treat acoustic spaces. These two elements were decided upon in part due to existing analogues with studio treatments for musical recording, source/receiver distance in this case would refer to mic placement, while absorption coefficient related to the treatment of walls via the installation of reflective and absorbent materials on studio walls [67]. There are similar analogues in the treatment of larger concert hall environments.

This studio treatment analogy is a key part of how a semantic interface would hypothetically operate, many semantic design techniques for digital audio effect interfaces over the years have used live recording terminology in order to convey elements of their effects intuitively; in drawing from these developments there emerges a defined scope on what physical factors within a auralisaiton model are appropriate for this line of study. While this work involves working with simulations, the aim is to replicate live spaces rather than abstract soundscapes, meaning the physical elements within this work need to relate to real spaces and how they are assessed. In addition to this, the scope of this work focuses on acoustic

treatment rather than architectural design, modifying elements of a space with a defined physical geometry.

For as long as the field of research has existed, room treatment in the context of acoustics has been predominately focused on the properties of the surfaces within a defined space, this is in part due to many theoretical aspects of acoustics being intrinsically linked with materials science concepts, going back to studies from the 1930's [68].

In this study, the definition of this property is restricted to the absorption coefficient of the primary wall surfaces of an acoustic space (front, rear, side walls), these surfaces are likely to have the largest surface areas of the space and would be the surfaces most likely to be treated with materials The modification the absorption coefficent for these surfaces aims to emulate the installation of more absorbent or reflective materials on said surfaces. Ideally this factor would account for where in particular treatment materials should be installed on a surface, which would involved a more involved investigation of reflections in the space which may differ from room to room. In order to derive a broader associative link between a perceptual term and this factor it was assumed that the absorption coefficient variance would apply to the entirety of the surface area of each wall.

Source/receiver distance isn't a defined aspect of how a space is designed, but how it is used, and how objects within it are oriented; a key aspect of acoustic design and a factor distinct from absorption coefficient. A 'receiver' in this case can be described as an listener or as a recoding device such as a microphone, in the former case the source receiver distance dictates where a performance area and audience section is within a space; seating within concert halls may be placed further back, at an elevation, or at a different angle depending on the ideal perceived acoustic environment desired. In the case of a receiver as recording equipment, source/receiver distance relates the concept to mic placement, with three dimensional distance of recording equipment from a sound source dependent on the instruments being recorded, and the desired timbral qualities of the recorded sound.

Much like with the approach towards selecting perceptual factors, there are drawbacks to this method of selecting physical properties of acoustic environments that should be acknowledged. The limitations on the amount of room properties selected for this work due to time constraints means that there is the chance that more appropriate room properties for investigations around associations with the chosen perceptual terms may have been

overlooked. Much like with perceptual factors, the ideal case is that a larger array of room properties are selected for analysis so that perceptual factors are defined by a much greater amount of room acoustic properties and therefore leading to more effective semantic interfaces. Practically speaking, introducing a larger number of room acoustic properties for each perceptual term to be tested against greatly increases the scale of the work needed, so this needs to be taken into account for future work.

In this study, the scope of this element is restricted to a single dimension, the forward/backward x dimensional distance between a source and receiver, as it was decided that two and three dimensional expressions of distance were too complex to study for the time frame of this testing. It is worth mentioning however, that apparent source width (ASW), described in ISO3382 as the fraction of the energy arriving from lateral directions [22], forms a key component of perceived reverb within an acoustic space, and would be a key element to investigate in further work around this topic.

In summary, the hypothetical semantic interface for this project work was conceptualised as a three schema representation, with the mathematical associative model for linking room acoustic elements with subjective descriptors acting as the interface between the external and conceptual layer, and the auralisation simulation used to acoustically model the 3D space of the room being analysed acting as the interface between the conceptual and internal layer. These two interfaces will be the focus of the project design. This project will involve three room acoustic properties, EDT, T30, and C80; with each property being semantically associated with both perceptual brightness and perceptual closeness, meaning six relationships in total will be investigated.

Chapter 4

Auralisation Tests

4.1 Introduction

This chapter outlines the experimental work undertaken in this project to investigate the associative relationships between elements of an acoustic space and the measurable properties of sound within it. This was done through modelling acoustic spaces with varying room properties via auralisation, with acoustic parameters being measured through the analysis of the resultant impulse response (IR), generated within each space.

This work was iterative, and over time elements of the experiments were modified as findings emerged around the capabilities and limitations of the software used, as well as the assumptions made around prior methodologies. Each of these experiments is discussed individually, with some discussion around the upsides and downsides of each approach, as well important takeaways to note for further work around this topic.

Ultimately, this section outlines both early experiments focused on a model of a real space, and later experiments focused on a simple example space provided by the library within the auralisation software, with a presentation of and brief discussion around the data from each experiment. This section outlines three iterations of the experiment:

- The first experiment outlined measures values from a series of receivers placed within a live space, the National Centre of Early Music; with wall surfaces being modified over the course of the experiment.

- The second experiment outlined functions the same as the first experiment, with four receivers being placed within an acoustic environment. However, this time the environment is not a real space, but an example room model provided by the modelling software with more basic geometry and smaller size. This space was changed to provide more reliable and understandable results to analyse.
- The third experiment is an extension of the second experiment, within the same example space 15 receivers were placed 1m behind each other towards the back of the space, and a greater array and range of wall surfaces were used.

This section presents data from each experiment with brief discussion, as well as wider observations on how the methodology behind the experiment evolved over time, and some practical takeaways for future experiments to derive similar associations between room elements and acoustic parameters.

4.2 Methodology

4.2.1 Aim of Experimental Work

The overall aim of this experimental work was to derive proportional relationships between a series of elements in an acoustic space that can be modified, the independent variables, and the resultant acoustic measurement parameters, the dependent variables. These proportionalities will outline how a measurement parameter can be described as the result of multiple elements within an acoustic space. These derivations will form part of the overall mathematical model for this project.

4.2.2 Variables and Null Hypothesis

ISO 3382-1 outlines a series of practical guidelines for measuring these parameters, as well as considerations for margins of error, defined in the standard as just noticeable differences (JND). The necessary conditions for reliable for impulse response measurement were used to guide the work of generating appropriate impulse responses from auralisation software.

The independent variables in this experimental work are the elements of a space that are being modified. In this work two variables were modified; the absorption coefficient of the materials of the main walls within the space, and the one dimensional distance from source to receiver. As detailed in section 3.3, these variables were decided upon with a focus on the use case of treating an existing space to a specification, rather than designing a new space from scratch. Therefore the question in altering sound in a space concerned said space having an already defined geometry. The two variables are slight abstractions of the two main practical approaches and end user can apply within a space to change the perceived qualities of sound; treating the walls of the space so that they are more or less reflective, and changing the position of a receiver (representative of either an audience in a live space or a microphone in a recording context) in relation to the source.

In this experiment, and in the wider project as a whole, source/receiver distance is defined as a one dimensional scalar measurement along the X axis with an azimuth of 0 as opposed to two dimensional [X, Y] measurements of distance. Within ISO3382-1 documentation, and associated literature around this topic, distance measurements are defined in terms of an array of microphones around a sound source, with an invariant scalar distance but placed at different angles around the source; while in this experiment distance is defined in straight line away from the source, with microphones being places along said lines. The primary reason for this change is significantly reduce the complexity of this area of analysis. Going back to the initial scope of this work, the aim for this research is to outlines methodologies and approaches towards developing semantic associations between semantic terms and elements of an acoustic space, it was decided in considering a more abstract definition of distance without the incorporation of angular elements that these associations could be derived within the time frame of the project, and these associations could inform investigations into two dimensional or event three dimensional measurements of distance.

Each of the independent variables was investigated individually. This lead to an experiment focused on two null hypotheses.

1. The first independent variable being investigated is the one dimensional scalar source receiver distance on the x axis. With this there was a focus on how distance effected the measure of C80 clarity. The null hypothesis for this element of investigation can

be described as *'The distance of a receiver relative to the source within any defined acoustic environment has no effect on measured C80 clarity'.*

2. The second independent variable being investigated is the broad band absorption coefficient of the primary side and back wall surfaces of an acoustic environment, with a focus on how modifying this variable leads to changes in decay time (EDT/T30). The null hypothesis for this element of investigation can be described as *'The absorption coefficient of the primary side and back wall surfaces will have no effect on the resultant EDT and T30 reverberation time measurements'.*

As the project developed it became clearer that the null hypotheses that were initially established were in fact already provable through existing acoustic theory. Through formulae it can be shown that reverberation time is proportional to the average absorption coefficient of a space, with the average absorption coefficient being the mean coefficient for each boundary surface within the space. This is shown in the equation below, where RT60 is reverberation time, V is the volume of the space in meters cubed, S is the total surface area of the space in meters squared, and alpha is the average coefficient:

$$RT_{60} = \frac{0.041V}{S\alpha} \tag{4.1}$$

The second null hypotheses can be proven through the principle of a diffuse field, where the energy density is the same on all points of the volume of a field [69]. The initial direct field is more prevalent than the resultant reverberating field at short distances from the source, with the reverberating field becoming more prevalent at greater distances from the source.

Nevertheless it was decided that there was still value in these investigations as demonstrations of the methodology behind this aspect of semantic interface development. Analysis of these hypotheses can provide a useful comparison between theoretical principles derived from literature and practical examples of how these properties manifest within a commercial auralisation program. Any practical considerations gleamed from these investigations will prove useful for future work around this topic.

4.2.3 Auralisation Modelling

This work was done with ODEON, a commercially available acoustic modelling program with a wide range of features suitable for this work [70]. In ODEON, one can import 3D geometry into the program and assign material properties to each of its surfaces, as well as providing the ability to places sources and receivers in the environment to model performers and listeners (DIAGRAM). ODEON will then use a hybrid ray tracing and imaging based solution to model how sound generated by the defined sources propagates through the space and is perceived by the receivers.

The nature of proprietary software, especially with software as complex as ODEON, is that it's hard to get a full grasp of the techniques used in modelling the three dimensional space, generate the resultant impulse responses, and measuring said responses in order to attain values. ODEON is proprietary software. and in terms of how it renders impulse responses and calculates acoustic measurements can be treated as a black box model; however the caveats of a lack of access to the specific techniques this software implements are outweighed by the softwares usefulness for the context in which it is being used, in building an measuring from virtual models of acoustic environments.

It's worth noting how doing this work within modelling software is only an approximation of measurement methodology within a real space, with a lack of information around the assumed environmental conditions for said space. The methodology for deciding an appropriate space defined ISO 3382-1 is based around practical observations of variability relating to elements in a room such as humidity and air temperature; air conditions and temperature can be modified in the software before simulations, but once again it is unknown how these variables effect the simulated acoustic measurement, not the equations used in said measurements.

Another thing to consider was the complexity of the ray tracing used within the program. Even in the ideal case, where a system can model millions of rays reflecting off wall surfaces at a given moment, it will only be representative of actual wave propagation. Ray tracing techniques often have lower accuracy at low frequencies due to the fact that typical methods fail to account for diffraction effects, which are more prominent in lower frequencies, in comparison to wave based techniques [71]. Therefore this research and its resultant data needed to be interrogated with this in mind, knowledge of the limitations of ray tracing

techniques necessitates the need for a constrained range of frequencies that can be analysed with confidence. In cross referencing auralisation models with real world examples of IR measurement there were some discrepancies between modelled and actual results, sub 250Hz results were unreliable across all parameters, a known limitation of existing ray tracing techniques, and therefore these results were not analysed

While ODEON has internal tools that generate measured values of the parameters being investigated. It was decided that due the 'black box' nature of techniques implemented in ODEON, an external analysis was to be performed on impulse responses generated by the auralisations instead.

Room impulse responses from each auralisation where fed into a series of MATLAB scripts, that would generate energy decay curves and calculate values of ISO3382-1 variables from the raw data of the imported audio files. These MATLAB scripts were provided by the Virtual Acoustic Team at Aalto University and are adaptations of the 'AcMus - Room Acoustic Parameters' toolbox created by Bruno S. Masiero, these scripts formed the basis of an open source acoustics software developed by the University of São Paulo.

In these scripts, bandpass filtering is used and the noise floor is accounted for in order to garner more accurate results from these calculations. The scripting being fully accessible allowed for ease of troubleshooting when anomalous values, values that exhibited significant deviation from the expected order and range of the parameter values, were generated.

4.3 Experiment Methodology

Firstly, the sources and receivers within the space were positioned. In every experiment, the source is of semi-omnidirectional directivity with a power of 9.0dB and is placed 1.5m above the Z plane with zero rotation or elevation, meaning it is pointed directly forward and towards the receivers. ISO3382-1 outlines a recommended height of 1.5m for sources involved in live measurement. Once the source was placed in an appropriate spot within the environment, with enough space for initial waves to propagate without surrounding objects causing unintended early reflections, an array of receivers are then placed within the space The receivers contained the same y and z co-ordinates as the source but moving further back along the x axis away from the source. All of these parameters were used for

every permutation of the experiment. In the first two experiments, four receivers were placed along the axis from 2m to 8m away from the source in uniform intervals of 2m. In the third experiment a different approach for receiver placement was taken, with 15 receivers being placed 1m away from each other in the same linear fashion as before.

The materials list of the space was then modified. In ODEON, the materials list defines the materials of every surface within the space in terms of their absorption coefficients along the frequency bands. For these experiments the wall surfaces to each side and in the back of the space had their materials changed to these uniform example materials. This material would be changed and the auralisation repeated in order to attain IR measurements across a range of absorption coefficients. In the first two experiments, materials of 20% (0.2), 40% (0.4), 60% (0.6), and 80% (0.8) absorption were used. In the third experiment, materials from 0% to 100% increments of 10% (0.1) were used.

For each receiver in each simulation, impulse responses were generated as B-Format .wav files. B-Format is an audio format that contains four channels rather than the typical two channel stereo approach; W, the omnidirectional sound pressure for a sound, and X, Y, and Z, the first order harmonics of a sound along those axis'. The generated impulse responses were then fed into the scripts within the MATLAB toolkit, where the B-Format impulse responses were transformed into two channel stereo files via transforming data within the W channel with data in the Y channel to decode the left and right stereo channels. This is achieved using the equation below, where X is the X channel of the B-Format data array and Y is the Y channel of that array.

$$L=X+0.707*Y \tag{4.2}$$

$$R=X-0.707*Y \tag{4.3}$$

The resultant stereo files are then analysed in MATLAB. For each room auralisation, The accuracy of measurements within an auralisation could be assessed via comparisons of measured T30 values to predicted T30 values. Within each room auralisation T30 should remain unchanged as the receiver moves further back along the space, as T30 is hypothetically a measure the reverberant impression of the entire space, and should remain invariant no matter where in the space it is measured. In this experiment, source-receiver

distance changes within each room auralisation while surface absorption coefficient changes between room simulations.

For this experiment, frequency bands of 250Hz, 500Hz, 1KHz, 2KHz, and 4KHz were used for all experiments. When running data from the impulse responses generated in ODEON, it became noticeable that there was a large amount of variability in the lower frequency bands within the measured parameters, and this often lead to unreliable or implausible values in comparison to higher frequencies. As a result sub 250Hz bands were not accounted for. Within the field of acoustic assessment, low frequency values have a tendency to be less reliable that mid or high frequencies. In ISO3382-1 a frequency range of 500-1000Hz is recommended for single number frequency averaging.

There were also problems in the experiment relating to EDT measurements. These variabilities didn't have any particular explanation in the theoretical underpinnings of this work, nor the practical work within the experiments themselves. It was deducted that this was due to the fidelity of the IR files themselves, the inaccurate measurements can be attributed to the fact that the early stages of the resultant energy decay curves showed peculiarities, suggesting that the rendering of the early parts of the impulse response was less accurate than required for this work. Documentation around ODEON suggests that impulse responses generated within the software are primarily used as a reference, for a user to get a general impression of what a space sounds like. This could mean less fidelity for resultant IR files, leading to inaccuracies with positionally variant measurements like EDT.

4.4 Experiment # 1 - National Centre of Early Music

4.4.1 Background

Initially the model used for these experiments was a model of a real world location, that location being the National Centre for Early Music (NCEM); a performance space based in St Margaret's Church in York, with an area of 3600m^2 and XYZ dimensions of $24.61 \times 13.56 \times 11.24m$. This particular auralisation experiment builds off existing literature investigating the acoustic properties of St Margarets Church through using three dimensional spatial models [72].

The church itself has been retrofitted and has had its acoustics redesigned by ARUP Acoustics via a series of panel absorbers along the walls and drapes on the roof. These elements can be configured to allow for more or less reflected surfaces, allowing for three distinct configurations. A configuration for music recitals where no wall panels are used, one for larger musical performances where 75% of the panels are used, and one for lectures and public speaking where all the panels are used. Examples of panel placements within the space are shown in figure 4.1:

a)

b)

c)

FIGURE 4.1: The National Centre for Early Music, a) front performance area, b) ground of audience area, c) absorption panel placements and roof in audience area

With the NCEM, there is an interesting case study of the practical implementation of the contextual modification or treatment of an acoustic space, with the sound absorptivity of the walls being increased or decreased depending on whether the space is being used in a musical or non-musical context. Implicit in these configurations is a value judgement on the importance of reflections on sound quality as perceived by a listener, and how that depends on the sound source. It was believed in the initial stages of this work that the NCEM would be a good case study to use for experiments. Via the OPENAir program there is a large amount of data available related to this space, including live IR measurements of the psychical space and their associated ISO3382-1 measurement values. Moreover, in working with a real model

There a few key geometric features of the space worth noting. Firstly, the space contains a series of pillars running through the centre, a resultant architectural feature of a retrofitted church. This provides an interesting quirk regarding sound propagation within the space. These pillars often cause significant reflections of their own from the initial propagating sound waves of omnidirectional or semi-omnidirectional sound sources. The geometric complexity of these pillars, owing to their shape being a circular combination of long and thin rectangular surfaces, means that reflections are often chaotic and random.

Similar observations can be gleamed when focusing on other elements within the space. The insides of the door and window frames throughout the space are not smooth surfaces, but a combination of small geometric shapes, within the room's acoustic model this leads to the primary walls of the space not being uniform, with small reflections within window and door frames having to be accounted for. Overall, the acoustic model of this space contains 969 unique surfaces and 1327 corners. The nature of the pillars within the space and other related small internal surfaces means that there needed to be consideration regarding the placement of a linear array of receivers for the experiment.

The layout of sources and receivers in the ODEON model is shown in figure 4.2:

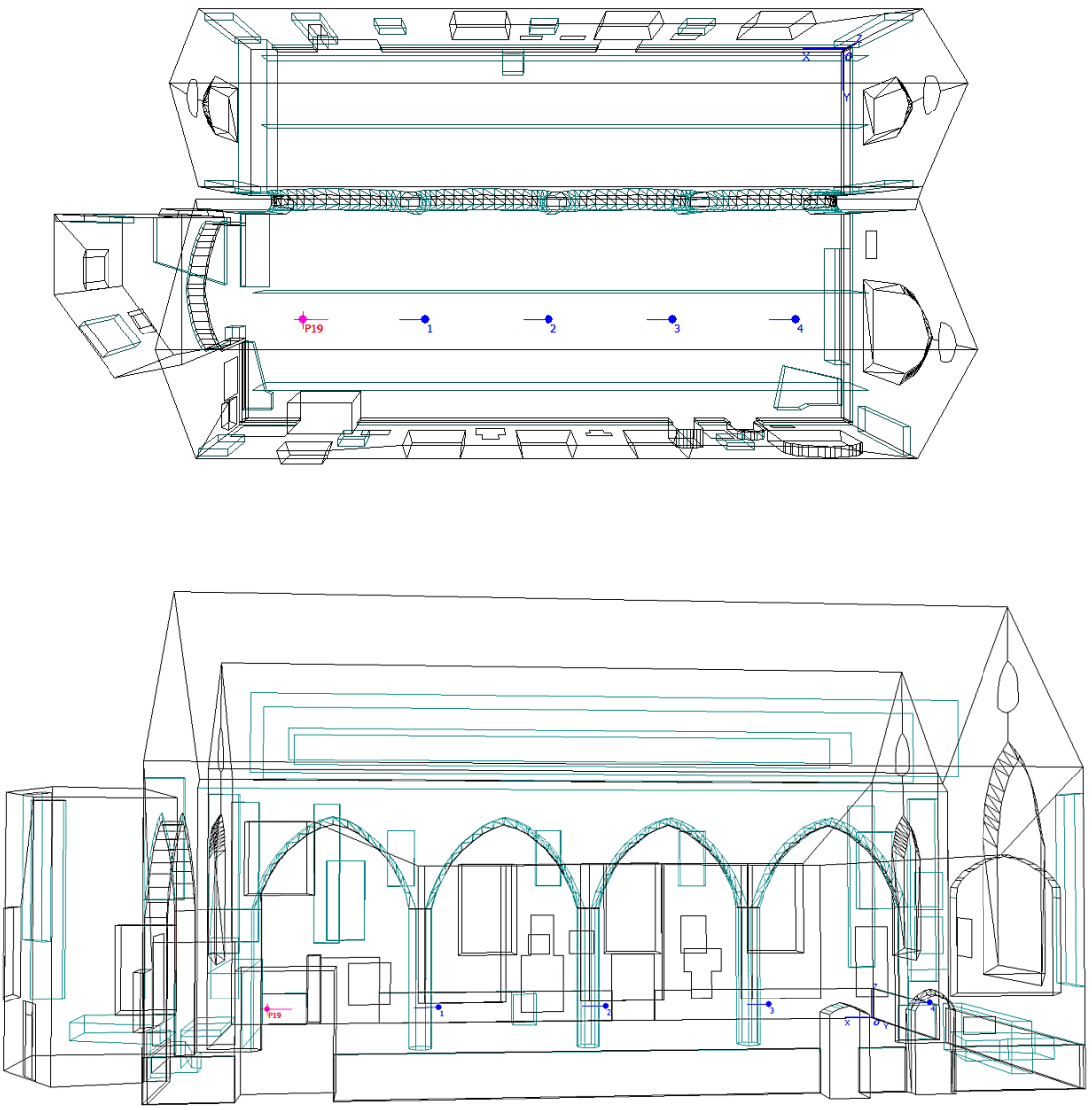

FIGURE 4.2: Source & Receiver Arrangement for Experiment 1

4.4.2 Results

As previously discussed, simulations were ran at wall surface absorption coefficients of 20%, 40%, 60%, and 80%. In each simulation the impulse response at each receiver node was exported and analysed. The following results show the measured values of EDT, T30, and C80 at each receiver node for all four simulations. Recordings were measured at frequencies

of 500, 1000, and 2000Hz, previously outlined as the important frequencies to observe in the context of this work. The results of this experiment can be seen in appendix B.1

Measurements for EDT are shown in figure 4.3:

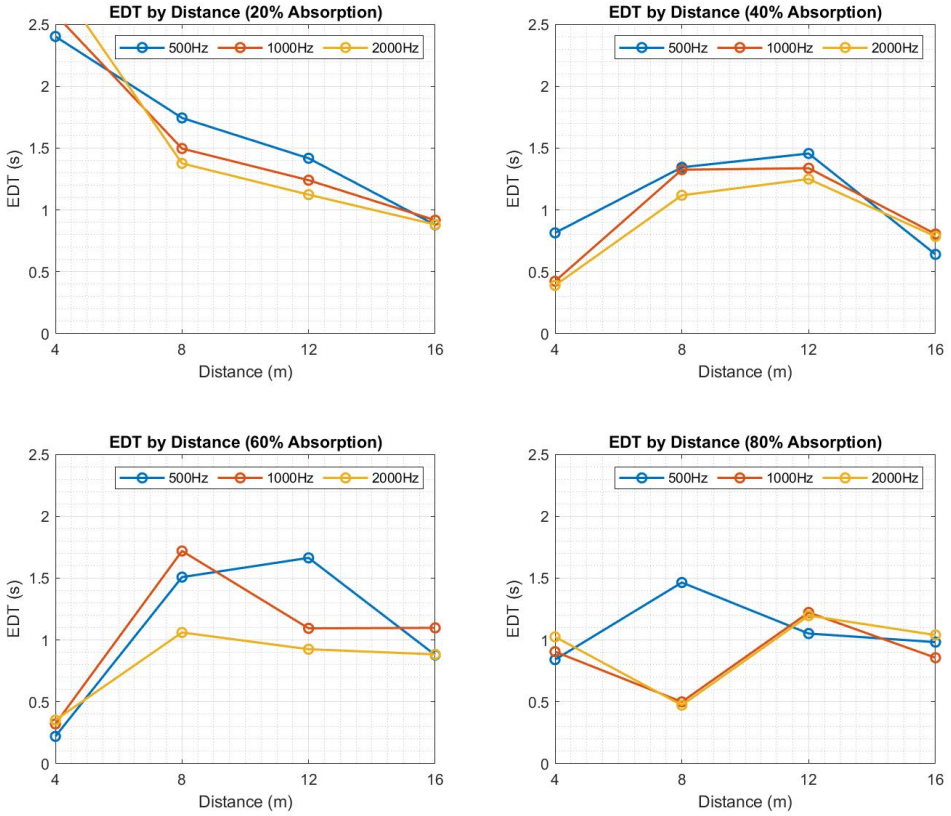

FIGURE 4.3: EDT Measurements for Experiment 1 (NCEM)

It is clear from initial observations that there is no correlation between EDT and receiver distance across the 40%, 60%, and 80% absorption simulations, with the 20% absorption experiment showing negative correlation. In addition, there is no visible repeating trend as absorption coefficient increases. In this environment there is no indication that the modification of either of the independent variables factors into measurements of EDT at all.

Measurements for T30 are shown in figure 4.4:

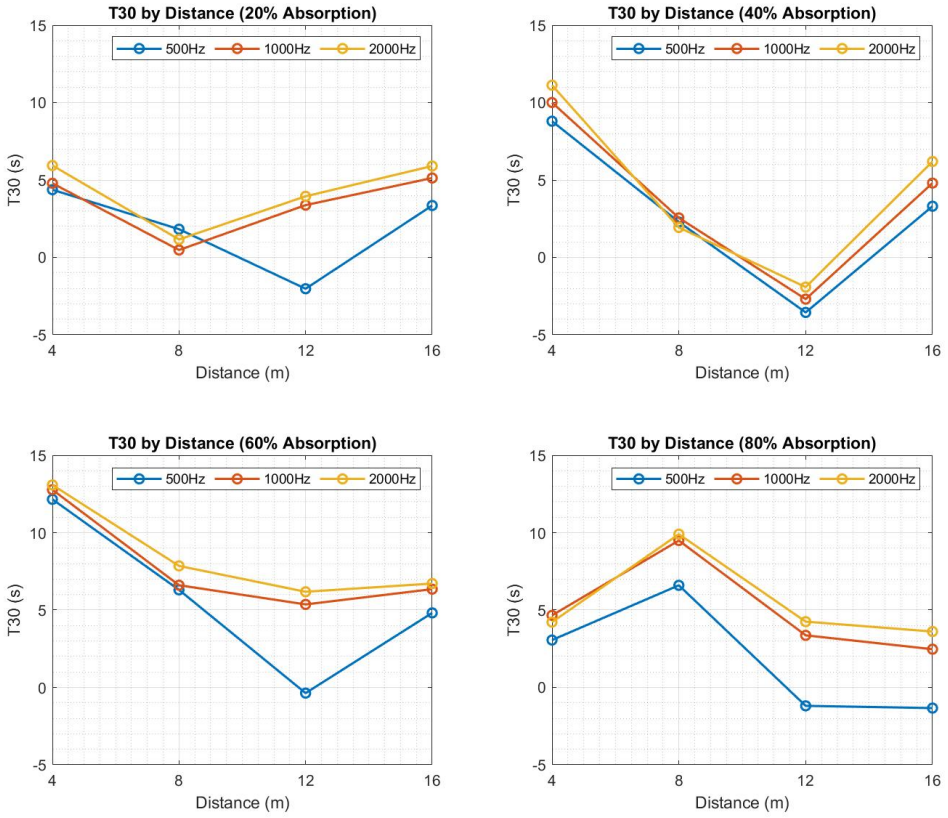

FIGURE 4.4: T30 Measurements for Experiment 1 (NCEM)

From observing these measurements, there is trend of decreasing T30 from 4m to 12m, and a slight increase from 12m to 16m, across each simulation barring the 80% simulation, which contains an anomalous measurement to the trend at the 8m receiver. These trends are also relatively uniform throughout the measured frequency bands, which indicates a reliable set of results.

Measurements for C80 are shown in figure 4.5:

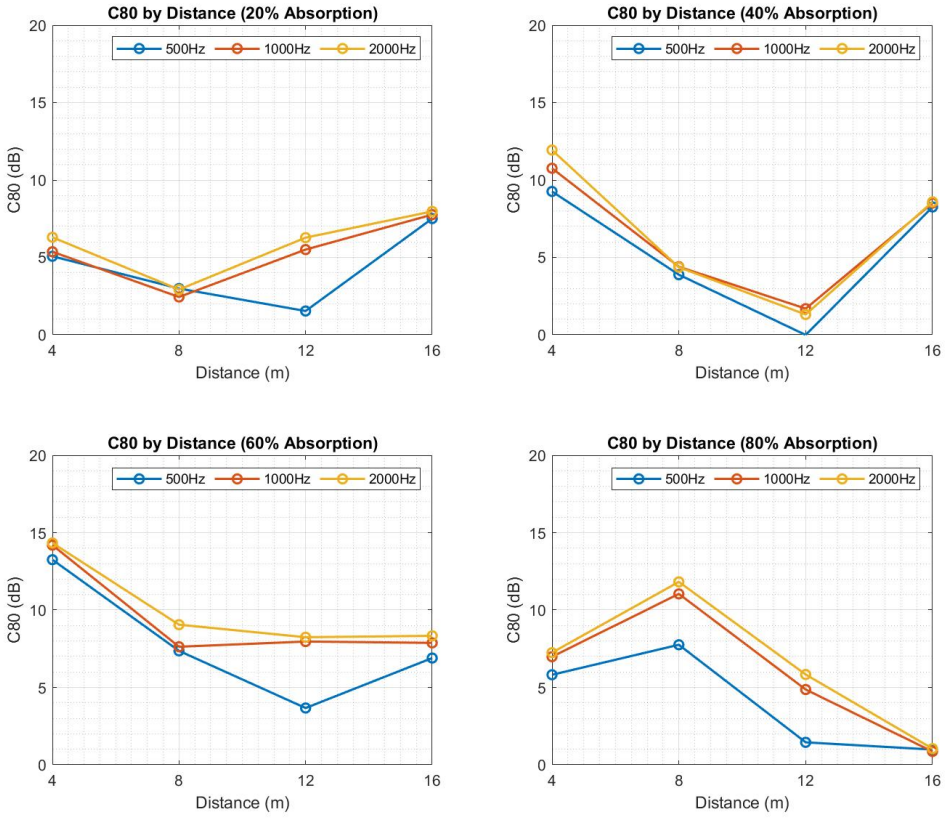

FIGURE 4.5: C80 Measurements for Experiment 1 (NCEM)

In observing these results, there is a trend of decreasing C80 up to the 12m receiver, with a slight increase from that point. There is a strong similarity in the general trends for measured C80 and measured T30 in these experiments, including the same anomalous receivers in the same experiment (8m and 16m at 80% Absorption).

4.4.3 Discussion

Due to the limited amount of receivers used, analysing the results of this experiment involves the observation of general trends relating to the dependent variables measured (EDT, T30, C80), conclusions therefore will be broad descriptions of how independent variables effect dependent variables, and will inform the more detailed and considered experiments later in this section.

Future experiments iterated on this methodology significantly due to the various constraints placed on the effectiveness from factors previously outlined (lack of receivers, etc), but there still important points to be drawn in analysing the data generated from this individual experiment.

There was a stark contrast in the uniformity of trends for C80 and T30 in comparison to EDT. This is most likely due to a number of factors. As previously stated, EDT is defined as positionally dependent in comparison to T30, which hypothetically is position invariant. As distance changes along the x axis for each graph in figure 4.2, one can see that the variability from positional change is non uniform, there is no positive or negative correlation at all. It could be said instead that EDT measurements are unreliable as the position of a receiver changes, this point is further reinforced through observing the differences in values amongst frequency bands in comparison to T30, which is also a measurement of decay time.

However, this unreliability within EDT results may also be due to issues with the analytical method itself. In observing the visual schroeder curve provided in the MATLAB IR analysis toolbox, there were abnormal results seen in the early regions of the curve, an example of this can be seen in figure 4.6, where the schroeder curve of the 4m receiver at 80% absorption is shown. It can be observed that in the initial stages of the curve there is a steep drop off, which is not typical for decay curves of this type. This could point to the methodology of analysis used in the toolbox providing unreliable values at the early regions of the curve, or it could suggest abnormalities that could be the results of the acoustic environment itself.

These abnormalities indicate that the analysis toolbox has problems rendering the early regions of the schroeder curve, from which EDT is measured. This would explain why EDT results seem more unreliable than T30 results, which are drawn from a later region of the curve. From initial analysis of this first experiment it was hard to discern the exact significance of this problem on results in comparison to other factors, but it was significant as these experiments were designed with EDT as a key dependent variable to be measured.

While there are common trends of decreasing C80 and T30 with distance, this often doesn't follow on across all the receivers in each simulation. T30 for instance, increases significantly at 16m for 20, 40, and 60% simulations, with similar trends for C80 in those

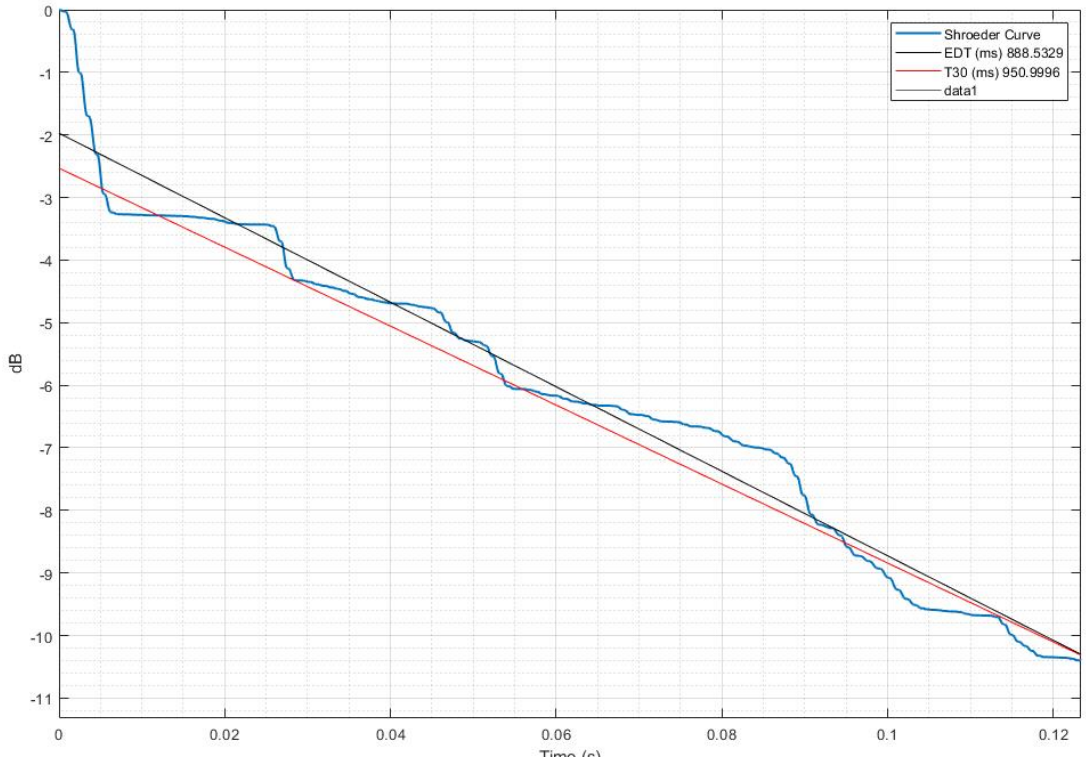

FIGURE 4.6: Schroeder Curve Example for 4m receiver for 80% Absorption

same simulations. In analysing how these results came to be it is worth understanding the layout of the receivers and the geometry of the room itself. The complexity of the geometry within the NCEM means the initial waves propagated from the omnidirectional source were subject to unpredictable early reflections, this is shown in figure 4.7.

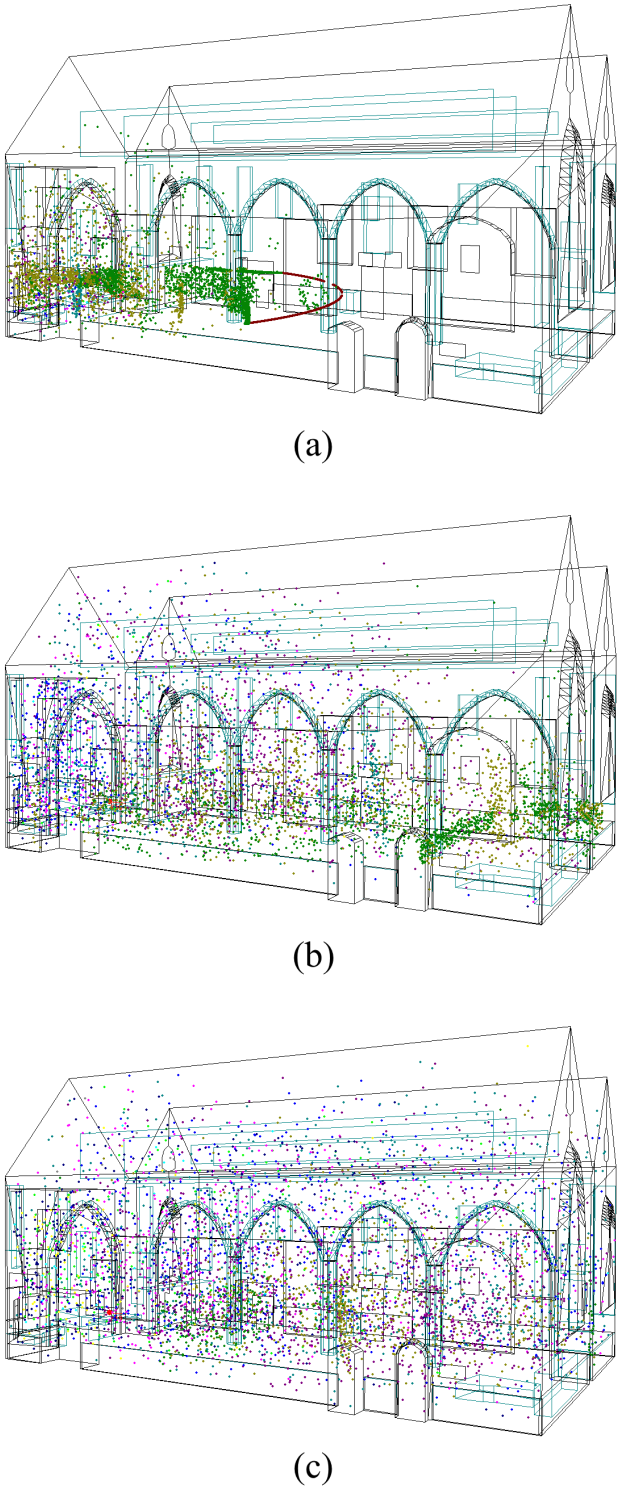

FIGURE 4.7: Visualisation of Early Reflections From Source in Experiment 1, a) at 28ms, b) at 63ms, c) at 90ms

In figure 4.7, one can see that the receivers were placed one behind the other parallel to a series of structural columns, each column formed a series of thin rectangular surfaces. The column structures are geometrically complex in themselves, and close enough to the receivers that reflections from them would meaningfully contribute to measurements. These reflections could factor in to anomalous, non trending measurements in this experiment.

The receivers were placed the way they were in this experiment since a large amount of space along a single, non z dimension was required to get a cleat sense of how distance factors in to acoustic measurement. The rectangular layout of the space meant that receivers were laid out behind each other along the longest axis, and in this case the influence of other architectural elements was unable to be mitigated. Relating to the wider context of this work. This point highlights one of the key difficulties in doing this modelling work in real spaces, since the scope of this work involved modifying elements within the space not the geometry of the space itself, fundamental structural elements such as column pillars in a cathedral structure may undesirably factor into how receivers and sources are arranged within the model of that space; this leads to a trade off between the range and amount of desired receivers, and a lack of interference from large structural elements providing room reflections.

4.5 Experiment # 2 - ODEON Example Space

4.5.1 Background

While early iterations of this experiment focused on models of real world spaces, complications during auralisations lead to a understanding of a key trade off in this work. Working in real spaces provides more authentic results that theoretically would provide more relevant insights to room treatments and modifications within real environments; however, these results are unpredictable with high amounts of variability, and there are a number of factors that could potentially lead to these results that are outside of the scope of the experiment.

A solution to this was to use a more abstract example space of more simple geometry to perform the experiments in instead; this change meant that key observations in resultant data pertained less to actual practical room treatment methodology and more towards

more general conclusions, but a more uniform environment meant that results were more reliable, and the resultant conclusions were more robust.

For a less complex acoustic environment to work in, the example room from ODEON's built in library was used. Modelled as a simple abstraction of a concert hall, the space consisted of a semi rectangular geometry, with a flat floor at the front of the space representing the front stage and an floor inclining up towards the back of the space representing the audience area.

With an area of 1268m^2 and XYZ dimensions of *22x16x10*m this space is much smaller than the NCEM, and is closer to a studio space in scale than a large performance space. The floor of the example space consists of two surfaces, the smaller surface is described as the 'podium floor' and is completely flat, while the second surface is described as the 'main audience floor' and is at an incline. The 'main audience floor' meant to emulate the sloped audience area of a concert hall or similar environment. The default material for this surface in ODEON is described as 'Empty chairs, upholsted with leather cover'.

The same methodology from Experiment #1 was applied in the example space, with the rear and side walls being modified and the receivers being placed in a linear array. The reasoning behind this was that the ability to directly compare results from this experiment with the experiment prior would lead to informative conclusions on how the geometric complexity of room models effects resultant auralisations and measured values.

The incline of the space meant that considerations needed to be made regarding the Z plane, therefore all sources and receivers were raised to prevent significant issues. In repeated testing the effects of receivers being closer to the ground as they moved further back towards the rear wall had a negligible impact on results regarding the two hypotheses in this experiment. The angular direction was rotated further downwards for each receiver being placed further away from the source and towards the wall, meaning receivers further across the room were tilted more downwards towards the source. The directional angle of each receiver is proportional to its Z distance from the source (since the main floor is an incline, this will increase as receivers are placed further away from the source). The arrangement of the source and receivers in this experiment is shown in figure 4.8, where one can observe the increasing Z position of the receivers as they are placed further back

in the space, as well as the receivers being pointed downward at a slight angle towards the source to compensate this.

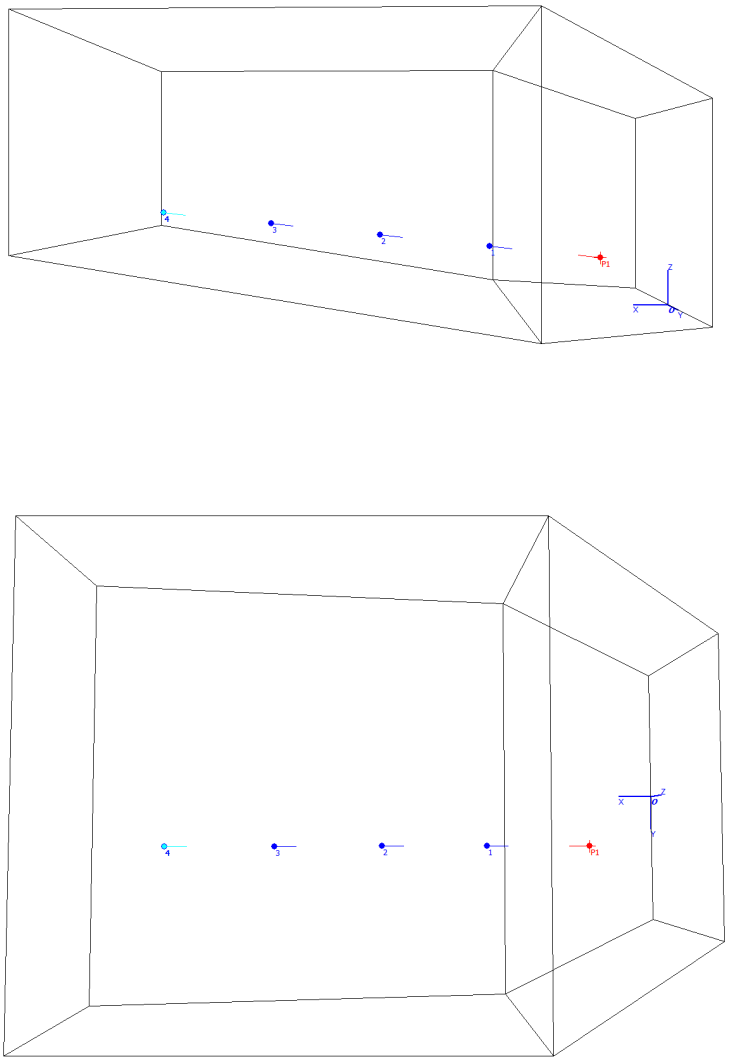

FIGURE 4.8: Source/Receiver placement for Example Room used in Experiment # 2 (Example Space)

4.5.2 Results

The results of this experiment are presented in the same format as the first experiment, values of C80, EDT, and T30 are shown for each of the receivers (4m, 8m, 8m, 16m) in

each simulation (absorption coefficent values of 20%, 40%, 60%, and 80%). The results of this experiment can be seen in Appendix B.2

Measurements for EDT are shown in figure 4.9:

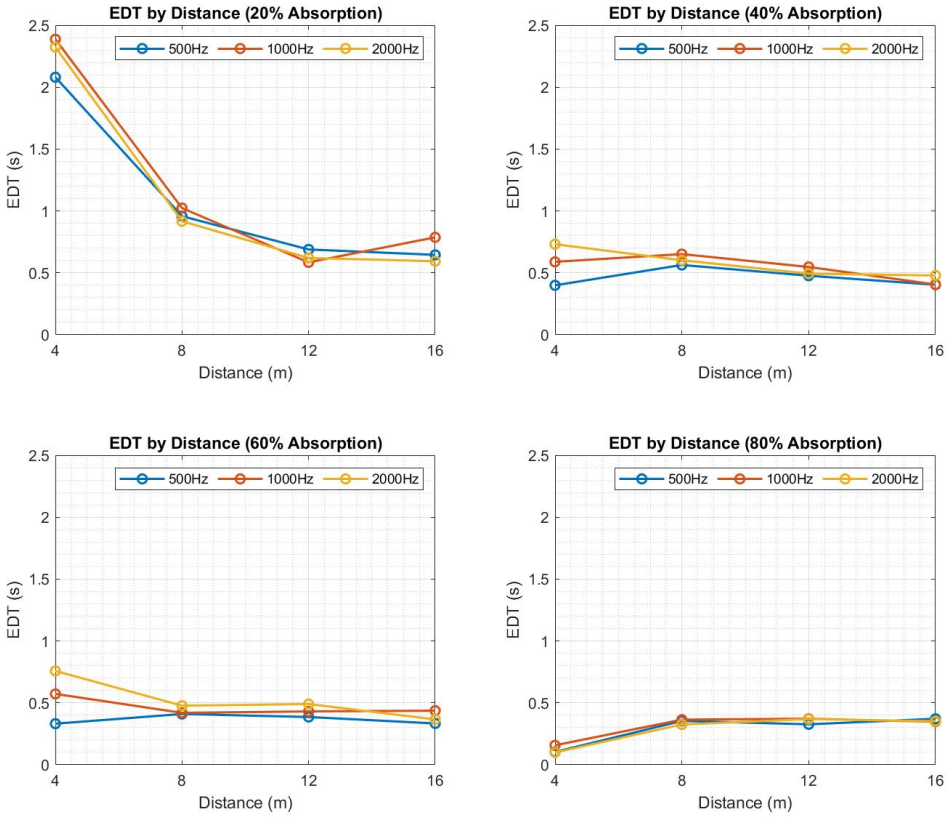

FIGURE 4.9: EDT Measurements for Experiment 2 (Example Space)

EDT measurements are more uniform across different frequencies, for each simulation there was decreased variability across frequencies for receivers at a further distance away from the source. This is in comparison to the first experiment which contained larger variability across frequencies for all parameters.

Across the simulations there is little to no variability for EDT as distance changes, with the exception being the 20% absorption experiment, which showcases an almost inverse proportional decrease in EDT as distance increases. Nor is there any significant variability between experiments as absorption coefficient changes outside of the 20% result, with no discernable linear trend with increasing absorption (between graphs). This suggests that in this environment, with it's simple geometry, the perceived reverberence doesn't significantly change as a result of either of the room elements tested.

Measurements for T30 are shown in figure 4.10:

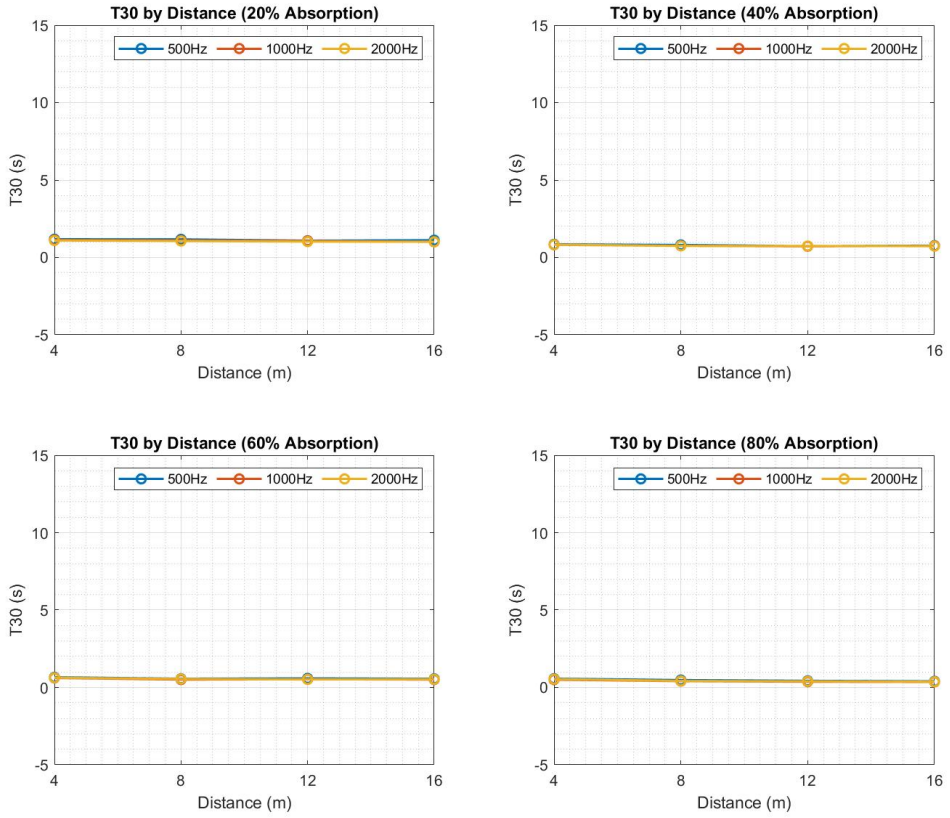

FIGURE 4.10: T30 Measurements for Experiment 2 (Example Space)

Accounting for errors in measurements, there is zero variability from both distance and absorption coefficient for T30, in each simulation T30 is measured at around 1s at all receivers. Comparing the T30 results to EDT measurements, it can be observed that there is a greater amount of variability in results from changing distance, as well as a greater variability between frequency bands, although for both experiments these variabilitys are negligible.

Drawing back to the theoretical principles behind EDT and T30, it is expected that T30 would be invariant across distances, as it is stated in literature that T30 is spatially invariant while EDT is not. This is also why it is expected for EDT variability across distance to be greater, but even the variabilitys shown within the EDT results are slight, suggesting that the simple room geometry provides less complex reflection patterns that are picked up by receivers and therefore more uniform results.

Measurements for C80 are shown in figure 4.11:

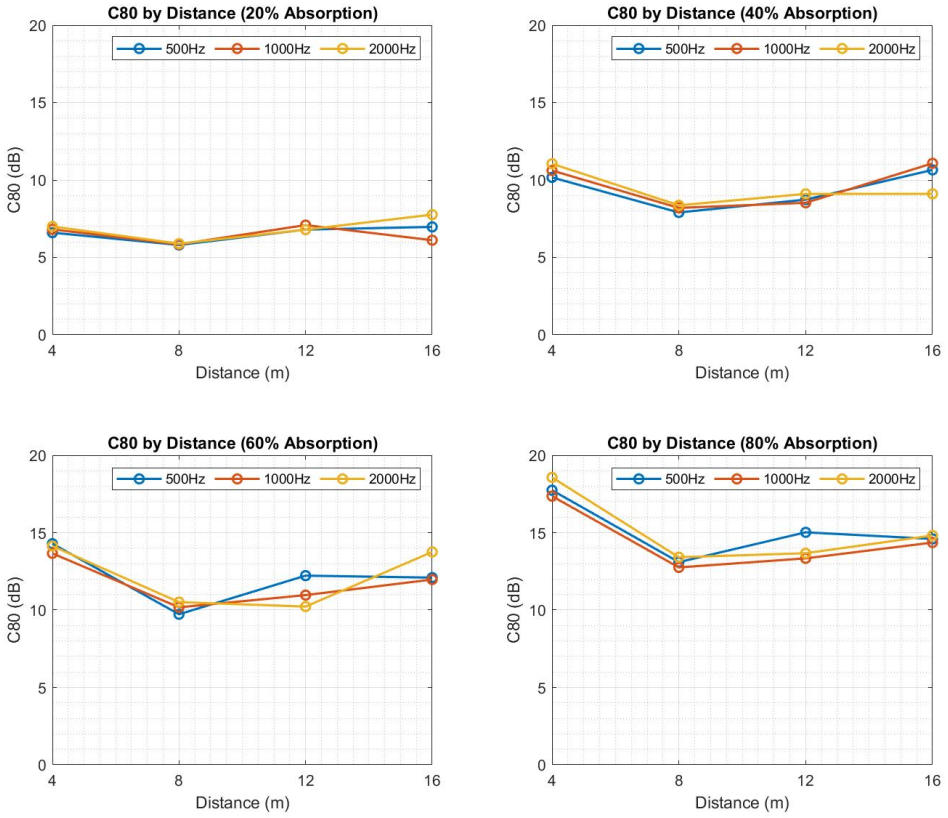

FIGURE 4.11: C80 Measurements for Experiment 2

Similar trends for C80 variability as receiver distance increases are shown in each simulation, C80 slightly decreases from 4m to 8m, and gradually increases from 8m onward. For each simulation, as absorption coefficient increases, so do the measured values of C80 at a receiver. While measurements vary in frequency for C80, these differences are small. These results imply that C80 increases as source receiver distance changes, but the trend is so slight that this not a reliable observation. In contrast, it is shown from these results that C80 increases as absorption coefficient increases, and this correlation is more distinct and verifiable relative to the amount of information this experiment provides.

4.5.3 Discussion

Across all variables measured, there was significantly less variation across receiver distance in comparison to the first experiment. In accounting for the differences in acoustic environments, it can be postulated that the simpler geometry factored into the more uniform results across distance. In this environment there was reduced variability across the frequency spectrum for measured values, this means that any variabilitys observed across distance or across absorption coefficient are more reliable in the frequency range relevant to the measurements being analysed. In future work it can be assumed in an ideal case there would be zero variability across the frequencies outlined, and therefore measurements froma single frequency are sufficient for results.

Certain EDT values, such as the measurement at 4m for the 20% absorption simulation, significantly deviated from the norm to the point where it can be deemed as an outlier result. In comparison to measurements of T30, EDT results contain more of these outliers, in addition to non uniform variabilitys across the frequency spectrum, this provides further evidence to the point discussed around experiment 1 in section 4.4.3, that rendering issues in the MATLAB analysis tools used to measure values factored in to these discrepancies

Figure 4.12 shows a visualisation of the reflections of early source waves within the example space. A notable aspect of this visualisation is that at point (b), 63ms into the simulation, there is the accumulation of reflections on the floor surface of the space approximately around the 16m receiver, this accumulation is due to the sloped main audience floor surface, and the fact that the nature of the way source and receivers were placed in the space means that direct signal from the source hit the floor surface at an angle. However, observations into measurements at the 16m receiver showed no notable difference in measurement than expected, measurements in that position were also uniform across frequency; suggesting that the random undesired reflective elements in the space, such as the sloped floor, factored a lot less than the intentionally modified room surfaces

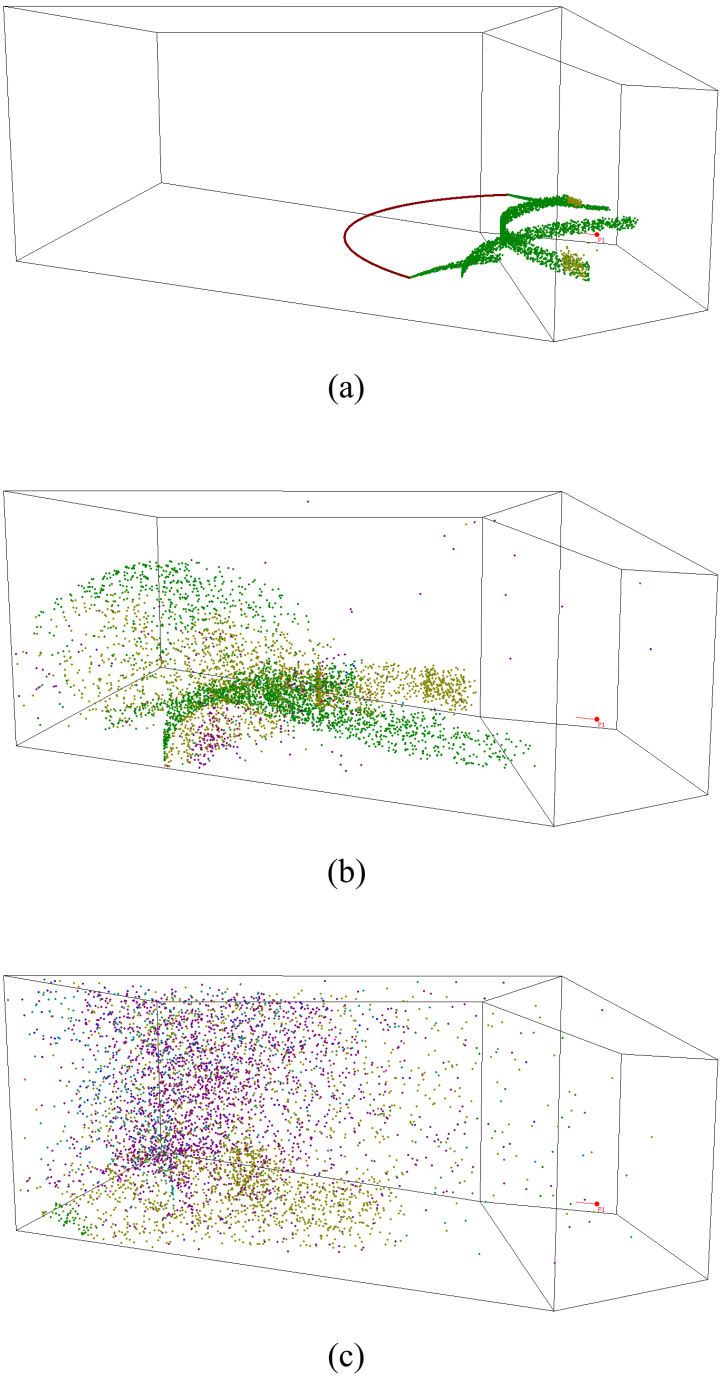

FIGURE 4.12: Visualisation of Early Reflections From Source in Experiment 2, a) at 28ms, b) at 63ms, c) at 90ms

4.6 Experiment # 3 - ODEON Example Space (Extended Scope)

4.6.1 Background

From observations and conclusions for experiment 1 and 2 it became clear that the initial scope and specifics of the experiment methodology was insufficient for informative analysis relating to the two null hypothesis and wider context of this experimental work. The amount of receivers was too few to gain an understanding of how parameters change across the entirety of a space, in experiment # 2 the reflections from the sloped main audience floor may have contributed to variabilitys in measurements, but this could not be perused further without more data points. The overall aims of this experiment are to derive the strength of correlations between independent (source/receiver distance, absorption coefficient) and dependent (EDT, T30, C80). Therefore a significantly large amount of receivers were needed in further experimental work in order to generate data that allows correlative analysis to be reliable and detailed.

A third experiment was designed in the same example space as experiment 2. The reason the example room was used over the real life space used in experiment 1 or a different real life environment, was that the reliability of results meant that observations and further analysis would be robust; an assumption drawn from conclusions of experiment 2, where a space of similar geometry led to more reliable results across frequencies. It was decided that the associative relationships between independent and dependent variables that could be derived were worth the trade off of less informative conclusions from analysis pertaining to a real world context.

In comparison to the four receivers spaced 4m apart for the first two experiments, this third experiment involved 15 receivers each spaced 1m apart from each other along the x axis, this layout is shown in figure 4.13. As a result of observations in experiment 2, it was decided that the receivers would not vary on the z axis, with the height of the source and receivers being increased to account for the sloped surface and its reflective effects. The simulation in this experiment was ran for three different values of absorption coefficient, 30%, 50%, aand 70%; not as a means to also investigate varying absorption coefficient with

these same simulations like the previous two experiments, but to investigate the reliability of these results. The trends that these measurements display across distance are expected to remain the same even if the magnitude of the measurement values is effect. The key data was measured at 50% absorption, the midpoint between a fully reflective and fully absorbent surface, with 30% and 70% measurements used solely as a point of comparison.

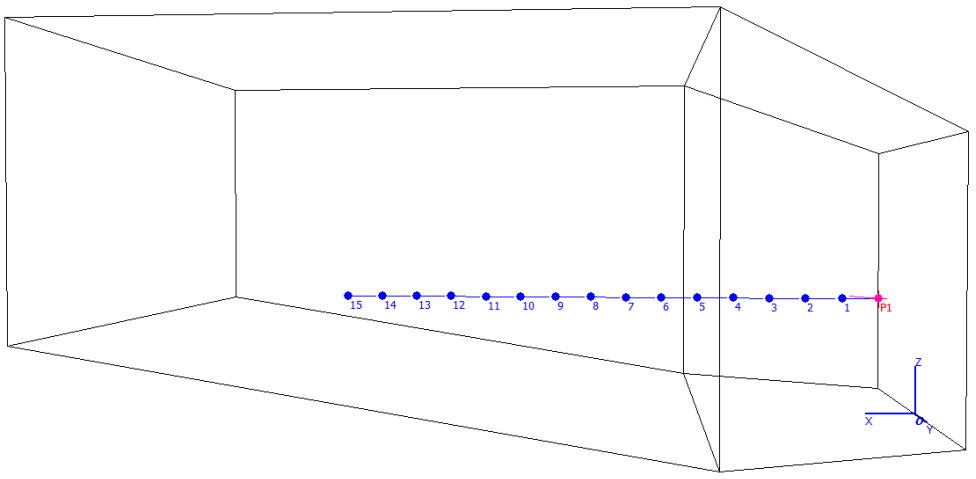

FIGURE 4.13: Source/Receiver placement for Example Room used in Experiment # 3

Absorption coefficient was investigated with a separate series of simulations in this experiment. A similar style of auralisation as shown in figure 4.13. was carried out, however only values from a single receiver of fixed distance 4m were measured. The simulation was run for absorption coefficient values from 0% to 100% in 10% intervals.

In previous experiments, the SPL values of each receiver in each simulation were not considered, it became apparent that it was valuable to also measure SPL values in this permutation of the experiment due to the valuable information that could be gleamed by comparing the change in SPL and the observed measurements. SPL is an objective quantification of the relative volume of sound picked up by a receiver, correlation with measurement trends could provide insight into how the raw volume of a sound is effected by the specified room modifications. Effects on volume are not within the scope and context of this work as initially outlined but could provide valuable observations in general.

As a more in depth and informed experiment, there needed to be considerations for margins of error. ISO3382-1 states appropriate error ranges for each measurement in this test, described as Just Noticeable Difference (JND). If the difference between two values falls within this range then it is stated to have a negligible amount of variability, and it can be assumed any difference within the margins of error can be disregarded.

As stated in the discussion around experiment 2, the results from that previous experiment signified that in ideal spaces (such as the ODEON example space) results would show little to no variability across the appropriate range of frequencies (between 500Hz and 2000Hz) . Therefore for this experiment, in each simulation a single set of results were recorded, this set being the average of recorded values at 500Hz and 1kHz as outlined in the ISO3382-1 standard.

It was necessary to partition the auralisations into two separate series of simulations in order to obtain insights into each independent variable separately, the aim of this experiment was to derive a correlative association between each independent variable and each dependent variable, how source/receiver distance and absorption coefficient effects measurements of EDT, T30, and C80. Because of this the ideal framing of results from this experiment is a series of six analyses between each room element and each measurement; the relative weight of each measurement as a result of varying distance or absorption being derived from these analyses.

4.6.2 Results - Source/Receiver Distance

Initially, values of sound pressure level (SPL) were calculated for each receiver, SPL acts as the measurement of acoustic power via the logarithmic result of the ratio between the pressure from a sound source compared to a reference, weighting curves that map the perceived loudness of the human ear across a range of frequencies are used as the reference. The results of these calculations are shown in figure 4.14; these results were, instead of being measured via the MATLAB analytical toolkit, measured directly within the ODEON program itself, since the toolkit did not have functionality for measuring SPL, the values shown in the figure are the values of the SPL(A) parameter across varying distance for constant 50% absorption. SPL(A) uses the IEC 61672:2003 standardised A weighting curve [73].

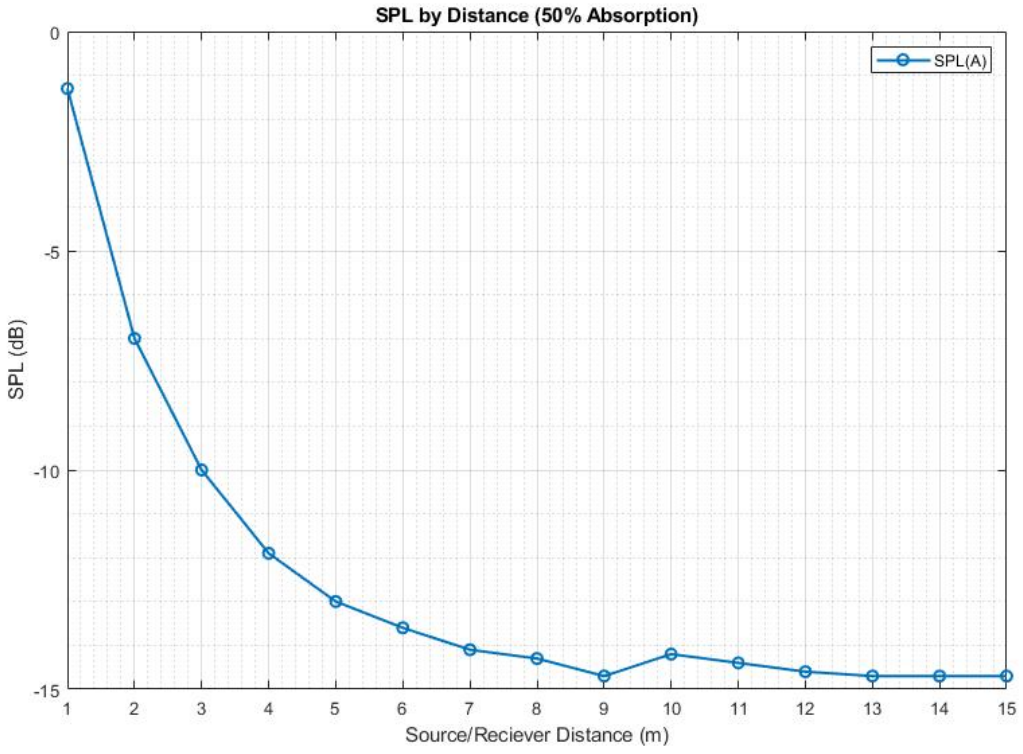

FIGURE 4.14: SPL Results for Single Absorption, Varying Distance Auralisation in Experiment 3

It can be observed that SPL undergoes an exponential decrease as source/receiver distance increases, demonstrating a potential inverse proportionality.

The measurements of EDT in the single absorption coefficient, varying distance auralisations are shown in figure 4.15, while comparisons between all the simulations (30, 50, 70%) and sound pressure level are shown in figure 4.16:

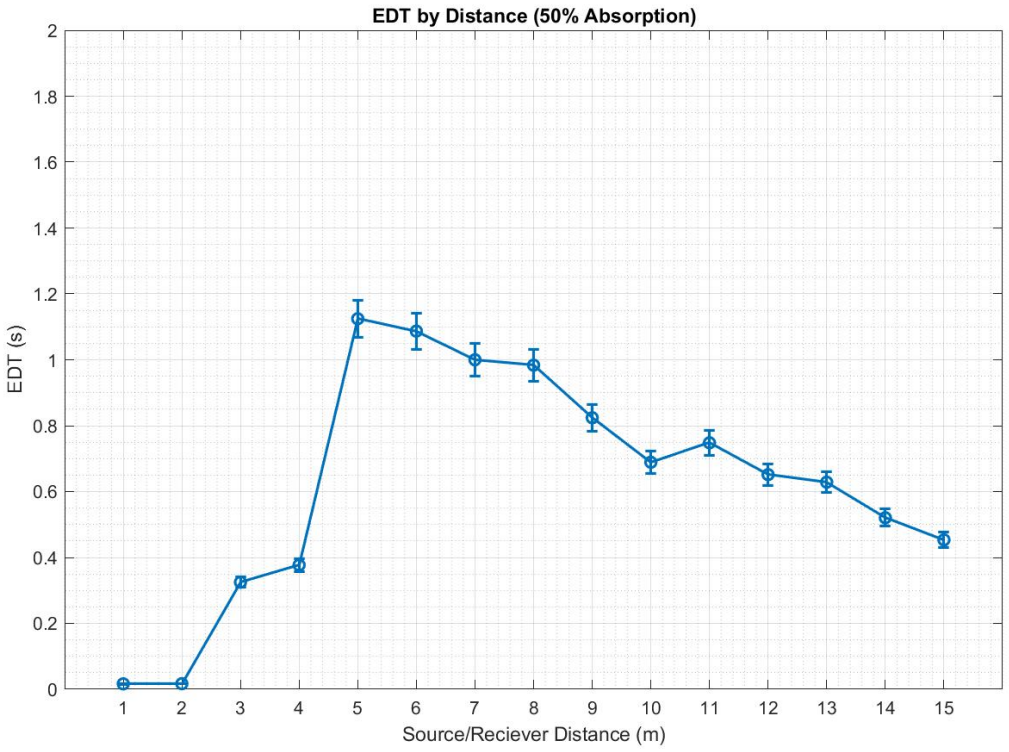

FIGURE 4.15: EDT Results for Single Absorption, Varying Distance Auralisation in Experiment 3

It is clear that the rise and fall of EDT as distance increases is on the whole not uniform, there are distinct spikes at 3m, 5m, 7m, and 11m in comparison to the more subtle incremental changes from other receiver measurements, this suggests a lack of direct proportionality between EDT and source/receiver distance within the context of this experiment. These anomalous values could be the result in IR rendering or analysis errors as previously described in this section, it could be a factor of the relative unreliability of EDT measurements with varying distance, which as previously discussed in section 2.2 is an innate property of EDT as a parameter.

Looking at figure 4.16, the general rise and fall trend across distance is replicated in simulations with 30% and 70% absorption, with the magnitude of EDT values decreasing as absorption coefficient increases. The distinct spikes shown in the 50% simulation are not replicated in the other simulations, with spikes occurring at different receivers, lending

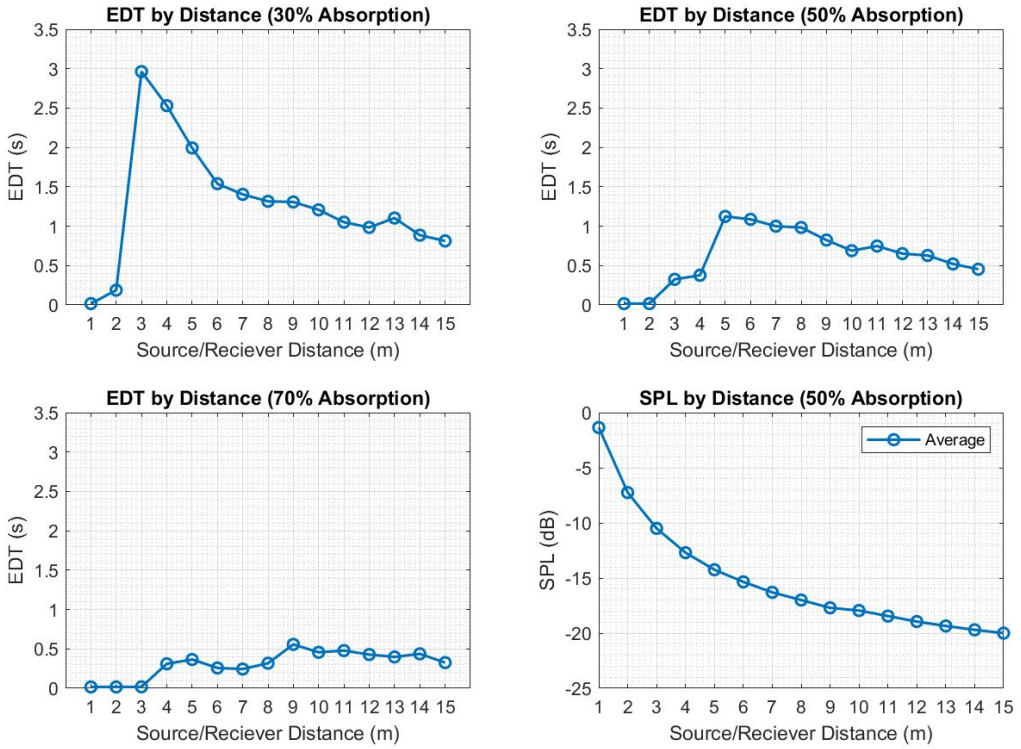

FIGURE 4.16: Comparisons Between EDT Results & SPL for Auralisations in Experiment 3

more evidence to the conclusion that EDT values are subject to random errors from the methodology used in this work.

A linear regression analysis was performed on this data, the result was a general equation with a single coefficient and y intercept to describe how distance factors into the measurement of EDT, this equation is shown below, where x is the source receiver distance:

$$EDT = 0.02355x + 0.4413 \tag{4.4}$$

The correlation coefficient between EDT and source/receiver distance was calculated to be *0.3002* .

There were attempts to create a polynomial regression for linear proportionality which fit the data, but lower order polynomials would not provide sufficient results. The decreasing rate of EDT decrease, from the sharp drop to little change after that point, implies an inverse exponential relationship between EDT and absorption. It was therefore decided to

investigate exponential proportionality. A linear regression was performed on the natural log of EDT instead, the result $ln(y) = ax + b$ would be in the form $y = be^{ax}$. Due to this, the EDT equation in terms of source/receiver distance and absorption coefficient can be written as shown

$$EDT = 0.02355d - e^{0.0635a} \tag{4.5}$$

However, a small regression gradient of 0.0635 means that the exponential $e^{0.0635} = 1.065$, meaning that this exponential can be considered as simply 1 within a reasonable level of precision. This means that through the regression of the natural log of results a linear relationship between EDT and absorption has been established. The expression than therefore be rewritten as shown below

$$EDT = 0.02355d - 1.065^{a} \tag{4.6}$$

$$EDT \propto d$$

$$EDT \propto -e^{a}$$

The measurements of T30 in the single absorption coefficient, varying distance auralisations are shown in figure 4.17, while comparisons between all the simulations (30, 50, 70%) and sound pressure level are shown in figure 4.18:

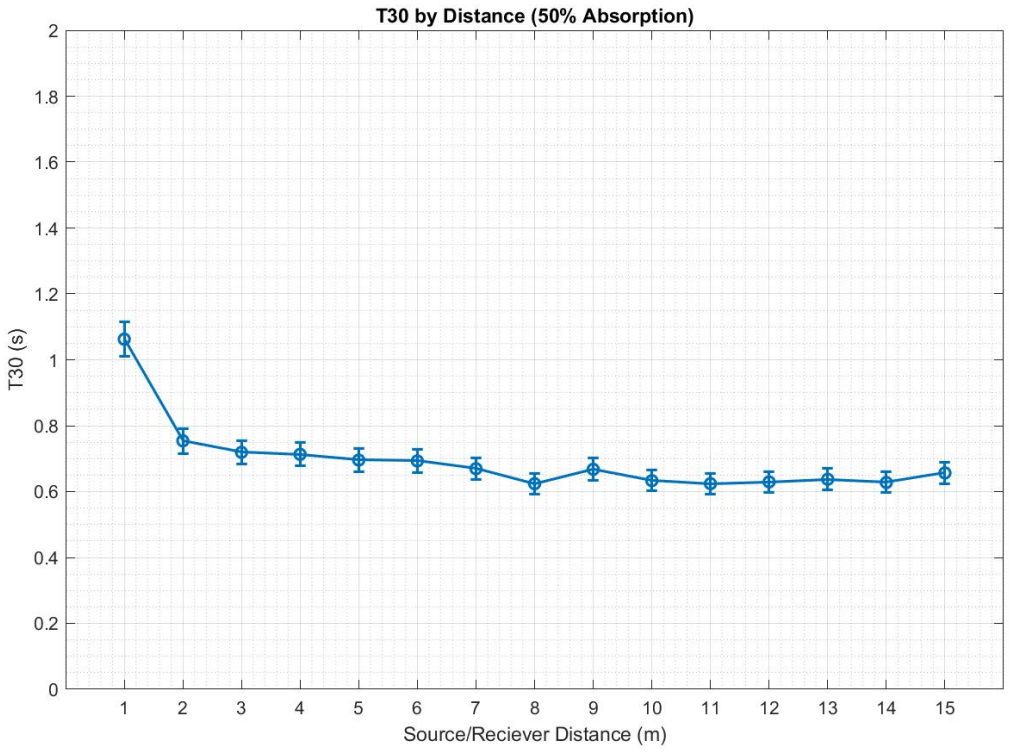

FIGURE 4.17: T30 Results for Single Absorption, Varying Distance Auralisation in Experiment 3

Through initial observations of figure 4.17 an inverse exponential like curve can be observed, however this only shows results for the 50% simulation. Using figure 4.18 to compare results of simulations with 30% and 70% coefficients it can be observed that outside of the receiver at 1m there is little to no change in T30 value at all, this point is remphaised through looking at the error bars in figure 4.17, which show that all measurements past the 1m receiver fall within the JND error range, meaning there is noticeable difference across the 2-15m range. The anomalous 1m value changes it's level of deviation from the measurement trend as absorption coefficent increases, meaning this receiver is producing less reliable results, it was therefore assumed that this receiver could be disregarded in further analysis as a true anomalous value.

A linear regression analysis was performed on this data, the result was a general equation with a single coefficient and y intercept to describe how distance factors into the

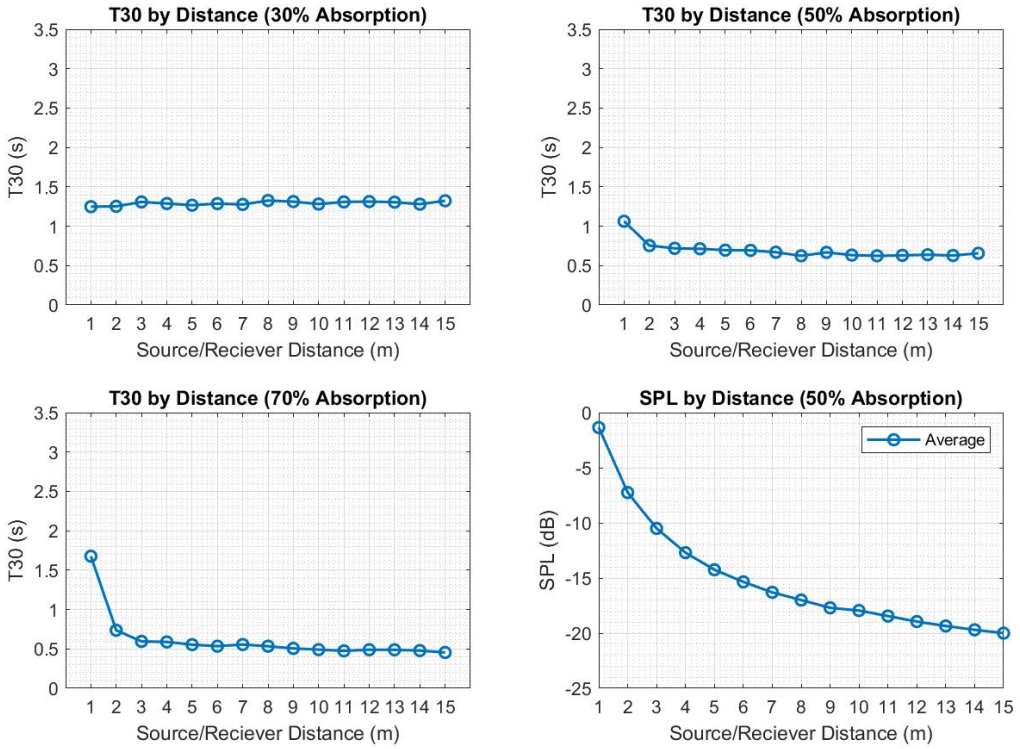

FIGURE 4.18: Comparisons Between T30 Results & SPL for Auralisations in Experiment 3

measurement of T30, this equation is shown below, where x is the source receiver distance:

$$T30 = -0.016711x + 0.82742 \tag{4.7}$$

In this case, the coefficient is so small that the overall effect of x is stated as negligible, therefore it can be stated that source/receiver distance does not have any proportional relationship with T30.

The correlation coefficient between T30 and source/receiver distance was calculated to be *-0.6814* .

The measurements of C80 in the single absorption coefficient, varying distance auralisations are shown in figure 4.19, while comparisons between all the simulations (30, 50, 70%) and sound pressure level are shown in figure 4.20:

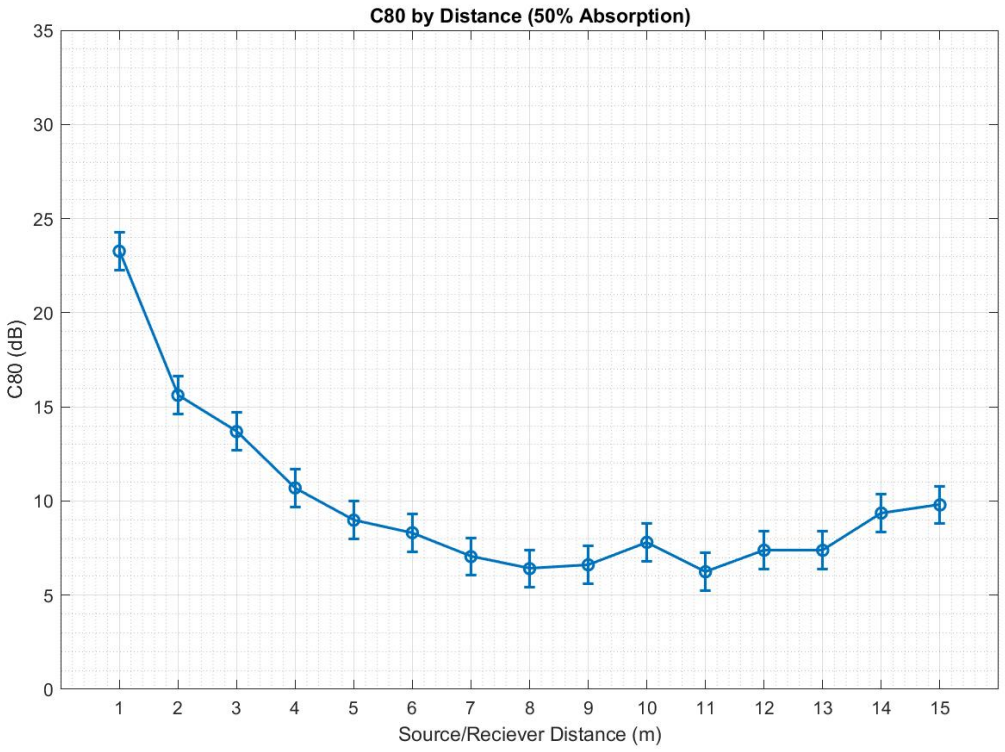

FIGURE 4.19: C80 Results for Single Absorption, Varying Distance Auralisation in Experiment 3

It can be seen that C80 values display a exponential like decrease over distance. It can be noted that the drop in C80 across 2m to 8m is on the whole beyond margin of error, and can therefore be stated as a significant observable decrease in C80, beyond this range there are fluctuations and readings that lie frequently within the JND range, suggesting that beyond 8m C80 is not really effected by distance.

As with EDT and T30, a linear regression analysis was performed on this data, the result was a general equation with a single coefficient and y intercept to describe how distance factors into the measurement of C80, this equation is shown below, where x is the source receiver distance:

$$C80 = -0.66551x + 15.229 \tag{4.8}$$

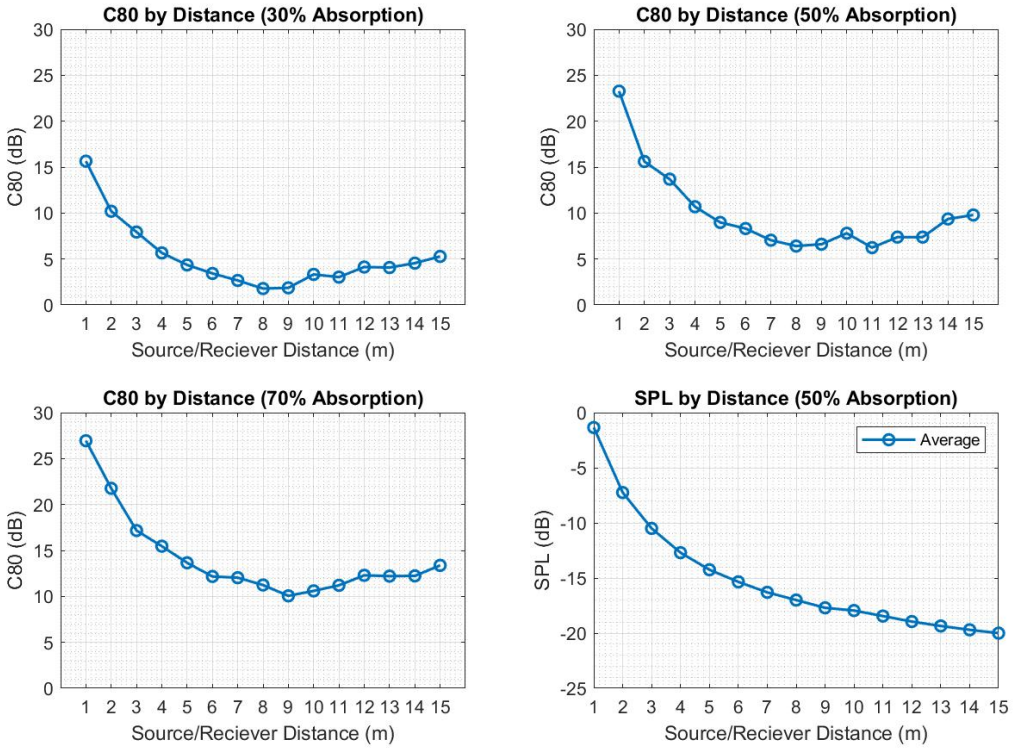

FIGURE 4.20: Comparisons Between C80 Results & SPL for Auralisations in Experiment 3

It can be noted from this regression equation that the distance coefficient is negative (C80 decreases as distance increases) and that the coefficient is a larger proportion of the y intercept than both EDT and T30

The correlation coefficient between C80 and source/receiver distance was calculated to be *-0.6521* .

4.6.3 Results - Absorption Coefficient

The measurements of EDT in the single distance, varying absorption coefficient auralisations are shown in figure 4.21:

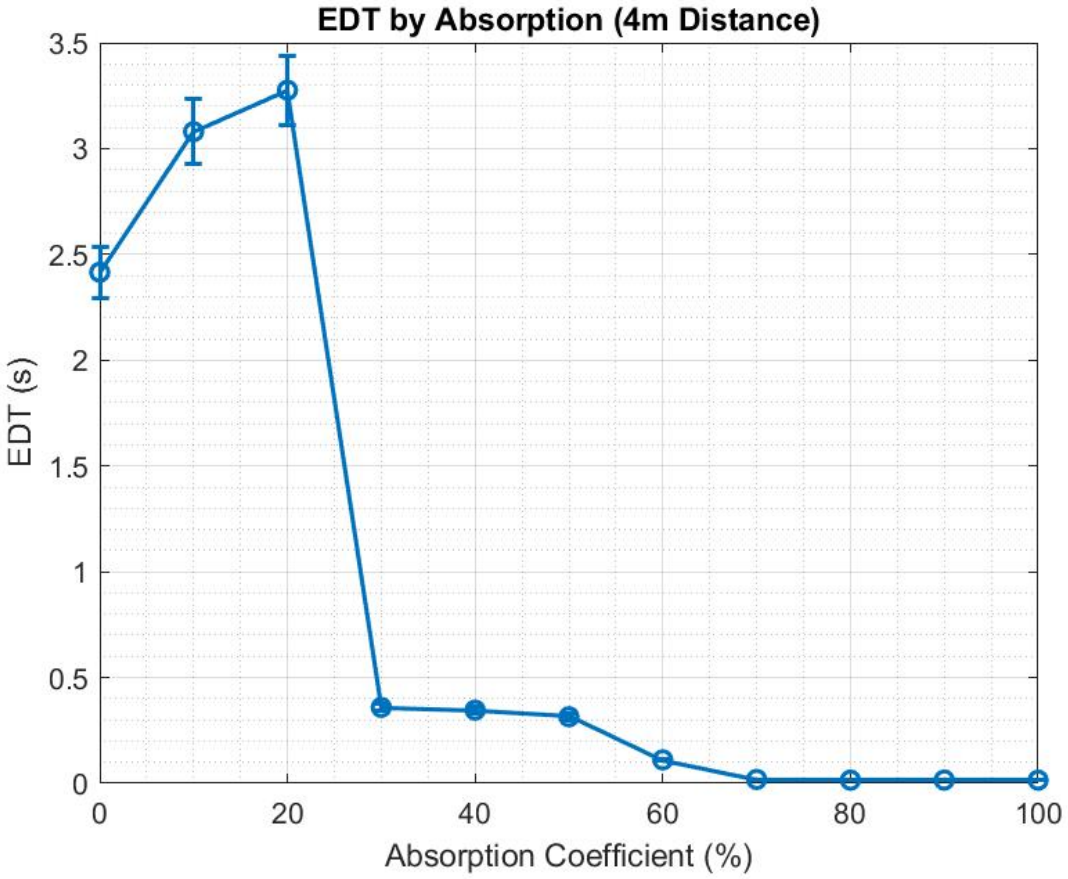

FIGURE 4.21: EDT Results for Single Distance Auralisation in Experiment 3

There is a distinct exponential decay shown as the absorption coefficient increases, the error bars show that this decay occurs beyond margins of error. In contrast to how EDT was evaluated with distance, the trend shown in figure 4.21 is more uniform demonstrating the fact that EDT is a relatively unstable parameter to measure with varying distance

A linear regression analysis was performed on this data, this equation is shown below, where x is the source receiver distance and e is the error:

$$EDT = -0.031781x + 2.4951 + e \tag{4.9}$$

The correlation coefficient between EDT and absorption coefficient was calculated to be *-0.7994* .

The measurements of T30 in the single distance, varying absorption coefficient auralisations are shown in figure 4.22:

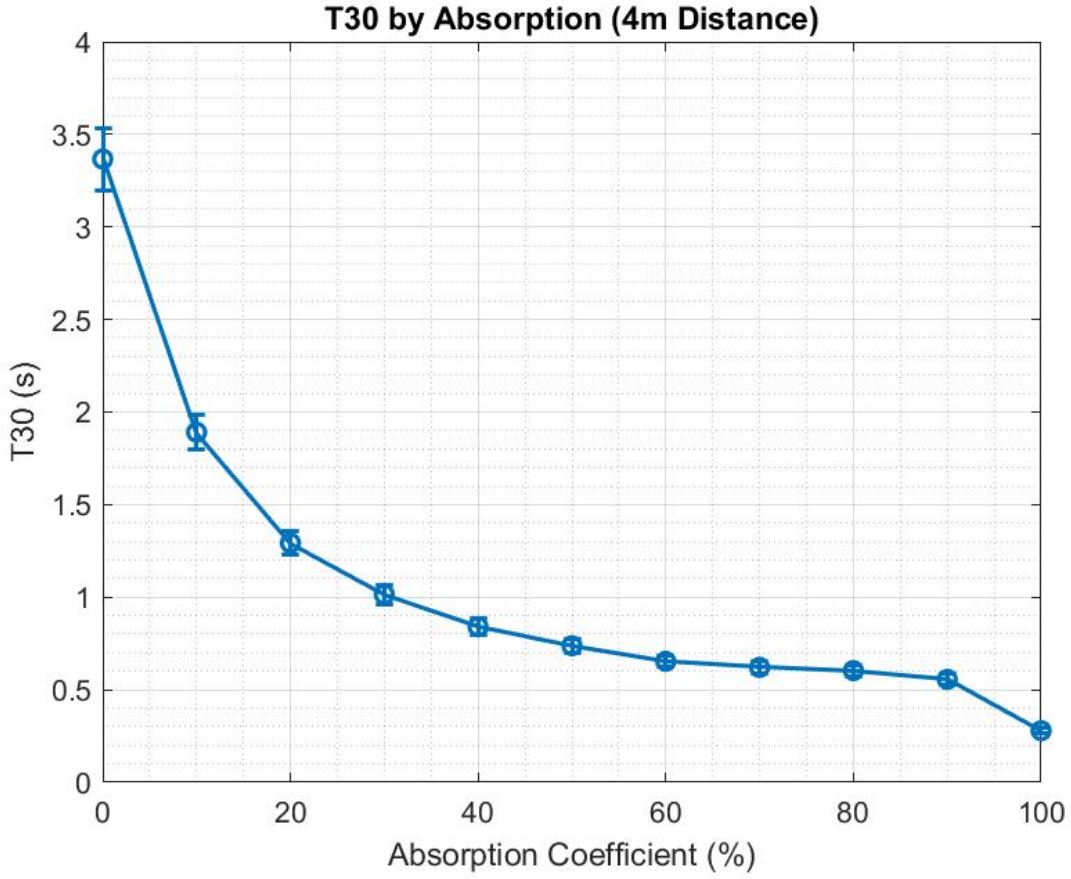

FIGURE 4.22: T30 Results for Single Distance Auralisation in Experiment 3

The general trend of the T30 measurements is almost exactly the same as EDT, the significance in this shared trend is that this is contrast with the results when distance was the independent variable. Both EDT and T30 are different types of decay time measurement, and they share similar derivation methodologies. It can be concluded from this data that both EDT and T30 are effected significantly by a change in absorption.

A linear regression analysis was performed on this data, the result was a general equation with a single coefficient and y intercept to describe how distance factors into the measurement of T30, this equation is shown below, where x is the source receiver distance and e is the error:

$$T30 = -0.021658x + 2.1608 + e \tag{4.10}$$

The correlation coefficient between T30 and absorption coefficient was calculated to be *-0.8214* .

A similar approach to EDT was undertaken with T30, which also displays a distinct exponential decay as absorption coefficient increases as shown in figure 4.22. Doing a linear regression with the absorption coefficients and log of resultant T30 values as previously calculated with EDT gives the result shown below

$$T30 = -e^{0.0190a} + 0.8009 \tag{4.11}$$

Since distance does not factor into T30, the overall equation can be displayed as shown.

$$T30 = -e^{0.0190a} \tag{4.12}$$

However, a small regression gradient of 0.0635 means that the exponential $e^{0.0190} = 1.0192$. This means that through the regression of the natural log of results a exponential relationship between EDT and absorption has been established. The expression can therefore be rewritten as shown below:

$$T30 = -1.0192^{a} \tag{4.13}$$

$$T30 \propto -e^{a}$$

The measurements of C80 in the single distance, varying absorption coefficient auralisations are shown in figure 4.23:

A linear regression analysis was performed on this data, the result was a general equation with a single coefficient and y intercept to describe how distance factors into the measurement of C80, this equation is shown below, where x is the source receiver distance:

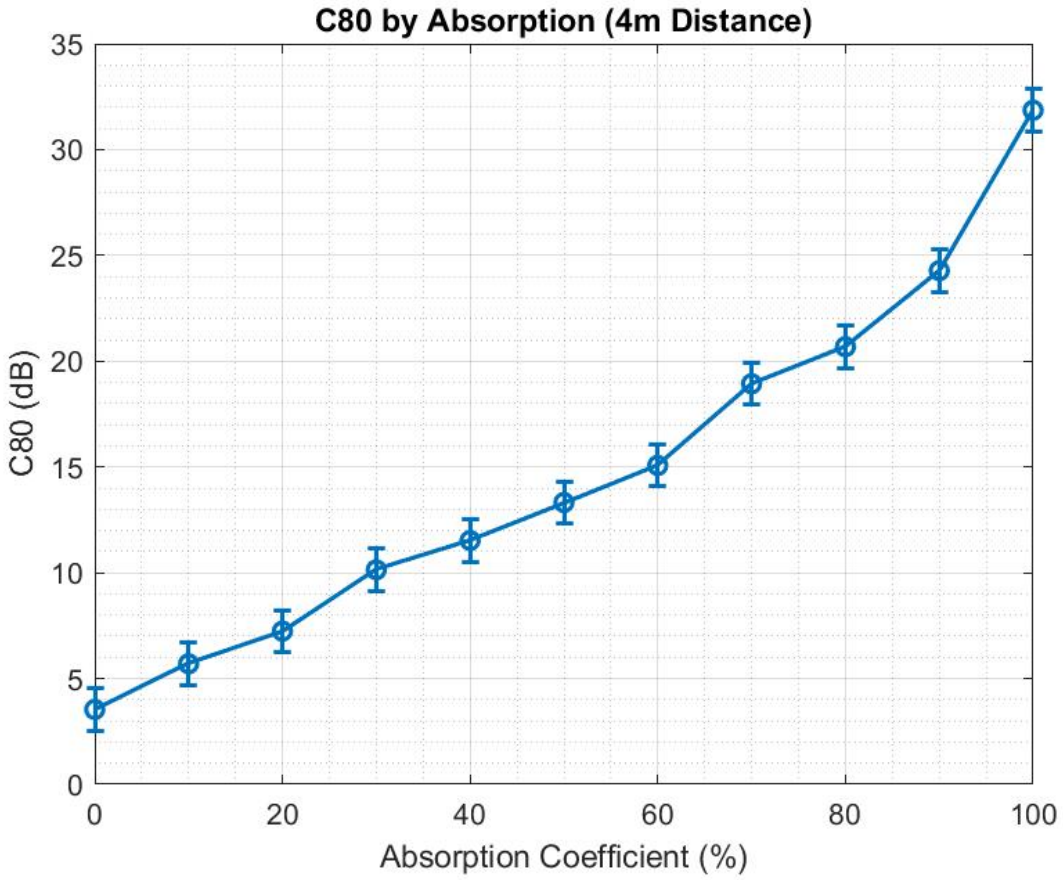

FIGURE 4.23: C80 Results for Single Distance Auralisation in Experiment 3

$$C80 = 0.25217x + 2.1444 \tag{4.14}$$

The correlation coefficient between C80 and absorption coefficient was calculated to be *0.9762*.

For C80, a similar series of analyses were performed to EDT and T30. It is shown in figure 4.19 that there is a distinct exponential decay in C80 as distance increases, therefore a linear regression was performed with he natural log of C80 values to give the resultant equation shown below

$$C80 = -e^{0.0535d} + 2.6478 \tag{4.15}$$

Figure 4.23 shows a more linear relationship between C80 and absorption coefficient, so

therefore the linear regression performed earlier would still be sufficient. Similarly to EDT and T30 for absorption coefficient, the exponential expression for C80 over distance can be expressed as $e^{0.0535} = 1.0550$, meaning the whole C80 expression can be rewritten as shown:

$$C80 = -1.0550^{d} + 0.2522a \tag{4.16}$$

$$C80 \propto -e^{d}$$

$$C80 \propto a$$

4.6.4 Discussion

It is worth evaluating the results and derived regression equations in relation to the two null hypothesis states at the beginning of this chapter. The first null hypothesis was 'The distance of a receiver relative to the source within any defined acoustic environment has no effect on measured C80 clarity'. It can be seen in figure 4.20 that there is an inverse exponential proportional relationship between C80 and distance, this provides evidence for the null hypothesis not being correct; data gathered in this experiment outlines that C80 can be decreased by moving the receiver further back and vice versa.

The aim of this experiment was to draw from the methodologies for the first 2 experiments and expand on them in order to derive approximate polynomial equations for how C80, EDT and T30 are effected by varying source/receiver distance and absorption coefficient. The regression analysis used in this work is a simplified rudimentary approach, it can be seen in the results that there are explicit exponential decay curves that don't fit well in standard linear regression analysis, but on the surface these equations can provide insight into how one element of room modification impacts these measurements in comparison to another. And in also drawing from previously stated observations, general conclusions can be drawn about the significance of each room modification on each measurement.

For EDT, in a general sense it can be noted from the regression equations 4.3 and 4.6 that EDT increases with increased source/receiver distance and decreases with increasing absorption coefficient. However, as outlined in the results shown in figure 4.15, the distance factor is inaccurate due to a potential large range of factors, and previous experiments

showed large amounts of non-uniform variability in results. Therefore the conclusion that EDT has any proportionality with source/receiver distance is less reliable than desired. Nevertheless, despite the general unreliability of EDT results by distance, figure 4.15 also shows a linear decrease in EDT from 5-15m, so this conclusion cannot be entirely discounted.

In contrast, there is little to no proportionality shown as EDT varies with absorption coefficient, shown in figure 4.21. Attempts to perform a linear regression with this dataset produce sub optimal results, which can be attributed to the single sharp decrease in EDT between 20 and 30% absorption, with little changes in EDT both before and after this point.

Exponential proportionality calculations show that EDT is not affected significantly by distance, but is inversely exponentially proportional with absorption coefficient. From the observation of plots for EDT data it can be concluded that this expression is unreliable, with plots for both changing distance and changing absorption failing to exhibit the distinct exponential decay expected for the measurement value. It was concluded that the practical implementation of these associations towards deriving an overall semantic representation for the perceptual terms in this research work should be undertaken with this conclusion in mind.

For T30, results shown in figures 4.17 and 4.18, as well as the results from the first two experiments, demonstrate that there is little to no effect on T30 from distance, a conclusion reinforced by the underlying theory. Therefore in this analysis it is discounted, with a sole focus on the effects of absorption coefficient instead. For absorption, a large correlation coefficient value (-0.6814) lends evidence that there is a proportionality shown, but through investigating results this assertion becomes to unreliable to pursue further.

For both EDT and T30, it is shown that there is an inverse exponential proportionality between measured values and absorption coefficient, but no proportionality between measured values and distance, disproving the first null hypothesis outlined in section 4.2.2.

Calculations for C80 show exponential proportionality between C80 and source/receiver distance and linear proportionality between C80 and absorption coefficient, with the latter point disproving the second null hypothesis stated in section 4.2.2.

Despite C80 only being a few orders of magnitude more than the decay times in terms of a typical numerical value, polynomial regression equations for this value contain coefficients for both distance and absorption many orders of magnitude higher than coefficients for wither other variable. This suggests that C80 is more effected than other parameters as these spaces are modified. The wider implications from this conclusion are that the decay times (EDT and T30), while evaluated as one of the most important aspects of received and measurable reverb in existing literature, do not factor into the context of this research work as much as C80 (clarity) does.

Chapter 5

Perceptual Acoustic Testing - Listening Test

This chapter outlines the experimental work undertaken in researching how perceptual factors relating to audio, in this case the terms 'brightness' and 'closeness', are affected by changes in audio stimuli, specifically relating to sound sources with differing acoustic qualities; in this case, stimuli with varying EDT, T30, and C80 measurements. In order to derive these conclusions, work was undertaken on a listening test to generate perceptual data from participants relating to the strength of the described terms.

This chapter contains the following sections:

Section 5.1, Background. This section involves discussion around the wider research work in perceptual sensory testing, providing context for the specific implementation of perceptual testing principles in this listening test; it also outlines the MUSHRA listening test methodology, which this test draws key elements from.

Section 5.2, Aim, Variables, and Null Hypotheses. This section outlines the purpose of the listening test as a whole, as well as outlining and explaining what specific variables will be measured and changed in this test, and stating the null hypothesis for which the results from this test will be compared against.

Section 5.3, Listening Test Development. This section involves discussion around the process of developing the format, stimuli, and statistical analytical framework of this

listening test, in going over this process useful conclusions which can be applied to further work in this topic are discussed.

Section 5.4, Methodology. This section provides a description of the format and practical implementation of the listening test, and the structure of the listening test itself.

Section 5.5, Results and Discussion. This sections outlines the results from the listening test, in addition to a series of analyses on the results to evaluate the validity of the previously stated null hypothesis, as well as to evolute the effectiveness of the methodology used and to point out interesting conclusions that be drawn from the data.

Section 5.6, Conclusion. This section summarises the listening test work and results and derives the proportionalities between the previously outlined independent and dependent variables, in order to being this work within the context of the larger mathematical model.

5.1 Background

In order for the prospective semantic interface outlined in this project to function, the perceptual terms of brightness and closeness would need to have an association with an element of room modification. While the auralisation experiments were designed to lead to derived expressions for acoustic measurements EDT, T30, and C80, in terms of these room modifications, a perceptual listening test was created and distributed to generate data around how 'brightness' and 'closeness' were perceived by users in terms of the aforementioned acoustic measurements.

Subjective listening tests are typically used in research to outline the desirability of and preference for certain acoustic characteristics within a sound. The listening test developed in this research work is designed to assess how the variability of an objective factor (be that EDT, T30, or C80) in a set of sound stimuli effects the qualitative subjective perception of a sound's 'brightness' or 'closeness' , subjective interpretations of objectively measurable aspects of sound. The overall methodology of this work was designed to not only draw from the subjective testing methodologies both in the context of audio and outside of it, but to also draw from measurement and objectivity based audio test methodologies, building from the literature outlined in section 2.4.

The methodology of the listening test used in this work draws heavily from the MUSHRA listening test methodology, but importantly is not a form of MUSHRA test in itself, the primary difference being that this test is being used in a subjective analytical context while a conventional MUSHRA test focuses on observable measurable aspects of acoustics and how they are perceived by experts in the field. The broad scope of this work involves the development interfaces where objective elements within a space can be modified through subjective means by a non-expert end user. Therefore it was decided that in gathering data to derive expressions for perceptual terms, a listening test needed to be developed for both expert and non expert end users. Data for how non experts perceive 'brightness' and 'closeness' is still valuable in the research's wider context even in the event that non expert results differ from expert results.

5.2 Aim, Variables, and Null Hypotheses

The primary aim for this listening test wass to investigate how different acoustic environments lead to changes in how sound is perceived by a listener. The specific end goal of this part of the research was the derivation of expressions for how the perceptual terms of 'brightness' and 'closeness' change as acoustic measurements of C80 clarity and EDT/T30 decay time change. It was important that this listening test was built around associations between subjective perceptual terms and objective acoustic measurements, following the overall project plan outlined in chapter **??**. Evaluating brightness against C80 means there is an understanding of what is happening within the acoustic environment if a correlation is found.

The independent variables used in this project were the subjective perceptual terms of brightness and closeness, and the dependent variables were varying values of EDT, T30, and C80 in each sound sample used in the test.

As this test contains 2 independent and 3 dependent variables, six null hypotheses can be outlined.

1. Perceived brightness of a sound will not have direct proportionality with changing EDT

2. Perceived brightness of a sound will not have direct proportionality with changing T30
3. Perceived brightness of a sound will not have direct proportionality with changing C80
4. Perceived closeness of a sound will not have direct proportionality with changing EDT
5. Perceived closeness of a sound will not have direct proportionality with changing T30
6. Perceived closeness of a sound will not have direct proportionality with changing C80

As previously discussed, the plan was to conduct listening tests with three different types of sound sample (percussive, melodic, and spoken), but instead of sound samples acting as independent variables, this work was imagined as three separate versions of the same test; with three different types of sound source used as stimuli, and the sound source type in each experiment being defined as a control variable. The variability of results between these three sound sources are not prioritised, and are instead used to cross reference results from the test and to derive interesting conclusions to how results are effected by this variable.

5.3 Listening Test Development

In developing the listening test for this project, a number of key questions needed to be tackled. There needed to be an understanding of how the sound sets within the test and the test itself would be structured and formatted, what stimuli would be used for sound sets in this test, and how the data generated from the test would be analysed. To achieve this, work was undertaken to investigate perceptual testing approaches outside the field of audio, as well as audio based methodologies such as MUSHRA, to develop a listening test methodology that fit the context of the aims and goals of this work and the variables being assessed.

Initially the structure of the listening test was designed as a type of MUSHRA test, primarily though the question structure, the scale used by participants to record results, and use of an hidden reference. The reference/stimuli set approach used in MUSHRA, where each

question featured a series of sound stimuli with one varying independent variable across the set being scored on the strength of a perceptual factor in comparison to a 'neutral' reference stimuli, was used as the main structure for sound sets in this listening test. Each of these 'questions' can be described as 'sound sets', literally meaning sets of sounds that a participant is given to assess and numerically score based on a certain factor.

It became clear that for the context of the overall research, the listening test would need to draw from the MUSHRA method while differing from it in a few key ways. The most important difference is that MUSHRA is designed around sensory interpreations of clearly quantifiable measurements, and the observation of those measurements by participants. In comparison, this listening test the assessment is built around less clearly defined subjective terminology. The null hypotheses imply that these subjective terms relate to objective measures, but in the test itself a participant is being asked to relate stimuli not to 'quality' but a defined perceptual term, and are therefore implicitly being asked about their interpretation of said term.

Another key difference is that MUSHRA relies on sourcing expert listeners; whereas for this work, due to both practical limitations and desire to have this work within the context of non experts interacting with the proposed final semantic interface, expertise was not prioritised when sourcing participants. Other aspects such as the scoring scale, and analysis framework also differed from MUSHRA, meaning that it was not entirely true that this test was MUSHRA, instead it drew influence from MUSHRA in its development.

In order to generate an appropriate set of sound stimuli for this listening test, a variety of factors were taken into account. In a basic sense, stimuli within this listening test will consist of a certain sound source within a certain acoustic environment. In a practical sense this can be acheived with pre-recorded sound files used in an online listening test environment via the convolution of a chosen sound source and a room impulse response of either a measured or emulated acoustic environment. Referring back to the aims and objectives of the listening test experiment, the sound source itself can be considered a control variable which does not change, while the impulse response it is convolved with is representative of the independent variables (EDT, T30, C80) that will be modified.

Through this conceptualisation of the sound stimuli there emerges a problem. It is impractical to create impulse responses to provide exactly defined values of EDT, T30, and

C80. These are measurements that are merely reflective of the environment itself and to to change the measurement values the environment itself must be modified in some way. The complexity of acoustic systems means it is difficult to map changes in the environment to uniform step changes in a particular measurement. While EDT, T30, and C80 were ultimately the parameters the results of this test were evaluated against, the practical design of the test considers the changing independent variables to be physical aspects of the environments that can be modified. As these physical aspects are changed then the acoustic measurements will also change.

Drawing from conclusions derived in the auralisation testing outlined in chapter4 4, it was known that EDT/T30/C80 have various correlations as the source/receiver distance and absorption coefficient with an acoustic environment also changed. With the knowledge of these proportional relationships, it was decided that for the practical development and deployment of this listening test, source/receiver distance and absorption coefficient would be the independent variables that would be changed, and that previous auralisation test results would be analysed to derive in what ranges for those independent variables there was a linear increase or decrease in EDT/T30/C80.

Evaluating the results shown in section 4.6.2 that show auralisation experiment results for changing source/receiver distance, there is a distinct linear decline in C80 between 2m and 8m. For T30 values are consistent with no irregularities outside of the 1m receiver anomalous result, and for EDT there is a linear decline from 4 to 15m. Since EDT results were unreliable across distance and T30 results did show variability across distance at all, the trends shown for it were de-prioritised in this context. It was decided as a result that a range of 2m to 8m was appropriate for the independent variable of source/receiver distance in the generated impulse responses. While values for C80 were not guaranteed to be at regular intervals, there was enough confidence that working within this range would lead to linear decreases in C80 as distance increased, therefore allowing for the assessment of how the strength of perceptual terms changes with changing C80.

Evaluating the results shown in section 4.6.3, showing results for absorption coefficient, one can observe that for EDT and T30 measurements there is a distinct decay from 10% to 70%, and for C80, there is a linear increase across the full 0-100% range. It was therefore decided that to ensure there is a linear trend of EDT and T30 decay times via changing absorption coefficient, IRs would be generated between the 10% and 70% ranges. As

previously stated when discussing how C80 related to distance, the values of EDT and T30 from generated IRs at uniform intervals of absorption coefficient may not generate results at uniform intervals, but there is confidence from earlier results that these decay time results will trend linearly and decrease as absorption increases.

In generating a series of impulse responses, IRs would be generated at a set number of receiver distances and a set number of absorption coefficients. It was decided as a trade off between the granularity of each sound set and the practical scope of the test that 4 ordinal categories were sufficient for each independent variable. For distance, values of 2m, 4m, 6m, and 8m were used. For absorption, values of 10%, 30%, 50%, 70% were used. An impulse response set was created, with IRs at every distance and at every absorption coefficient. The resultant full IR set can be visualised as shown in figure 5.1:

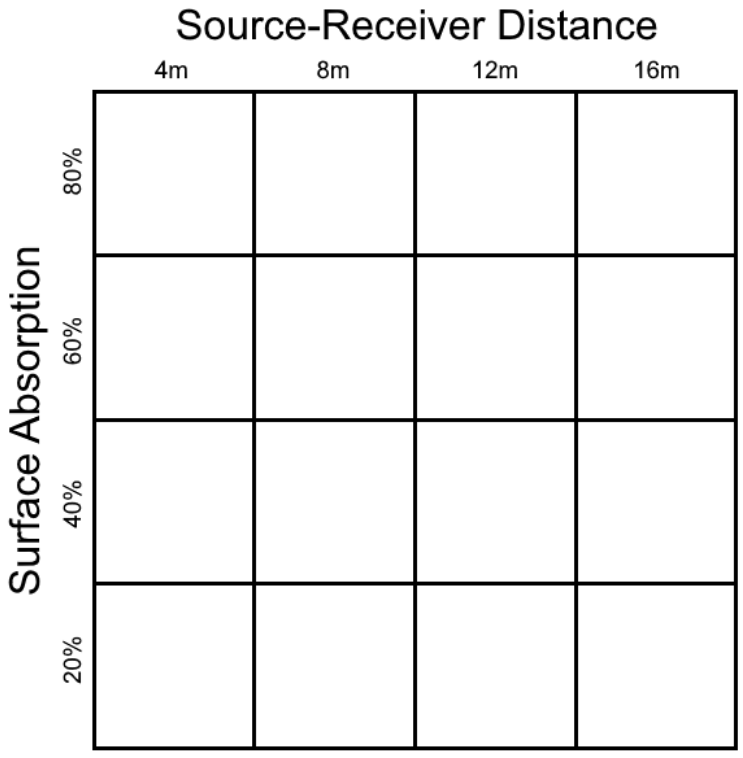

FIGURE 5.1: A visualisation of the impulse responses generated for the listening test

For each sound set, a single set of stimuli are assessed against a reference, in each sound set one variable will change across stimuli and one will remain the same, meaning a sound set could be using impulse responses of varying distance or varying absorption but not both within the same sound set. This can be visualised as shown in figure 5.2, where it can be observed that 8 groups of impulse responses can be created, leading to 8 sound sets in the test for each perceptual test.

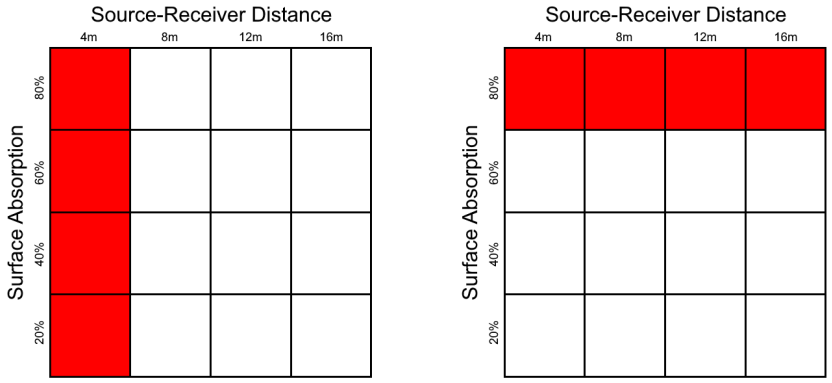

FIGURE 5.2: A visualisation of the sets of impulse responses used for each listening test sound set

To generate sound stimuli, the impulse responses needed to be convolved with an appropriate sound source. The MUSHRA methodology standard [54] outlines that a sound source, described as 'critical material' in the standard, needs to be considered as feasible broadcast material. It was decided upon that 'broadcast material' in the context of MUSHRA testing could be interpreted as performed live material within the context of this work. From this it was initially decided that a melodic source, a sound sample of singing, would be used as the source to be convolved with the impulse responses. However as the project developed it became clear that there could be usefulness in doing this same test for different types of sound source, not only a melodic sample, but also a percussive sample and spoken sample, in order to fully represent the types of sound typically heard within live performance and recorded sound contexts. Previous studies have indicated that spoken, melodic and percussive sounds do indeed lead to different perceptions of reverb in large environments [74].

5.4 Methodology

As previously stated, the overall listening test is in fact a combination of three variants of the same listening test, with each one assessing a different type of sound stimuli. However in the context of the actual listening test that was designed, data for all three of these variants was gathered in a single test. From the participants' perspective the listening test was an assessment of how three types of stimuli (percussive, melodic, speech) effected the strength of two perceptual terms (brightness, closeness). The dependent variables for this test (EDT, T30, C80) were aspects of the impulse responses used for the convolutions that generated the percussive, melodic, and speech sound samples used in sound sets. These impulse responses were the same for convolutions of each type of sound source, an IR at 2m and 40% absorption was used to generate the percussive 2m/40% sound file, the melodic 2m/40% sound file, and the speech 2m/40% sound file; so it is assumed that results for each of those sound sources are in fact results for the 2m/40% impulse response, and its values of EDT, T30, and C80 accordingly. EDT, T30, and C80 values fir the impulse responses convolved in this test are shown in Appendix D.

Typical listening test experiments involve normalising sound sources to a constant sound level in order to ensure that fluctuating volume does not factor into how participants assess individual samples. It was decided that for this specific listening test that this would not be the case, the primary reason for this being that the main differentiator for IRs of varying source/receiver distance was sound level. However, this aspect of the sound stimuli may also factor into the strength of a perceptual term rather than the distance itself (for example a sound could sound 'more bright' because of how comparatively loud it is rather than it's observed spatial position related to the receiver. It was decided that this was a worthwhile trade-off in order to not undermine the aforementioned source/receiver distance variable being assessed.

Time and COVID-19 constraints meant that certain compromises needed to be made in comparison to a typical listening test. It was decided that the listening test would be conducted remotely via an online test using the Qualtrics online survey software. This approached helped significantly in mitigating the logistical problems of organising in person testing in the middle of a pandemic, albeit with obvious drawbacks. With remote testing there is less explicit control over the listening conditions for participants, meaning the

room tests are conducted in and what type of headphones/speakers are used. It was hoped that the effects lack of a set of controlled listening conditions would be offset by the fact that subjective assessment undertaken in this listening test would be in comparison to a hidden reference, and the perceptual strength scores for each participant would be in comparison to their score for the hidden reference.

For example, if a listener A scored the hidden reference for a sound set at 0 and participant B scored it 1, then the relative strength of other sounds within a set for listener A would be how much more or less than 0 each sound scored, and for listener B it would be how much more or less than 1 each sound scored. The assumption being made here is that different listening conditions for participants and will effect every sound being assessed by the same amount, which is a simplification considering that two distinct perceptual terms are being assessed and three different types of sound stimuli are being used, but it was decided for the scope of this work that this assumption was appropriate.

Each participant was asked to input their age range range (24 or younger, 25-44, 45-64, 65 or older) and their experience level regarding acoustics (no experience, some experience, significant experience), for the purposes of this test undergraduate study would be enough for 'some experience' and working within the acoustics industry or working on academic research would be enough for 'significant experience'. It was decided that experience with acoustics would not be requested for participants with the reasons being twofold; firstly because unlike traditional standardised tests like MUSHRA the context around this test is about layman subjective definitions of perceived sound to be implemented in interfaces designed for general non-expert use, and secondly to increase the pool of potential participants for the test.

Out of the 16 completed responses to this listening test, 10 of the participants were in the 24 or under age range, while 5 participants were in the 25-44 age range and one participant was in the 45-64 age range. Likewise in this test 2 of the participants described themselves as having 'no experience', 7 described themselves as having 'some experience', and 7 described themselves has having a 'significant' amount of experience.

Participants were enlisted via open calls to the University of York's electronics department, as well as collogues within Audiolab, participants were advised to use headphones (although it was not stated that professional monitoring headphones were required) and undertake

the test within a single session within their standard home listening environment. Although there were 46 unique logs of participants starting the listening test, of these only 16 participants finished the listening test in full, meaning only these 16 responses could be used for analysis. There are two primary factors for this disparity. Firstly, this listening test was not paid and did not feature a reward for participation; secondly, this test was 45 minutes long, while other listening tests being distributed from other researchers in the department were typically in the 15-30 minute range. In future perceptual testing for this context these drawbacks need to be kept in mind, with potential benefits of stripping test designs to only what assessments are fully necessary for the design of a semantic interface.

The percussive sound source used in this listening test is a default live drum loop from the Ableton live 9 library. The melodic sound source was a short clip of female choral singing in anechoic environment, sourced from the OpenAIR library of anechoic sounds. The spoken sample was also sourced from OpenAIR, and features a female voice reading a short excerpt in a book. Each sound sample was six seconds long.

It was decided that at the start of each section there would be a calibration section where participants could listen to examples of melodic, percussive, and spoken sound files at low and high strengths of the perceptual factor for that section. For the first section there would be examples of sounds at low and high brightness, with the same being true for the closeness section. The aim of this was to provide a common understanding on the specific 'brightness' and 'closeness' being investigated in the test for both experts and non-experts.

- For brightness examples, the original sound sources received a +6dB boost for frequencies above 1kHz for high brightness (bright), and a -6dB cut for frequencies above 1kHz for low brightness (dull). This draws from timbral brightness definitions outlining high frequency flux as a integral component [61].

- For closeness examples, the original sound sources were convolved with IRs from the example space acoustic model used for auralisation experiments outlined in chapter 4, an IR from a reciever 2m away from the source was convolved for high closeness (close) examples, and 16m away for low closeness (far) examples. This draws from definitions of timbral closeness being associated with intimacy and the physical distance of a listener [62].

This 'calibration' was designed to give participants a broad idea of what each perceptual term was practically referring to, with definitions referring to existing literature which was used in the perceptual term selection process. This calibration was not intended to provide a qualitative anchor for which listener scores would be compared, instead the participant was trusted in their subjective interpretation of the perceptual term based on their understanding. This is a key aspect of this listening test work in comparison to existing systematic approaches like MUSHRA, that the context for this experimental work is towards the design of interfaces designed for non-experts facilitates perceptual listening testing aimed towards more layperson ideas of what perceptual terms mean.

During the listening test, participants judge sets of the same sound source in acoustic environments that vary according to the independent variable in comparison to a reference sound. As previously stated, the reference sound is in an environment where the independent variable is at its midpoint (5m for changing distance, 40% for changing absorption). For each sound set a participant is shown all the sounds within a set in addition to a hidden reference, and is asked to score each sound in the set in comparison to the control sound in terms how much more or less bright/close it is. Within a set there is always a duplicate of the reference sound that acts as an hidden reference. An example of what sound sets on the listening test look like is shown in figure 5.3.

Within each set, the order of sound clips will be randomised manually before being fed into the test software. The order of the stimuli for each sound set was shuffled beforehand as the test was written, meaning the order of the stimuli within a set is the same for each participant evaluating that sound set, but stimuli were not presented in ascending or descending order. The order of sound sets presented is randomly shuffled each time the test is undertaken, meaning the order of the sound sets presented is different for each participant doing the listening test. This is true for both sections of the test.

Finally, the 24 sound sets will be separated into 2 separate groups (for descriptive purposes these groups can be defined as set group A and set group B). The test is divided into two sections. In the first section 12 sound sets are evaluated for 'brightness', and in the second section 12 sound sets are evaluated for 'closeness'. A different set group is evaluated for each section (A for brightness then B for closeness, B for brightness then A for closeness) depending on which of the two tests is given to the listener. The flow of the listening test is shown in figure 5.4

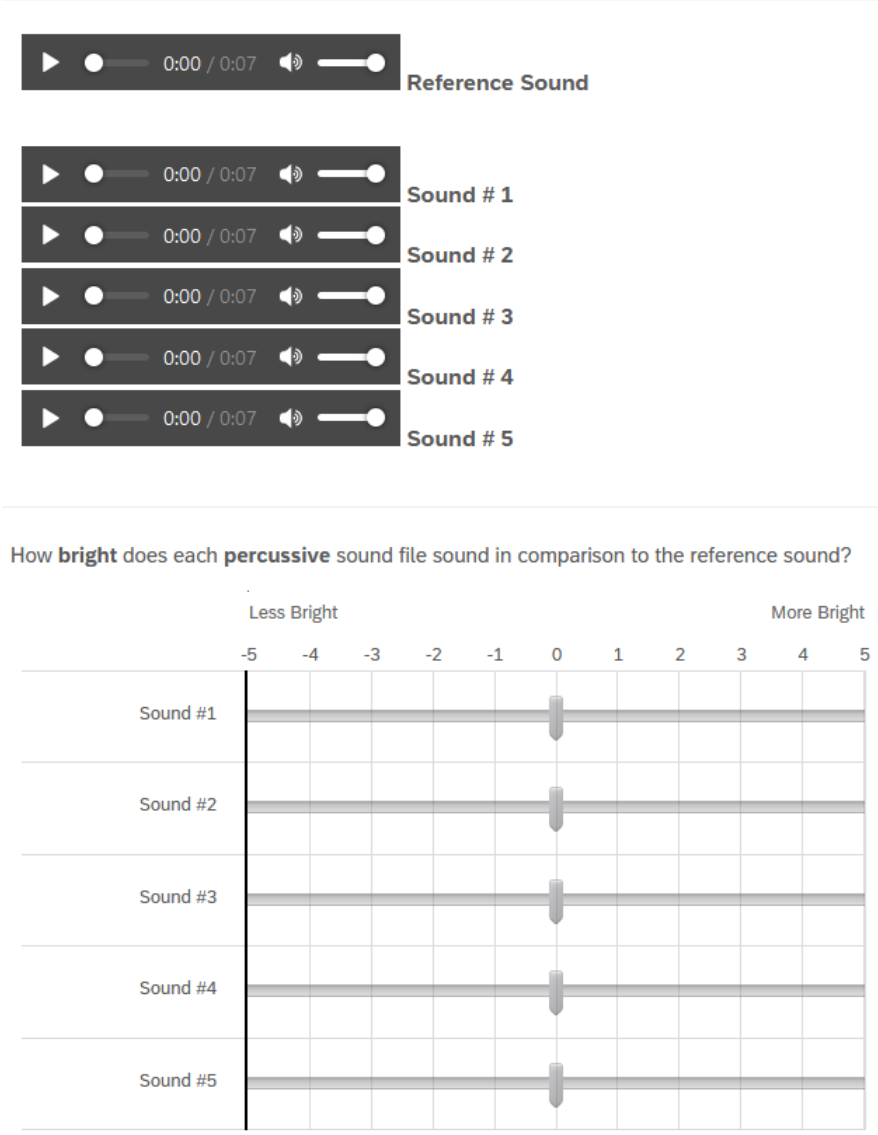

FIGURE 5.3: An example of a sound set in the listening test designed via the 'Qualtrics' software, the sound clips in this sound set are in a set non ordered pattern

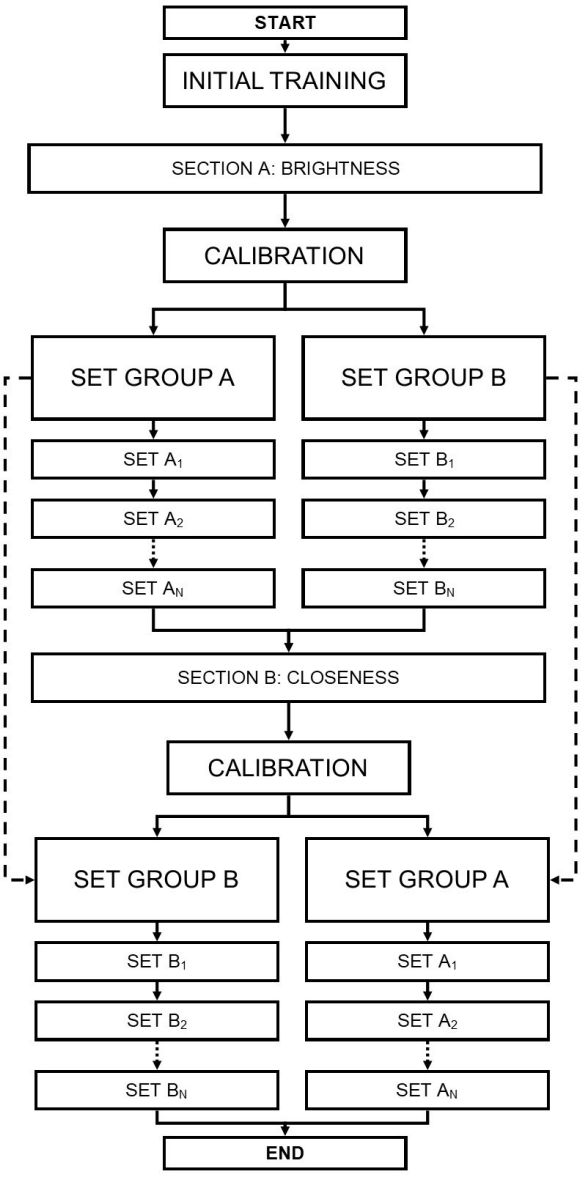

FIGURE 5.4: A block diagram for the structure and flow of the listening test, each of the main sound set blocks has the order randomised for every participant

For each set, each sound will be rated in comparison to the reference sound on a positive/negative discrete scale from -5 to +5, with scoring being in 0.1 increments, leading to 100 point scoring scale, in line with standard practice for MUSHRA and related methodologies. A simple visualisation of this scale is shown in figure 5.5.

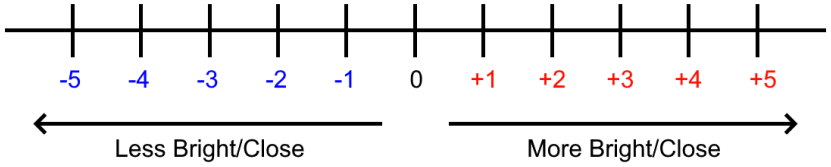

FIGURE 5.5: A diagram visualising the -5 to +5 scoring scale used within the test

This perceptual listening test underwent ethical approval from the University of York's Physical Science Ethics Committee.

5.5 Results and Discussion

The test was distributed online with no hard requirements for participants outside of the usage of headphones. The vast majority of participants did not fully complete the test. This meant that despite 46 participants responding to the test, only 16 complete responses were kept in the results, with the incomplete entries being discarded. Each sound set will have been assessed by 8 of these 16 participants, with half assessing the sound sets in set group A and half for set group B. The full results for each sound set can be seen in Appendix F.

In this section the results from each sound set in the test are shown and discussed, with the results then being interpreted in a number of ways. The results are firstly replotted with the EDT/T30/C80 values of the convolved impulse responses of each stimuli. These resulting plots will show the effects of the acoustic parameters on perceptual terms as per the aims of this test. In addition to this ANOVA analyses were performed on groupings of sound sets from the test. In these ANOVA analyses, the relative effects of both distance and absorption are evaluated for a perceptual term within the same ANOVA analysis. For each ANOVA one variable will be changing within each sound set and the other will change between sound sets. For example, a two way ANOVA analysis for perceptual brightness over distance sound sets evaluates brightness scores for all sound sets where distance is

an independent variable; for this ANOVA the change in distance within each sound set is considered as the one variable and the change of constant absorption coefficient between different sound sets is considered as the second variable. Further work was undertaken to evaluate the hidden reference values for responses, deriving averages per participant and averages per sound set in order to derive greater understanding on the effectiveness and validity of this methodology, and if there is a significant effect on results due to flaws in this listening test approach. Two way ANOVA results are displayed in tables within Appendix G

5.5.1 Listening Test Response Averages

The following plots show the mean value of responses for each stimuli set in the listening test plotted against the independently modified variable (distance, absorption) for drums, singing and speech sound sets utilising that set. From these plots a general overview of the trends of results can be observed. Results were broken down into four groups, Brightness over distance, brightness over absorption, closeness over distance, and closeness over absorption. For each category there are 4 sound sets in the test that assess that relationship, with the other independent variable acting as a control in the sound set. For each sound set all stimuli were at a different control value for this other variable, meaning assessing brightness over distance involves sound sets where the control absorption was at 10, 30, 50, and 70%.

The average results for evaluations of the four sound sets for perceptual brightness over distance (all stimuli for each sound set were at a different constant distance) are shown in figure 5.6. A two away ANOVA was also generated for the the results of these four sound sets and the approximated averages and standard deviations for each distance from that are shown in figure 5.7.

A consistent downward slope can be observed for each sound set assessing brightness over distance in this listening test, indicating that there is a inverse linear proportionality between receiver distance and perceived brightness. This trend can also be observed in the ANOVA mean plot, generally the standard deviations between receiver positions do not overlap, indicating that this variability is significant. The result trend was consistent across different sound stimuli, with the average results being well within each others standard

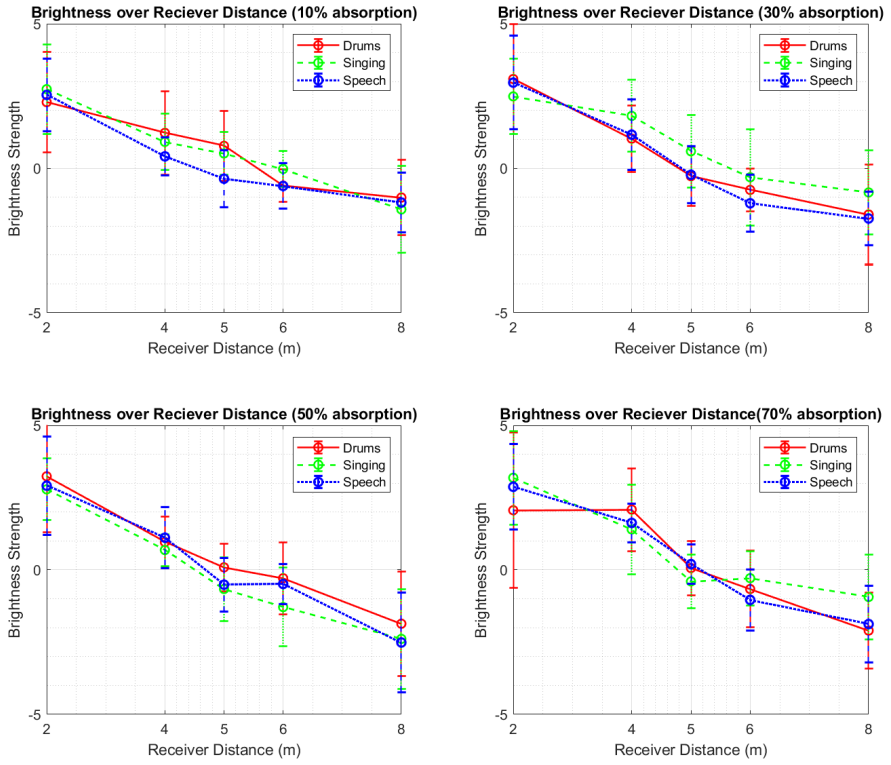

FIGURE 5.6: Mean brightness scores from participants for sound sets where distance was the independent variable

deviation range, indicating no tangible difference in perceived brightness across sound stimuli.

Appendix G.1 shows the two way ANOVA results table for the groups of sound sets where perceptual brightness was assessed with changing distance.. The p value for the distance variable is 0 for all stimuli, as this below 0.05 it can be stated that there is a significant difference in perceptual brightness as distance changes.

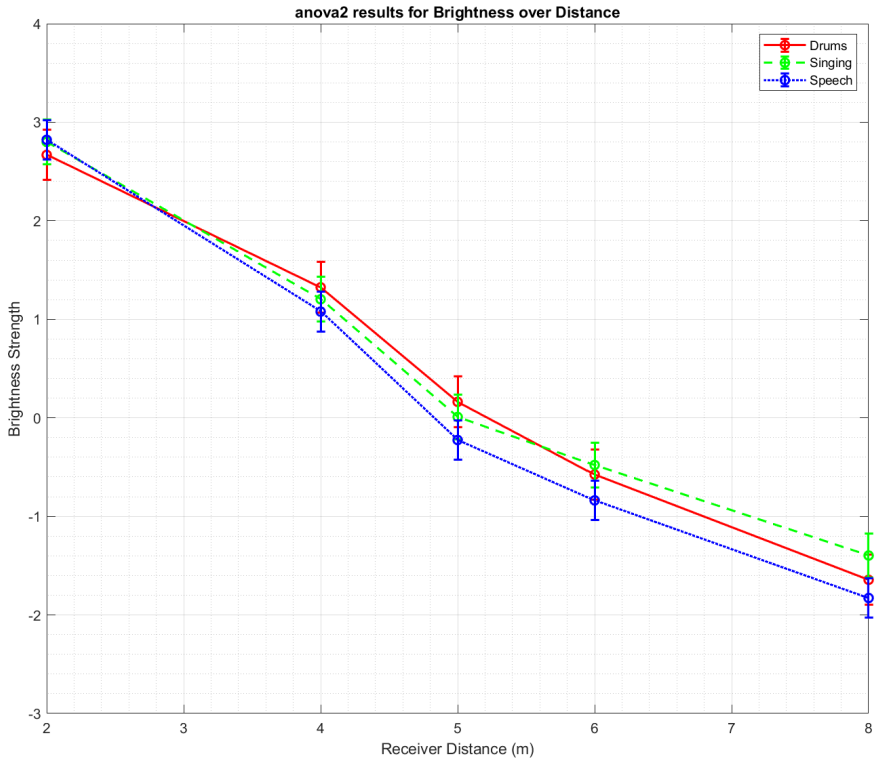

FIGURE 5.7: two way ANOVA approximated mean values for user responses assessing perceptual brightness strength over changing distance

The average results for evaluations of the four sound sets for perceptual brightness over absorption (all stimuli for each sound set were at a different constant distance) are shown in figure 5.8. A two away ANOVA was also generated for the the results of these four sound sets and the approximated averages and standard deviations for each distance from that are shown in figure 5.9.

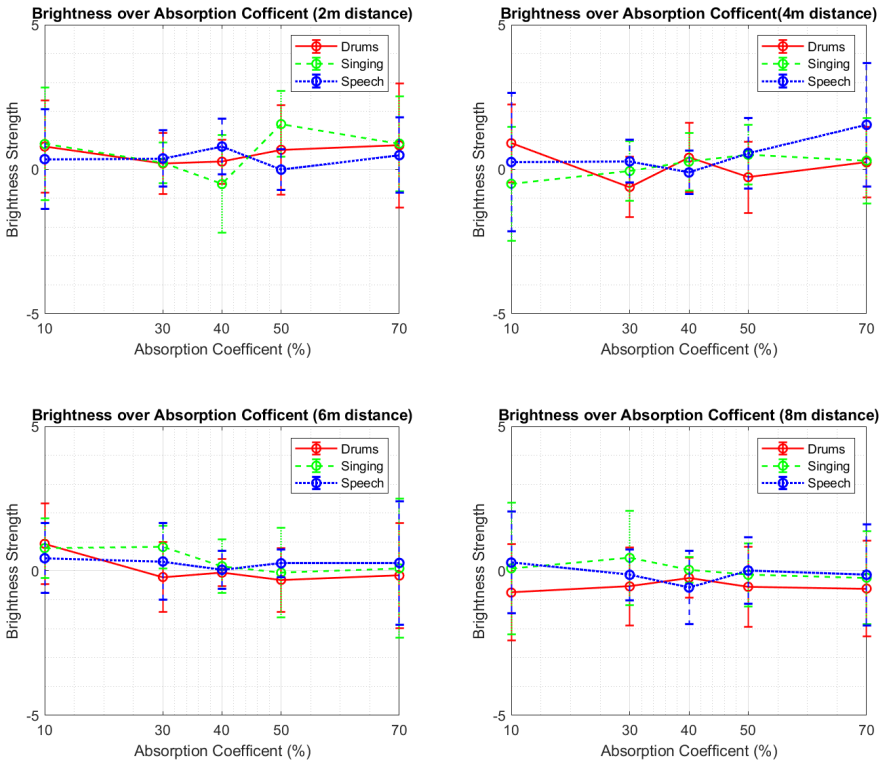

FIGURE 5.8: Mean brightness scores from participants for sound sets where absorption was the independent variable

These results proved little use to the derivation of any useful conclusions, the ANOVA standard deviations shown in figure 5.8 show that responses for each stimuli assessed in this group had high amounts of variability, while the plots of averages in each sound set shown in figure 5.9 show that participants on average measured values around 0, indicating that they did not discern any differences between stimuli. It could therefore be said from this analysis that there no strong evidence for any correlation between absorption coefficient and perceptual brightness. The two way ANOVA results also have a large amount of variability between types of sound source, in contrast with brightness over distance results shown in the previous plots.

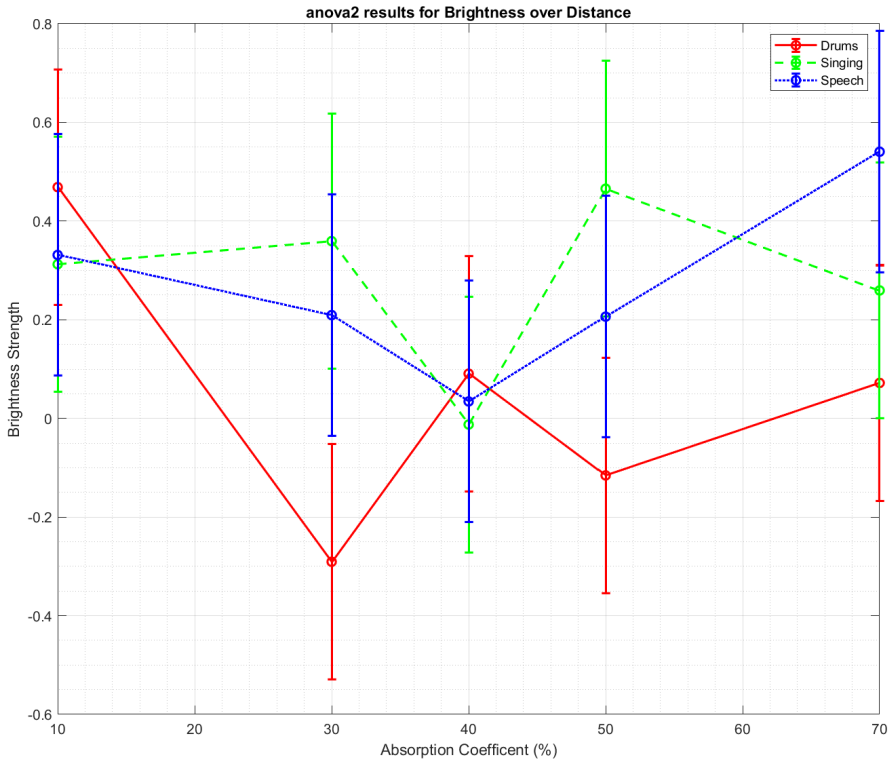

FIGURE 5.9: two way ANOVA approximated mean values for user responses assessing perceptual brightness strength over changing absorption

Appendix G.2 shows the two way ANOVA results table for the groups of sound sets where perceptual brightness was assessed with changing absorption. The p value for the absorption variable is 0.24 for the sound set group with percussive stimuli, 0.75 for melodic stimuli, and 0.67 for speech stimuli; as all of these values are above 0.05 it can be stated that there is no significant difference in perceptual brightness as absorption changes for these sound sets.

The average results for evaluations of the four sound sets for perceptual closeness over distance (all stimuli for each sound set were at a different constant distance) are shown in figure 5.10. A two away ANOVA was also generated for the the results of these four sound sets and the approximated averages and standard deviations for each distance from that are shown in figure 5.11.

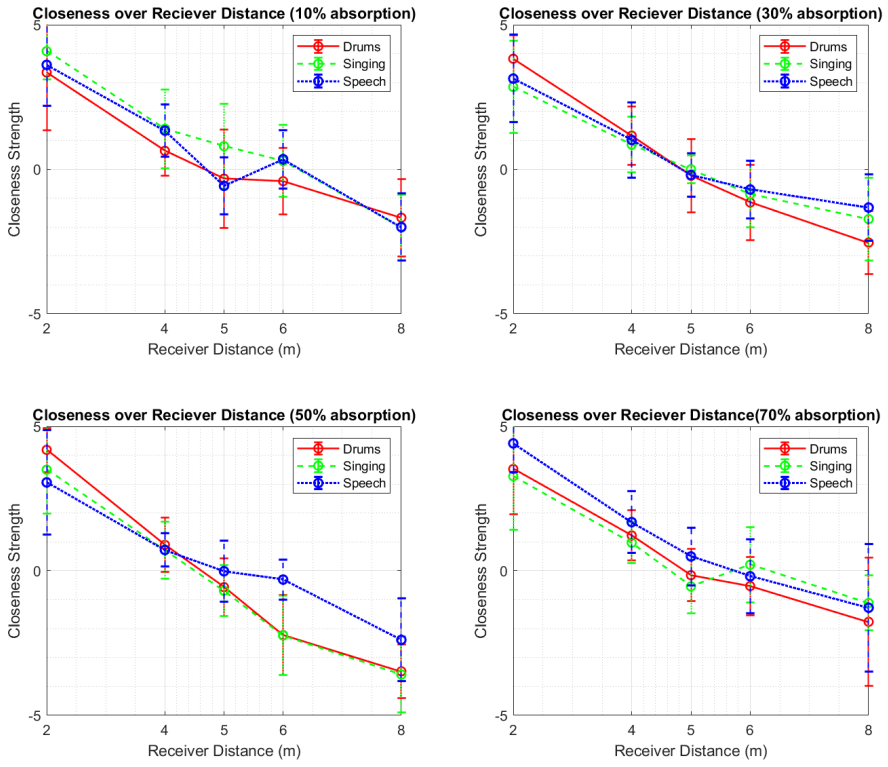

FIGURE 5.10: Mean closeness scores from participants for sound sets where distance was the independent variable

There is a distinct downward slope that can be observed in mean plots for each sound set and in the ANOVA approximated mean results, indicating that perceived closeness decreases as receiver distance increases. Comparing the three different sound stimuli shows a low amount variability between values indicating the robustness of this trend. The two way ANOVA mean plot also indicates this linear decrease, for each of the sound stimuli in the ANOVA plot, standard deviations show less overlap from receiver to receiver.

Appendix G.3 shows the two way ANOVA results table for the groups of sound sets where perceptual closeness was assessed with changing distance. The p value for the distance variable is 0 for all stimuli, as this below 0.05 it can be stated that there is a significant

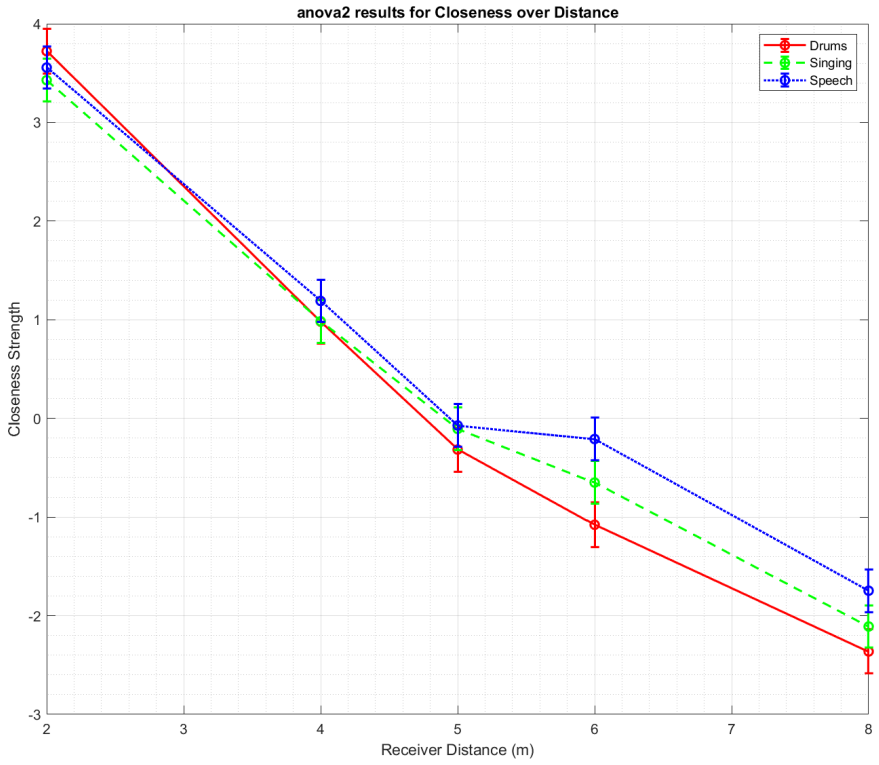

FIGURE 5.11: two way ANOVA approximated mean values for user responses assessing perceptual closeness strength over changing distance

difference in perceptual closeness as distance changes. It is also notable that the interaction p value for the set group with melodic stimuli is 0.02, also below the 0.05 threshold; this implies that the significance of change in perceptual closeness to due to changing distance is somewhat determined by changing absorption, in this case implying that changing the constant absorption coefficient within a sound set leads to more significant variance of closeness within the set.

The average results for evaluations of the four sound sets for perceptual closeness over absorption (all stimuli for each sound set were at a different constant distance) are shown in figure 5.12. A two away ANOVA was also generated for the the results of these four sound sets and the approximated averages and standard deviations for each distance from that are shown in figure 5.13.

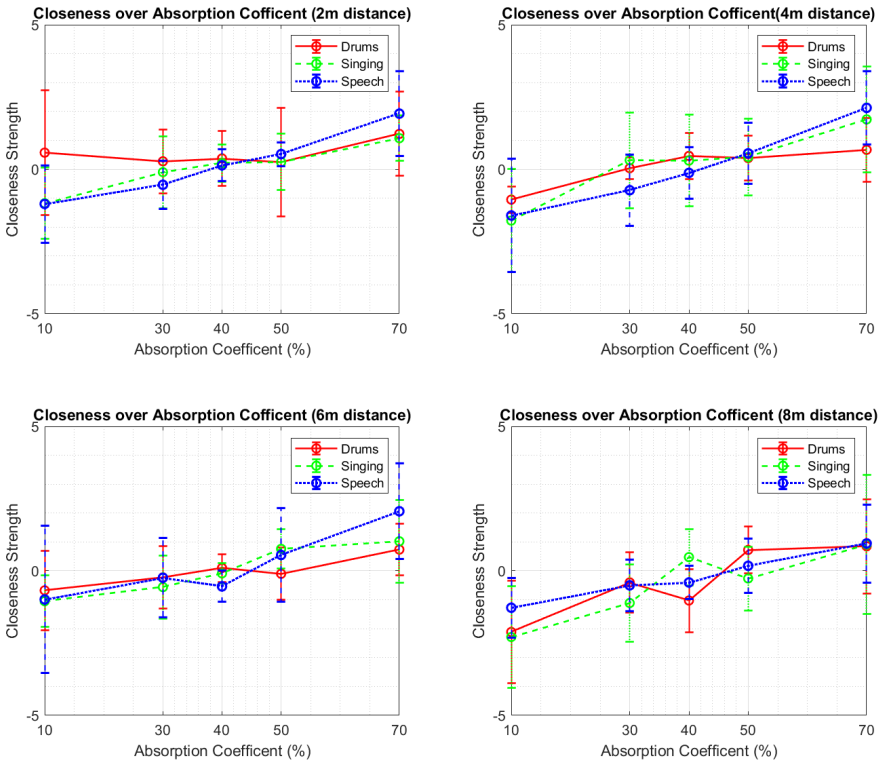

FIGURE 5.12: Mean closeness scores from participants for sound sets where absorption was the independent variable

The average response value plots show a slight increasing trend in perceptual closeness as absorption coefficient increases, but the amount of variability is significantly less compared to the values shown in the previous group (figure 5.10 and 5.11). The two way ANOVA plot shows a large amount of variability between values at each absorption coefficient, this is in comparison to the group of sound sets around closeness and distance outlined in the previous section. The approximated mean values for each stimuli show a greater linear proportionality between closeness and absorption coefficient.

Appendix G.4 shows the two way ANOVA results table for the groups of sound sets where perceptual closeness was assessed with changing absorption. The p value for the distance

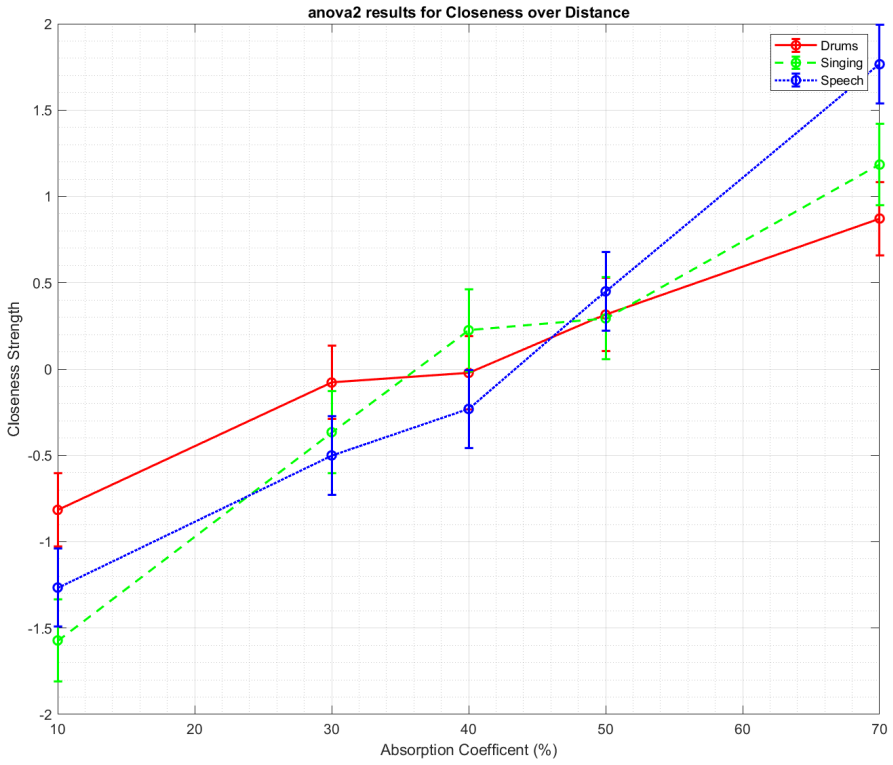

FIGURE 5.13: two way ANOVA approximated mean values for user responses assessing perceptual closeness strength over changing absorption

variable is 0 for all stimuli, as this below 0.05 it can be stated that there is a significant difference in perceptual closeness as absorption changes.

5.5.2 Mapping Listening Test Results to Acoustic Measurements

Drawing from work in the previous section, the previously plotted figures displaying response averages for changing distance and absorption were mapped onto the measured values of EDT, T30, and C80 for each stimuli. As previously stated in this chapter, plots of these acoustic measurement do not contain x values at regular interval due to the nature of impulse response generation process. Nevertheless they can provide clear observations of the relationship between the two perceptual terms and the acoustic measurements, assessing the validity of the null hypothesis outlined at the start of this chapter. In this section brightness is assessed in relation to each acoustic measurement, with the same process undertaken for closeness.

Figures 5.14 and 5.15 shows the average responses for all sound sets evaluating brightness, with the values of the independent variables of the stimuli in each sound set changed to the EDT values of those stimuli, essentially mapping changes in EDT to changes in perceptual brightness for the eight impulse responses sets used in this listening test. Figure 5.14 shows brightness over EDT plots for all sound sets where distance was changed, while figure 5.15 shows plots over changing absorption. Note that the y axis for each of the plots in 5.14 and 5.15 is not normalised and changes with each plot for best fit.

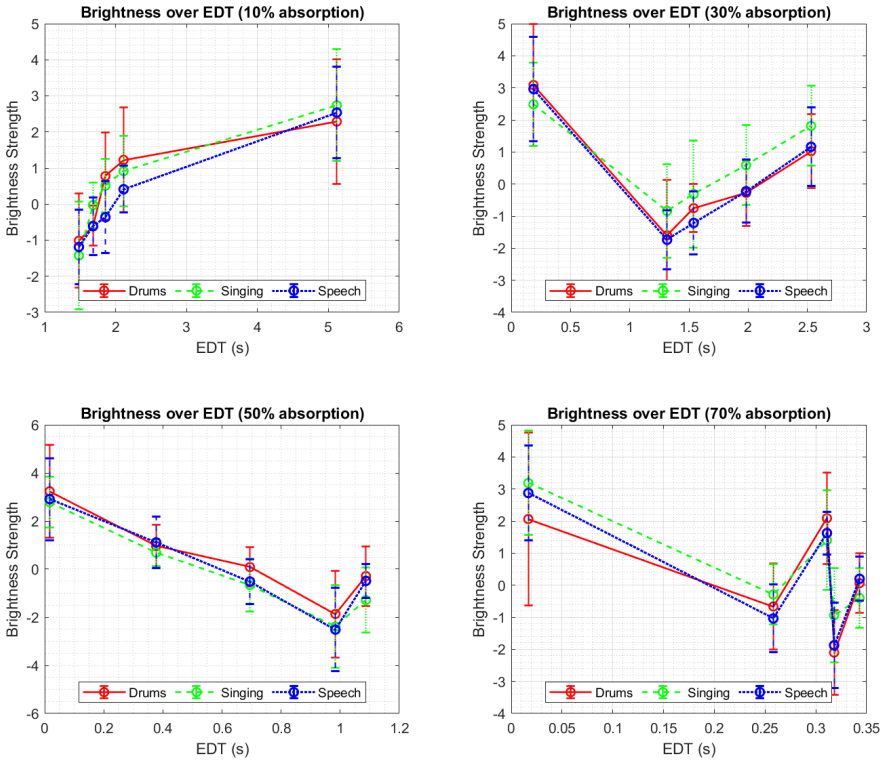

FIGURE 5.14: Mean average results for user responses assessing perceptual brightness strength over EDT variability in changing distance sound sets

The changed scale of plots due to rescaling for EDT values outlines that there is little to no mean variability for sound stimuli of perceptual brightness from increasing EDT. This is best observed in figure 5.14, where the average do not significantly vary from zero, effectively demonstrating that participants do not observe any brightness change as EDT changes. In figure 5.15, with sound sets where absorption coefficient was the cause of changing EDT, the resultant EDT values on the x axis are more uniform; resultantly the plots show the lack of correlation more clearly, within each subplot in the figure there is no trend or pattern of brightness as EDT increases, and therefore the variabilitys shown seem

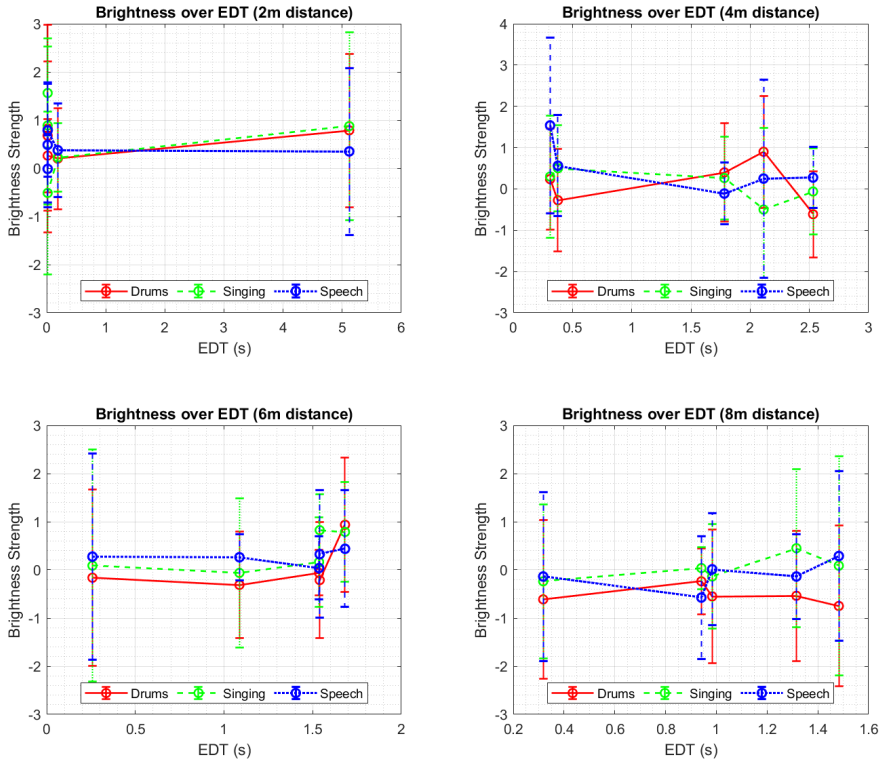

FIGURE 5.15: Mean average results for user responses assessing perceptual brightness strength over EDT variability in changing absorption sound sets

more chaotic and random. From observation, it can be stated that there is insignificant correlation between measurements of perceptual brightness and EDT.

A simple linear regression was performed on each of the sounds within the eight sound sets shown in figures 5.14 and 5.15, averaging the regressions for each of the sound stimuli types (melodic, percussive, speech) results in the equation $y = 0.0749x + 0.1294$. Figure 5.16 shows the plot of this regression, showing regression plots for each of the sound stimuli types individually (melodic, percussive, speech) in addition to the overall mean average regression, all sound sets for the stimuli type (4 sound sets for each type) were analysed, leading to 32 data points for each plot.

The plots show that there the regression plots for each of the types of sound stimuli share a small positive gradient, with the positive gradient from the speech sound set regression not being large enough to be observed within the plot. The relatively small gradient from mean regression provides evidence to the fact that there is no correlation between brightness and EDT, the validity of this regression is slightly dampened by the presence of EDT values

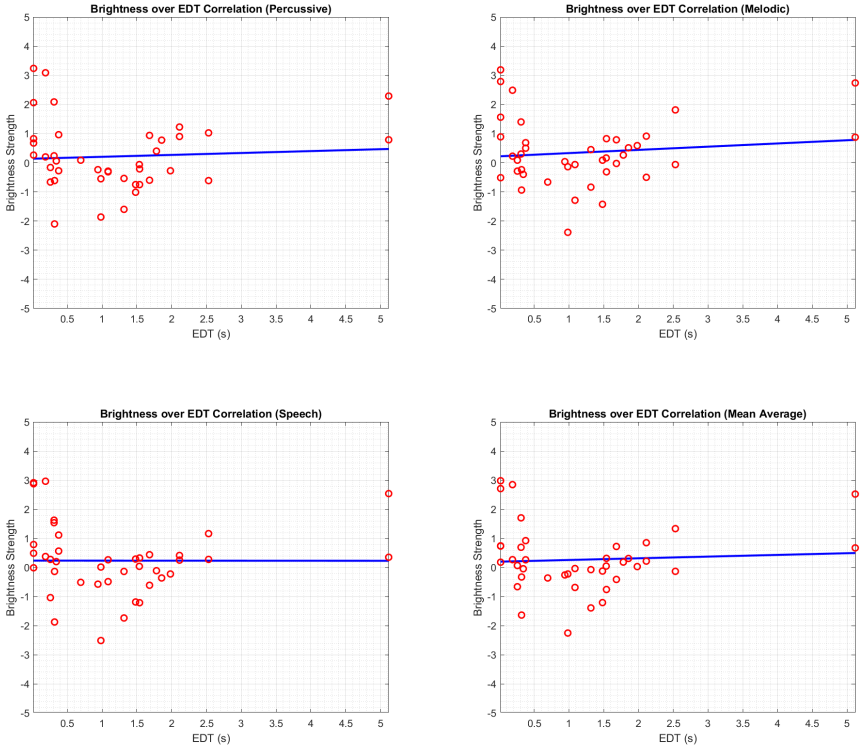

FIGURE 5.16: Linear regression plot for average scores for perceptual brightness over EDT values of the sound stimuli

shown that have a EDT of significantly larger than 2.5s, but even within the range of most values there is a high variability in perceived brightness. This provides evidence that the null hypothesis "Perceived brightness of a sound will not have direct proportionality with changing EDT" was correct.

Figures 5.17 and 5.18 show the average response values for the same eight sound stimuli, assessing Brightness over T30. Figure 5.17 shows T30 variability as distance changes, while figure 5.18 shows variability as absorption changes.

Similar observations and conclusions to the ones discussed with the previous brightness/EDT can also be drawn from these two figures. These similarities are due to the nature of both measurements being forms of decay time measurements. Like the previous group of average responses shown in figure 5.14, the results shown in figure 5.17 show that for sound sets where the sound stimuli varied by receiver distance there was little deviation from zero, indicating the no differences were observed between sounds in these sound sets as EDT values increased. For changing absorption coefficient. A similar trend to the EDT over

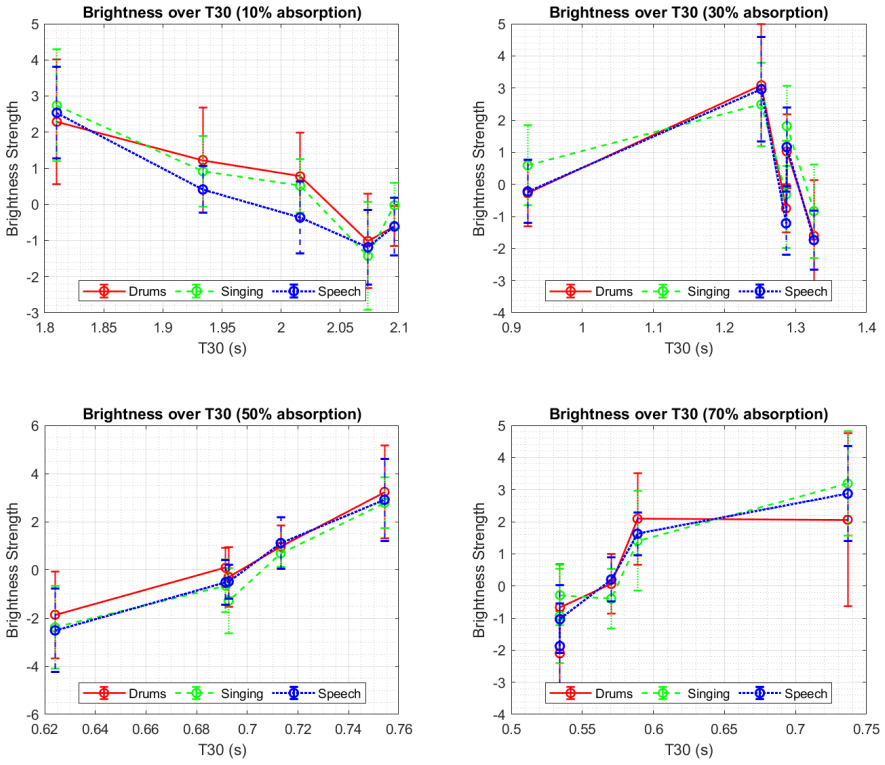

FIGURE 5.17: Mean average results for user responses assessing perceptual brightness strength over T30 variability in changing distance sound sets

distance plots shown in figure 5.14 can be observed in figure 5.17, with random variability and non correlation between distance and EDT. While the 50% and 70% absorption results show a slight general increase across EDT values, there are too many results anomalous from the rest of the data within each sound set that there is not sufficient evidence to indicate any relationship between EDT and brightness.

This point can be emphasised through performing a regression analysis similar to the one demonstrated in figure 5.16, resulting in the equation $y = 0.2256x - 0.0285$. The regression plot for perceptual brightness over T30 values of the sound stimuli used in listening test sound sets is shown in figure 5.19.

The regression equation for T30 provides a larger gradient than the equivalent EDT regression equation, indicating that T30 is a much stronger factor relating to changing brightness than EDT, although observing the plots of mean perceptual brightness scores for both EDT (figures 5.14 and 5.15) and T30 (figures 5.16 and 5.17) show similarities in result trends. In addition, much like for the EDT plots shown earlier, speech sound sets exhibit

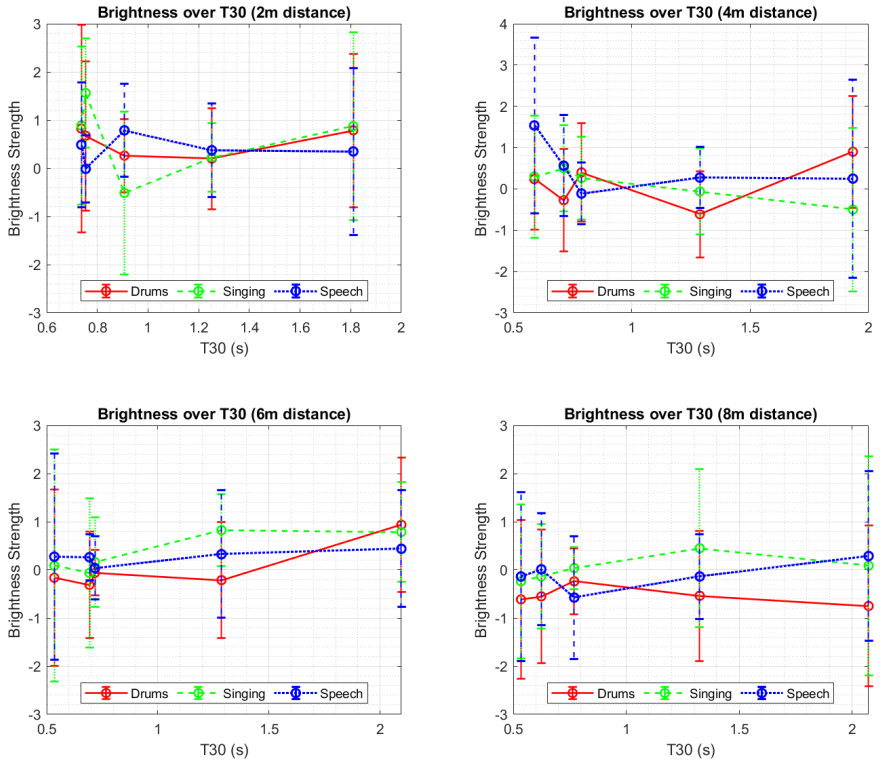

FIGURE 5.18: Mean average results for user responses assessing perceptual brightness strength over T30 variability in changing absorption sound sets

this positive gradient less than sets with other stimuli types, affirming the similarities between EDT and T30 in this context. This means that while is evidence towards the null hypothesis of "Perceived brightness of a sound will not have direct proportionality with changing T30" being untrue, that specific conclusion would not be as reliable as desired.

Figure 5.20 and 5.21 show the average response values for the eight sound sets evaluating brightness, over values of C80, with figure 5.20 showing results for sound sets where distance was changed, and figure 5.21 showing the same for sound sets where absorption coefficient was changed.

Much like previous results for EDT and T30, there is little deviation from zero for C80 when the sound stimuli varies by distance as shown in figure 5.19; even though C80 increases in a linear fashion, as shown on the plots. However, when looking at the results for changing absorption shown in figure 5.19, there is a distinct increase in brightness as C80 increases across all four sound stimuli sets shown. This provides some evidence that there is a relationship between increasing C80 and perceptual brightness.

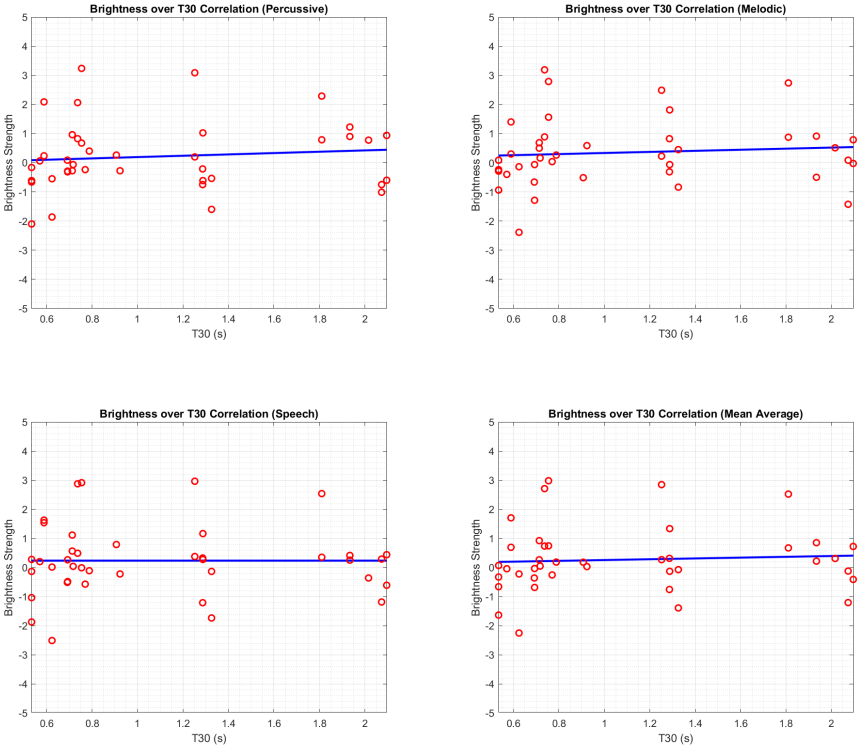

FIGURE 5.19: Linear regression plot for average scores for perceptual brightness over T30 values of the sound stimuli

A regression analysis like the ones performed for EDT and T30 shows the trend in brightness over C80 across all of these sets of sound stimuli, with the resultant equation of $y = 0.0893x - 0.5367$. The resultant plot is shown in figure 5.22

This regression shows a distinct positive correlation between perceptual brightness and C80, although it worth noting that the proportionality is small. This positive correlation is exhibited similarly with all types of sound stimuli. This analysis provides evidence disproving the null hypothesis "perceived brightness of a sound will not have direct proportionality with changing C80".

In deriving a definition of perceptual brightness through the results of this listening test, decay time EDT proved to be negligible in terms of affecting the brightness of a sound across the range of values within stimuli. C80 and T30 shown to have a positive collection with brightness scores, albeit with only small increases across the range of values within the stimuli. That EDT, a measurement relating to 'perceived' reverb, proved to be less of a factor than T30, a measurement relating to 'measurable' reverb, provides some evidence

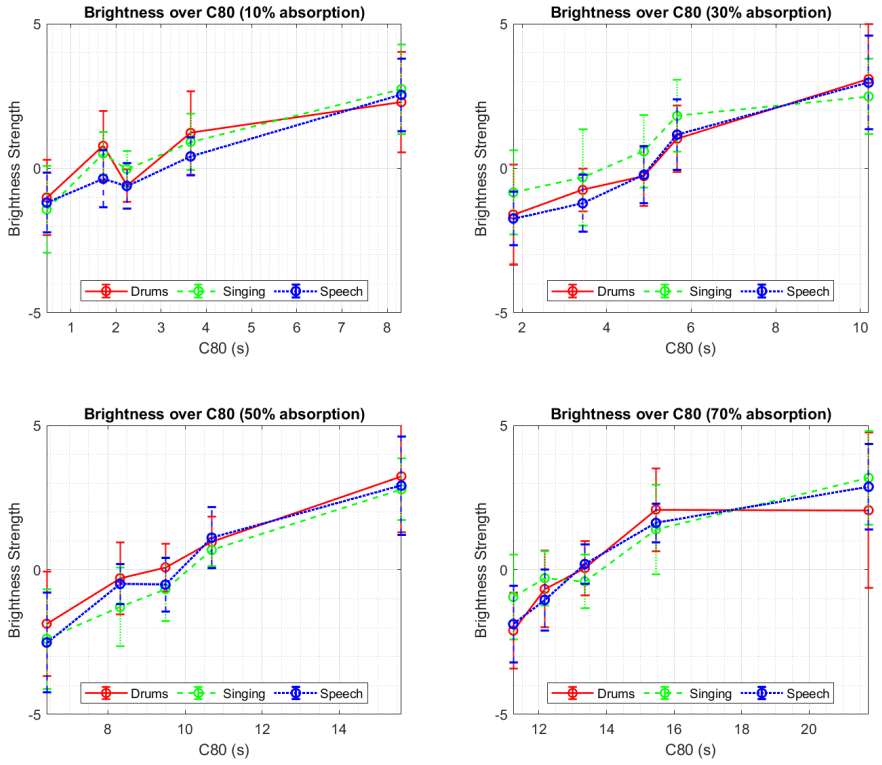

FIGURE 5.20: Mean average results for user responses assessing perceptual brightness strength over C80 variability in changing distance sound sets

to the ineffectiveness of EDT measurements in effecting how a listener perceives brightness within a space. C80 being a factor while being a measure of early over late arriving energy, provides evidence that clarity factors into how a listener perceives brightness within a space, as well as indicating the relevance and reliability of measurements performed at later times on the energy decay curve.

In evaluating the result plots and regression plots for perceptual brightness scores the following conclusions can be reached:

- Perceptual brightness is shown to have no distinct correlation with measured values of EDT
- Perceptual brightness is shown to linearly increase with measured values of T30
- Perceptual brightness is shown to linearly increase with measured values of C80

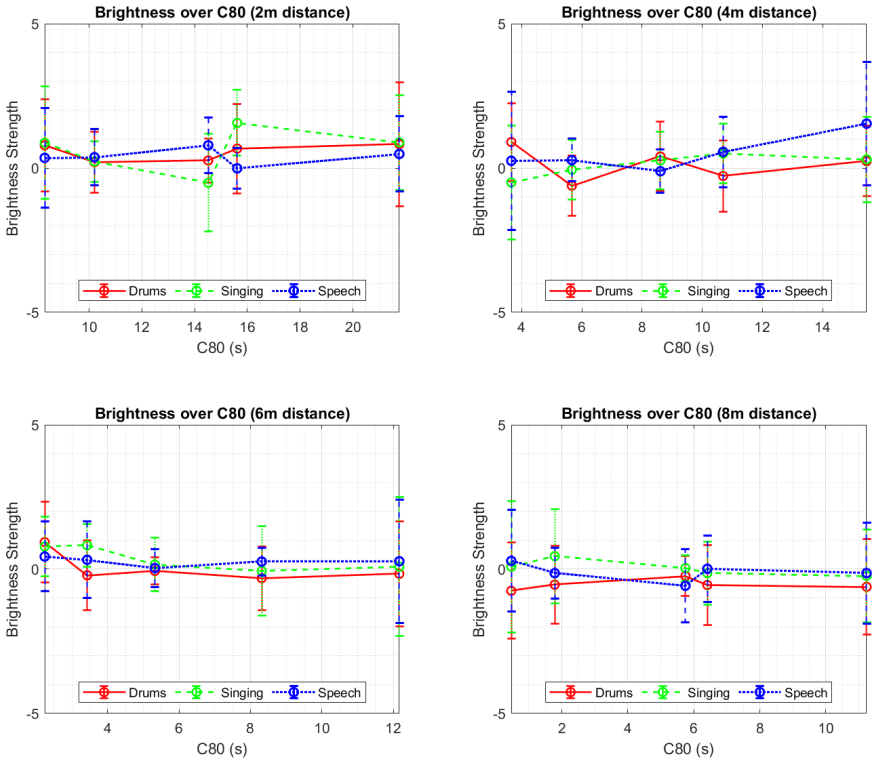

FIGURE 5.21: Mean average results for user responses assessing perceptual brightness strength over T30 variability in changing absorption sound sets

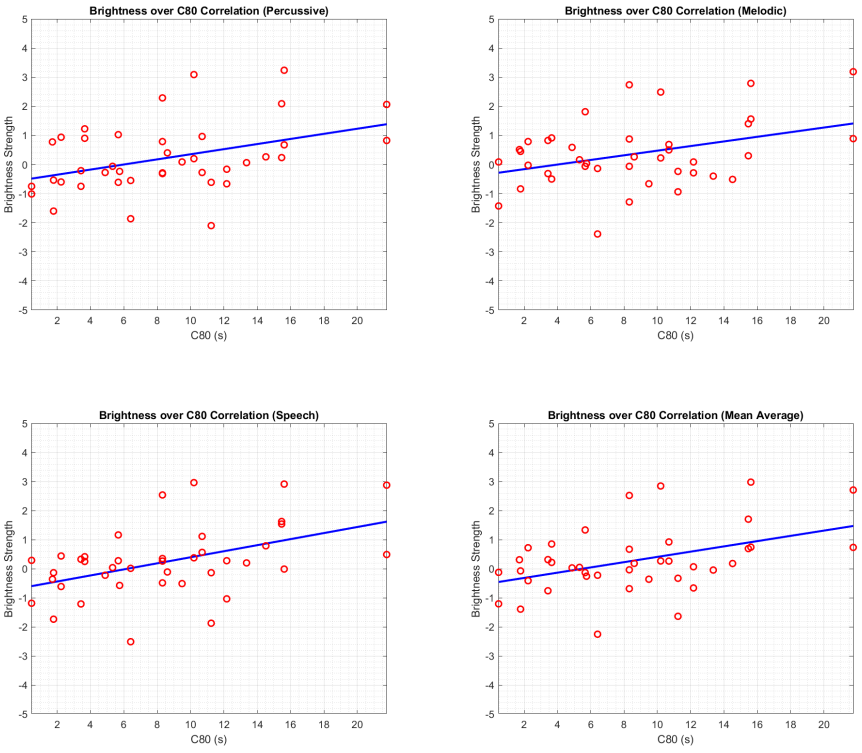

FIGURE 5.22: Linear regression plot for average scores for perceptual brightness over C80 values of the sound stimuli

The same analysis that was performed for brightness was also undertaken for perceptual closeness. Figures 5.23 and 5.24 show the plots of average response data for each of the eight sound stimuli sets in sound sets evaluating closeness over EDT, with figure 5.23 showing closeness/EDT plots for sound sets where distance was changed and figure 5.24 showing plots for sound sets where absorption were changed.

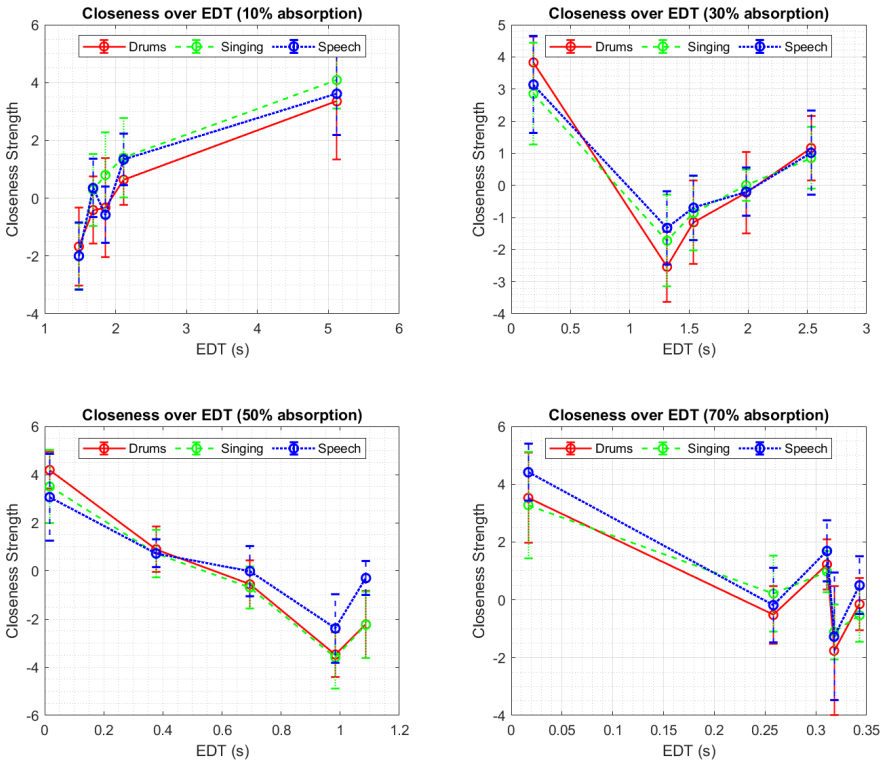

FIGURE 5.23: Mean average results for user responses assessing perceptual closeness strength over EDT variability in changing distance sound sets

Initial analysis of these plots proved inconclusive, for each individual plot and set of stimuli there is a distinct trend across EDT values, but these correlations differ significantly between sound sets. Therefore linear regression was performed on the values in each of the 8 sets analysed, leading to the equation $y = -0.0448x + 0.1800$. A regression plot is shown in figure 5.25

Even discounting results that deviated significantly from the regression plot line, it can be observed that the values themselves have high non uniform variability between points, indicating there is no distinct relationship shown between EDT and perceptual closeness score averages. This is true across all three types of sound stimuli, with melodic and speech

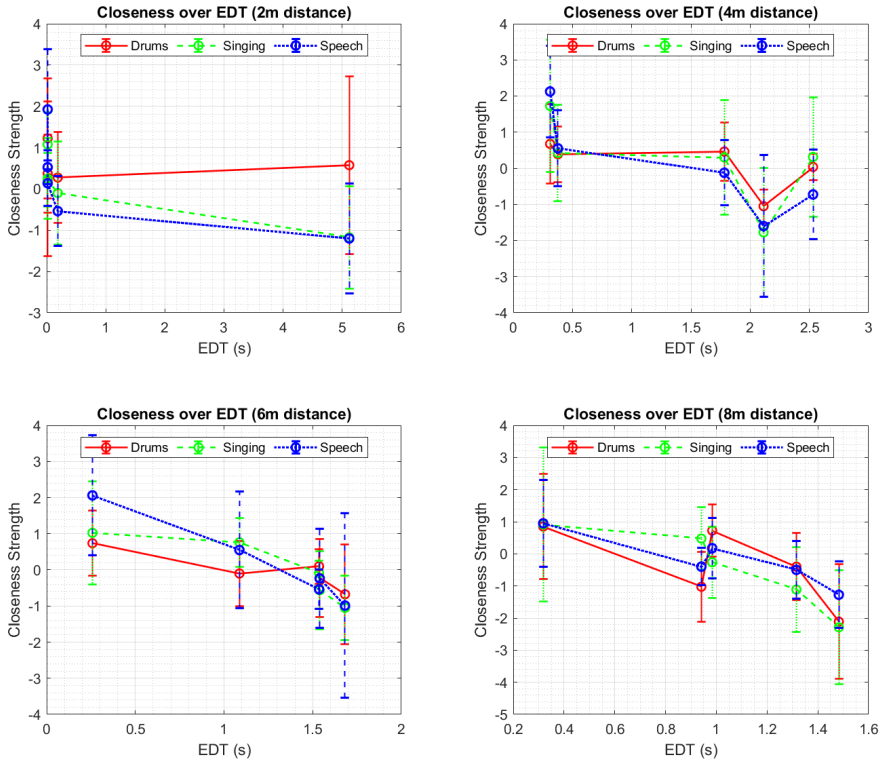

FIGURE 5.24: Mean average results for user responses assessing perceptual closeness strength over EDT variability in changing absorption sound sets

sound sets exhibiting a greater amount of negative correlation than percussive sound sets, due to the erratic scattering as EDT increases these negative gradients should be taken with some scepticism however. The scattering of actual average score values suggests that the results at high EDT values shown to the right of the plot skew the regression towards having a negligible negative gradient and little to no variation from zero. The non uniform variation in results for perceptual closeness is similar to the results of regression analysis for brightness, suggesting that the results of this test show that EDT does not significantly factor into either perceptual term. Relating this to the null hypothesis stated at the beginning of this chapter, this evidence implies the statement "perceived closeness of a sound will not have direct proportionality with changing EDT" is true.

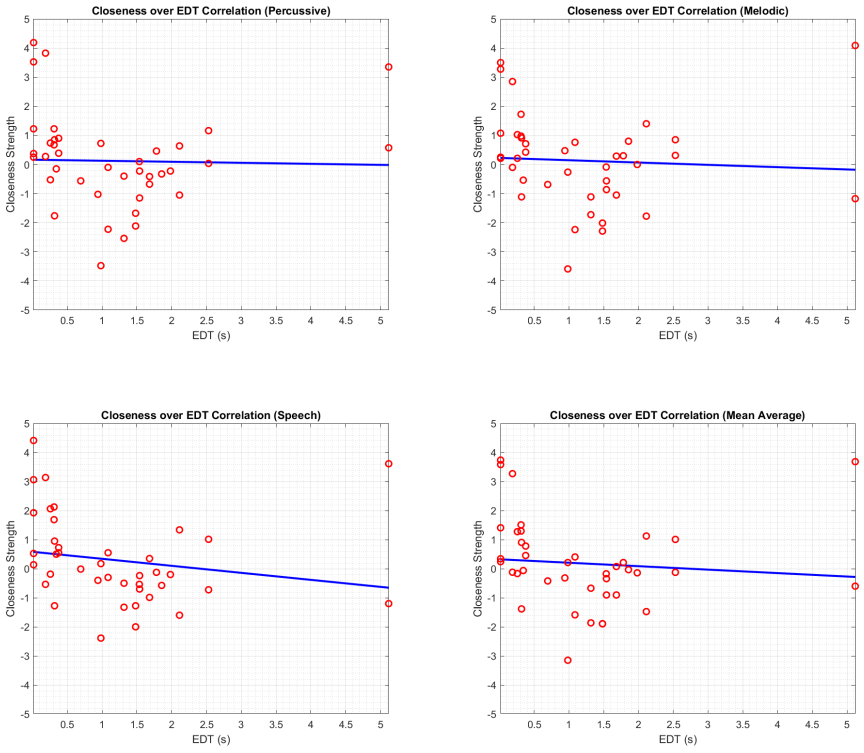

FIGURE 5.25: Linear regression plot for average scores for perceptual closeness over EDT values of the sound stimuli

Figures 5.26 and 5.27 show plots of average response data for the eight closeness sound sets over values of T30 of each sound stimuli in each sound sets. Figure 5.26 shows all the sound sets where distance was changed across the sound stimuli set, while figure 5.27 shows the sound sets for when absorption was changed.

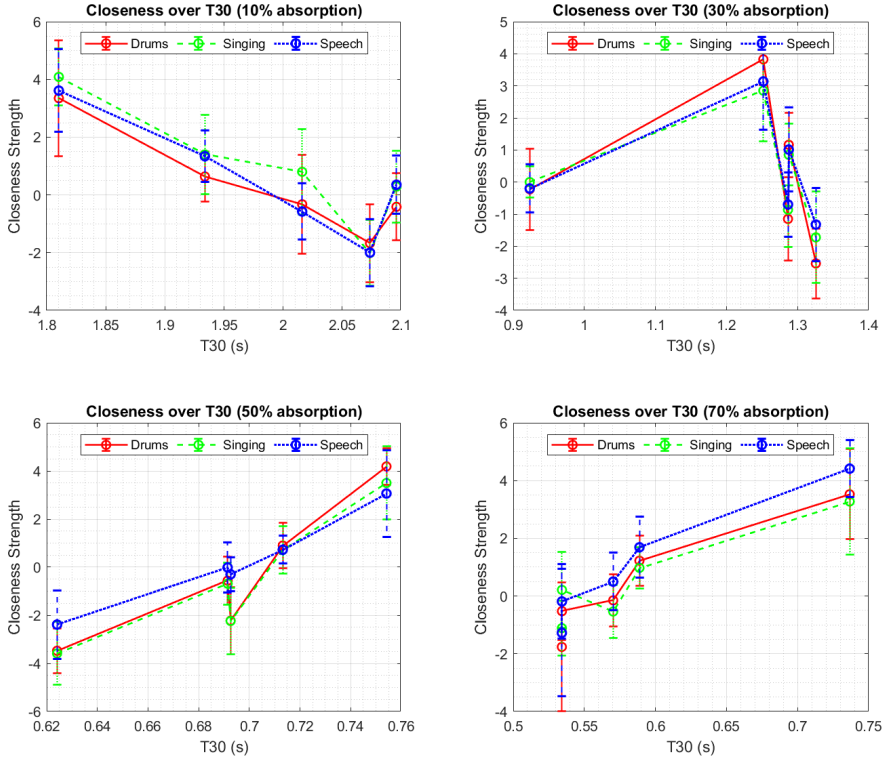

FIGURE 5.26: Mean average results for user responses assessing perceptual closeness strength over T30 variability in changing distance sound sets

Similar to EDT results, a regression analysis was performed for all the sound stimuli T30 values and perceptual closeness scores, resulting in the equation $y = -0.3323x + 0.4909$. The results of this are shown in figure 5.28.

There are two general conclusions to be drawn from the regression plots regarding closeness and T30, the negative correlation graident shown for all types of sound stimuli, and the drastic changes in EDT in both directions between data points which is also seen throughout all types of sound stimuli. In comparison to the closeness over EDT plot shown in figure 5.25 there is a more distinct negative gradient that can be observed in the resultant mean average equation.However, much like closeness and EDT discussed earlier, the diffuse and erratic scattering of data points along the T30 axis suggests a lack of uniform correlation

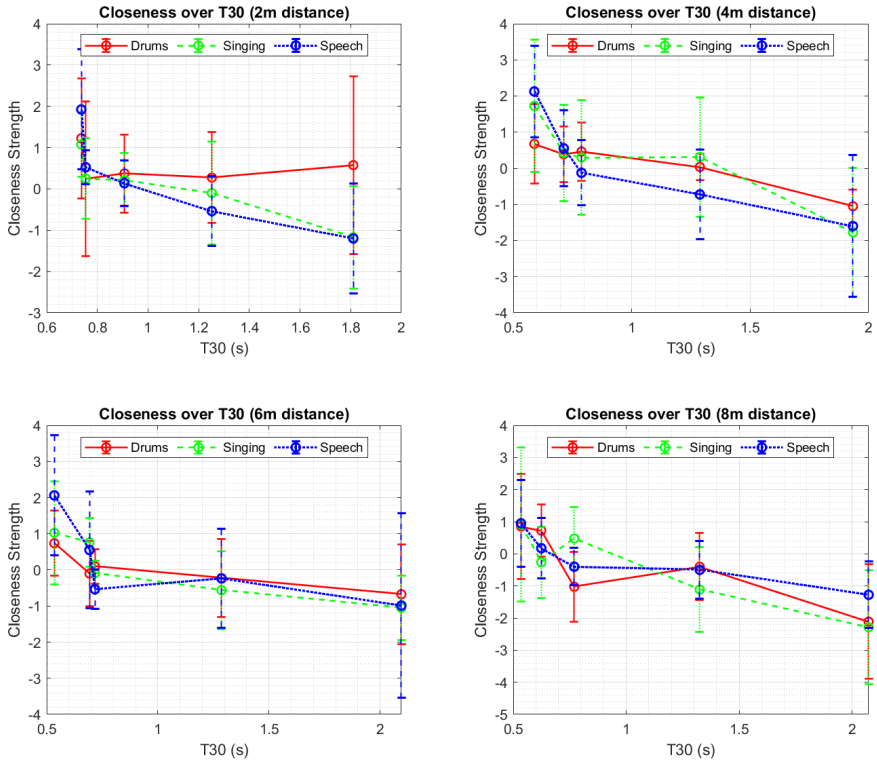

FIGURE 5.27: Mean average results for user responses assessing perceptual closeness strength over T30 variability in changing absorption sound sets

either positive or negative. Ultimately, the size of the mean average gradient equation for closeness is by a large magnitude (6x the amount) larger than for closeness and EDT, suggesting that even though there are large variances in certain cases form data point to data point, there is a clearly observable negative trend in closeness strength as T30 increases, an inverse proportionality. This conclusion ultimately reasons that the null hypothesis that the "T30 has no impact on perceptual closeness" is false.

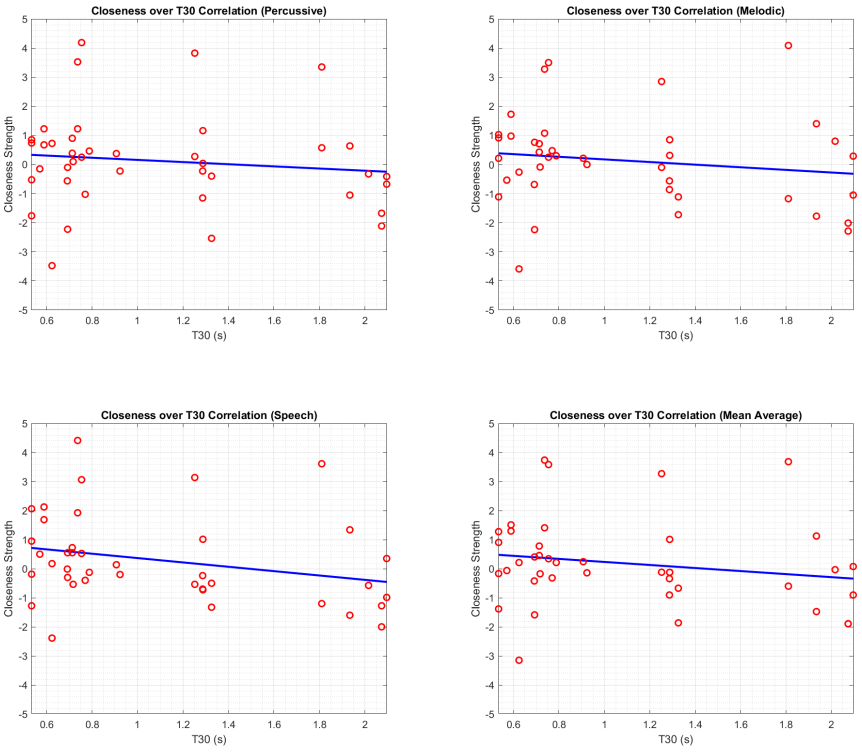

FIGURE 5.28: Linear regression plot for average scores for perceptual closeness over T30 values of the sound stimuli

Figures 5.29 and 5.30 show plots of average response data for the eight closeness sound sets over values of C80 for each the sound stimuli assessed in the sound sets, with figure 5.29 showing all sound sets where distance was modified between stimuli within the sound set, and figure 5.30 showing the same but with absorption being modified in the sound set.

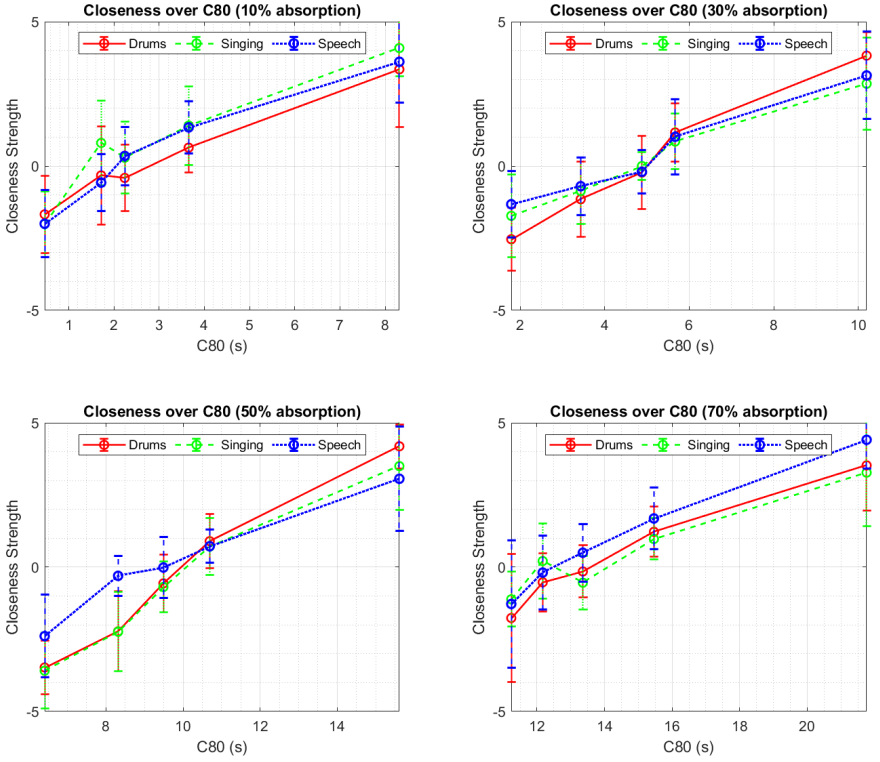

FIGURE 5.29: Mean average results for user responses assessing perceptual closeness strength over C80 variability in changing distance sound sets

As before with values of EDT and T30, a regression analysis was performed for the C80 values in the eight sound sets assessing closeness, resulting in the equation $y = 0.1588x - 1.2172$. A regression plot is shown in figure 5.31:

The gradient from this regression is distinct enough for there to be similar conclusion as there was when C80 was assessed in comparison to brightness. The regression plot shows a clear positive correlation between closeness and C80, the most clear and prominent of any regression plot generated in this analysis, and this positive gradient is similar across all types of sound stimuli. Therefore this data provides evidence to disclaim the null hypothesis "perceived closeness of a sound will not have direct proportionality with changing C80".

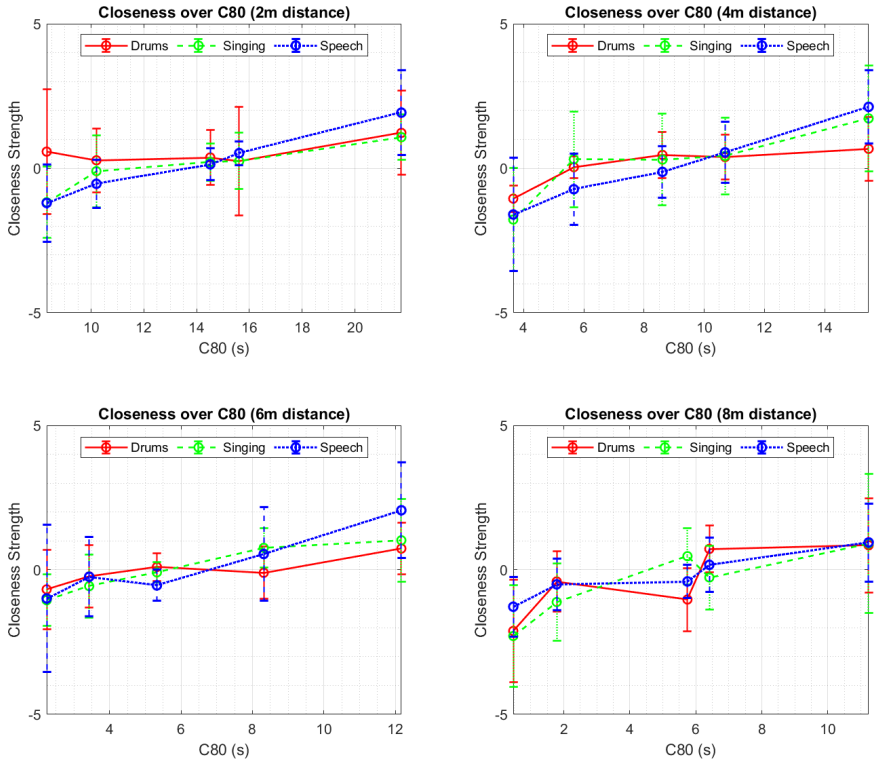

FIGURE 5.30: Mean average results for user responses assessing perceptual closeness strength over C80 variability in changing absorption sound sets

In evaluating the result plots and regression plots for perceptual closeness scores the following conclusions can be reached:

- Perceptual closeness is shown to have no distinct correlation with measured values of EDT
- Perceptual closeness is shown to linearly decrease with measured values of T30
- Perceptual closeness is shown to linearly increase with measured values of C80

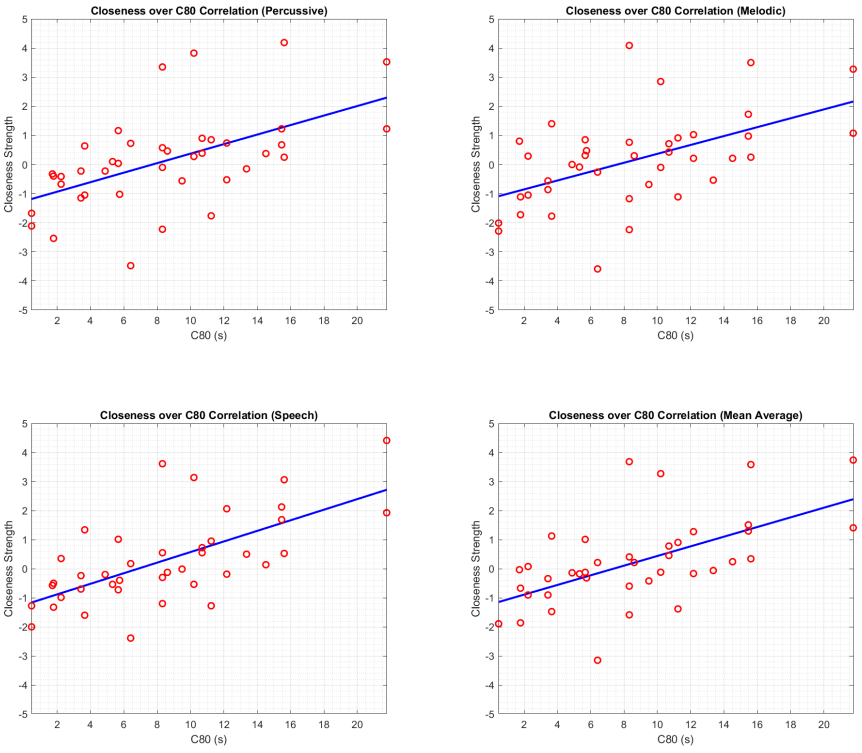

FIGURE 5.31: Linear regression plot for average scores for perceptual closeness over C80 values of the sound stimuli

5.5.3 Hidden Reference Analysis

The design of this listening test involved the usage of a hidden reference, a copy of the reference sound stimuli used to assess the quality of individual responses. Ideally a participant should give a score of zero for each hidden reference, since there is no change from the original file. This was implemented in MUSHRA as a means to assess the expertise of the participants. In a move away from strict objectivity based testing, where a skill of a participant can be quantified, the hidden references were not used as a means for testing the participants but the validity of the test methodology and the practical test itself. Participants were asked to state their experience level before the listening test began, stating either no experience, some experience, or a high amount of experience with acoustics.

An analysis was performed using these hidden reference scores for each sound set to determine averages for each participant, this analysis of hidden reference values allowed a comparison to be made about deviations from zero in these scores for the three different experience levels. This involved taking the absolute or modulus values of the hidden reference scores, effectively the deviation from the expected value, and calculating an average for each participant for both brightness and closeness sound sets.

Average hidden reference score modulus values for each participant are shown in table 5.1.

Generally it can be observed that on average a participant would score hidden references at around 0, with no differences for brightness and closeness sound sets, this indicates that the usage of sound stimuli within this test methodology is working as intended, and results for the rest of the sound stimuli used can be considered more valid.

The average hidden reference score modulus values for each experience level are shown in table 5.2.

This table shows that there is no distinct difference between levels of expertise. The vast majority of participants had at least some levels of experience, but even the two participants with no experience answered within a tolerable range around zero.

This same analysis of hidden references can also be undertaken for the age range groups of participants, shown in table 5.3.

Participant #	Brightness Average	Closeness Average	Total Average	Standard Deviation
1	0.26	0.09	0.08	0.74
2	0.18	0.23	0.20	1.29
3	0.24	0.08	0.16	0.77
4	0.12	0.04	0.04	0.48
5	0.08	0.00	0.04	0.20
6	0.48	0.35	0.42	0.96
7	0.46	0.21	0.33	0.70
8	0.56	0.03	0.26	0.73
9	0.28	0.13	0.07	0.63
10	0.53	0.19	0.36	1.81
11	0.23	0.54	0.16	0.81
12	0.15	0.32	0.23	1.34
13	0.03	0.19	0.08	0.67
14	0.02	0.66	0.34	1.55
15	0.01	0.00	0.00	0.02
16	0.21	0.54	0.38	1.26
Average	0.24	0.23	0.20	0.87

TABLE 5.1: Mean average hidden reference score modulus values per each participant in the listening test

Experience Level	No. of Participants	Bright Avr	Close Avr	Total Avr
None	2	0.28	0.19	0.22
Some Amount	7	0.26	0.21	0.21
Significant Amount	7	0.21	0.25	0.18

TABLE 5.2: Mean average hidden reference score modulus values per each level of experience in the listening test

Age Range	No. of Participants	Bright Avr	Close Avr	Total Avr
24 or Under	10	0.21	0.15	0.18
25-44	5	0.30	0.40	0.35
45-64	1	0.23	0.54	0.38

TABLE 5.3: Mean average hidden reference score modulus values per each level of experience in the listening test

The table shows that much like with experience level there was no large difference between age ranges in the score they assigned to the hidden reference in each sound set. A noticeable observation is that the average modulus scores for both brightness and closeness increased with age range, however further analysis of this is beyond the scope of this experiment and there is not a large enough spread of participants in all age groups to pursue this observation further.

5.6 Summary and Conclusion

The aim of this listening test was to assess the effect that the acoustic measurements of EDT, T30, and C80 had on the perceptual terms of 'brightness' and 'closeness'. In focusing on how each measurement effects each perceptual term individually, this test can be seen as the investigation of six primary hypotheses, stated at the beginning of the chapter. Analysis undertaken on the results of the listening test focuses on these individual hypotheses, and each hypothesis was interrogated individually in order to derive definitions of 'brightness' and 'closeness in terms of EDT, T30, and C80.

A listening test based off both the principles of MUSHRA and elements from perceptual sensory testing in non audio field was designed, where participants would listen to sets of sound stimuli in comparison to a 'neutral' reference sound and input a score based on either how 'bright' or how 'close' each stimuli sounded in comparison to the reference. The sound stimuli was generated through the convolution of a selected series of IRs with known EDT/T30/C80 values and three types of sound source; a percussive drum sound, a melodic singing sample, and a spoken word sample. Analysis of results showed no meaningful differences in scoring for brightness or closeness for these three types of sound source, so in further analysis the three convolutions with each IR were assumed to be equivalent.

The test was designed so that each sound set contained an hidden reference, a duplicate of the reference sound, to allow the assessment of the subjective accuracy of scoring for each sound set and for each participant. Through analysis it was discovered that on average the eight participants per sound set scored the hidden reference value at around zero, the desired value. In addition, there were not distinct scoring differences between participants of different experience levels. nor with the two perceptual terms.

Linear regressions were performed on the series of sounds used in sound sets for each section of the test. With sound stimuli used for the brightness section of the test being used to assess brightness and similarly for closeness. Each of the sound stimuli used in testing is a convolution between a sound source and an impulse response, and the EDT/T30/C80 values of these impulse responses were plotted against the average brightness/closeness scores. Results for each IR were derived via taking a mean of results of sound stimuli from the percussive, melodic and spoken convolutions of said IR. For example the average brightness score for the impulse response at 2m and 10% absorption is the mean of results from drums, singing, and speech stimuli at 2m and 10%. From each regression plot, observations are made on the size and direction of the resultant gradient, as well as how well the regression fits to the scatter plot of derived averages.

The linear regression plot for perceptual brightness over EDT can be expressed as the equation $y = 0.0749x + 0.1294$

This regression analysis shows that there is no correlation between EDT and perceptual brightness, as the gradient for the regression is insignificant as the scattered plot shows high amounts of non linear variability as EDT increases. Therefore it can be said that *EDT has no effect on brightness.*

The linear regression plot for perceptual brightness over T30 can be expressed as the equation $y = 0.2256x - 0.0285$.

The regression plot shows a positive gradient with an intercept around zero, indicating that the increase of T30 leads to relatively small increases of perceptual brightness strength, while the scatter plot shows a large amount of variability for comparatively small increases in T30. Therefore it can be said that *T30 has a positive correlation with brightness.*

The linear regression plot for perceptual brightness over C80 can be expressed as the equation $y = 0.0893x - 0.5367$.

The regression plot shows that the derived line of best fit trends with the scatter plot, and within the range of C80 values from the sound stimuli there is a distinct positive gradient. Therefore it can be said that *C80 has a positive correlation with brightness.*

A linear regression plot for perceptual closeness over EDT can be expressed as the equation $y = -0.0448x + 0.1800$.

The regression plot shows the regression does not trend with the scatter plot of average values, which show high levels of variability for small increases in EDT, so while there is a negative gradient in this regression it is deemed as too small deviating too much from the scatter plot to be deemed meaningful. Therefore it can be said that *EDT has no correlation with closeness.*

A linear regression plot for perceptual closeness over T30 can be expressed as the equation $y = -0.3323x + 0.4909$.

The regression plot shows a similar negative gradient to the closeness over EDT plot, but with a greater slope and greater trend with actual results across the range of values in this set of sound stimuli. Therefore it can be said that *T30 has a slight negative correlation with closeness.*

A linear regression plot for perceptual closeness over C80 can be expressed as the equation $y = 0.1588x - 1.2172$.

This regression plot shows a positive slope for the line of best fit which maps very closely to the majority of actual values sampled in the regression, therefore it can be said that *C80 has a positive correlation with closeness.*

Thusly, the conclusions from the test can be compared to the hull hypotheses as shown:

1. Perceived brightness of a sound has no direct proportionality with changing EDT, proving the null hypothesis
2. Perceived brightness of a sound has a linear proportionality with changing T30, disproving the null hypothesis
3. Perceived brightness of a sound has a linear proportionality with changing C80, disproving the null hypothesis
4. Perceived closeness of a sound has no direct proportionality with changing EDT, proving the null hypothesis
5. Perceived closeness of a sound has an inverse linear proportionality with changing T30, disproving the null hypothesis

6. Perceived closeness of a sound has a linear proportionality with changing C80, disproving the null hypothesis

Chapter 6

Conclusions and Further Work

The aim of this chapter is summarise conclusions from both the auralisation testing and semantic listening test experiments undertaken in this project, outlining the key takeaways from each; as well as the ultimate results for perceptual brightness and closeness descriptions as defined at the beginning of this thesis and relating to the initial hypothesis stated in section 1.2 of this thesis, as well as related discussions around the effectiveness and usefulness of this project work and potential further applications for it. This chapter consists of the following sections:

Section 6.1 - Derivation of Semantic Associations for Brightness and Closeness This section outlines the development of semantic representations for perceptual brightness and closeness to be implemented within a hypothetical semantic audio interface, presenting conclusions derived from auralisation and listening test experiments.

Section 6.2 - Review of Research Question This section is a discussion of results in relation to the initial hypothesis, and whether the overall project work was successful according to said hypothesis.

Section 6.3 - Key Conclusions from Experimental Work. This section will involve discussion around the key results from this project work; in the ODEON experiments, listening test experiment, and the overall derivation of expressions for perceptual brightness and closeness

Section 6.4 - Recommendations for Further Work. This section is a summary of useful takeaways from the testing and analysis work for this project, identifying elements of the methodology and implementation for this work that went well as well as less succesful aspects towards further work around this topic.

Section 6.5 - Significance of Research and Further Applications. This section involves discussion around how the theoretical and methodological aspects of this work can be potentially applied beyond the scale of this particular project and field of research in general towards wider academic and commercial applications of semantic interfacing for acoustic environments

6.1 Derivation of Semantic Expressions for Brightness and Closeness

Auralisation tests, outlined in Chapter 4, provided expressions of acoustic measurements EDT, T30 and C80 in terms of adjustable acoustic elements of a space; while semantic listening tests, as outlined in Chapter 5 produced associative descriptions for the perceptual terms and the aforementioned acoustic measurements.

Results show that the primary factor for perceptual brightness out of the ones analysed for this work is source/receiver distance, with absorption acting as a secondary factor. Perceptual brightness has a inverse proportionality with both absorption and distance, meaning that small changes to these variables produce little discernable differences, but these differences scale up in larger environments.

As stated in previous chapters, brightness is often associated with the reverberation of an acoustic space. Therefore, out of source/receiver distance and the absorption coefficient of surfaces within the space, prior knowledge would suggest that absorption coefficient would be a more the prominent factor. However, the opposite is observed within these experiments.

Inversely to brightness, source/receiver distance is the primary contributor to perceptual closeness, increasing distance leads to exponentially decreasing closeness. Absorption also contributes to closeness, leading to exponential increases as the absorption coefficient is

increased. As previously stated for brightness, exponential proportionality indicates that small changes for lower values of absorption and distance lead to relatively smaller changes in closeness. In smaller environments changing the source/receiver distance will lead to a less noticeable impact in closeness than a larger space.

Within existing literature closeness is associated with distance and proximity, therefore the source/receiver distance being the main factor of the two tested for perceptual closeness is in line with both what acousticians and non-experts think about what being close is in a practical sense. The quantification of this relationship between receiver distance and perceptual closeness is the novel output of this experimental work, that source/receiver distance is the primary factor for closeness is simply an affirmation of reliability of the methods and practices demonstrated in this thesis.

In conclusion, it can be said that in order to increase the brightness of a sound within an acoustic environment the absorption coefficient of the wall surfaces of a space must be decreased by a substantial amount, although the effect of changing source/receiver distance must also be considered when implementing this factor into a semantic interface; and that in order to increase the closeness of a sound within an acoustic environment the receiver should be brought closer to the source, although the absorption coefficient of wall surfaces must also be considered for semantic implementation of closeness.

6.2 Review of Research Question

The original hypothesis stated at the beginning of this thesis was *'Principles derived from semantic audio can be used to develop intuitive interfaces for auralisation models in order to fit the subjective needs of end users; which in itself can inform the intelligent optimisation of acoustic environments for the same purpose'*. In evaluating the conclusions from this work and methodologies for the experiments undertaken in this project, the general conclusion is that techniques from semantic audio proved effective in experimental and analytical work to derive general expressions for perceptual brightness and perceptual closeness.

While there are aspects of this research work that still remain unclear due to compromises made for this project and uncontrollable extrinsic factors effecting practical assessments

during the project period, the overall results of this work show a clear association the selected perceptual terms and the selected elements of room modification; through this research a there is a much clearer idea of how brightness (through increasing reflectiveness of surfaces) and closeness (through decreasing reflectiveness of surfaces and decreasing distance to the sound source) can be induced through changing the acoustic environment. Hypothetically this work can be scaled up significantly, allowing for a more granular investigative approach for even more elements of room modification and even more perceptual terms. The validity of this hypothesis can only be defined in this research work in terms of theoretical background and experimental work, since the practical implimentation of the perceptual terms in a proof of concept model was beyond the scope of this work.

6.3 Key Conclusions from Experimental Work

The overall goal of this work was to develop methodologies for designing interfaces where end users could interact with virtual models of acoustic spaces via semantic terminology. In order to achieve this the research work aimed to, through theoretical principles and data generated from experiments, investigate how subjective perceptual elements of sound could be affected through the modification of an acoustic space.

A small scale example example of this semantic design work was developed through the investigation of the perceptual terms brightness and closeness, and how these terms could be effected by changing the relative distance of the position of a sound receiver in comparison a sound source, expressed through single dimension source/receiver distance; as well as the level of reflectiveness of the primary wall surfaces, expressed as the absorption coefficient of wall surfaces.

This research work has lead to the conclusion that in an acoustic environment, perceptual brightness will increase exponentially with large decreases absorption coefficient. As a term used in a prospective semantic interface, increasing the brightness parameter would lead to the wall surfaces of a space being modified to become more reflective, within the context of room treatment this would mean the installation of reflective panels along the wall surfaces of a room.

Similarly, perceptual closeness will both increase exponentially with increasing absorption coefficient and increase exponentially with decreasing source receiver distance. It was derived that absorption is a more significant factor than distance in this regard, and it would take a many orders difference larger change in distance than change in absorption in order to change the perceived closeness of a sound within a space. This means that for a semantic interface, increasing the closeness parameter would lead to the wall surfaces being modified to become more absorbent, practically this would involve the usage of high absorption panels on wall surfaces. For closeness the receiver, whether that be a recording microphone or a human listener, is moved closer towards the source. The results imply that changing the absorption coefficient by 10% is for significant for perceived closeness than changing the source/receiver distance by 1m.

This implication that the reflectiveness of surfaces, at least in small and medium sized environments, factor more into 'closeness' than distance runs counter to the common association of closeness with distance. This therefore indicates that on a broader scale perceptual terms are not only defined by a large amount of factors within the sound itself, but also factors that do not fit a typical listeners definition of that perceptual term. It therefore outlines the need, if this work is to be scaled up and expanded upon, in thorough semantic testing of terms rather than referencing their common definitions for implementation within interfaces. In analysing all the factors that make up perceptual closeness rather than simply defining it as how 'close' a sound is to a listener.

In designing the research project this way there were a number of concessions made towards the level of detail the proposed work would operate with regarding the two room modification factors. For example, source/receiver distance was only in a single dimension, not accounting for lateral effects from a sound source (Y axis) or the influence of elevation on perceived sound (Z axis). This was a deliberate choice that allowed for an appropriate abstraction of distance modification that fit within the timeframe and scale of this research work, and the relative simplicity of measured and calculate results allowed conclusions around this factor to be clear and concise.

Similarly for absorption, this modification factor did not account for localised variances in absorption coefficient, either through different materials in different surfaces or through varying reflectiveness on a single surface. In the auralisation experimental work outlined in chapter 4, working with a model of the National Centre for Early Music and the resultant

instability of the auralisation measurements with said model, highlighted how increasing the geometric complexity of a space will lead to more complex unpredictable wave propagation within an acoustic environment; meaning it is harder to discern and quantify elements in an environment that define the perceptual of semantic factors of sound for users. Working with less complex environments with a smaller amount of more uniform surfaces allowed analytical work to produce much clearer conclusions.

The constraints for how room modification elements were quantified however, mean that clear cut approaches in how to modify a space to induce perceived 'brighter' or 'closer' sound become less feasible. Conclusions from this work outline that brightness is expected to exponentially decrease with increased absorption coefficient, but it is vague in terms of what that means in the actual modification of a space; increasing the reflectiveness of a certain surface over another, or even on a certain part of a surface via reflective panels, could lead to greater or less brightness increases. Moreover, real world materials do not have uniform absorption coefficient across frequencies, and the material of a surface is not the only factor for how reflective or absorbent said surface is. Take for example how wedge shaped acoustic foam can be optimised through adjusting the angle of wedges via finite element analysis [75].

The auralisation experiments lead to expressions for EDT, T30, and C80 in terms of distance and absorption. In these experiments it became apparent that EDT was an unreliable parameter to measure in the changing environments in the acoustic model. From theoretical principles it had already been established that EDT was sensitive to changes in distance, and practically measurements showed that these changes lead to chaotic non uniform variance. EDT and T30 are both decay time measurements, but T30 proved to be more reliable and a more significant factor for the perceptual terms, as shown through results from the listening test experiment outlined in chapter 5. Through the listening test results it was concluded that decay time T30 and energy ratio C80 factored into both perceptual brightness and closeness, emphasising the importance of those two acoustic measurements.

Listening test results were made less reliable due to the small sample size, and the online remote implementation of the test. Anomalous results for each sound set and in each ANOVA analysis were less likely to be made known due to the lack of data. Although the analysis of anchor value results indicated that all participants in the test were scoring

appropriately in a general sense, the conclusions from plotted results and regressions would be more reliable if they drawing from a larger pool of data

6.4 Recommendations for Further Work

In the context of the designed aim for this work to lead to verifiable quantifications of perceptual terms for usage in semantic audio interfaces, the scale and scope of this work was limited. An ideal semantic interface would allow for perceptual terms that are defined by a much greater range of variable factors within an acoustic space, and the perceptual terms being quantified would representing a greater range of aspects of timbre within a sound. Therefore the primary takeaway for further research work around this topic would be to increase both the number of elements in an acoustic space being investigated and the number of perceptual terms to be quantified for usage in said semantic interface.

If this research work was to be scaled up, it would still be important to take considerations into account when selecting perceptual terms for the semantic interface, as discussed in section 3.3. Distinct and relevant perceptual factors are crucial towards the development of practical semantic interfaces that non experts can utilise. The selection criteria process allowed for much smoother project development than would have happened otherwise, and emphasised the importance of the early stages of project development being cross referenced with the initial hypothesis of this research work, as well as the general aims and goals of this work.

For auralisation work, an approach where the techniques used to render impulse responses were freely available for analysis would prove beneficial, as it would allow more streamlined troubleshooting for errors or anamolus results when performing auralisation simulations or analysing the results of simulations. In this research work, auralisation experiments initially proved difficult to interpret and analyse due to the chaotic readings for EDT potentially being the result of a loss of information rendering impulse responses earlier in the energy decay curve.

There is also the question of what is the most appropriate simulated environment to perform auralisations in towards the aim of analysing impulse responses for acoustic measurements. The experimental work outlined in this project went through an iterative process as a live

environment with complex geometry was eventually replaced with a simple hypothetical room model with simple geometry; and all the spaces used emulated relatively small scale listening environments rather than large scale ones like concert halls. For further research work, it is recommended that further investigations be undertaken in order to determine the right type of environment for these types of auralisations to be run in. Ideally prior research and methodology development work would lead to experimental frameworks that can account for the complex reflections of real environments like the National Centre for Early Music.

In further listening test work, it is of great importance that the test is designed to be reasonably completed by participants. This means making a decision for how many questions it is appropriate to ask and why those questions are being asked. In deciding to assess melodic, percussive and spoken sound samples; the listening test for this project became three times larger than it needed to be, as the type of sound sample lead to little variance in results. The listening test had a estimated completion time of 45 minutes and out of 46 respondents only 16 answered all of the questions on the test, leading to small sample size despite high initial turnout.

In moving this work towards practical implementations of semantic interfaces for acoustic modelling applications, considerations need to be made not only for how perceptual terms are defined, but how perceptual terms overlap and interact with each. In this research work the expressions for brightness and closeness were considered individually, however in moving towards the implementation of these terms, accounting for how these two terms relate to each other will allow for a more accurate definition of perceptual terms for the end user. This is also true with room modification elements, where investigating the effects of distance and absorption being changed simultaneously would prove valuable. However this drastically increases the complexity of the overall semantic interface model if a greater number of perceptual terms and room modification elements were used.

6.5 Significance of Research and Further Applications

The aim of this research work was to develop methodologies toward the development of a semantic audio interface, where subjective perceptual term acted as parameters that could be changed in a scalar fashion, and a aspects of a virtually rendered acoustic model would

change towards increasing or decreasing the strength of stated perceptual term of a sound within it. These prospective interfaces could allow users to make sounds within a modelled space brighter or less bright via the modification of a 'brightness' variable.

The immediate use case for a practical example of this interface would be for allowing non experts who desire acoustic environments to sound a certain way, either virtual acoustic environments to be used in VR/AR and related spatial listening applications or real acoustic environments such as recording studios or concert halls, to design spaces to produce desired effects; allowing for virtual environments to have desired acoustic qualities and allowing for the knowledge of how to treat an acoustic environment in the real world to induce those qualities. This would empower non expert designers and creatives to tackle room treatment and acoustic design on their own through semantic interfaces, in the way that they would have normally interfaced with acoustic and audio engineers for sound in a live space; in a similar fashion to how digital audio effects lead to the same thing happening for sound recording and mixing.

In allowing non experts to conceptualise an acoustic environment in terms of perceptual terms that were selected due to their understandability and relevance, prospective semantic audio interfaces could also act as potential educational tools. Allowing users to hear how for example, sound in a room can be made to sound brighter or less bright. If a non expert could interact with an interface and be able to hear in real time the difference of the sound as it's 'brightness' changes then they would be able to more easily understand what 'brightness' is in the context of audio.

Machine learning, the study of algorithms that automatically improve through repeated data inputs, has the potential to draw from this semantic audio interface work and apply it in a wider variety of novel applications. A key aim for research around the field of semantic data, in the context of information science, is to make concepts and the relationship between them machine readable, so that independent agents (AI) have the means to interpret and act upon what has typically been non quantifiable data. In the context of this work and semantic audio as a whole, this involves presenting expressions for subjective concepts around the timbre and related characteristics of audio in a way that's interpretable by a machine learning system. The field of music information retrieval, discussed in section 2.3.3, contains many examples of this type of work.

This project work involves the derivation of mathematical expressions of perceptual brightness and closeness in terms of modifiable elements within an acoustic environment. In quantifying these subjective perceptual terms in a machine readable format, and through expanding resultant semantic audio interfaces to incorporate more perceptual terms and granular modifications within a space (lateral positional movement of receiver, positional aspects to surface absorption), machine learning programs could accommodate more complex semantic models for perceptual terms. Existing musical information retrieval (MIR) frameworks around perceptual factors for musical timbre semantically describe the subjective perception of sound by a listener at many orders of magnitude greater than in this research work; and these can be incorporated into semantic interfaces driven by an algorithm based methodology rather than simple testing like in this research project, meaning that semantic interfaces could interact with a greater variety of elements within an modelled acoustic environment, and that perceptual factors in the interface could be more well defined.

Existing literature in the field of MIR has investigated how machine learning can lead to the procedural generation of audio content towards a desired emotive response through mapping these responses to mid level timbre features such dissonance and rhythmic complexity. In a similar fashion, work on semantic interfaces for acoustic environments could potentially lead to optimisation algorithms for treating an acoustic environment, either a virtual model or a real space, toward a desired emotional response from listeners; drawing from similar principles where instead of emotive responses being mapped to timbre features they are mapped to perceptual factors, with these perceptual factors being mapped to adjustable aspects of an acoustic space like as described in this research work. An automated algorithmically driven system like the one described would be able to automatically treat an acoustic environment towards the desired emotional response of end users for sound within it.

Appendix A

MATLAB IR Analysis Toolkit Code

All MATLAB .m files for the IR analysis toolkit are attached in the accompanying Appendix A folder, this folder contains:

- **acousticParams.m** - The primary script for the toolkit, allowing for an input of a data matrix of a single of multichannel audio file as well as the desired paramaters and frequency bands of measurement. The resultant output will provide measurements of the input file for the selected parameters and bands, in the form of a results table.
- **APTScript.m** Front end script which takes input in the form of a filepath and provides graphical plots as outputs, as well as a spectogram generated using
- **bandFilter.m** - Filters octave bands according to user input. Subscript called by acousticParams.m
- **energyCalc.m** - Calculates energy ratio based acoustic measurements, C50 and C80 clarity, D50 and D80 definition, and CT spectral centroid. Subscript called by acousticParams.m
- **parseArgs.m** - Interprets user preference inputs from acousticParams.m
- **reverberationCalc.m** - Calculates reverberation based acoustic measurements, EDT early decay time and T20, T30, and T40 decay time. Subscript called by acousticParams.m

- **spectroPres.m** - Generates spectogram for input audio file
- **subaxis.m** - Creates MATLAB subplots with their own x and y axis. Subscript called by acousticParams.m

Appendix B

Auralisation Experiment Results

B.1 Experiment #1 IR Measurements

EDT Measurements (Absorption 20%)			
	500Hz	**1000Hz**	**2000Hz**
4m	2.401	2.580	2.858
8m	1.743	1.496	1.376
12m	1.418	1.240	1.124
16m	0.880	0.917	0.882
EDT Measurements (Absorption 40%)			
	500Hz	**1000Hz**	**2000Hz**
4m	0.815	0.424	0.390
8m	1.345	1.325	1.118
12m	1.456	1.337	1.250
16m	0.641	0.806	0.786
EDT Measurements (Absorption 60%)			
	500Hz	**1000Hz**	**2000Hz**
4m	0.221	0.322	0.353
8m	1.508	1.719	1.061
12m	1.663	1.094	0.925
16m	0.879	1.098	0.884
EDT Measurements (Absorption 80%)			
	500Hz	**1000Hz**	**2000Hz**
4m	0.841	0.906	1.025
8m	1.464	0.500	0.473
12m	1.052	1.222	1.197
16m	0.983	0.857	1.040

T30 Measurements (Absorption 20%)			
	500Hz	**1000Hz**	**2000Hz**
4m	4.372	4.788	5.940
8m	1.821	0.476	1.158
12m	-2.029	3.377	3.946
16m	3.352	5.129	5.904
T30 Measurements (Absorption 40%)			
	500Hz	**1000Hz**	**2000Hz**
4m	8.802	10.009	11.127
8m	2.287	2.552	1.932
12m	-3.560	-2.700	-1.921
16m	3.306	4.801	6.196
T30 Measurements (Absorption 60%)			
	500Hz	**1000Hz**	**2000Hz**
4m	12.156	12.777	13.085
8m	6.303	6.604	7.852
12m	-0.370	5.364	6.180
16m	4.816	6.349	6.713
T30 Measurements (Absorption 80%)			
	500Hz	**1000Hz**	**2000Hz**
4m	3.060	4.655	4.219
8m	6.597	9.493	9.901
12m	-1.193	3.365	4.250
16m	-1.336	2.469	3.611

B.2 Experiment #2 IR Measurements

C80 Measurements (Absorption 20%)			
	500Hz	**1000Hz**	**2000Hz**
4m	5.070	5.362	6.298
8m	2.985	2.440	2.919
12m	1.533	5.504	6.274
16m	7.506	7.761	7.969
C80 Measurements (Absorption 40%)			
	500Hz	**1000Hz**	**2000Hz**
4m	9.258	10.759	11.936
8m	3.880	4.396	4.360
12m	-0.014	1.685	1.307
16m	8.256	8.509	8.587
C80 Measurements (Absorption 60%)			
	500Hz	**1000Hz**	**2000Hz**
4m	13.253	14.194	14.336
8m	7.361	7.627	9.053
12m	3.667	7.963	8.242
16m	6.896	7.877	8.333
C80 Measurements (Absorption 80%)			
	500Hz	**1000Hz**	**2000Hz**
4m	5.816	6.984	7.245
8m	7.755	11.043	11.821
12m	1.449	4.862	5.833
16m	0.983	0.857	1.040

Appendix C

Listening Test Survey

The qualtrics survey of the listening test that was designed and distributed to participants can be found at `https://york.qualtrics.com/jfe/form/SV_0jNldhDjm0rvE34`. There is also a printed version of the survey in the accompanying Appendix C folder.

Appendix D

Listening Test IR Measurements

D.1 Listening Test IR Set Measurements (EDT)

Distance/Absorption Coefficient	10%	30%	Hidden Reference (40%)	50%	70%
2m	5.118	0.188	0.017	0.017	0.017
4m	2.113	2.531	1.782	0.377	0.311
Hidden Reference (5m)	1.857	1.983	X	0.694	0.343
6m	1.682	1.540	1.536	1.087	0.258
8m	1.483	1.315	0.941	0.984	0.318

D.2 Listening Test IR Set Measurements (T30)

Distance/Absorption Coefficient	10%	30%	Hidden Reference (40%)	50%	70%
2m	1.810	1.252	0.907	0.754	0.737
4m	1.934	1.288	0.788	0.713	0.589
Hidden Reference (5m)	2.016	0.923	X	0.691	0.570
6m	2.096	1.287	0.717	0.693	0.534
8m	2.073	1.326	0.770	0.624	0.534

D.3 Listening Test IR Set Measurements (C80)

Distance/Absorption Coefficient	**10%**	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
2m	8.314	10.201	14.516	15.614	21.761
4m	3.652	5.665	8.610	10.692	15.462
Hidden Reference (5m)	1.717	4.878	X	9.490	13.357
6m	2.238	3.428	5.324	8.315	12.168
8m	0.467	1.783	5.746	6.409	11.237

Appendix E

Listening Test Sound Files

All .wav files attached in accompanying Appendix E folder, this folder contains:

- **Anechoic Sources** - Contains the .wav files for the percussive, melodic, and speech sources that were convolved to generate the sound stimuli as described in section 5.4
- **Brightness & Closeness Examples** - Contains the example .wav files used for the calibration part of each section in the test, as described in section 5.4
- **Impulse Responses** - Contains .wav files for each IR used in the listening test
- **Sound Stimuli Sets** - Contains the convolved sound files used in questions in the listening test, each stimuli set contains files for a single question, with the control variable being displayed in the folder name (Drums40% contains all percussive sound stimuli with a constant absorption coefficient of 40%)

Appendix F

Listening Test Results

F.1 Listening Test Results - Brightness

The following shows the scores entered by participants for each of the brightness questions in the listening test. Each table shows results for all questions where one variable (distance of absorption) was the same constant value while the other changed, and each table contains subtables for each type of stimuli (melodic, percussive, speech). Each row shows results for each participant.

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Distance 2m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
2	0.5	-0.8	-0.2	-1.4
1.9	1.2	0	4.1	0.9
-1.2	-1.6	0	1.1	2.9
-2	-1	0	1	5
2	1	0	0	0
1.9	0.5	0.2	-1	-1.1
1.4	-0.2	1.7	-0.2	-0.3
0.3	1.2	1	0.6	0.6
Melodic Set (Control Distance 2m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
1.2	-0.1	0.1	0.5	0.3
4.3	-0.9	-4.5	2.6	1.1
1.9	0.8	0.2	1.2	-0.3
-0.5	-0.4	0	1.7	0.2
1.5	0.6	-0.9	1	-0.5
-2.4	1.3	0	2	1.8
0	0	0	0	0
1	0.5	1	3.5	4.5
Speech Set (Control Distance 2m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-0.1	0.3	0.1	0	0.9
3.8	0.6	2.4	-1	-0.5
1	1.5	2.1	0	0.5
-0.8	-1.5	0.8	0.3	-1.8
-0.4	1.6	0.6	0.6	0.6
1.3	0	-0.2	1	2.2
0	0	0	0	0
-2	0.5	0.5	-1	2

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Distance 4m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
0.8	0	0.3	0.3	-0.5
2.1	-0.7	-1.1	-2.2	-1.4
0.6	-0.1	0	0.2	-0.3
-0.5	0	0	-0.2	0.5
-0.8	-0.5	0.4	0.2	-0.3
2	-3.1	3.1	-2	1.4
0	0	0	0	0
3	-0.5	0.5	1.5	2.5
Melodic Set (Control Distance 4m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
0.4	0.2	-0.3	-0.3	-0.4
3.3	-1.2	0	2.1	-2.3
-1.9	0.7	-0.5	-1.2	2.4
-3	-2	-1	1	2
-1	0	0	1	0
-2	1	1.4	-0.3	-0.4
0.7	0.8	0.5	0.7	0.6
-0.5	0	2	1	0.5
Speech Set (Control Distance 4m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
2.2	0	0.8	-0.4	-1.7
2.3	0.6	0	1.8	2.9
-1.1	-0.6	-0.4	0.5	0.4
-4	-1	-1	1.5	5
3	1	0	0	0
1.6	0.7	-1.1	-1.6	0.5
-1	1	-0.2	0.7	2.2
-1	0.5	1	2	3

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Distance 6m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-0.3	0	0.1	-0.2	0.2
2.6	0.4	-0.2	1.5	-2.8
0.2	1	0.7	0	-1.7
1.3	1	0.1	-2.1	-0.8
-1	0	-0.2	0.3	-0.5
2.2	-2.1	0	-0.5	3.3
0	0	0	0	0
2.5	-2	-1	-1.5	1
Melodic Set (Control Distance 6m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
1.5	0.7	2	1.2	1.3
1.3	1	-0.5	-1.9	-3.8
0.7	-0.4	0	-1	-1.4
2.4	1.2	0.5	0	-0.6
0.3	1	0	-0.7	-0.6
1.1	1.1	-1.2	-1.1	1.3
0	0	0	0	0
-1	2	0.5	3	4.5
Speech Set (Control Distance 6m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
1.2	-0.2	0	0.4	-0.4
1.3	1	0	-0.4	-2.3
-1.4	1.4	1.2	0.1	0.5
-1.5	-0.5	-0.5	0.5	5
1	2	0	1	0
1.3	-2.2	-0.9	0.3	-1.3
1.3	1	0.7	0.6	0.4
0.3	0.1	-0.2	-0.4	0.3

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Distance 8m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
0.8	0.7	-0.2	-0.5	-0.4
-2.1	-2.6	0.1	-1.5	-3.4
-3.1	-0.5	-1.8	0.1	0.9
-2.5	-0.5	-0.2	1.5	2
1	1	0	-1	-1
-1.2	-2.5	0.5	-3.1	-2
0.5	0.3	-0.4	0.5	-0.3
0.6	-0.2	0.1	-0.4	-0.7
Melodic Set (Control Distance 8m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
0.4	1.2	0.2	-0.2	-0.4
2.3	1.2	-0.5	-1.4	-2.3
-2.1	0.2	0	0.6	-0.2
-4	-2	-0.1	-0.1	3
3	0	0	0	-1
-0.4	1.5	-0.1	-1	-1.6
1	3	-0.2	2	0
0.5	-1.5	1	-1	0.6
Speech Set (Control Distance 8m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-0.9	-1	0	-1.7	
2.4	1.2	-2.3	-0.8	-3.5
1.6	0.8	0	0.3	-0.5
2.5	-0.8	0	0	0.6
-0.5	0.5	0.8	0.5	1
0.1	-0.9	-2.6	2.1	1
0	0	0	0	0
-1.5	-1	0.5	-2	2

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Absorption 10%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
4.6	2.2	-0.2	0	-0.5
1.1	-0.7	2.1	-1.4	-1
4.2	1.2	0.6	-0.1	-0.5
3.6	2.4	1.4	0	1.4
2.8	-0.7	-1.3	-1	-1.5
1	3.1	2.1	-1	-3.1
0	0.3	0	-0.3	-0.9
1	2	1.5	-1	-2
Melodic Set (Control Absorption 10%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.6	1.8	2	1	0.5
4.2	1.7	0	-1.1	-3.2
1.9	1.2	0	0	-2.9
4	2	-0.1	-0.1	-2.5
5	1	0	0	-1
2.1	-0.6	1	-0.6	-2.3
0.3	0.5	0.9	0.2	0.5
1.8	-0.3	0.3	0.4	-0.5
Speech Set (Control Absorption 10%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.2	1	0.5	-0.2	0.5
3.6	0	0.3	-0.4	-1.4
1.2	0	-1.5	-1.1	-1.4
2.5	1	0	0	-1
5	1	0	0	-2
1.2	0.9	-2.2	-1.4	-2.9
2	-0.7	-0.5	-2	-1
2.6	0.1	0.5	0.2	-0.3

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Absorption 30%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
0.8	-0.5	0.2	-0.2	1.4
5	0.8	0	-1.9	-2.9
1	1.2	0.3	-0.6	-2.7
5	3	-0.5	-1.5	-2.5
4	1	0	-1	-1
4.2	1.1	-2.7	-0.8	-3.5
0.7	-0.4	0.5	-0.5	0.4
4	2	0	0.5	-2
Melodic Set (Control Absorption 30%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.6	1.2	0.2	-0.1	-0.2
4.3	3.6	2.5	-0.4	-1.7
2	0.5	0	-1.6	-2.3
3.7	3	1.3	2.4	2
2	1.4	0	-0.6	-1
2.3	2.3	-1.3	-3.2	-2.5
0	0	0	0	0
3	2.5	2	1	-1
Speech Set (Control Absorption 30%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.7	-0.5	0.5	-0.3	-1.3
3.5	3.1	-2.5	-2.4	-1.8
3.8	1.7	0	-1.5	-1.5
5	0.5	0.4	-1.1	-2.4
4	0.5	-0.2	-0.6	-3
1.2	2	0.5	-2.8	-1.4
0	0	0	0	0
3.5	2	-0.5	-1	-2.5

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Absorption 50%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
0.8	0.2	0.4	1	0.8
5	1.6	-1.3	0	-3.5
1.6	1.5	-0.1	-2.9	-3.5
5	2.5	0.5	-1	-2
5	0	0	0	-3
3.9	1.3	-0.8	0.2	-2.7
0.6	0.3	1	0.9	1
4	0.3	1	-0.5	-2
Melodic Set (Control Absorption 50%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
1.5	0.8	-1.4	-0.9	-2.3
3.8	-0.4	0.8	-1.4	-3.2
2.7	1.2	-2.8	-4	-5
3.5	1	0	-1	-1.5
4	1	0	0	-3
2.8	1	-1	-2.2	-3.1
1	0.8	-0.2	0.4	1
3	0.1	-0.7	-1.2	-2
Speech Set (Control Absorption 50%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
3.2	0.7	-0.1	0	-0.5
1.3	-0.5	0.6	-1	-3
4.5	1	-0.2	-1.1	-4
3.9	2.5	-1.2	0.8	-4
3.5	1	0.3	-0.5	-1
1.9	2.2	-1.5	-1.1	-4.1
0	0	0	0	0
5	2	-2	-1	-3.5

Listening Test - Brightness Scores (-5 to 5)				
Percussive Set (Control Absorption 70%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
3.5	1.7	0.1	0.1	-1.8
4.1	2.5	-1	-0.8	-2.5
3.6	2.2	0.3	-0.7	-1.5
2	0	-1.4	-1.3	-0.6
3.5	1	0.5	-1	-2
3.2	4.1	1.6	2.1	-2.2
0.6	1.2	-0.1	-1.2	-1.2
-4	4	0.5	-2.5	-5
Melodic Set (Control Absorption 70%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
3.3	1.2	1.2	0	1.2
5	3.6	-1.8	1	-3.3
2.7	0.7	-1	0	-1.7
4.1	-0.1	-0.1	-0.9	0.4
2.4	1	0.3	-2	-0.8
3	0.8	-0.8	-0.9	-2.3
0	0	0	0	0
5	4	-1	0.5	-1
Speech Set (Control Absorption 70%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
1.2	0.8	-0.1	-0.2	0.4
4.7	1.7	0.5	-0.1	-2.6
1.4	2.7	0	-1.3	-2.4
5	2	1	-0.5	-1
4	2	1	0	-1
2.4	1.9	-0.3	-1.7	-2.5
2.3	0.9	0.5	-1.4	-1.9
2	1	-1	-3.1	-4

F.2 Listening Test Results - Closeness

The following shows the scores entered by participants for each of the closeness questions in the listening test. Each table shows results for all questions where one variable (distance of absorption) was the same constant value while the other changed, and each table contains subtables for each type of stimuli (melodic, percussive, speech). Each row shows results for each participant.

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Distance 2m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-1.3	-1	0	0	1.8
3.2	1.1	-0.6	-2.3	-0.7
-1.1	-0.5	-0.1	0.2	0.6
3.2	1.9	2.5	1.1	0.5
1	0.3	0	0.5	2
2.1	1.4	0.7	-1.5	1.6
0	0	0	0	0
-2.5	-1	0.5	4	4
Melodic Set (Control Distance 2m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-2.4	-1.2	-0.5	-1	0.2
-3	-1.2	0.1	-1.3	2.3
-1.8	-1.1	0	1	1.8
-1.5	0	0	1	1.5
0	0	0	0	0
-0.3	-0.4	-0.3	0.2	0.7
0.6	0.6	1.4	0.7	0.9
-1	2.5	1	1.4	1.2
Speech Set (Control Distance 2m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-1.8	-0.8	0.2	0.5	
-3.2	-1	0.5	0	5
-0.4	0.3	1	1	1.1
-1.5	-1	0	1	3
-0.5	0	0	0	1
-1.4	-0.9	-0.3	0.6	2.1
1.2	0.8	0.5	0.6	1.4
-1.4	-0.7	0.2	0.8	1.3

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Distance 4m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-0.8	-0.2	0	0.2	1
-1.2	0	1.5	2.1	2.9
-1	0.2	1.4	0	0.9
-2	0.5	1	0.5	1
-1	0	0	0	0
-0.7	0.1	-0.7	-0.5	-0.1
-1.2	-0.7	0.7	0.3	-0.8
-0.5	0.4	-0.2	0.5	0.5
Melodic Set (Control Distance 4m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-3.1	-2.1	-0.7	0	1.1
0.7	1.5	2.9	-1.5	-1.2
-1.8	-0.8	0	-0.3	1.1
-2.8	1.4	-0.1	1.2	3.8
0	-1.5	-0.5	1	2
-3.2	2.5	-1.7	0	4
0	0	0	0	0
-4	1.5	2.5	3	3
Speech Set (Control Distance 4m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-1.7	-0.8	-0.3	0	2.2
-0.9	-0.6	1	-0.8	2.1
-1.9	0	0	0.5	1.4
-3.8	-3	-1.3	1.9	1.8
1.8	1	0	-0.5	1.8
-4.3	-1.9	-1.4	1.8	3.7
0	0	0	0	0
-2	-0.5	1	1.5	4

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Distance 6m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
2.2	-2	0	-0.2	1
-1.6	0	1	-2	1.8
-1	0.8	0	0.2	-0.5
-2	1	0	1	2
-1	-0.5	0	0	1
-1.7	-1.3	0.6	0.4	0.5
-0.7	0.8	-0.3	-0.6	0.4
0.4	-0.6	-0.5	0.4	-0.3
Melodic Set (Control Distance 6m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-2	-1	0	0.2	0.4
-1.2	-2.4	0	1.8	3.9
-1	0	-0.4	0.5	0
-1	-0.5	0	0.5	2.5
-1	0	0	0	0
-2.3	-1.7	-0.1	0.4	0.8
0.6	0.9	-0.7	1.7	-0.1
-0.5	0.2	0.5	1	0.7
Speech Set (Control Distance 6m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-2.5	-1.5	0	1	1.8
4.1	2.1	-0.9	-2.3	0.9
0	0	0	1.5	4.8
-3	-2.2	-0.7	0	1.2
0	0.5	-0.3	-0.5	1
-3.5	-1.3	-1.4	1.7	2.8
0	0	0	0	0
-3	0.5	-1	3	4

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Distance 8m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-2.5	-1.1	0	1.3	1.2
0.6	1.1	-1.1	0.7	0.1
-2.9	-0.6	-1.6	0.6	1.5
-3.8	-2.3	-2.2	-0.7	0.8
-1	0	0	0.6	0.5
-4.3	-0.8	-2.8	1.8	4.2
0	0	0	0	0
-3	0.5	-0.5	1.5	-1.5
Melodic Set (Control Distance 8m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-2.8	-1	0	0	2.8
-4.3	-2.7	-0.5	-1	-3.6
-0.3	0	0	0	0
-4.3	-1.8	1.5	1.9	1.1
-0.5	-2	-0.5	-0.5	0
-3.1	-2.4	1.3	-0.5	3
0	0	0	0	0
-3	1	2	-2	4
Speech Set (Control Distance 8m, Varying Absorption)				
10%	**30%**	**Hidden Reference (40%)**	**50%**	**70%**
-1.4	-0.8	0.2	0	2
-2.6	-1	-1.6	1.5	2.1
-1.4	-2.2	-0.1	-1.3	-0.6
-2	-0.5	0	0.5	2.5
0	0	0	0	1
-1.9	0.7	-0.4	0.3	1
-1.4	0.3	-0.5	1.2	1
0.5	-0.5	-0.8	-0.8	-1.4

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Absorption 10%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
3.4	1.2	0.4	-1.4	-1.8
5	-1.3	-4.4	0	-3
3.8	0.9	0.3	-1.5	-3.2
5	1	0	-1	-1.5
5	1	0	0	-2
3.7	0.7	1	-0.8	-2.4
-0.7	1.4	-0.5	2.1	0.7
1.6	0.2	0.6	-0.7	-0.2
Melodic Set (Control Absorption 10%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
4.3	1.1	-0.9	-0.6	-1.5
2.6	0.4	-0.6	-1.6	-3.4
2.7	0.3	-0.5	-0.7	-1.7
4.2	2.4	1.8	1.1	-2.5
5	3	1.6	0.5	-3
5	3.5	2	1.6	0.2
3.9	0	0	0	-2.7
5	0.5	3	2	-1.5
Speech Set (Control Absorption 10%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
3.9	1.9	-0.4	0	-1.7
1.4	1.6	-2.7	1.4	-4.2
2.9	0.8	-1.1	0	-1.6
1.9	-0.4	-0.4	0.9	-2.1
5	0.7	-0.5	1.6	-1.9
5	1.9	0	-1.6	-2.5
3.8	2.2	0	0	0
5	2	0.5	0.5	-2

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Absorption 30%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
5	1.3	0	-1.1	-2.3
4.2	1.2	-0.8	0.7	-2.8
3	1	-1.7	-1.7	-2
4.1	1.8	2.5	0	-3.9
3	1	0	-0.5	-1
2.8	-1	-1.3	-2.1	-2.9
4.5	1.5	0	-1	-1.4
4	2.5	-0.5	-3.5	-4
Melodic Set (Control Absorption 30%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
1	-1	-1	-1.2	-1.2
5	2	0.4	0	-2
1.9	1.1	0	-2.7	-3.8
4	1	0	0	-1
5	2	0	-1	-2
2.4	0.5	-0.3	-2.2	-3.3
1.5	0.4	0.6	0.7	0.8
2	0.8	0.3	-0.5	-1.3
Speech Set (Control Absorption 30%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.8	-0.8	0.2	-0.4	-0.5
4.1	2.5	0	-0.3	-1
3.8	1	-1.6	-2.3	-3.1
3.5	1.5	0	-0.5	-1
5	3	0	-1	-2
4.1	0.9	-0.2	-0.6	-2
0.8	-0.3	0.9	1.1	0.7
1	0.3	-0.9	-1.6	-1.7

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Absorption 50%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
4.3	1.2	0	-0.9	-3.5
2.6	1.8	1	-3	-1.9
4.5	0.1	-1.5	-2.7	-3.4
3.9	1.7	-0.9	-4.8	-4.1
4	1	0	-1.2	-3.1
5	1.5	-2.1	-1	-5
4.2	0.9	0	-1.2	-2.8
5	-1	-1	-3	-4
Melodic Set (Control Absorption 50%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
0	0	0.9	-1	-2.5
3.5	-1	-1.2	-3.6	-2.8
4.6	0	0	-1.2	-2.3
4.2	1.4	-1.4	-3.8	-5
4.1	1.9	-1.6	-2.2	-4.2
2.9	1.6	-1.2	-3.1	-4.9
4.2	0.8	0	0	-2
4.5	1	-1	-3	-5
Speech Set (Control Absorption 50%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
0.2	-0.4	-1.2	0.2	-0.5
5	1.6	2.1	0	-5
4.1	0.7	-0.3	0	-3
3.5	1	0	-0.5	-3
5	1	0	-1	-2
3.7	0.9	-1.1	-1	-2.9
1	0.5	0.7	0.9	-0.8
2	0.5	-0.3	-1	-1.9

Listening Test - Closeness Scores (-5 to 5)				
Percussive Set (Control Absorption 70%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.8	-0.2	1.8	-0.4	-1.2
5	2.7	0	1.5	-1.8
3.3	0.7	-1	-0.6	-3
5	1.5	0	-1.5	-3
5	2	0	0	-2
3.9	0.9	-0.9	-1.7	-3
0.5	1.2	-0.9	-1.1	3.4
2.7	1	-0.2	-0.4	-3.5
Melodic Set (Control Absorption 70%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
2.2	-0.5	-0.2	0.2	-0.8
4.3	1.6	-2.3	0	-1.8
4.3	0.8	-0.3	0	-1.9
5	1.5	-0.5	-0.5	-1.5
5	1	0	0	-1
3.6	1.3	-1.5	-0.3	-2.2
-0.4	1.6	0.6	3.3	0.7
2.2	0.5	-0.1	-1	-0.4
Speech Set (Control Absorption 70%, Varying Distance)				
2m	**4m**	**Hidden Reference (5m)**	**6m**	**8m**
5	1.9	-0.2	-1.2	-2.4
2.2	0.1	1.2	-2.3	-4.5
5	0.8	0	1.5	-0.5
5	3.2	2.5	1.3	0.5
5	1	-0.5	-0.5	-3
4.1	1.6	0	-0.8	2.6
4	1.9	0	0	-0.9
5	3	1	0.5	-2

Appendix G

Listening Test Two Way ANOVA Results

G.1 Brightness (Constant Absorption, Changing Distance Questions)

BRIGHTNESS					
Group of Percussive Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Distance (Varies Within Each Sound Set)	357.952	4	89.4881	42.84	0
Absorption (Varies Between Sound Sets)	1.638	3	0.5462	0.26	0.8531
Interaction	23.839	12	1.9866	0.95	0.4987
Error	292.465	140	2.089		
Total	675.895	159			
Group of Melodic Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Distance (Varies Within Each Sound Set)	337.718	4	84.4294	51.72	0
Absorption (Varies Between Sound Sets)	20.114	3	6.7047	4.11	0.0079
Interaction	17.077	12	1.4231	0.87	0.577
Error	228.53	140	1.6324		
Total	603.439	159			
Group of Speech Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Distance (Varies Within Each Sound Set)	416.5	4	104.125	80.13	0
Absorption (Varies Between Sound Sets)	1.448	3	0.483	0.37	0.7737
Interaction	17.65	12	1.471	1.13	0.3394
Error	181.931	140	1.3		
Total	617.529	159			

G.2 Brightness (Constant Distance, Changing Absorption Questions)

BRIGHTNESS					
Group of Percussive Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Absorption (Varies Within Each Sound Set)	10.27	4.00	2.57	1.41	0.24
Distance (Varies Between Sound Sets)	24.06	3.00	8.02	4.40	0.01
Interaction	13.19	12.00	1.10	0.60	0.84
Error	255.50	140.00	1.83		
Total	303.02	159.00			
Group of Melodic Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Absorption (Varies Within Each Sound Set)	4.09	4.00	1.02	0.48	0.75
Distance (Varies Between Sound Sets)	8.14	3.00	2.71	1.27	0.29
Interaction	28.27	12.00	2.36	1.10	0.37
Error	300.18	140.00	2.14		
Total	340.68	159.00			
Group of Speech Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Absorption (Varies Within Each Sound Set)	4.48	4.00	1.12	0.58	0.67
Distance (Varies Between Sound Sets)	8.58	3.00	2.86	1.49	0.22
Interaction	14.52	12.00	1.21	0.63	0.81
Error	268.32	140.00	1.92		
Total	295.91	159.00			

G.3 Closeness (Constant Absorption, Changing Distance Questions)

CLOSENESS					
Group of Percussive Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Distance (Varies Within Each Sound Set)	687.34	4.00	171.83	106.94	0.00
Absorption (Varies Between Sound Sets)	10.85	3.00	3.62	2.25	0.09
Interaction	28.26	12.00	2.36	1.47	0.14
Error	224.95	140.00	1.61		
Total	951.39	159.00			
Group of Melodic Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Distance (Varies Within Each Sound Set)	548.02	4.00	137.01	91.50	0.00
Absorption (Varies Between Sound Sets)	41.09	3.00	13.70	9.15	0.00
Interaction	38.45	12.00	3.20	2.14	0.02
Error	209.63	140.00	1.50		
Total	837.20	159.00			
Group of Speech Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Distance (Varies Within Each Sound Set)	501.98	4.00	125.49	84.07	0.00
Absorption (Varies Between Sound Sets)	14.63	3.00	4.88	3.27	0.02
Interaction	15.05	12.00	1.25	0.84	0.61
Error	208.98	140.00	1.49		
Total	740.63	159.00			

G.4 Closeness (Constant Distance, Changing Absorption Questions)

CLOSENESS					
Group of Percussive Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Absorption (Varies Within Each Sound Set)	48.54	4.00	12.13	8.41	0.00
Distance (Varies Between Sound Sets)	17.82	3.00	5.94	4.11	0.01
Interaction	29.37	12.00	2.45	1.70	0.07
Error	202.09	140.00	1.44		
Total	297.82	159.00			
Group of Melodic Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Absorption (Varies Within Each Sound Set)	132.26	4.00	33.07	18.36	0.00
Distance (Varies Between Sound Sets)	9.61	3.00	3.20	1.78	0.15
Interaction	16.21	12.00	1.35	0.75	0.70
Error	252.19	140.00	1.80		
Total	410.28	159.00			
Group of Speech Sets					
Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Absorption (Varies Within Each Sound Set)	166.90	4.00	41.73	25.32	0.00
Distance (Varies Between Sound Sets)	3.85	3.00	1.28	0.78	0.51
Interaction	8.89	12.00	0.74	0.45	0.94
Error	230.69	140.00	1.65		
Total	410.33	159.00			

Bibliography

[1] V. Rawool, *Hearing conservation: In occupational, recreational, educational, and home settings*. Thieme, 2011.

[2] I. L. Vér and L. L. Beranek, *Noise and vibration control engineering: principles and applications*. John Wiley & Sons, 2005.

[3] R. Erickson, *Sound structure in music*. Univ of California Press, 1975.

[4] E. Ballestero, P. Robinson, and S. Dance, "Head-tracked auralisations for a dynamic audio experience in virtual reality sceneries," in *24th International Congress on Sound and Vibration*, 2017.

[5] V. Alluri and P. Toiviainen, "Exploring perceptual and acoustical correlates of polyphonic timbre," *Music Perception*, vol. 27, no. 3, p. 227, 2010.

[6] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[7] www.openairlib.net, *Innocent railway tunnel*. [Online]. Available: `https://www.openair.hosted.york.ac.uk/?page_id=525`.

[8] D. S. Brungart and W. R. Rabiowitz, "Auditory localization in the near-field," Georgia Institute of Technology, 1996.

[9] F. R. Moore, "A general model for spatial processing of sounds," *Computer Music Journal*, vol. 7, no. 3, pp. 6–15, 1983.

[10] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *124th AES convention*, Citeseer, 2008, pp. 17–20.

[11] W. G. Gardner, "3d audio and acoustic environment modeling," *Wave Arts, Inc*, vol. 99, 1999.

[12] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, 1968.

[13] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *Journal of the Audio Engineering Society*, vol. 47, no. 9, pp. 675–705, 1999.

[14] M. Cobos, S. Spors, J. Ahrens, and J. J. Lopez, "On the use of small microphone arrays for wave field synthesis auralization," in *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio*, Audio Engineering Society, 2012.

[15] T. Lokki, L. Savioja, R. Väänänen, J. Huopaniemi, and T. Takala, "Creating interactive virtual auditory environments," *IEEE Computer Graphics and Applications*, no. 4, pp. 49–57, 2002.

[16] G. M. Naylor, "Odeon—another hybrid room acoustical model," *Applied Acoustics*, vol. 38, no. 2-4, pp. 131–143, 1993.

[17] J. H. Rindel, "The use of computer modeling in room acoustics," *Journal of vibroengineering*, vol. 3, no. 4, pp. 219–224, 2000.

[18] R. A. Tenenbaum, F. O. Taminato, and V. S. Melo, "Fast auralization using radial basis functions type of artificial neural network techniques," *Applied Acoustics*, vol. 157, p. 106 993, 2020.

[19] S. Bilbao, "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1524–1533, 2013.

[20] L. Tronchin, F. Merli, M. Manfren, and B. Nastasi, "Validation and application of three-dimensional auralisation during concert hall renovation," *Building Acoustics*, p. 1351010X20926791, 2020.

[21] Y. Jing and N. Xiang, "On boundary conditions for the diffusion equation in room-acoustic prediction: Theory, simulations, and experiments," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 145–153, 2008.

[22] J. S. Bradley, "Review of objective room acoustics measures and future needs," *Applied Acoustics*, vol. 72, no. 10, pp. 713–720, 2011.

[23] ——, "Using iso 3382 measures, and their extensions, to evaluate acoustical conditions in concert halls," *Acoustical science and technology*, vol. 26, no. 2, pp. 170–178, 2005.

[24] M. Aretz and R. Orlowski, "Sound strength and reverberation time in small concert halls," *Applied Acoustics*, vol. 70, no. 8, pp. 1099–1110, 2009.

[25] J. S. Bradley and G. A. Soulodre, "The influence of late arriving energy on spatial impression," *The Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2263–2271, 1995.

[26] H. Haas, "The influence of a single echo on the audibility of speech," *Journal of the Audio Engineering Society*, vol. 20, no. 2, pp. 146–159, 1972.

[27] M. Vorländer and M. Guski, "Suggestions for revision of iso 3382," in *Proceedings of the DAGA*, 2016.

[28] T. Lokki, "Tasting music like wine: Sensory evaluation of concert halls," *Physics Today*, vol. 67, no. 1, p. 27, 2014.

[29] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 835–849, 2011.

[30] H. T. Lawless and H. Heymann, *Sensory evaluation of food: principles and practices*. Springer Science & Business Media, 2010.

[31] T. Lokki, J. Pätynen, S. Tervo, S. Siltanen, and L. Savioja, "Engaging concert hall acoustics is made up of temporal envelope preserving reflections," *The Journal of the Acoustical Society of America*, vol. 129, no. 6, EL223–EL228, 2011.

[32] D. Ko and W. Woszczyk, "Virtual acoustics for musicians: Subjective evaluation of a virtual acoustic system in performance of string quartets," *Journal of the Audio Engineering Society*, vol. 66, no. 9, pp. 712–723, 2018.

[33] M. West, *Developing high quality data models*. Elsevier, 2011.

[34] P. Blackburn and J. Bos, "Computational semantics," *Theoria: An International Journal for Theory, History and Foundations of Science*, pp. 27–45, 2003.

[35] J. F. Sowa, "Semantic networks," *Encyclopedia of Cognitive Science*, 2006.

[36] O. LASSILA, "Resource description framework (rdf) model and syntax specification, w3c," *http://www. w3. org/TR/REC-rdf-syntax/*, 1999.

[37] H. N. Schifferstein, B. M. Kudrowitz, and C. Breuer, "Food perception and aesthetics-linking sensory science to culinary practice," *Journal of Culinary Science & Technology*, pp. 1–43, 2020.

[38] R. Izhaki, *Mixing audio: concepts, practices, and tools*. Routledge, 2017, pp. 412–429.

[39] T. Carpentier, M. Noisternig, and O. Warusfel, "Twenty years of ircam spat: Looking back, looking forward," 2015.

[40] Z. Rafii and B. Pardo, "Learning to control a reverberator using subjective perceptual descriptors.," in *ISMIR*, 2009, pp. 285–290.

[41] T. Li and M. Ogihara, "Toward intelligent music information retrieval," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 564–574, 2006.

[42] E. Savage, *Sponsors*, 2020. [Online]. Available: `https://www.ismir2020.net/sponsors/`.

[43] L. Ferreira and J. Whitehead, "Learning to generate music with sentiment.," in *ISMIR*, 2019, pp. 384–390.

[44] S. Chowdhury, A. Vall, V. Haunschmid, and G. Widmer, "Towards explainable music emotion recognition: The route via mid-level features," *arXiv preprint arXiv:1907.03572*, 2019.

[45] *Fast - overview*, 2014. [Online]. Available: `http://www.semanticaudio.ac.uk/overview/`.

[46] G. Fazekas, T. Wilmering, and M. Sandler, "A knowledge representation framework for context-dependent audio processing," in *Audio Engineering Society Conference: 42nd International Conference: Semantic Audio*, Audio Engineering Society, 2011.

[47] T. Wilmering, G. Fazekas, and M. B. Sandler, "Aufx-o: Novel methods for the representation of audio processing workflows," in *International Semantic Web Conference*, Springer, 2016, pp. 229–237.

[48] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. D. Reiss, "Safe: A system for extraction and retrieval of semantic audio descriptors," 2014.

[49] T. Lokki, "Tasting music like wine: Sensory evaluation of concert halls," *Physics Today*, vol. 67, no. 1, p. 27, 2014.

[50] H. T. Lawless, H. Heymann, *et al.*, *Sensory evaluation of food: principles and practices*. Springer, 2010, vol. 2.

[51] S. Zielinski, F. Rumsey, and S. Bech, "On some biases encountered in modern audio quality listening tests-a review," *Journal of the Audio Engineering Society*, vol. 56, no. 6, p. 428, 2008.

[52] D. Ko and W. Woszczyk, "Virtual acoustics for musicians: Subjective evaluation of a virtual acoustic system in performance of string quartets," *Journal of the Audio Engineering Society*, vol. 66, no. 9, pp. 712–723, 2018.

[53] G. Vercellesi, M. Zerbini, and A. L. Vitali, "Objective and subjective evaluation mpeg layer iii perceived quality," in *2006 14th European Signal Processing Conference*, IEEE, 2006, pp. 1–5.

[54] B Series, "Method for the subjective assessment of intermediate quality level of audio systems," *International Telecommunication Union Radiocommunication Assembly*, 2014.

[55] M. Hammer and D. McLeod, "The semantic data model: A modelling mechanism for data base applications," in *Proceedings of the 1978 ACM SIGMOD international conference on management of data*, 1978, pp. 26–36.

[56] E. K. Clemons, "Design of a prototype ansi/sparc three-schema data base system," in *1979 International Workshop on Managing Requirements Knowledge (MARK)*, IEEE, 1979, pp. 689–696.

[57] A. Kuusinen and T. Lokki, "Wheel of concert hall acoustics," *Acta Acustica united with Acustica*, vol. 103, no. 2, pp. 185–188, 2017.

[58] P. Seetharaman and B. Pardo, "Crowdsourcing a reverberation descriptor map," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 587–596.

[59] A. Zacharakis, "Musical timbre: Bridging perception with semantics," Ph.D. dissertation, Queen Mary University of London, 2013, pp. 97–98.

[60] G. Darke, "Assessment of timbre using verbal attributes," in *Conference on Interdisciplinary Musicology. Montreal, Quebec*, sn, 2005.

[61] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?" *Acta acustica united with acustica*, vol. 92, no. 5, pp. 820–825, 2006.

[62] J. R. Hyde, "Discussion of the relation between initial time delay gap (itdg) and acoustical intimacy: Leo beranek's final thoughts on the subject, documented," in *Acoustics*, Multidisciplinary Digital Publishing Institute, vol. 1, 2019, pp. 561–569.

[63] M. Sarkar, C. Lan, and J. Diaz, "Words that describe timbre: A study of auditory perception through language," *MIT Media Lab*,

[64] N. Kaplanis, S. Bech, T. Lokki, T. van Waterschoot, and S. Holdt Jensen, "Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3562–3576, 2019.

[65] M. Wozniewski, Z. Settel, and J. R. Cooperstock, "A framework for immersive spatial audio performance.," in *NIME*, vol. 6, 2006, pp. 144–149.

[66] A. Kuusinen and T. Lokki, "Auditory distance perception in concert halls and the origins of acoustic intimacy," in *Ninth International Conference On Auditorium Acoustics, Paris, France, October 29-31*, Institute of Acoustics, 2015, p. 152.

[67] N. Gupta, "An analysis of acoustic treatment on recording studio," 2019.

[68] C. F. Eyring, "Reverberation time in "dead" rooms," *The Journal of the Acoustical Society of America*, vol. 1, no. 2A, pp. 217–241, 1930.

[69] H. Nélisse and J. Nicolas, "Characterization of a diffuse field in a reverberant room," *The Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3517–3524, 1997.

[70] C. L. Christensen, "Odeon, a design tool for auditorium acoustics, noise control and loudspeaker systems," in *Proceedings of Reproduced Sound 17: Measuring, Modelling or Muddling*, 2001, pp. 137–144.

[71] S. Siltanen, T. Lokki, and L. Savioja, "Rays or waves? understanding the strengths and weaknesses of computational room acoustics modeling techniques," in *Proc. Int. Symposium on Room Acoustics*, 2010.

[72] A. Foteinou, "Perception of objective parameter variations in virtual acoustic spaces," Ph.D. dissertation, University of York, 2013.

[73] C. Meyer-Bisch, "Measuring noise," *Medecine Sciences: M/S*, vol. 21, no. 5, pp. 546–550, 2005.

[74] I. Frissen, B. F. Katz, and C. Guastavino, "Effect of sound source stimuli on the perception of reverberation in large volumes," in *Auditory Display*, Springer, 2009, pp. 358–376.

[75] Y. J. Kang and J. S. Bolton, "Optimal design of acoustical foam treatments," 1996.