Glasgow School of Art

Doctoral Thesis

The effect of early reflection distribution on perceived stage acoustic conditions

Author:
Iain Laird

Supervisors:
Dr. Paul Chapman,
Dr. Damian Murphy

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Digital Design Studio, Glasgow School of Art

Submitted: August 2016

Declaration of Authorship

I, Iain LAIRD, declare that the enclosed submission for the degree of Doctor of Philosophy and consisting of a written thesis meets the regulations stated in the handbook for the mode of submission selected and approved by the Research Degrees Sub-Committee.

I declare that this submission is my own work, and has not been submitted for any other academic award.

Student

Signed: _____

Date: _____

Supervisor Support

Signed: _____

Date: _____

"Life is full of surprises, but never when you need one."

Calvin / Bill Watterson
Homicidal Psycho Jungle Cat

Abstract

It is widely accepted that performing musicians adjust their technique according to the acoustic conditions they hear on stage. It is likely that a musician performing in favourable acoustic conditions will give a higher quality performance. However, preferred conditions for performers are comparatively less well understood than for audience members. This presents a significant challenge when attempting to design a successful auditorium. Stage acoustic conditions are commonly assessed in terms of the overall energy of early reflections, relative to the direct sound, and reverberation time. These parameters relate to two subjective attributes of high importance to performers. However, these parameters are independent of the spatial or temporal distribution of the reflected energy which, in auditorium acoustics, are known to influence the perception of sound. It is proposed that a similar effect is observed for soloist performers and that these aspects of the soundfield will influence the perceived quality of the acoustic conditions. This research aims to observe how the spatial and temporal distribution of early reflections varies for differing stage enclosures and to determine if these factors influence a soloist's impression of the stage acoustics. A detailed acoustic survey of eight concert hall stages has been undertaken to characterise how the spatio-temporal distribution of early energy varies under different circumstances. This includes musician related aspects such as position on stage and orientation in addition to venue related features, such as the geometry of the stage enclosure. Spatial soundfield measurement and analysis techniques are developed to enable the spatial and temporal characteristics of early reflections to be observed. A set of objective parameters are developed to formally characterise these observations. An interactive listening test allows experienced musicians to compare a series of virtual stage enclosures by playing their instrument. Test subjects rate each hall in terms of preference and in relation to specific subjective attributes. The listening test uses a real-time auralisation system to render the acoustic conditions of a concert hall, in controlled laboratory conditions. This auralisation is based on Spatial Impulse Response Rendering (SIRR) to accurately render stage acoustic conditions over a loudspeaker array. This research proposes new methods of measuring and assessing stage acoustic conditions which will aid in the design of future auditoria. In addition, this research demonstrates the use of more recent spatial audio techniques in stage acoustic laboratory experiments.

Acknowledgements

Firstly, I would like to thank my supervisors, Paul Chapman and Damian Murphy who supported me throughout my research. Both of whom provided huge encouragement, advice and patience. I would like to thank Paul for the kind words of encouragement and for helping me source equipment at short notice. In addition, I am particularly grateful for his understanding when I had commitments at Arup to fulfil. I am particularly grateful to Damian especially for the huge amount of feedback and advice he gave throughout this research as well as the numerous times he travelled to Glasgow for our meetings. Damian also kindly included me on research visits to the University of York and Aalto University which I found hugely inspiring and thoroughly enjoyable. Damian very much kept me on track and guided my development as a researcher.

I would like to thank fellow researchers at Aalto University and York University for your very warm welcome during visits. In particular, I would like to thank Jude Brereton, with whom I had many fascinating discussions about how to auralise stage acoustic conditions for a musician.

I would also like to thank all staff members and colleagues at DDS and GSA who helped me during this research, especially when helping me source equipment, finding participants or lending a sympathetic ear. I am grateful to Jessica, Hannan, Polina, Neng and Christine for all your encouragement in the last push. I am also particularly grateful to Daisy, Ronan, Avril, Dave, Steve and Mhairi for all your help throughout this research.

A big thanks to all the participants who took part in the pilot tests and the musicians who took part in the main listening test. Many of the participants travelled from far and wide and all worked very hard during each session. I am extremely grateful for your help. I would also like to thank all the staff at each of the venues I surveyed as part of this research. I would also like to thank staff at the Royal Conservatoire of Scotland who went the extra mile getting me access to venues and distributing my call for participants. Additionally, thanks to the Mitchell Library, Scottish Music Centre, Glasgow Orchestral Society, Glasgow Philharmonia, Glasgow Studio Orchestra, RSNO and SCO who very kindly forwarded on my call for participants.

A great deal of thanks go to Arup for supporting this research and for their generous loan of equipment and the use of the SoundLab to undertake this research. In particular, I would like to thank everyone at the Glasgow office for all your support, encouragement and huge patience throughout this research. In particular, Oli Atack, Francesca Coppa, Martin Butterfield, Stephanie Harper, Luke Robertson and Cesar Bustos. I am especially grateful to Martin and Oli for helping me source and transport equipment and putting me in touch with venues and orchestras.

I would also like to thank Ian Knowles, Helen Butcher, Jeremy Newton, Rob Harris and Jo Webb for all your kind words of encouragement and patience throughout this research. I am particularly grateful to Ian who loaned me a copy of Gade's thesis which was a fantastic help at exactly the right time! I would also like to thank Anne Guthrie and Terence Caulkins with whom I had many interesting discussions regarding stage acoustic auralisation.

A special thanks have to go to Seb Jouan, without whom I would never have had the opportunity to research this fascinating area. It was conversations with Seb that were the main inspiration

for this research. I am particularly grateful to Seb for his encouragement which ultimately led to me working on some exciting projects with Arup.

I would also like to thank friends and family who have supported me thoughout this research. I am grateful to my parents and the rest of the Lairds for their encouragement and understanding during tough times. I would also like to thank Ray, Aideen, Fiona, James, Sarah Daly, Sarah Humble, Lucy and Peter who kept me sane all these years.

Most of all I would like to thank Debra Choong who accompanied me from start to finish on this incredible challenge. She listened to every problem, helped me work through every hiccup and was there every time it got too much. She understood when I had to work late, came along to conferences and even helped me during dry runs of my listening tests. Without a doubt, this wouldn't have been possible without her. Thank you so much.

Contents

List of Figures

List of Tables

Abbreviations

SIRR	**S**patial **I**mpulse **R**esponse **R**endering
DirAC	**Dir**ectional **A**udio **C**oding
SDM	**S**patial **D**ecomposition **M**ethod
FFT	**F**ast **F**ourier **T**ransform
STFT	**S**hort **T**ime **F**ourier **T**ransform
RMS	**R**oot **M**ean **S**quare
ILD	**I**nteraural **L**evel **D**ifference
ITD	**I**nteraural **T**ime **D**ifference
MDAP	**M**ultiple **D**irection **A**mplitude **P**anning
VBAP	**V**ector **B**ase **A**mplitude **P**anning
HRTF	**H**ead **R**elated **T**ransfer **F**unction
IACC	**I**nter **A**ural **C**ross Correlation **C**oefficient
ASW	**A**pparent **S**ource **W**idth
FOA	**F**irst **O**rder **A**mbisonics
HOA	**H**igher **O**rder **A**mbisonics
WFS	**W**ave **F**ield **S**ynthesis
EDT	**E**arly **D**ecay **T**ime
HATS	**H**ead **A**nd **T**orso **S**imulator
MDPREF	**M**ulti **D**imensional scaling of **PREF**erence data
PREFMAP	**P**reference **M**apping
PCA	**P**rincipal **C**omponent **A**nalysis
MRV	**M**ean **R**esultant **V**ector
VST	**V**irtual **S**tudio **T**echnology
KDE	**K**ernel **D**ensity **E**stimate
ABRSM	**A**ssociated **B**oard of the **R**oyal **S**chools of **M**usic
FDL	**F**requency **D**omain delay - **L**ine

Symbols

T_{30}	Reverberation time
C_{80}	Clarity
D_{50}	Definition
ST_{early}	Early objective support (20-100ms)
ST_{late}	Late objective support (100-1000ms)
ST_{total}	Total objective support (20-1000ms)
$ST_{early,mod}$	Modified ST_{early} (10-100ms)
t_{mean}	Mean time of arrival
t_{σ}	Standard deviation of time of arrival
θ_{mean}	lateral angle of mean resultant vector
ϕ_{mean}	elevation angle of mean resultant vector
$Spread$	Spatial spread of early reflections
c	Speed of sound ($344ms^{-1}$)
T_s	Centre time
CS	Stage Clarity
EEL	Early Ensemble Level
G	Strength
RR_{160}	Running Reverberation

Publications

- Laird, I., Chapman, P., Murphy, D. (2011) *'Development of a Virtual Performance Studio with Application of Virtual Acoustic Recording Methods'*, Proceedings of the 130th Audio Engineering Society Convention, London, UK
- Laird, I., Chapman, P., Murphy, D. (2012) *'Energy-based calibration of virtual performance systems'*, Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12), York, UK
- Laird, I., Chapman, P., Murphy, D. (2012) *'Spatialisation accuracy of a Virtual Performance System'*, Proceedings of the Joint Baltic-Nordic Acoustics Meeting, Odense, Denmark
- Laird, I., Chapman, P., Murphy, D. (2014) *'Comparison of Spatial Audio Techniques for use in Stage Acoustic Laboratory Experiments'*, Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany

Dedicated to Gemma Balfour
and to Gregor 'Monster' Morris

Chapter 1

Introduction

"A concert hall is the extension of my musical instrument. I always focus my mind on the audience area and feel how my performance sounds there. (Ob. -A)" (Ueno et al., 2005)

The acoustic conditions experienced by performing musicians can have a significant impact on how they perform. For some musicians, the acoustics of the venue are of such importance that it is considered an integral part of their instrument's sound. Even subconsciously, the on-stage acoustic conditions can have an observable impact on the musician's technique (Kalkandjiev and Weinzierl, 2013, Ueno et al., 2005), influencing their approach to articulation, tempo or intonation. It follows that the acoustic conditions experienced by the performer can directly affect the performance as heard by the audience. For the audience to hear a performance at its best, it is critical that the acoustic conditions experienced by the performer provide an assistive environment.

However, the preferred acoustic conditions for musicians are currently not well understood which can make the design of stage enclosures very challenging and often the focus of remedial work in newly commissioned performance spaces (Kahle, 2013). As a result, musicians are often required to endure poor acoustic conditions which can adversely affect the quality of their performance.

Therefore, the overriding aim of stage acoustic research is to understand how different aspects of the acoustic response influence a musician's subjective impression of a venue. By studying the relationship between musician preference and objective acoustic parameters, it may be possible to refine the stage enclosure design to provide a more assistive environment for their performance. This is the fundamental motivation of this thesis.

Gade's pioneering work in this field uncovered many essential facets of how musicians perceive concert hall acoustics (Gade, 1982). Through interviews and laboratory-based listening tests, it was found that one of the primary concerns of a soloist was the level of support they received from the hall; simply, how much effort they had to exert to get a sufficient response from the hall. Gade discovered that the subjective attribute of *"support"* was linked to the overall energy of early reflections, expressed in relation to the direct sound. Musicians tended to favour playing in stage acoustic conditions with higher levels of early reflections. This led to the development

of the *Objective Support* parameters which remain part of the standardised method of measuring and assessing concert halls (International Organisation of Standardisation, 2009). Research by Ueno and Tachibana (2003), also found that factors such as reverberation time, energy of early reflections and the energy of a late arriving reflection can significantly affect a performer's impression of a venue.

Gade also demonstrated that musicians of different instrument families perceive stage acoustic conditions in different ways (Gade, 1982). It was found that the threshold of perception of early reflections varied greatly between wind and string instruments. The likely cause of this was attributed to the difference in how each instrument masks reflections from a particular direction. Additionally, it was shown that the presence of early reflections could be perceived by musicians in different ways; where some musicians could detect the presence of reflections via changes in timbre. This suggests that different musicians performing in the same acoustic conditions may perceive the space in different ways.

During a performance, musicians experience a high degree of stress where their attention may shift to other performance-related tasks. For example, the musician may shift their focus to the physical act of playing their instrument during a challenging passage or listen for a particular instrument when waiting for a cue. Guthrie (2014) referred to this as the musician's *'Cognitive Load'* which may also affect what aspects of the acoustic conditions are audible to musicians. Due to these significant masking effects, it is not fully understood how different acoustic conditions affect a musician's impression of a space. It is important to fully understand these phenomena so that the performance space can be designed to assist the musician rather than hinder them.

From an audience perspective, it is widely recognised that the subjective impression of a performance can be influenced by the distribution of early reflections. Their frequency content, amplitude, temporal and spatial distribution can influence the audience's perception of envelopment, clarity, loudness and timbre of a performance (Robinson et al., 2013a). It is possible that the distribution of early reflections received on stage has a similar subjective effect on performing musicians, however the significant masking properties present during a performance may cause these effects to be inaudible.

The stage enclosure is largely responsible for providing early reflections which support the efforts of the musician, however, many musicians refer to an additional feeling of *'projection'* or *'bloom'* which arrives from the auditorium (Kahle, 2013). It has also been suggested that the temporal and spatial distribution of early reflections may have an influence of the perceived *'quality of support'* from the stage (Miranda Jofre et al., 2013). While a venue may provide an adequate level of early energy to support the performer, the delivery of this energy may affect how it is then perceived. This suggests that the spatial or temporal distribution of early reflections could be a salient factor related to musician preference. If these are found to be salient factors then additional acoustic parameters may be required to assess or design future stage enclosures.

1.1 Hypothesis and aims

This research aims to determine how the spatial or temporal distribution of early reflections influences a performer's impression of the stage acoustic conditions. Of specific interest, is how the distribution of early reflections influences a musician's subjective impression of the performance space and how this affects musician preference towards stage acoustic conditions. The hypothesis that guides this thesis is stated as follows:

In the context of a performing soloist, the preferred acoustic conditions on stage are strongly dependent on the spatio-temporal distribution of early reflections in addition to their overall level relative to the direct sound.

In order to prove or disprove this hypothesis, it will be necessary to carry out an interactive listening test with musicians to observe their subjective reaction in different acoustic environments. This will be achieved using a novel interactive auralisation system which will allow musicians to instantaneously compare different acoustic environments in controlled laboratory conditions. In addition to this, an acoustic survey of local venues will be carried out to observe how the distribution of early reflections varies on different stages.

This hypothesis is supported by the following work:

1. The creation of a stage acoustic measurement protocol that considers the temporal and spatial delivery of early reflections.
2. Development of objective acoustic parameters which characterise the spatial and temporal distribution of early reflections.
3. Evaluation of different stage acoustic conditions in eight performance spaces.
4. The creation of an interactive virtual stage environment to determine the subjective effect of early reflection distribution.
5. The study of musician responses to controlled acoustic conditions via interactive listening tests.

Specifically, the listening tests will aim to determine:

- If musician test participants can detect variations in spatial or temporal distribution of early reflections.
- Which perceptual attributes, related to the stage acoustic conditions, are influenced by these variations.
- If the spatio-temporal distribution of early reflections influence musician preference patterns towards stage acoustic conditions.

1.2 Contributions to the field

This work offers a number of novel contributions to the field of stage acoustic research:

- An implementation of parametric analysis techniques to observe the spatial and temporal distribution of early reflections from a measured impulse response.
- An objective analysis of how stage acoustic conditions are influenced by hall-related and musician related variables.
- Accurate interactive auralisation of stage acoustic conditions for a soloist in a laboratory environment.
- A deeper understanding of how the distribution of early reflections influences specific subjective attributes for a musician.

1.3 Organisation of thesis

Chapter 2 reviews the fundamental concepts of physical sound propagation in performance spaces in addition to relevant aspects of auditory perception and psychoacoustics. This chapter also reviews the current methods of assessing auditorium and stage acoustic conditions.

Chapter 3 describes how acoustic conditions are measured and summarise existing methods for analysing the properties of stage acoustic conditions. This chapter introduces a novel method of measuring the stage acoustic conditions such that the temporal and spatial distribution of the acoustic response can be studied in detail. These methods are used to assess the stage acoustic conditions of eight performance spaces.

Chapter 4 reviews the most recent methods of analysing the spatial and temporal aspects of a room acoustic response. It also describes the methods used to evaluate the early reflections measured on stage.

Chapter 5 presents the analysis of eight concert performance spaces and describes how the acoustic conditions vary due to aspects of the stage enclosure and the performer.

Chapter 6 introduces the idea of an interactive stage acoustic auralisation system which allows musicians to perform on virtual stages. This chapter reviews different auralisation methods and discusses their appropriateness for this research.

Chapter 7 presents the results of two pilot tests that were undertaken during the development of the interactive auralisation system. These tests provide initial justification to conduct the main experiment in addition to testing the effectiveness of the auralisation system.

Chapter 8 describes the design of the interactive auralisation system used in this research and how the system is calibrated and operated.

Chapter 9 describes the main subjective test to determine if the spatio-temporal distribution of early reflections is audible to a performing musician. Further tests describe how this influences

the perception of the performance space in terms of known attributes and if there are broad preferences in terms of optimal stage acoustic conditions.

Chapter 10 summarises the overall findings of this research and discusses future avenues of research relevant to the field of stage acoustics.

Chapter 2

Fundamentals

When a musician performs on stage, it is widely recognised that they will adjust their technique to suit a particular venue (Ueno et al., 2007). Some musicians perform this act subconsciously whereas others will adjust their technique by listening for specific properties of the reflected sound and reacting accordingly. The process is analogous to a person gradually acquiring a skill by repetitively making an action with a tool and refining their actions based on their observations of the consequences (Ueno and Tachibana, 2005). On repetition of the task the person gradually begins to understand the consequences of various actions or techniques and is able to adapt the action in subtle ways to achieve the desired outcome. This process, referred to as *'Tacit Knowing'* (Ueno and Tachibana, 2005), differs from explicit learning as the human subject may not be able to describe the process by which the skill was acquired.

From a musician's perspective, a performance can be viewed as a complex feedback loop where the physical adjustments to their technique are based on how the musician perceives the direct sound and reflected sound in relation to their prior experience and training. Figure 2.1 shows a schematic representing the contributing factors in this process in the context of a solo performer. It shows how the physical action of playing the musical instrument creates the direct sound which is heard immediately by the musician. The direct sound propagates out into the concert hall and is modified and replicated as reflections which propagate back to the musician. The resulting sound heard by the performer is a summation of the direct sound and reflected sound from the performance space.

The impression of the direct and reflected sound from the venue are thought to be influenced by various masking effects. As shown in Figure 2.1, the proximity of the musical instrument may cause auditory masking of the reflected sound. In addition, the musician may attend to specific aspects of what they hear, causing other aspects to be cognitively masked. The decision to adjust their technique is then based upon a comparison of what they hear with previous experience or musical intent.

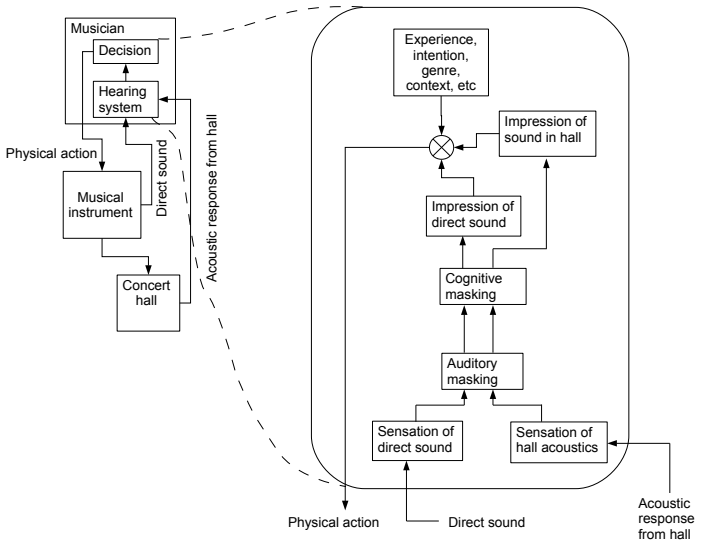


FIGURE 2.1: *A flow chart describing how sound the sound generated by a performer is modified by the concert hall and is heard by the musician. What the musician hears is subject to various masking phenomena and is interpreted by the auditory system. The musician then decides, based on previous experience, how to adapt the sound to produce a desired effect. Adapted from (Ueno and Tachibana, 2005)*

It is feasible that the decision-making process varies over time as a musician acclimates to a space after an exploratory period. In this period, the musician may deliberately exaggerate phrasing or articulation to excite the venue in different ways, listening to the reaction throughout and gradually adjusting until they are satisfied.

Overall, this research is concerned primarily with how the stage acoustic response influences the musician's impression of the hall. Therefore it is necessary to first review the physical processes which modify the sound returning from the concert hall and the psychological mechanisms by which they are interpreted. This chapter will describe the fundamental aspects of each part of this process, specifically, how sound is generated, how it propagates through the hall and how it is sensed and perceived by the musician. It will also review basic concepts of auditorium and stage design and current literature on the perception of acoustic conditions from a performers point of view. Finally, the existing approaches for evaluating stage enclosures will be reviewed.

2.1 Sound generation and propagation

The most fundamental aspect of musical performance is the generation of sound by the musician's instrument and its subsequent propagation through the concert hall. Understanding how the instrument's sound is modified by the concert hall is an important factor in determining optimal conditions for a performer. This section will describe some of the basic principles behind sound generation and propagation.

2.1.1 Free-field sound propagation

Sound is generated when mechanical vibrations of an object disturb the particles of the medium around it. The particle vibrations in the medium can be described either by particle velocity or by local fluctuations in the pressure of the medium (caused by the associated changes in medium density). The local increase and decrease of pressure due to sound are referred to as compression and rarefraction respectively. When sound propagates through a reflection-free area it is termed as propagating in *'anechoic'* or *'free-field'* conditions (Bies and Hansen, 2003).

In free-field conditions, a vibrating point source will radiate sound equally in all directions causing a pressure wave to propagate out from the source as air particles shift in relation to one another. The Helmholtz wave equation, derived from Newton's laws of motion, describes the behaviour of sound pressure in a fluid medium as a function of time and space. This can be expressed as the partial differential equation (2.1). This describes the pressure variations in a homogenous fluid with zero viscosity (Williams, 1999).

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{2.1}$$

where $\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$ and $p(x, y, z, t)$ is an infinitesimal variation of acoustic pressure from the equilibrium value. c is equal to the speed of sound in air. The speed of sound in air varies with aspects such as temperature and humidity. In this thesis, unless otherwise stated, the speed of sound is assumed to be $c = 344ms^{-1}$.

Sound intensity describes the mean energy flow of a sound wave that is transported through an area (Bies and Hansen, 2003). The instantaneous intensity is defined by equation (2.2) where $p(t)$ is the sound pressure and $u(t)$ is the particle velocity. As the particle velocity is a vector quantity, the sound intensity also describes the direction of the flow of energy.

$$I(t) = p(t)u(t) \tag{2.2}$$

As the pressure wave radiates away from the source, the energy is spread over an increasing spherical area. This causes the intensity to decrease with the inverse square of the distance from the sound source, as shown in Figure 2.2.

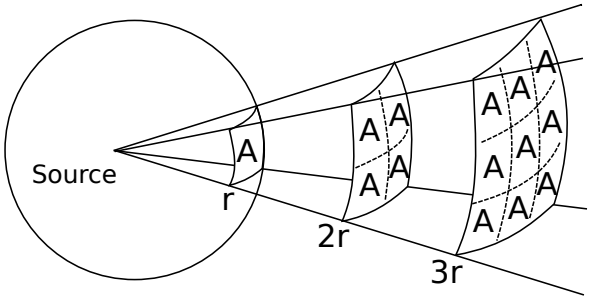


FIGURE 2.2: *As sound propagates spherically away from a sound source the energy is spread over an increasing area therefore the intensity reduces with distance from the sound source. Adapted from (Everest and Pohlmann, 2009)*

This is expressed in equation (2.3) which determines the intensity, I, of a sound source radiating with a sound power, W, measured at a radius r. As the spherically radiating sound wave reaches a large radius from the sound source, its curvature can be considered to be negligible and thus the sound wave can be treated as a plane wave.

$$I = \frac{W}{4\pi r^2} \tag{2.3}$$

Sound intensity can also be expressed as a decibel ratio which is convenient given the wide range of intensities sound can take. For instance, the quietest audible sound (threshold of hearing) is approximately $10^{-12}W/m^2$ whereas a sound loud enough to be painful for a human is approximately $100W/m^2$ (Everest and Pohlmann, 2009). It is also the case that humans perceive the loudness of a sound logarithmically. A 10dB increase in sound intensity level will be judged as being twice as loud. A doubling of sound intensity level (i.e. +3dB) is often quoted as a just noticeable increase in loudness. The sound Intensity level, L_i, is described by equation (2.4), where I_{ref} is $10^{-12}W/m^2$ and corresponds to the human threshold of hearing.

$$L_i = 10log_{10}\left(\frac{I}{I_{ref}}\right) \tag{2.4}$$

Similarly, the energy output of a sound source can be defined by its sound power level, L_w, as shown in equation (2.5). The reference, W_{ref}, is also $10^{-12}W/m^2$.

$$L_w = 10log_{10}\left(\frac{W}{W_{ref}}\right) \tag{2.5}$$

This quantity defines the energy output regardless of the environment it inhabits. As sound pressure is proportional to the square of sound intensity, the sound pressure level, L_p, can also

be expressed in this way, as shown in equation (2.6) (Howard and Angus, 2001).

$$\begin{aligned} L_p &= 10log_{10}\left(\frac{p^2}{p_{ref}^2}\right) \\ &= 20log_{10}\left(\frac{p}{p_{ref}}\right) \end{aligned} \tag{2.6}$$

As sound propagates outward into the concert hall it reduces in intensity due to distance attenuation but also due to energy losses related to heat exchanges with the volume of air (Kuttruff, 1979). In general, the attenuation due to air varies with temperature and humidity and is frequency-dependent, with absorption generally increasing with frequency.

As the distance between the musical instrument and the performer's ears is typically very small, the sound pressure level experienced by the musician, from the direct sound alone, will be much louder than other sounds heard on stage. Musical sounds often produce unweighted sound pressure levels between 40dB and 100dB (Fletcher and Rossing, 1998).

2.1.2 Sound source spectrum

The sound from musical instruments is often spectrally complex due to the coupling of many different resonant bodies contributing sinusoidal components of different frequencies to the overall sound. For instance, a bow drawn across the strings of a violin generates vibrations which travel into the body and then radiate into the air.

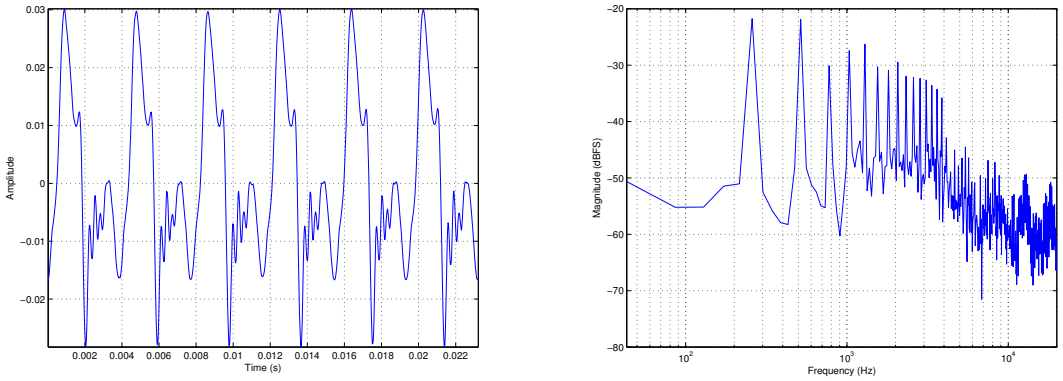
Many tonal musical sounds consist of a series of harmonically related sinusoidal components referred to as the fundamental and overtones (or harmonics). The frequency and relative amplitude of the harmonics determines the perceived timbre of the instrument and is primarily how a note of identical pitch played by two different instruments can produce very different sounds (Fletcher and Rossing, 1998). Noise-like or transient sounds can produce a randomly varying time domain signal comprised of a large number of non-related harmonics.

Experienced musicians are able to control the relative amplitude of these harmonics through various physical techniques. It is also possible for musicians to modulate the frequency of the note using a technique called vibrato. Musical sounds can also be characterised by their amplitude envelope. Broadly speaking, most note envelopes consist of an onset phase, a steady state phase and a decay phase.

The spectral content of sound can be observed via Fourier analysis which allows a complex signal to be decomposed into a series of weighted, sinusoidal components, $e^{-i\omega t}$, and viewed in the frequency domain (known as the Fourier Series). The Fourier Transform of a continuous complex waveform $p(t)$ is defined in equation (2.7) (Williams, 1999):

$$P(\omega) = \int_{-\infty}^{\infty} p(t)e^{-i\omega t}dt \tag{2.7}$$

Figure 2.3 shows an example of a complex sound analysed in the frequency domain. This example shows the time and frequency domain representations of a sustained note played on an Alto Saxophone. It can be seen from the frequency domain representation that the signal is comprised of numerous peaks, indicating a large number of harmonics.



(a) Time domain

(b) Frequency domain

FIGURE 2.3: *Time domain 2.3(a) and frequency domain 2.3(b) representations of a sustained note played on an Alto Saxophone. The complex time domain signal comprises a rich harmonic spectrum shown by peaks in the frequency domain.*

Considering only the fundamental harmonic, musical notes have a wide frequency range from around 27Hz to 4100Hz although most instruments typically occupy only parts of this bandwidth. The additional harmonic content typically extends this range much higher in frequency (Everest and Pohlmann, 2009). In the field of Room Acoustics, sound is generally considered between 20Hz and 22kHz as this is the commonly quoted range of frequencies audible to humans.

2.1.3 Sound source directivity

It is well established that musical instruments do not radiate sound equally in all directions. It has been observed that an instruments directivity will vary due to a multitude of factors including playing technique, note fundamental and note transition (Otondo and Rindel, 2004). The musical instrument radiation pattern will fluctuate randomly throughout the period of a steady note. Often, longer term measurements are used to produce a more generalised view of the radiation characteristics of particular instruments. Figure 2.4 shows a diagram of how the directivity characteristics of a trumpet vary with frequency. At low frequencies the sound appears to radiate equally in all directions whereas at higher frequencies the sound tends to radiate more directly out of the trumpet bell.

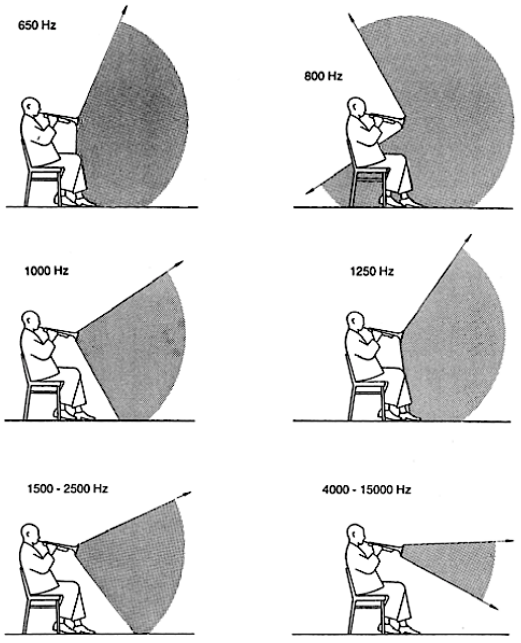


FIGURE 2.4: *Figure from Meyer (Meyer, 2009) demonstrating how the radiation characteristics of a trumpet vary with frequency. In this example the trumpet becomes significantly more directional at high frequencies*

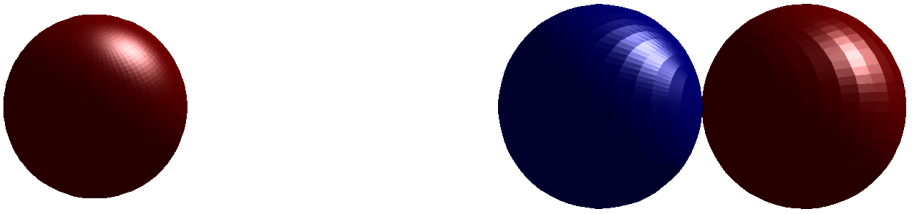
A simple definition of the radiation characteristics of a sound source is the directivity factor Q which is the ratio of the intensity at a given angle, I_θ, to the average intensity, I_{av} (Kinsler et al., 1982).

$$Q = \frac{I_\theta}{I_{av}} \tag{2.8}$$

The directivity index, D, is shown in equation (2.9) which expresses the directivity factor in decibels (Smith et al., 1996).

$$D = 10log_{10}(Q) \tag{2.9}$$

The simplest radiation characteristic is a monopole or 'point source' where sound radiates equally in all directions from a pulsating sphere. This source has a directivity factor of $Q = 1$. When a point source is located on a reflecting surface the spherical propagation is restricted in direction. A point source located on a single flat plane has a directivity factor of $Q = 2$ (or $D = +3dB$). If the source is located at the boundary of two perpendicular planes then directivity factor increases to $Q = 4$ (Smith et al., 1996). A dipole source can can be likened to an oscillating plane which at any particular time will generate a positive pressure on one side and a negative pressure at the other (Fletcher and Rossing, 1998). More complex directivity patterns can be derived using arrangements of monopoles, known as multipoles. Examples of a monopole and dipole are shown in Figure 2.5.



(a) Monopole (b) Dipole

FIGURE 2.5: *Monopole 2.5(a) and dipole 2.5(b) directivity functions are the basic building blocks of more complex radiation patterns. Images generated using code from (Wiggins, 2004).*

Musical instrument radiation patterns are often considered as a weighted sum of multipoles and as such can be approximated using the exterior expansion for the wave equation (Menzies and Al-Akaidi, 2007). By defining the multipoles as spherical harmonics of increasing order, the pressure in the far field at a particular angle of orientation can be considered using equation (2.10):

$$p_{far} = \frac{e^{-ikr}}{r} \sum_{m,n} Y_{mn}(\theta, \delta) O_{m,n}(k) \tag{2.10}$$

Where $O_{m,n}$ represents the weighting coefficient of the corresponding spherical harmonic $Y_{m,n}$. A spherical harmonic decomposition of a musical instrument's radiation pattern can be achieved by recording in free-field conditions using a spherical array of microphones.

The radiation characteristics of a musical instrument are important for both the musician and a distant listener as it will determine the relative amplitude of harmonic content of the sound that both are exposed to, thus it is an integral part of the instrument's timbre (Otondo and Rindel, 2004). As a simple example, the sound from a piano in a free-field environment will have a noticeably different timbre when it is heard from the side when compared to the sound heard at the keyboard.

2.2 Sound in enclosed spaces

When sound propagates through a concert hall, the sound energy is reflected, scattered or absorbed by all the surfaces in the room and attenuated further by the air as it propagates. As shown in Figure 2.6, at any location in that space, the sound will consist of the direct sound and delayed and attenuated versions of the direct sound. The superposition of reflected sounds tends to influence how the sound source is perceived by a listener and so it is important to consider how reflections propagate through a space.

Consider a source signal that consists of an impulsive sound. The signal received at any location will consist of the direct sound and reflected sound and is known as the *impulse response*

(Kuttruff, 1979). The impulse response is commonly used to assess the acoustic conditions of enclosed spaces such as concert halls.

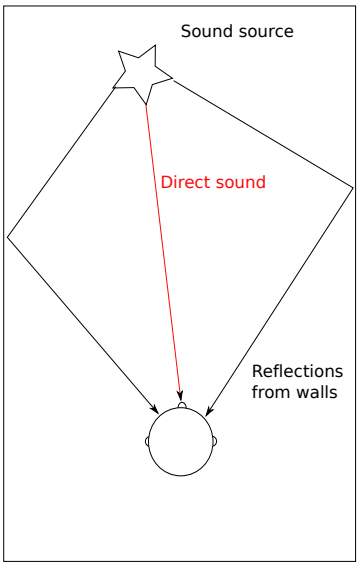


FIGURE 2.6: *Sound in an enclosed space consists of the direct sound (shown in red) and reflections from the enclosing walls*

An impulse response measured in a room consists of three major parts as shown in Figure 2.7. The direct sound, where the sound travels directly to the receiver, occurs after a very short time delay (depending on the distance between source and receiver). Shortly after the direct sound, early reflections begin to arrive and are quite sparsely populated in time. Early reflections tend to have encountered only a small number of surfaces. As time increases the reflections tend to increase in arrival density to the point where individual reflections can no longer be distinguished. At this point, the reflected sound can be likened to an exponentially decaying noise signal and is referred to as the reverberant tail. The temporal density of reflections, N_r, arriving in any given time instant can be predicted using equation (2.11) (Kuttruff, 1979).

$$\frac{N_r}{dt} = 4\pi \frac{c^3 t^2}{V} \tag{2.11}$$

where V is the volume of the space. From this it can be seen that the mean density of reflections increases quadratically with respect to time.

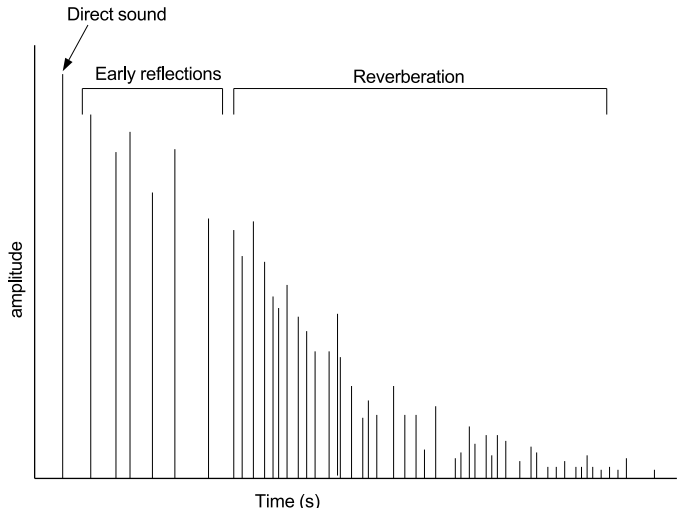
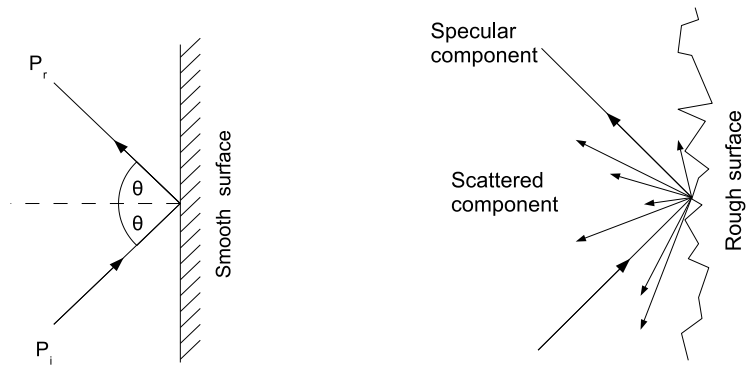


FIGURE 2.7: *A room impulse response typically consists of the direct sound, followed by early reflections where arrivals increase in density with time resulting in the reverberant part of the room response.*

Each section of the impulse response is widely recognised to influence the perceptual impression of sound generated in a space or provide different subjective cues to the listener. A successful performance space will have been designed to control and balance each section of the impulse response to produce a specific subjective effect for both the audience and musicians. The individual parts of the impulse response are discussed in more detail in the following sections.

2.2.1 Early reflections

Early reflections arrive within a short time frame after the direct sound (broadly considered to be within the first 80-100ms relative to the direct sound). As sound energy encounters a surface, it is modified as it loses energy to the surface via friction. The reflected sound is dependent on the construction of the surfaces the sound has encountered. Reflections are usually considered either specular (if reflected off a smooth surface) or diffuse (if reflected off a rough surface) (Kuttruff, 2007). Reflected sound may have encountered only a single surface before being heard (known as first-order reflections) or may have encountered many surfaces (n^{th} order reflections).



(a) Specular reflection (b) Diffuse reflection

FIGURE 2.8: *Specular reflection from a flat planar surface 2.8(a) and diffuse reflection from a rough, corrugated surface 2.8(b) showing scattering of reflected sound wave.*

A specular reflection (as shown in Figure 2.8(a)) occurs when sound energy encounters a flat surface which is large compared to the wavelength of the sound. The reflected sound travels away from the surface at a symmetrical direction from the incident sound as a result of Snell's Law (equation (2.12)) where c_i is the speed of sound in each medium and θ_i and θ_r are the incident and reflected angles respectively (Everest and Pohlmann, 2009). The reflected sound has very similar temporal characteristics to the incident sound but is typically attenuated due to friction with reflecting surface. In performance spaces, reflectors are often angled to control the direction in which sound travels around the venue.

$$c_1 sin(\theta_i) = c_2 sin(\theta_r) \tag{2.12}$$

Diffuse reflections (as shown in Figure 2.8(b)) occur when sound encounters a surface that is rough (or corrugated) in comparison to its wavelength. Reflected sound is dispersed temporally and spatially after it has encountered the surface. As the sound encounters a diffusing surface, the Huygen's principle shows that a set of secondary sources are created on the surface each of which each radiate in a hemispherical manner. The interaction of these secondary sources produces different interference patterns at different frequencies which determine how the diffusing surface reacts at different frequencies (Cox and D'Antonio, 2004).

Diffusion can be caused unintentionally by irregularities, shaping or ornamentation of particular surfaces or can be deliberately designed and controlled by using diffusers which use specific patterns of indentations to scatter sound in particular ways. This often proves an effective way of treating problematic reflections without the use of acoustic absorption which can (sometimes undesirably) significantly reduce the energy within the room response. It is often the case that a reflection from a surface consists of both specular and diffuse components where the diffuse component propagates in non-specular directions (Vorlander, 2008). Diffuse reflections contain the same energy spread over a finite time period and resemble a noise-like burst with a much lower peak than a specular reflection (Robinson et al., 2013b).

Specular reflections exhibit similar temporal properties to the incoming sound and will encounter a frequency-dependent reduction in energy as a consequence of the surface absorption coefficient. Absorption is typically caused by conversion of sound energy to heat through a frictional process. The amount of energy reflected by the surface can be described by its reflection coefficient, R, shown in equation (2.13) (Kuttruff, 2007).

$$R = \frac{Z\cos\theta - Z_0}{Z\cos\theta + Z_0} \tag{2.13}$$

Where $Z_0 = \rho_0 c$ denotes the characteristic acoustic impedance of air and Z denotes the impedance of the boundary. θ denotes the angle of incidence. Similarly, the amount of energy that is absorbed by the surface can be described by its absorption coefficient, α, as shown in equation 2.14. The absorption characteristics of any surface are frequency dependent. An absorption coefficient of 1 indicates sound is completely reflected while a value of 0 indicates incoming sound energy is completely absorbed.

$$\alpha = 1 - |R|^2 \tag{2.14}$$

Specular reflections that arrive shortly after the direct sound can result in *'comb-filtering'* which can cause timbral colouration (Halmrast, 2001). This phenomenon occurs due to constructive and destructive interference and is characterised by a regular series of peaks and troughs in the frequency response, as shown in Figure 2.9. The magnitude of the peaks and troughs is dictated by the amplitude of the reflection whereas the spacing between them is determined by the delay time of the reflection. Audible comb filtering generally only occurs up to a particular time delay after which the reflection is perceived as a separate event.

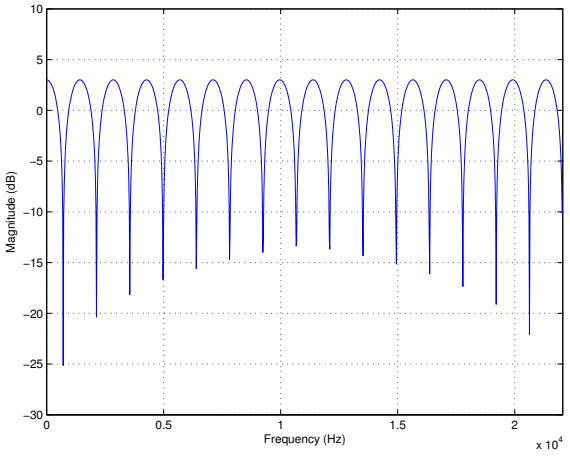


FIGURE 2.9: *A plot showing the comb-filtering effect of a single reflection closely following the direct sound*

There are numerous sources of early specular reflections on concert hall stages including the floor, music stands and (depending on their construction) the stage enclosure walls and reflectors (Halmrast, 2007). When numerous early reflections occur within a short time period or a single reflection is dispersed over time, the spacing between the peaks and troughs become much less

regular, reducing the perceived colouration. Often concert hall venues utilise a mix of both specular and diffuse reflections to produce a desired subjective effect.

The temporal, spatial and amplitude distribution of early reflections are widely considered to provide important subjective cues for concert hall acoustics including feelings associated with spaciousness, clarity and the size and loudness of an orchestra. Therefore a large number of acoustic parameters have developed in order to assess these reflections. These parameters will be discussed in detail later in this chapter.

2.2.2 Reverberation

As reflections increase in arrival density over time, there is a point at which individual arrivals are sufficiently dense that the temporal and spectral structure of individual reflections can no longer be distinguished and the resulting sound can be considered as exponentially decaying noise. This is known as the reverberant tail of the impulse response.

Reverberation is typically considered spatially diffuse and has an amplitude which tends to decay exponentially with time. An ideal, diffuse soundfield occurs if the energy density is equal at all points in the room, i.e. it is spatially uniform. In practice, an ideally diffuse soundfield does not occur as there is always a net energy flow in the direction of the walls (Lindau et al., 2010). A diffuse soundfield occurs when a large number of plane waves with random phases are present in the room. This often transpires when a large number of reflections, that have interacted with many surfaces, are received at a point in the venue.

The reverberant decay is typically characterised by the reverberation time, T_{60}, which is defined as the time taken for the sound pressure in a room to decay by 60dB after a broadband noise source active in the space has been switched off. The reverberation time of enclosed spaces can be predicted by the well known Sabine relationship shown in equation (2.15).

$$T_{60} = \frac{0.161V}{\sum S_i \alpha_i} \tag{2.15}$$

where V is the volume of the room, α_i and S_i denote the absorption characteristics and surface area of the i^{th} surface in the room respectively.

The instantaneous energy density of a soundfield (energy per unit volume) can be calculated using equation (2.16) where ρ is the density of the medium and $Z_0 = \rho c$ is the acoustic impedance.

$$E(t) = \frac{1}{2}\rho \left[\frac{p^2(t)}{Z_0^2} + u^2(t) \right] \tag{2.16}$$

The ratio of energy contributing to the overall transfer of energy can be used to estimate the diffuseness of a soundfield. This can be expressed as a ratio of intensity to energy shown in (2.17) which varies between 0 (ideally diffuse soundfield i.e. no net transfer of energy) and 1

(completely non-diffuse soundfield).

$$\psi = \frac{\|\langle I(t)/c \rangle\|}{\langle E(t) \rangle} = \frac{2Z_0 \|\langle p(t)u(t) \rangle\|}{\langle p(t)^2 \rangle + Z_0^2 \langle u(t)^2 \rangle} \tag{2.17}$$

The transition point between the early reflections and reverberant decay is known as the *'mixing time'* and is defined as the point at which the soundfield becomes fully diffuse and can be described more easily by statistical methods (Defrance and Polack, 2008, Lindau et al., 2010). This transition point is often utilised by acoustical parameters which use a ratio of early to late arriving energy to describe various psychoacoustic effects. The mixing time can be predicted from the room volume using the equation (2.18) (Lindau et al., 2012).

$$t_{mixing} = \sqrt{V} \tag{2.18}$$

Numerous methods have been derived to measure or predict the mixing time which are in common usage amongst acousticians. However, the mixing time may be considered both as an objective parameter (which varies with physical parameters including room volume) and also as a subjective parameter which may not give the same value as the objective parameter (Defrance and Polack, 2008, Lindau et al., 2010). Many models are based on the volume, surface area and reverberation time of the room. It is more common however for a fixed time delay to be used to distinguish the early reflections from the reverberation. These values often vary dependent on the specific situation but are commonly quoted as 50ms, 80ms or 100ms after the direct sound.

2.3 Psychoacoustics

Sound can be interpreted both as a physical phenomenon (sound event) and a perceptual phenomenon (auditory event) (Blauert, 1996). In the previous section, it was discussed how sounds can be modified as they propagate through an enclosed space. These modifications can influence how sound is perceived in a space and so it is important to understand this relationship in order to design a successful performance space. This section will discuss the mechanisms by which humans can sense and perceive sound, specifically in reference to the complex acoustic environment inhabited by a performer.

2.3.1 Human hearing system

The human hearing system transforms physical sound waves into a series of electrical impulses (sensation) which can then be interpreted by the brain (perception). The conversion of sound energy into electrical impulses is performed by a series of organs as shown in Figure 2.10.

The vibratory motion of air particles due to sound waves cause a sympathetic movement of a membrane in the ear known as the tympanic membrane, transferring sound into a mechanical motion. This motion is transmitted through an arrangement of bones (referred to as the Ossicles - Malleus, Incus and Stapes) to the base of the cochlea whose primary function is to transfer

mechanical motion into a series of electrical nerve impulses (Bies and Hansen, 2003, Howard and Angus, 2001).

The Ossicles are held in place with muscle tissue (Stapedius and Tensor tympani muscles) which can contract in response to high intensity sound to attenuate the propagation of sound into the cochlea. This action, known as the *acoustic reflex*, acts as a way of protecting the delicate hair cells of the inner ear. However, it is thought that this reflex was more likely evolved in humans to reduce masking from internal sounds such as chewing or vocalising (Moller, 2006).

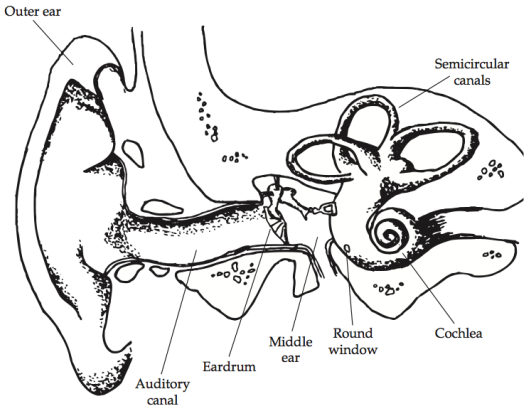
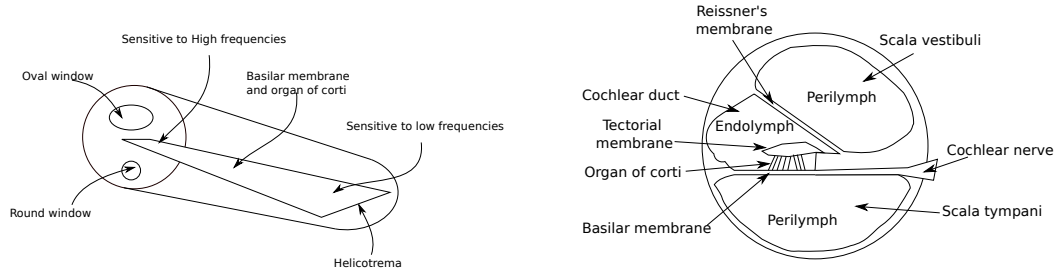


FIGURE 2.10: *A cross section of the human hearing system. This arrangement of organs allows sound energy to be transferred to a series of nerve impulses which are interpreted by the brain. Reproduced from (Everest and Pohlmann, 2009)*

The Cochlea, as shown in Figure 2.11, is a bony, spiral-shaped cavity, approximately 3cm long, filled with an incompressible fluid (perilymph). The Stapes (stirrup shaped bone of the ossicles) is connected to an oval membrane at the one end of the cochlea. The mechanical motion of the Stapes on the oval window causes movement in the fluid (with sympathetic movement of the round window at the other end of the cochlea). The arrangement of the Ossicles serves as an impedance matching device which allows efficient transfer of motion into the fluid of the Cochlea (Moore, 1997). The motion of this fluid disturbs the Basilar membrane which in turn displaces hair cells known as the Organ of Corti. This displacement allows the transfer of calcium and potassium ions around each hair cell causing cell polarisation which results in the generation of an electrical signal which is sent through nerve fibres to the brain. The Basilar membrane is stiffer at the basal end which causes the peak displacement of the membrane to be frequency dependent. This ensures that sound of certain frequencies activate only particular hair cells thus allowing a human to discern sounds of different frequencies. The audible bandwidth for a healthy human being is often quoted as being from 20Hz to around 20kHz however many people display less sensitivity at high frequencies with age (Presbyacusis (Bies and Hansen, 2003)) or with prolonged exposure to high intensity sound.



(a) profile of Cochlea (unrolled)

(b) section of Cochlea

FIGURE 2.11: *Section 2.11(b) and profile 2.11(a) of the Cochlea (unrolled for clarity) showing the arrangement of the round and oval window and the Basilar membrane and Organ of Corti. Adapted from (Kuttruff, 2007)*

The electrical signals generated by the hair cells of the cochlea are passed to the central auditory system where the signals are processed via a number of different processes that drive how humans perceive sound. These are thought to consist of a number of pattern recognition and stream segregation phases. This allows signals which have similar temporal, spatial or spectral information to be grouped together or separated as appropriate (Thiele, 1980).

An association model, proposed by Thiele (1980), suggests that the processing of auditory events is based on associating patterns in the signals received at the ears with previously observed signals from memory. This model also proposes that signals received at both ears are processed simultaneously on two different cognitive levels. The first level aims to ascertain the egocentric location of the sound source and the second, higher level of processing, aims to identify the sound source. The higher levels of processing are considered to be multi-modal whereby data from all senses are contributing rather than working in isolation.

Sound can also be sensed via bone conduction (Bies and Hansen, 2003) where vibrations travel through the body, directly vibrating the fluid in the inner ear. When a person makes a vocal sound, the vibrations also travel through tissue and bone directly to the cochlea. This can contribute to hearing one's own voice as different to when it is recorded and played back. Most musical instruments are operated very close to the body, for example, a violin is held between the shoulder and jaw of the musician. It is likely that a portion of the direct sound heard by the musician is via bone conduction.

The human perception of loudness is known to be frequency-dependent with the highest sensitivity peaking at approximately 3kHz-4kHz and the lowest sensitivity occurring at low frequencies. Figure 2.12 shows the widely known equal-loudness contours (Everest and Pohlmann, 2009). These curves show the required sound pressure level for a pure tone to appear at the same level of another pure tone at a different frequency. The frequency dependence of human hearing is often reflected in engineering measurements using a standardised A-weighting filter which attenuates low and high frequency components of a recorded signal (International Organisation of Standardisation). The A-weighting filter has a frequency response approximately equal to the inverse of the 40-phon equal-loudness curve.

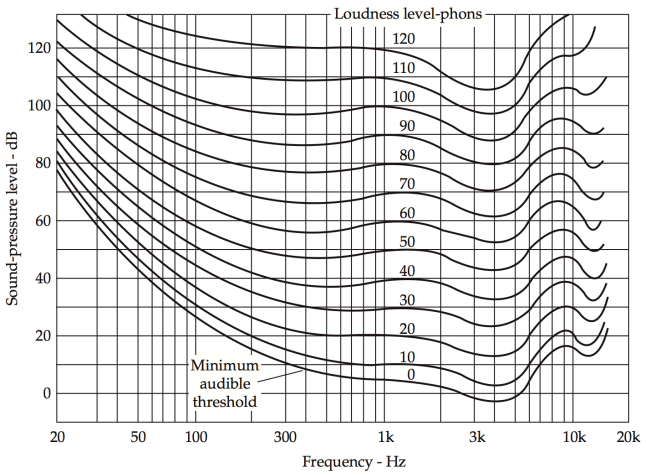


FIGURE 2.12: *Equal-loudness contours of the human hearing system. These contours were derived from Robinson and Dadson. Each curve represents the level a pure tone must be played at in order to appear at the same level as a different pure tone. Notice the the peak in sensitivity occurring at 4kHz and the lack of sensitivity at low frequencies. Reproduced from (Everest and Pohlmann, 2009)*

When stimulated with vibrations of a single frequency, the Basilar membrane tends to disturb clusters of hair cells which gives rise to a limitation in the frequency resolution of the human hearing system. The frequency resolution of the human hearing system is often modelled as a series of overlapping bandpass filters known as 'critical bands'. If two sounds occur within the same critical band then a summation of the two events is often heard. If one sound is louder than the other, the quieter sound will be masked. This is known as spectral masking.

One widely used model is the Equivalent Rectangular Bandwidth (ERB) scale where the spectrum is represented as a series of 42 overlapping critical bands. The bandwidth of each ERB critical band can be calculated using equation (2.19) where f_c is the centre frequency of each band (Glasberg and Moore, 1990). In acoustic engineering disciplines, the audio spectrum is often assessed in octave or third-octave bands (Bies and Hansen, 2003) where an octave refers to a doubling of frequency.

$$ERB(f) = 24.7(4.37f_c + 1) \tag{2.19}$$

The human hearing system also has a limited temporal resolution. The consequence of this is that a sound can be masked by another that is in close temporal proximity (Moore, 1997). This is referred to as temporal masking. A sound can be masked in this way if it precedes another sound (backward masking) or if it follows another sound (forward masking). The degree of forward masking has been shown to decrease exponentially as the time delay between sounds is increased but can persist for up to 100ms (Moore, 1997).

In some conditions, the spatial arrangement of sound sources can determine how efficiently they mask each other. For instance, when two people are talking simultaneously, it is easier to understand one of them when they are spatially separated. This process is known as *spatial unmasking* (Brungart et al., 2001) and is important in understanding why the same acoustic conditions can be judged differently by musicians of different instrument families. It is logical to

assume that the directional characteristics of an instrument will partly determine how audible reflections from a given angle of arrival are.

For a performing musician, the direct sound from the instrument will be perceived both as a nearby acoustic emission and also via bone conduction (depending on the operation of the instrument). The acoustic reflex may be activated which may attenuate airborne sound from the instrument. The dominance of the direct sound is likely to mask reflected sounds from the venue if they arrive during the production of a tone. The effect of temporal masking is likely to vary depending on musical aspects such as the length of the sound produced by the musician. For some instruments, due to the way they are played, the level at each ear may be significantly different. Furthermore, the extent of auditory masking can be influenced by higher level cognitive functions for example, if the listener is distracted by participating in another task. The combined effect of this is a masking pattern which varies temporally, spectrally and spatially. This may result in some musicians benefitting more from a particular early reflection pattern than others (Gade, 2010).

2.3.2 Sound source localisation

A central aspect of the human auditory system is its ability to resolve the direction of arrival of a sound. When a sound arrives from a lateral angle, the sound will arrive at the ipsilateral ear before the contralateral due to the additional distance to that ear, as shown on the right of Figure 2.13. As the angle of incidence increases so to does the time difference or Interaural Time Difference (ITD). As sound can diffract around the head only at low frequencies, the ITD is the dominant localisation cue at low frequencies (typically below approximately 700Hz). At higher frequencies, the sound does not diffract around the head and causes a level difference between the ears, due to shadowing effects. This cue is referred to as the Interaural Level Difference (ILD) which is dominant at high frequencies typically above 700Hz (Everest and Pohlmann, 2009). This is shown on the left of Figure 2.13.

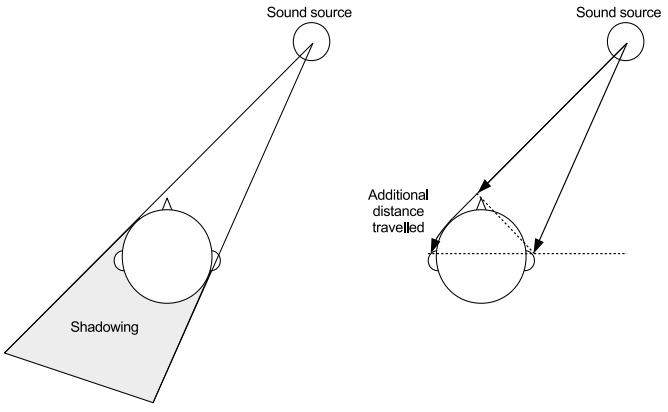


FIGURE 2.13: *Localisation via Interaural level difference (left) and Interaural time difference (right). ILD tends to operate at high frequencies whereas ITD tends to operate at low frequencies.*

The ILD and ITD models however cannot account for the ability to localise elevated sounds as these cues will remain constant with elevation. A further method of sound source localisation are monoaural effects caused by reflections of sound of the shoulders, head and complex structure of the pinna. This is depicted in Figure 2.14 where the reflected sounds sum at the ear drum and cause comb filtering at different frequencies resulting in a direction-dependent change in spectrum. These minute fluctuations are interpreted by the brain to assist in resolving the direction of arrival. This effect is highly individualised due to the differences in physical shape of all human beings. The amount of colouration imposed by the interaction with the head and the body is known as the head-related transfer function (HRTF) which is dependent on the angle of incidence of the sound (Everest and Pohlmann, 2009).

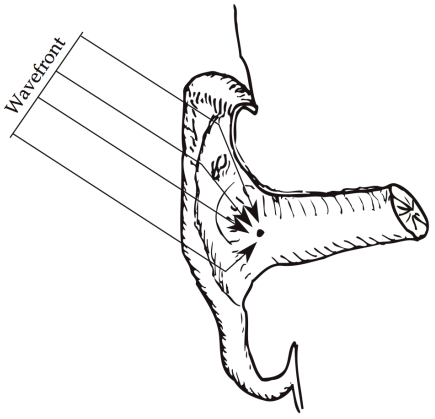


FIGURE 2.14: *A wavefront interacting with the Pinna. The wavefront encounters different regions of the pinna and are reflected towards the auditory canal. The delays caused by the different paths produce specific comb filtering patterns which are interpreted by the brain and are associated with a particular source angle. Image reproduced from (Everest and Pohlmann, 2009)*

A further method of resolving directional cues is via minute head movements which produce modulations in all the afore mentioned cues which can then be interpreted by the brain. These minute actions can minimise the possibility of front-back confusion and can help resolve the direction of sound sources when they occupy the so called *'cones of confusion'*, areas where sources arriving from different angles can produce the same ILD or ITD (Moore, 1997).

Interaural coherence cues are thought to be used for the perception of space. The Interaural Correlation coefficient (IACC) (2.26) is a measure of the similarity of signals reaching both ears as a function of time delay (Everest and Pohlmann, 2009). The dissimilarity between the signals reaching the ears correlate with spatial effects such as Apparent Source Width (ASW) where early lateral reflections give rise to a perceived widening of a sound source.

$$IACC(\tau) = \frac{\int_{t1}^{t2} x_1(t)x_2(t+\tau)dt}{\sqrt{\int_{t1}^{t2} x_1(t) \int_{t1}^{t2} x_2(t)}} \tag{2.20}$$

With the fundamental mechanics of the human perception of sound now covered, it is possible to discuss the impact of different sound stimuli. In particular, the subjective effect of different acoustic conditions on the perception of a sound source.

2.3.3 Precedence Effect

The *Precedence effect* is an auditory phenomenon where a sound followed by an identical sound after a short delay is perceived by a listener as a single auditory event (Wallach et al., 1949). By increasing the delay between the lead and lag signals past what is known as the echo threshold the fused event separates out into a perceived primary sound and echo. When the sounds are fused, the perceived spatial location is dominated by the leading sound as shown in Figure 2.15.

The echo threshold varies from person to person and also with the type of stimulus. For instance, transient signals have an echo threshold often of 10ms or lower whereas signals that start with a transient followed by a steady state (music/speech) produce a longer precedence window on average around 30ms (Lokki et al., 2011). In concert hall acoustics where the delayed sound is formed from a number of lower amplitude reflections, the precedence window will be elongated further.

In concert hall acoustics, this is a significant effect as the arrangement of early reflections can determine if the reflections are fused with the primary event or perceived as echoes. In the former case, fused reflections can contribute to perceived spatial or level effects as discussed in the previous section.

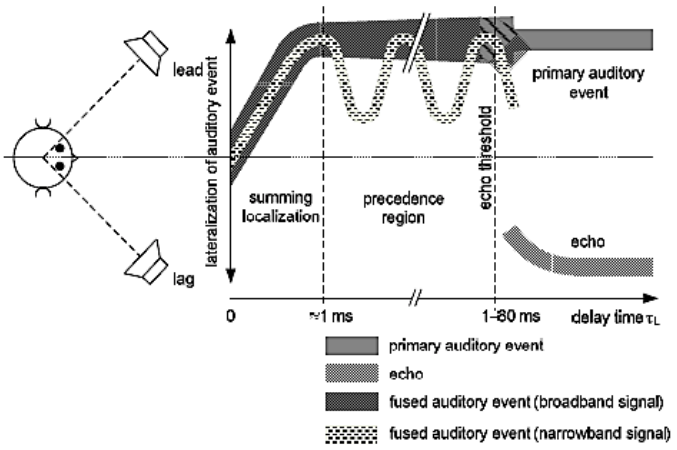


FIGURE 2.15: *A diagram illustrating the precedence effect where early reflections are perceptually fused with the direct sound until the echo threshold where they are perceived as two separate events. Image reproduced from Blauert and Jonas (2005)*

The Precedence effect exhibits build-up behaviour upon repetition of the stimulus pair where the echo threshold is longer than if the stimulus was played only once (Blauert and Jonas, 2005). The Clifton effect is an observation whereby the precedence effect appears to break down whenever

the auditory scene changes (Blauert and Jonas, 2005). It was shown that the echo threshold was shorter for a new auditory scene, which could be interpreted as an analysis of the new scene from scratch. It has been observed that old precedence effects can persist if the auditory scene is reset within a certain time frame.

2.3.4 Auditory processing

The human hearing system has the ability to perceive sounds as a complete entity but also focus in on the specific constituents. This ability is commonly referred to as the *'Cocktail Party Effect'* whereby it is possible to focus in on and understand one person talking in a large noisy group of people. Another example of this is in enclosed spaces, the sound reaching a listener consists of the direct sound and also sound reflected from multiple surfaces. Yet in most cases, the listener can accurately discern the location of the sound source in the presence of multiple competing reflections. Auditory Scene Analysis (ASA) is a proposed model explaining how this is possible. This theory relies principally on Gestalt principles where humans tend to group objects together by aspects such as proximity, similarity, symmetry and continuity (Thiele, 1980). These aspects apply to sound where, for instance, two spectrally different (but spatially identical) sounds that have the same amplitude envelope may be perceived as arising from the same source.

This ability is often applied sub-consciously but can also be invoked by experienced listeners who can actively focus on different parts of a sound. This ability is central to musicians in many different contexts. For instance, a musician approaching a challenging passage in a piece may prioritise the physical actions involved in playing correctly over any aspects of the acoustic conditions on stage. This has been referred to previously as *Cognitive Load* (Guthrie et al., 2013) and is a specific occurrence of *in-attentional blindness* (Lennox and Myatt, 2011) that can be observed in other areas where subjects can fail to react to a stimulus when they have been asked to complete simple tasks. Furthermore, if a musician is presented with challenging acoustics, they may focus more on the direct sound from their instrument and suppress the influence of their surroundings. If a musician is playing in an orchestra then often they are shifting between listening to the sound of their own instrument and listening out for cues from other instruments.

Kahle (2013) discusses the benefits of addressing room acoustic quality by considering the notion of stream segregation. It is suggested that room and source presence are perceptual factors that can be assessed initially in terms of loudness and then expanded to include other aspects such as frequency balance and spatial distribution. This appears to correspond with findings from many studies that suggest that different elements of the response serve different functions to the musician and are assessed differently. It also seems to serve as a simple explanation of a musician's sensation of the instrument and room gradually becoming one as balances form between source and room presence. Dammerud (2009) similarly reports that some players take advantage of the reverberant sound from the auditorium for intonation, balance, articulation and timing whereas other players prefer working on the immediate sound and the early sound on stage.

Gade (1982) also cites a similar phenomenon describing how during experiments a musician's concentration on acoustic aspects can vary with repetition. For instance, if they are asked to

repeat a simple phrase as a source signal to a virtual stage simulation, the physical act of playing this phrase becomes automatic allowing them to concentrate more fully on specific aspects of their acoustic environment. In these experiments, very simple phrases were chosen so musician test subjects would not tire easily on multiple repetitions of a complex piece.

In auditorium acoustics, the sound source is generally located at a distance to the receiver (exocentric) whereas in stage acoustic research the source and receiver are often collocated (Egocentric) or nearly collocated (pseudo-egocentric). The notion of an egocentric sound source is similar to the ability of some humans to echolocate to build up a detailed impression of their surroundings (Halmrast, 2013). It has been previously demonstrated that humans are able to make accurate judgements about the size of the space they inhabit by listening to the effect of their own vocalisations (Yadav et al., 2011).

In summary, it is evident that the acoustic environment experienced by the musician involves a number of complex physiological and psychological processes. This makes it difficult to predict what parts of the acoustic response can be heard by a musician and how various objective attributes influence the subjective characteristics of the sound. This presents a significant challenge in designing a concert hall with the desired acoustic attributes for a performer. The following section will review the broad acoustic design concerns for concert halls including some common designs for stages and auditoria.

2.4 Concert hall acoustics

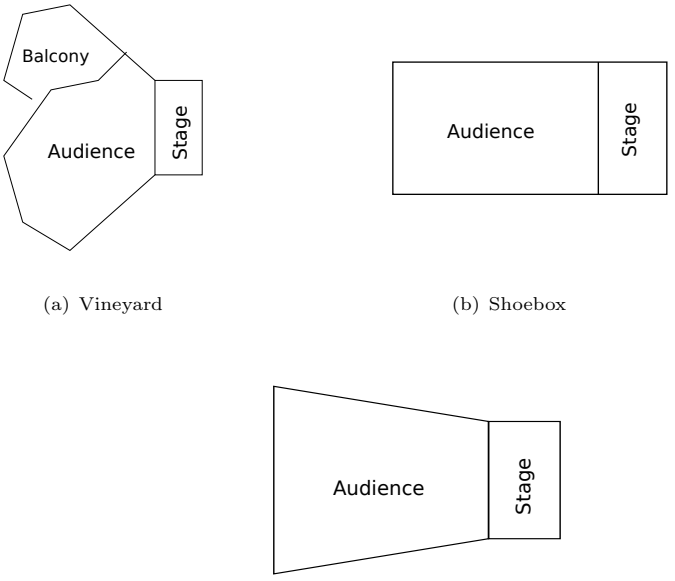
Over many years, a number of popular designs for performance spaces have evolved, some accidentally and others, more recently, by design as understanding of acoustic propagation and auditory perception have increased (Essert, 1997). The design of a performance space is often influenced by many seemingly unrelated factors including the business plan of the venue, the aesthetics and functional requirements of the building and also the type of music expected to be performed. A concert hall with world-renowned acoustics may be appropriate for symphonic orchestral music but may not provide an appropriate atmosphere for solo recital or popular music performance. This section will review some of the common concert hall designs including different approaches to stage enclosure design.

2.4.1 Auditorium layout

The general shape of the auditorium is a primary concern for both architects and acousticians as it can influence both the aesthetic and practical usage of the hall, dictating for example, sight lines and venue capacity but also how the sound propagates through the venue. Figure 2.16 shows three of the most common concert hall geometries namely, the *vineyard layout*, the *shoebox layout* and the *fan layout*. Each of these concert hall layouts are known for providing various advantages and disadvantages in terms of audience layout but also early reflection patterns.

Many of the most highly regarded concert halls are referred to as “shoebox” halls owing to their symmetrical cuboid shape. Most shoebox halls feature high ceilings and balconies around the

perimeter of the "receiving" end of the hall. These halls generally promote strong early lateral reflections which can improve the perceived loudness and size of the source on stage (apparent source width effect). Furthermore, strong lateral reflections can promote feelings of spaciousness and envelopment for the audience (Rossing, 2007). The Musicvereinsaal (Vienna, Austria) and the Grand Hall at Glasgow City Halls (Glasgow, UK) are examples of shoebox configurations.



(a) Vineyard

(b) Shoebox

(c) Fan

FIGURE 2.16: *Plans of typical concert hall configurations*

Fan-shaped halls are influenced by designs for early cinemas and theatres and are known to provide optimal sight lines from the audience to the stage. Their shape can be used to reflect sound from the rear of the hall. However, the splayed angle of the walls can redirect lateral early reflections resulting in a perceived lack of clarity (Everest and Pohlmann, 2009). An example of a fan shaped hall is the Liverpool Philharmonic hall.

Vineyard style halls are typical of modern 20th century concert halls and in general feature subdivided, terrace-like seating areas which surround the stage. The walls which surround each part of the seating area help to promote early reflections. The Berlin Philarmonie is a famous example of a vineyard style concert hall (Rossing, 2007).

Recital halls are designed for much smaller ensembles (or soloists) and normally cater for much smaller audiences (typically fewer than 1000 seats). Often, recital halls are shoebox shaped with a relatively small stage. Recital spaces often tend to be designed to a lower reverberation time (typically between around 1.4s-1.7s) (Barron, 2009). Smaller recital spaces are often required to accommodate ensembles of differing sizes ranging from small orchestras to solo performers. In order to cater for many differing situations, some variability is designed into the space to adjust

the acoustic conditions for different settings. This can include elements such as reversible panels (where one side is reflecting and the other absorbing) or movable curtains.

It is increasingly common for some venues to be designed as multi-purpose halls which allow a much more varied concert programme and hence a higher profit margin. Multi-purpose halls often feature removable seating and sound reinforcement systems. Multipurpose halls can feature operable elements or removable curtains in order to adapt the acoustics to the type of performance. In some cases, the acoustics can be enhanced artificially through the use of reverberant chambers or digital reverberation enhancement systems.

The audience seating area can have a significant impact on the acoustics of a concert hall as human beings act as fairly efficient acoustic absorbers. The arrangement and size of the audience is considered carefully at the beginning of many concert hall designs. The audience area can either be flat or sloped (*'raked'*) in profile. A raked seating arrangement is often used to improve the sightlines for the audience (Everest and Pohlmann, 2009). In drama theatres often a steep rake is used in contrast to concert halls which typically make use of much shallower angles in order to avoid reducing the overall volume of the space. In some concert halls the seats themselves are carefully designed so that the acoustic conditions are not too drastically affected by the number of seats occupied for a performance. Many concert hall spaces also make use of balconies which are arranged around the three walls of the audience area and in some cases extend around the stage so that audience surround the performers.

2.4.2 Stage enclosure

The design of stage enclosures is typically an inherent part of the overall design of an auditorium and so can similarly be influenced by a multitude of acoustic and non-acoustic factors. The overall dimension of the stage is often determined by the type of music (and hence the number of musicians) that is programmed. A symphony orchestra can have in excess of 100 members and so a stage will require a footprint large enough to accommodate them comfortably.

In concert halls, it is common for the stage to occupy the same volume as the auditorium in contrast to some opera houses or theatres where the stage is framed by a proscenium arch or features a flytower above the stage for scenery. In many theatres, the orchestra are situated in an orchestra pit located below, and at the front edge of, the stage. In concert halls, the stage may also be elevated above the level of the front stalls seating or level with the front row of the audience as is more often the case in smaller recital venues.

The stage enclosure geometry can be considered from the perspective of both the audience and musicians. For example, consider a stage enclosure featuring splayed reflecting side walls, as shown in Figure 2.17(a). This design may help to reflect sound from an ensemble out towards the audience but may reduce the reflected energy received by the musicians on stage.

The geometrical arrangement of reflecting surfaces will determine the angle and time of arrival of reflections on stage while the material characteristics of each surface will determine the diffuseness, frequency content. The relative loudness of each reflection can be determined by all of the afore mentioned factors in addition to the location of the musician on stage and the instrument's

radiation characteristics. Stage enclosures can also feature diffusing or non-diffusing surfaces depending on the requirements of the venue. Figure 2.17 considers a number of basic stage enclosures shapes seen in various concert halls. Figure 2.17(d) is typically found in shoebox halls such as the Musicvereinsaal in Vienna. The stage shaped shown in Figure 2.17(b) is very similar to the Musiikkitalo in Helsinki. Figure 2.17(c) is similar to the Hatch recital Hall, University of Rochester. Some venues feature variations upon these shapes, for example the Wigmore Hall, London features a cylindrical stage shape. The Walt Disney Concert Hall in Los Angeles features splayed side walls that are curved outward towards the front stalls.

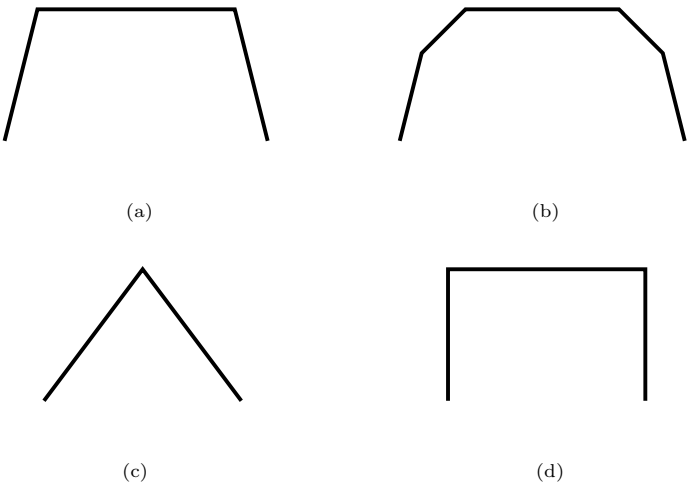


FIGURE 2.17: *Plans of typical concert hall stage shapes*

Numerous concert hall stages feature suspended overhead reflectors (as depicted in Figure 2.18) which are either permanent or operable in order to change the configuration of the stage (Rossing, 2007). The overhead reflector can comprise of an array of planar or curved elements with numerous gaps for lighting elements. The overhead reflectors are typically installed for ensemble or orchestral performance to aid communication between musicians. This is achieved by providing an uninterrupted reflection path between musicians. The angle of orientation is carefully designed so that both musicians and audience benefit.

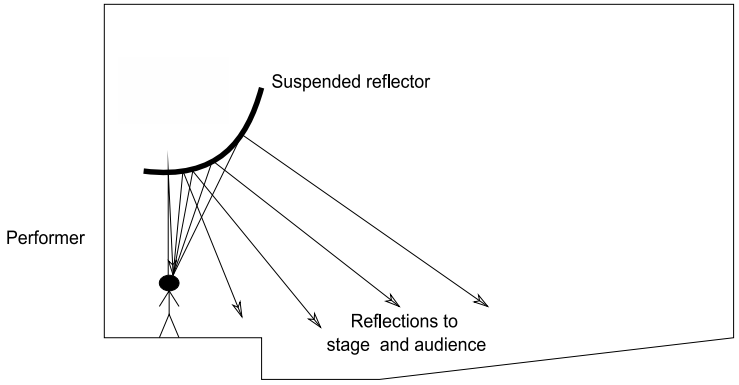


FIGURE 2.18: *Suspended overhead reflectors can provide early reflections to both the stage and audience areas*

On some stage enclosures, the upper portion of some reflecting surfaces are tilted inwards towards the hall on all sides, promoting early reflections from above the lateral plane. For example, in the Queens Hall in the Royal Library, Copenhagen (Gade, 2010). Elevated reflectors have been demonstrated to assist orchestras in terms of their ease of playing together. This is due to the early reflections allowing distant musicians to be heard more clearly than if the sound were left to propagate through the orchestra.

A common feature of some stage enclosures are lightweight areas of flooring which can be raised or lowered (risers). These are primarily for use with orchestras which elevate some musicians over others. Risers can also contribute to the low frequency register of Celli and Double Bass instruments that partially rely on a resonating floor as part of their sound (Dammerud, 2009).

Additional reflectors are sometimes placed on stage behind a performer or small ensemble, often if the stage provides insufficient early reflections for the performers. These reflectors are typically used to modify the acoustic response of the stage enclosure by adding specular or diffuse reflections from behind the performer. A commercial example of this is the VAMPS (Variable Acoustic Modular Performance System) (Cox and D'Antonio, 2004, RPG) which is shown in Figure 2.19. This particular variable reflector can be set to provide both diffuse and specular reflections as required. Other passive variable reflector systems have been investigated previously such as the triangular *'retro-reflectors'* developed by Tuominen et al. (2013) which provide a similar function to stage monitors in popular music performance.

FIGURE 2.19: *Image of the VAMPS reflector system (Cox and D'Antonio, 2004, RPG). This arrangement of reflectors is designed to provide specular or diffuse early reflections to performers from the stage.*

Recently, it has been suggested digital signal processing could be used to augment the existing stage acoustic response of a venue to suit a particular scenario (Ko et al., 2013). Similar to an assisted reverberation system, microphones are used to capture the sound from a musician or ensemble which is processed and rendered over a surrounding loudspeaker array to emulate early reflections.

It is clear that the design of stage enclosures can greatly influence the acoustic response on stage and in the audience and therefore is often considered alongside the geometry of the rest of the auditorium to ensure that sound propagates around the room in a pleasing manner for both musicians and audience members.

2.5 Musician awareness and adjustment to acoustic conditions

As discussed previously in this chapter, the proximity of the musical instrument can produce a significant masking source for early arriving reflections. It is therefore crucial to understand what parts of the acoustic response produce an audible effect for performing musicians. A number of recent studies have been conducted to determine what aspects of stage acoustic musicians are audible for musicians of various instruments. Of specific importance to this research, the audibility of early reflections for soloists has been explored previously by Gade (1982).

It was shown that the threshold of perception of a single (overhead) early reflection decreased as the time delay relative to the direct sound increased. This can be seen in Figure 2.20 which shows the threshold perception for a single reflection at various time delays and for different families of instruments. In general, it can be seen that the threshold of perception of a single reflection decreases with time delay, suggesting that late arriving reflections are more easily detected than

early arriving reflections of equal energy. This follows as the masking properties of the direct sound are reduced as the reflection arrival time increases.

Gade (1982) also investigated the audibility of groups of simulated early reflections within a particular time interval. The horizontal lines in Figure 2.20 show the threshold of perception for 6 equally spaced reflections between 20ms and 100ms. It can be seen that the threshold of perception for groups of reflections also varies between different instrument families (strings and flutes). It was discussed that these differences were mainly related to differences in masking patterns created by different instruments and the differences in the way the instruments were held and operated.

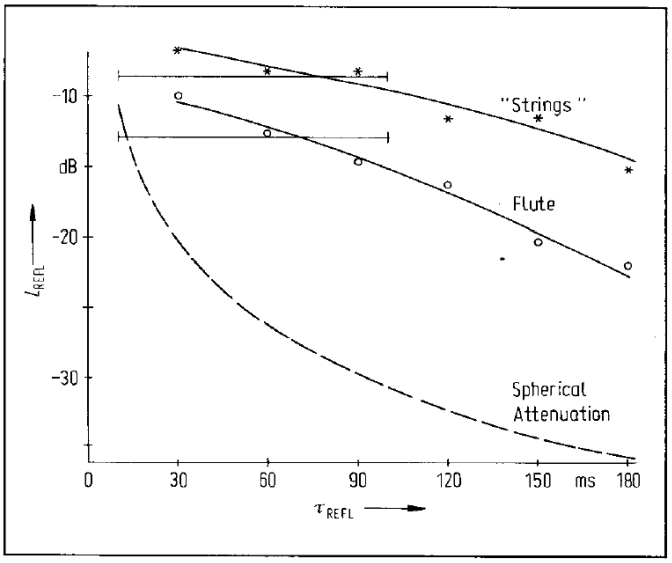


FIGURE 2.20: *A plot showing the threshold of perception for single reflections for varying delay times and for different instrument groups. Single reflections are easier to detect as the delay time is increased, however flute players appear to be more sensitive to reflections at all examined delay times. Reproduced from (Gade, 1982)*

However, Gade's research (Gade, 1982) also revealed that some of the test subjects could discern the presence of early arriving reflections due to the resulting change in timbre, as opposed to being able to discern specific reflection events. This is significant as it suggests that reflections arriving at certain time regions produce different subjective effects. Where early reflections from nearby surfaces may influence the perception of timbre, later arriving reflections may produce variations in perceived support or envelopment.

In addition, a study by Bermond and Davies (2001) aimed to explore the subjective effect of temporally diffuse reflections on performing musicians. It was demonstrated that musician test participants could discern the difference between soundfields consisting of specular reflections and those with diffuse reflections and the effect is mainly perceived in terms of *"harshness"*, diffuse soundfields were perceived as less harsh in timbre.

A laboratory experiment conducted by Ueno and Tachibana (2003) aimed to determine which aspects of a venue's acoustic response were preferable for musicians. The variables in the experiment were reverberation time, the level of early reflections and the level of a late reflection

arriving at 200ms delay relative to the direct sound. It was found that for soloists, a moderate reverberation time (in this case 1.9 seconds) was preferred by most of the test subjects. It was also found that the strength of early reflections contributed to the impression of room size. Furthermore, it was found that high levels of early reflections were disliked by musicians as it made the space sound too small. It was found that the weakest level of early reflections was preferred by most experienced musicians. It was also commented that the presence of early reflections didn't necessarily promote the perception of support.

A number of studies have indicated that the height to width ratio is an important factor for successful stages (Dammerud, 2009, Guthrie et al., 2013). There appears to be a general preference of musicians towards high and narrow stages for ensemble playing, presumably to reduce the propagation time of early reflections (Gade, 2010). Guthrie et al. (2013) found that judgements of preference and ease of playing are influenced not only by established omnidirectional parameters but also by the directional distribution of arriving energy. Overall it was found that narrow stages with hard, flat (non-scattering side walls) are preferred for horns whereas an increased width or scattering side walls is preferred for upper strings.

Many professional musicians are highly sensitive to the influence of acoustic conditions on the sound of their instruments and consider it as an integral part of their instrument's sound. It has been demonstrated that musicians will adjust their articulation and phrasing over time to suit the specific acoustic conditions they experience (Ueno et al., 2007). In one study (Ueno and Tachibana, 2005), musician test participants describe in detail how they approach performing in unfamiliar acoustic conditions. To begin with, they report getting a rough impression of the hall as they begin playing. They then fine-tune aspects such as dynamics, intonation, tempo and vibrato and vary the orientation of their instrument until they are satisfied with what they hear. Conversely, in the same study, some musicians reported that they do not actively change their technique in different acoustic conditions. However, by analysing the direct sound from the instrument during their performance it was found that these musicians were subconsciously varying their technique.

In a similar study (Kato et al., 2007), it was found that the differences in the recorded sound were statistically significant for performers in terms of aspects including tempo, vibrato rate, vibrato extent (in terms of frequency and intensity) and A-weighted sound pressure level. A further study by the same authors (Kato et al., 2008) found that there was a tendency to reduce the length of notes (or increase spacing between them) in a phrase when they were presented with a more reverberant acoustic environment.

Overall, the findings of these studies strongly suggest that the temporal distribution of early reflections has an effect on the musician's impression of the acoustics on stage. The results imply that early arriving reflections can influence the musician's impression of timbre, as can the temporal diffuseness of those reflections. The results also imply that the spatial distribution of early reflections can determine how audible reflections are to musicians of specific instruments. In addition, these studies have demonstrated the complex feedback loop that exists between the musician and the venue. Musicians will tend to adapt their technique based on what they hear back from the venue sometimes instinctively and sometimes by design. This acclimatisation

appears to happen when the musician enters the venue and is gradually refined during the performance.

2.5.1 Musician concerns and preferences

The preceding section suggests that distribution of reflections will strongly influence the sound perceived by the musician which will cause variations in their technique. It is therefore important to determine what aspects of the acoustic conditions are important to musicians and what is preferred. As noted by Gade (1982), a musician's requirements for acoustic conditions may vary with the context of the performance. Where a solo performer may be concerned with how their sound is propagating towards the audience, an orchestral performer might be more concerned with how audible other musicians are to assist in communicating with each other. Furthermore, different genres, musical instrument families and performance styles of music may require different characteristics for performance.

Several qualitative studies have been undertaken to understand what aspects of a hall's acoustic response are important to performing musicians. Gade (1982) found that many musicians referred to the acoustic response in very similar ways and categorised their main areas of concern which are summarised below:

Reverberance is mainly perceived during breaks or shift of tone played, since it sustains the tones just played. It binds adjacent notes together, can blur details in the performance and may give a sense of response from the hall.

Support makes the musician feel that (s)he can hear himself and that it is not necessary to force the instrument to develop the tone. It can be felt during the onset of tones and is therefore believed to be related to properties different from reverberance.

Timbre is defined as the influence of the room on the tone colour of the instrument and on the balance in level in different registers.

Dynamics describes the dynamic range obtainable in the room and the degree to which the room obeys the dynamic intentions of the player

Dammerud noted that for orchestras, preferred stages were described as having good **bloom** and **projection**. In interviews, Dammerud (2009) found that a violinist had referred to bloom as being how much reverberation, plus warmth, is present and projection being how the sound carries out into the hall. This is later more formally defined in his study as being the impression of hearing what the audience hear can be an element of communication and reassurance. Dammerud (2009) suggests bloom is similar to a sense of acoustic support and projection is analogous to the feeling of communication with the audience; both parameters appear to be linked to the level of late acoustic response. This appears to be linked with the degree the stage is acoustically exposed to the main auditorium. It was further discussed that a balance should be maintained between the strength of early reflections from the stage enclosure and later reflections and decay from the auditorium.

Additionally, musicians in orchestras and ensembles also consider **Ease of Ensemble** of crucial importance to a successful stage enclosure. This effect refers to how easy it is for the musicians in an ensemble to hear each other so that they can remain synchronised and get a fair impression of how loud other musicians are playing.

It is clear from this that, generally, a musician's preference to stage acoustic conditions is multidimensional. As will be discussed later in this chapter, most stage acoustic research appears to concentrate on Support, Reverberance and Ease of Ensemble. Currently, little is understood regarding how the acoustic conditions influence perceptions of Timbre and Dynamics.

2.5.2 Gesture

During a performance, a musician will often move around on stage while they are playing for both functional and expressive reasons. This movement has been noted to have audible effect on the received acoustic response at an audience position but may also affect the acoustical feedback received by the musician. Different classifications of gestures used by musicians when performing have been summarised (Wanderley, 1999). Figurative gestures are generally more associated with articulation technique i.e. vibrato, tremolo etc. Ancillary gestures are those which are produced by movement of the musician or musical instrument. Gestures can also be classified as being low or high in amplitude i.e. an orientation shift of 90° is seen as a high amplitude movement but the circling of a clarinet bell at the end of a phrase is seen as a low amplitude movement. It has been observed that for a remote listener, these movements create an audible modulation in the direct sound from the musical instrument as different radiation patterns are exposed to the listener.

A previous study by Cabrera et al Cabrera et al. (2010) explored this idea by measuring the acoustic response of various spaces with a head and torso simulator (HATS) which was angled at different orientations. A HATS system is a mannequin dummy with microphones positioned in the head. These systems usually have very realistically shaped pinnae which simulate the effect of an HRTF. It is used to record binaural sound. The measurements were then evaluated by obtaining the ILD, IACC and the room gain. This work appeared to show that any colouration of acoustic feedback present when the HATS simulator was rotated was unlikely to be audible. In contrast however, there is some qualitative evidence to suggest that some performers can hear differences in acoustic conditions caused by movement and consciously move around when adjusting to the acoustics (Ueno and Tachibana, 2005).

2.6 Objective acoustic parameters

Historically, the successful acoustic design of concert halls has developed iteratively over a long period of time by replicating successful building elements that by chance featured what was perceived as desirable acoustic conditions (Essert, 1997). Modern day acoustic design extends this principle by characterising the acoustic response of successful concert halls using a series of objective parameters and relating these parameters to various subjective impressions. By

utilising accompanying prediction methods, a concert hall can be designed to strict criteria. These design approaches can then be verified via objective measurement. The following sections will describe the basic acoustic parameters which are used in the design of performance spaces.

2.6.1 Reverberation time

One of the primary concerns regarding concert hall design is the length of the acoustic decay which is evaluated using reverberation time parameters (EDT, T_{20}, T_{30}, T_{60}). The reverberation time parameters are the most widely used in concert hall and building acoustics and are used extensively in the basic planning phase of auditorium design. Through accumulated experience, there are recommended ranges for reverberation time which have been deemed appropriate for various genres of music. It is defined in ISO 3382-1 (International Organisation of Standardisation, 2009) as the time taken for a steady state sound in a room to decay in sound pressure level by 60dB.

Reverberation time is based on a statistical approach to room acoustic measurement which considers only the monoaural decay of energy in the space. These parameters were traditionally measured by exciting a room with a pink noise source and measuring the level decay duration from the moment the source had been switched off. More recently, the reverberation time is extracted from the impulse response by deriving the Schroeder decay curve. This is obtained by backwards integration of the squared impulse response (International Organisation of Standardisation, 2009).

Early Decay Time (EDT) is commonly linked to perceived reverberance, for an audience member, and calculates reverberation time between a level of -5dB and -15dB using linear regression. Similarly, T_{20} and T_{30} calculate Reverberation time but over a level range of -5dB and -25dB and -5dB and -35dB respectively. EDT, T_{20} and T_{30} are computed using linear regression of a decay curve as shown in Figure 2.21. In this case, it can be seen that the EDT (0.23s) and T_{30} (1.65s) are very different which suggests a two-stage, non-linear acoustic decay. Similar analysis of the acoustic response at a larger distance from the sound source will produce EDT and T_{30} values which are very similar, implying a linear decay.

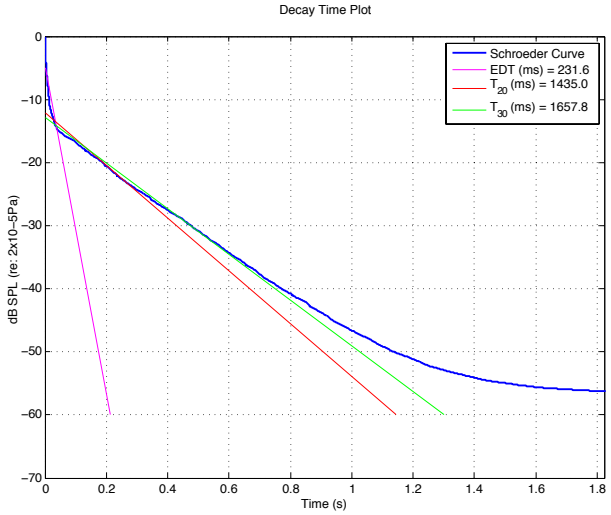


FIGURE 2.21: *Plot showing the Schroeder curve with lines showing how EDT, T_{20} and T_{30} are estimated.*

The reverberation time is a frequency-dependent parameter and varies mainly due to the physical construction of a space, such as room volume and the surface area of materials with different acoustic absorption characteristics. Different studies have used numerous different methods to obtain a single figure value including mid-frequency averages (International Organisation of Standardisation, 2009), 125Hz to 2kHz averages or simply quoting the mid frequency value as recommended by Beranek (2004).

The reverberation time in an auditorium is carefully designed to suit the size and usage of the space. For example, a short reverberation time is considered appropriate for speech to promote speech intelligibility. For orchestral music, the ideal reverberation time is much longer but can vary due to the specific genre of the performance. As an 'optimum' reverberation time is a highly subjective phenomenon, a range of appropriate values are generally aimed for. Figure 2.22 shows the optimal reverberation time against room volume (m^3).

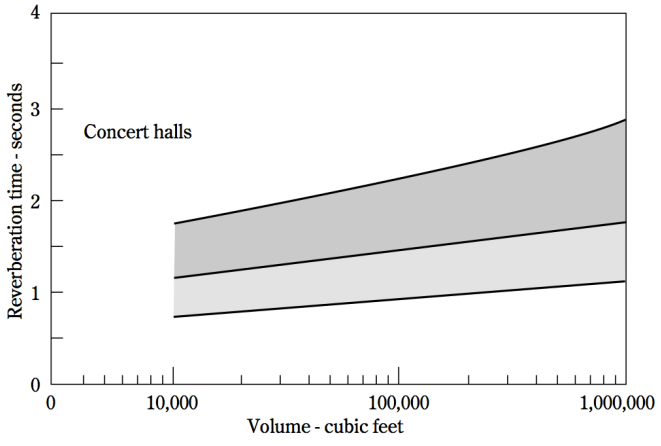


FIGURE 2.22: *Plot showing the 'optimal' reverberation time for concert halls against room volume. The lower shaded area refers to opera and chamber music whereas the darker, upper region refers to symphonic music. Plot reproduced from (Everest and Pohlmann, 2009)*

Bradley (2010) discussed how EDT and T_{30} values can be very similar for audience positions, implying a linear decay of sound pressure level. As the receiver is positioned closer to the sound source an abrupt drop in sound pressure level is followed by a more gradual decay. This will cause the EDT and T_{30} values to differ significantly where EDT would be significantly shorter than the T_{30}.

While the reverberation time is considered to be one of the most important factors in concert hall design, it has been shown that two concert halls with similar reverberant characteristics can have very different subjective characteristics. Typically, this is due to the amplitude, temporal and spatial distribution of early reflections. The early reflections can significantly influence the character of sound heard in a space in terms of its perceived proximity, size, loudness, spatial impression and timbre (Barron, 2009).

2.6.2 Clarity

The presence of early reflections can produce noticeable acoustic effects which influence how a sound source is perceived. One such effect is known as Clarity which is defined by Beranek (2004) as the degree to which a listener (seated in the audience) can distinguish sounds in a musical performance. The Clarity index, C_{80}, is an objective parameter which describes this sensation. It is an energy ratio of early to late sound where the transition between early and late is defined as 80ms. High values of C_{80} occur most often in less reverberant halls (indicating a high degree of clarity) while negative values of clarity are sometimes observed in large cathedral type spaces. Sounds playing in halls with low C_{80} are often described as sounding muddy (Beranek, 2004). Like the reverberation time parameters, C_{80} can be derived from the room impulse response and is frequency-dependent, measured in octave or third octave frequency bands or expressed as a single figure value. C_{80} is calculated using equation (2.21).

$$C_{80} = 10log_{10}\left(\frac{\int_{0ms}^{80ms} p^2(t)dt}{\int_{80ms}^{\infty ms} p^2(t)dt}\right) \tag{2.21}$$

Definition is a similar parameter (often abbreviated as D_{50}) used more often in the context of speech clarity. The main differences are in the integration times used to segregate early and late energy. It is generally well accepted that 50ms is a more suitable integration time for speech. D_{50} is typically expressed as a percentage.

$$D_{50} = 100\left(\frac{\int_{0ms}^{50ms} p^2(t)dt}{\int_{0ms}^{\infty ms} p^2(t)dt}\right) \tag{2.22}$$

A further parameter which indicates clarity is the centre time, T_s. The centre time calculates the *'centre of gravity'* of the impulse response i.e. the time at which the energy before is equal to the energy after. A low value of T_s indicates the energy balance is in favour of early sound, indicating increased clarity. The advantage of this parameter is that there is no sharp limit between early and late sound as with C_{80} and D_{50}.

$$T_s = \frac{\int_{0ms} tp^2(t)dt}{\int_{0ms}^{\infty ms} p^2(t)dt} \cdot 1000 \tag{2.23}$$

Acoustic parameters such as C_{80} consider the transition between early and late parts of the acoustic response as a fixed value (often 50ms, 80ms or 100ms). However, as discussed previously, the true mixing time for a room is known to be dependent on the volume of the room. Furthermore, the subjective mixing time is often different from those shown by measurements. Some authors argue that the mixing time should be used as the boundary point between early and late parts of the acoustic response in parameters such as C_{80} (Bradley, 2010).

2.6.3 Spatial impression

A listener's spatial impression is also of great importance in concert hall acoustics. Which can influence aspects such as the perceived size of the orchestra in addition to feelings of envelopment. The widening of the orchestra, from an auditory point of view, is referred to as the Apparent Source Width (ASW) effect whereas listener envelopment is defined as the impression of being surrounded by the reverberant soundfield (Rossing, 2007). Early investigations into this subjective effect found that the presence of early lateral reflections strongly influenced both of these i.e. early reflections arriving from side walls.

The Lateral Energy Fraction (LEF) was devised as a method of assessing apparent source width and listener envelopment which are both affected by the presence of strong lateral early reflections. LEF is defined in equation (2.24) as the linear ratio of lateral reflections (not including the direct sound) to early reflections arriving in all directions. LEF is measured using a coincident pair of microphones, one with omnidirectional polar pattern and the other with dipole (figure-of-eight) characteristics. The figure-of-eight microphone is oriented so that the lobes of the microphone are pointing towards the side walls of the auditorium. This allows reflections arriving from lateral directions to be distinguished from those arriving from other directions. High values of LEF indicate a high degree of spatial impression and tend to occur when reflections arrive from lateral directions. LEF correlates well with the ASW effect where sound sources can appear larger in the presence of strong, early lateral reflections.

$$LEF = \left(\frac{\int_{5ms}^{80ms} p_{fo8}^2(t)dt}{\int_{0ms}^{80ms} p_{omni}^2(t)dt} \right) \tag{2.24}$$

A listener's spatial impression in concert halls can also be attributed to differences in the time and level of arrival of early reflections at each ear. As discussed previously, the auditory system relies partially on the coherence of the signals received at each ear to determine such aspects of the soundfield. The impression of listener envelopment tends to increase as dissimilarities between the ears increase. This effect can be quantified using the Interaural Cross Correlation Function shown in (2.25). This determines a measure of coherence between the ears as a function of time delay, τ. The Interaural Cross Correlation Coefficient (IACC) represents the maximum

value of the IACF and is commonly used in concert hall design and assessment (Ando, 1998).

$$IACF(\tau) = \frac{\int_{t1}^{t2} x_1(t)x_2(t+\tau)dt}{\sqrt{\int_{t1}^{t2} x_1(t) \int_{t1}^{t2} x_2(t)}} \tag{2.25}$$

$$IACC_{t_1,t_2} = \underset{\tau}{max}|IACF(\tau)| \tag{2.26}$$

Where x_1 and x_2 are the signals at the left and right ears respectively.

The IACC is commonly measured using a dummy head system or can be derived from 3D soundfield measurements with the application of suitable HRTF responses.

2.7 Stage acoustic parameters

Acoustic parameters have also been derived to assess the acoustic conditions on stage. These parameters are related to specific concerns of musicians as discussed previously. For example, a musicians impression of being supported by the hall tends to become stronger when they perform in the presence of higher amplitude early reflections. The following sections will discuss the acoustic parameters related to stage acoustic conditions.

2.7.1 Stage clarity

Stage Clarity, CS, shown by equation (2.27) is a parameter which is functionally identical to audience clarity C_{80} with the exception that the measurement takes place one metre away from the sound source.

$$CS = 10log_{10}\left(\frac{\int_{0ms}^{80ms} p^2(t)dt}{\int_{80ms}^{\infty ms} p^2(t)dt}\right) \tag{2.27}$$

This parameter is used typically for representing the ease an ensemble can play together due to the importance of early reflections for communication with other musicians (Gade, 1989).

2.7.2 Ensemble

Early Ensemble Level (EEL) is another parameter which aims to evaluate how easy it is for musicians to play together. EEL was defined also by Gade (1989) as the ratio of energy contained in the early reflections and direct sound to the energy of the direct sound only (measured at the source position). Higher values of EEL, shown in equation (2.28), indicates the presence of strong early reflections arriving between 0-80ms which is thought to correspond to an increased ability to hear other musicians.

$$EEL = 10log_{10}\left(\frac{\int_{0ms}^{80ms} p^2(t)dt}{\int_{0ms}^{10ms} p^2(t)dt}\right) \tag{2.28}$$

EEL, like other stage acoustic parameters uses the direct sound, measured at the source position, as a reference value as it was recognised by Gade that the ability to hear each other will also be based on the level of direct sound in addition to early reflections. When performing as an orchestra, the direct sound is often attenuated by the presence of other musicians. In practice, the time region 0-10ms chosen for the direct sound also includes the floor reflection. Two microphones are required to measure this parameter, one close to the sound source, which measures the direct sound, and one at the position of interest, which measures the resulting acoustic response.

2.7.3 Objective Support

Objective Support (ST) is one of the most widely used parameters used to characterise stage acoustic conditions and was one of the first to appear in international standards (International Organisation of Standardisation, 2009). It is a quantity defined by Gade (1982, 1989) which relates fractions of energy within certain time intervals to the direct sound in impulse responses recorded on orchestra platforms. It was devised as an objective predictor of how well early reflections assisted the performers own efforts, this attribute described as *'support'*.

ST_{early} (ST1) (equation (2.29)) provides an energy comparison between the direct sound (as reference) and reflections arriving in the time period of 20ms to 100ms as shown in equation (2.29) The energy in each particular time interval is obtained using backwards integration. ST_{early} has an estimated Just Noticeable Difference (JND) of approximately 2dB (Hak et al., 2012) although this has yet to be confirmed through study. The assessment window for early ST_{early} will include reflections that have travelled a distance of between approximately 6.8m and 34.3m. The lower integration time requires that the measurement position must be at least 3.4m away from the nearest reflecting surface (excluding the floor) to be included within the analysis.

ST_{late} (ST2) (equation (2.30)) provides a similar comparison but over a longer (and later) time region of 100ms to 1000ms as in equation (2.30) and has been associated with the musician's perception of reverberance (International Organisation of Standardisation, 2009). Furthermore, an additional parameter ST_{Total} (ST3) (equation (2.31)) compares the energy between 20-1000ms with the direct sound.

All objective support parameters are measured using an omnidirectional microphone and omni-directional loudspeaker positioned 1m apart and at a height of approximately 1.5m above the stage floor. A single figure value of the objective support parameters is obtained by averaging between 250Hz and 2kHz and by averaging measurements over at least three positions on stage (Lautenbach and Vercammen, 2013).

$$ST_{early} = 10log_{10}\left(\frac{\int_{20ms}^{100ms} p^2(t)dt}{\int_{0ms}^{10ms} p^2(t)dt}\right) \tag{2.29}$$

$$ST_{late} = 10log_{10}\left(\frac{\int_{100ms}^{1000ms} p^2(t)dt}{\int_{0ms}^{10ms} p^2(t)dt}\right) \tag{2.30}$$

$$ST_{Total} = 10log_{10}\left(\frac{\int_{20ms}^{1000ms} p^2(t)dt}{\int_{0ms}^{10ms} p^2(t)dt}\right) \tag{2.31}$$

Giovannini and Arianna (2010) have collated results of numerous stage acoustic studies to compare how the results of objective support vary across different studies. These values are reproduced below in Table 2.1 with some additional studies to show typical values obtained. The widest range of values, reported in ISO3382-1(2009), is -24dB to -8dB for ST_{early} and -24dB to -10dB for ST_{late}.

Study	Number of Halls	ST_{early}(dB)	ST_{late}(dB)
ISO 3382-1 (2009)	-	-24 to -8	-24 to -10
Gade (1989)	19	-16.6 to -10.9	-
Chiang et al. (2003)	5	-15.9 to -9.0	-17.3 to -10.9
Jeon and Barron (2005)	1	-24.0 to -15.0	-
Dammerud and Baron (2007)	4	-17.1 to -12.5	-17.0 to -14.6
Giovannini and Arianna (2010)	4	-16.5 to -11.2	-17.5 to -11.4
Guthrie (2014)	10	-18.7 to -10.3	-19.3 to -9.7

TABLE 2.1: *Summary of stage support measures obtained by other studies. Adapted from (Giovannini and Arianna, 2010)*

Objective support is currently the only stage acoustic parameter that has been included in an international standard (International Organisation of Standardisation, 2009). However, in recent years, there has been increasing discussion regarding the objective and subjective relevance of objective support resulting in the development of numerous adaptions of this parameter, some of which are described below.

Wenmaekers et al. (2012) have proposed an adaption to both early and late objective support ($ST_{early,d}$ and $ST_{late,d}$) which makes the upper integration time variable depending on the source-receiver distance. Furthermore, the reference window (which includes the direct sound and floor reflection) is reduced to 10ms in length and is measured only one position away from any reflecting side walls or surfaces.

$$ST_{early,d} = 10log_{10}\left(\frac{\int_{10ms}^{103ms-delay} p_d^2(t)dt}{\int_{0ms}^{10ms} p_{1m}^2(t)dt}\right) \tag{2.32}$$

$$ST_{late,d} = 10log_{10}\left(\frac{\int_{103ms-delay}^{\infty ms} p_d^2(t)dt}{\int_{0ms}^{10ms} p_{1m}^2(t)dt}\right) \tag{2.33}$$

2.7.4 Strength

Strength, G, is also defined as the sound pressure level caused by an omnidirectional sound source on stage measured at a listener position in the hall, with reference to the sound pressure level at 10m distance from the same sound source measured in free-field conditions (International Organisation of Standardisation, 2009). It is related to the judgement of overall loudness,

typically from an audience perspective. As is evident in equation (2.34), Strength assesses the energy of the whole impulse response including the direct sound. In recent years, it has been proposed as a candidate for a stage acoustic parameter (Dammerud, 2009).

$$G = 10log_{10}\left(\frac{\int_0^\infty p^2(t)dt}{\int_0^\infty p_{10}^2(t)dt}\right) \tag{2.34}$$

Dammerud (2009) proposed an alternative measure based on Strength which correlated well with support. The measure $G_{Support}$ and G_{late} are equivalent to a time-windowed version of Strength which uses energy between 0-80ms and $80 - \infty$ms respectively. One of the advantages of this approach is that the region of interest of the impulse response is compared to the level of the direct sound only, while in Support, the reference term includes both the direct sound and the floor reflection and possibly any number of additional early reflections from music stands etc. This requires (sometimes prohibitive) arrangements to be made to ensure a suitable reflection free zone for each measurement which may not always be truly representative of the conditions experienced by the performer.

$$G_{Support} = 10log_{10}\left(\frac{\int_0^{80ms} p^2(t)dt}{\int_0^\infty p_{10}^2(t)dt}\right) \tag{2.35}$$

$$G_{late} = 10log_{10}\left(\frac{\int_{80ms}^\infty p^2(t)dt}{\int_0^\infty p_{10}^2(t)dt}\right) \tag{2.36}$$

One of the disadvantages of this parameter is that it relies on the prior measurement of a sound source at a 10m distance under free-field conditions. This ideally requires the availability of an anechoic chamber and relies on the equipment settings remaining constant between the free-field measurement and throughout the venue measurement. There has recently been an attempt to determine a method of obtaining the reference value for Strength in-situ by obtaining a source impulse response free of reflections by truncation before the arrival of any reflections (Linfors et al., 2013).

2.7.5 LQ_{7-40}

Another proposed parameter primarily developed for assessing acoustic conditions at the conductors position (van den Braak and van Luxemburg, 2008) as defined in equation (2.37).

$$LQ_{7-40} = 10log_{10}\left[\frac{\int_7^{40} p^2(t)dt}{\int_{40}^\infty p^2(t)dt}\right] \tag{2.37}$$

This measure is obtained with separate source and receiver positions on stage. As early reflections from the stage enclosure are known to be important for musicians playing together (40-80ms), this parameter encompasses their influence in contrast to the late reverberant sound ($80ms - \infty$

ms). While it bears some similarity to objective support, there is little subjective relevance to the situation that is being investigated in this study.

2.7.6 Running Reverberation

A further parameter known as *'Running Reverberation'* (Griesinger, 1995) or RR160 (equation (2.38)) assesses a ratio of energy in adjacent 160ms time intervals.

$$RR160 = 10log_{10}\left[\frac{\int_{160}^{320} p^2(t)dt}{\int_{0}^{160} p^2(t)dt}\right] \tag{2.38}$$

This parameter assumes that, while playing music, it is not possible for the musician to hear the details of the impulse response of the room, only aspects of reverberant level are assumed to be important cues. It is proposed as a parameter that correlates with the subjective impression of the level of reverberation during musical performance although few studies have been conducted to validate this. These assumptions do not correlate well with other studies which have found aspects of the early reflections to be important factors in the perception of stage acoustics for a performer; if they are consciously aware of them or not.

2.7.7 Geometrical parameters

Stage acoustics research is primarily concerned with producing optimal conditions for performers by relating their preference patterns to objective descriptions of the acoustic conditions measured on stage. Ultimately, however it is required that a stage enclosure be built which produces the desired acoustic characteristics and so it is useful to determine how the architectural design can affect the acoustic parameters and therefore the predicted preference of the musicians. Hyung et al. (2010) consider the effects of stage volume and absorption on acoustic parameters. This study made use of acoustic models and found that the stage volume ratio (i.e. the volume of the stage divided by the total volume of the hall) could accurately predict the level of early objective support measured on stage. It was found that decreasing the stage volume ratio resulted in a higher level of ST_{early}.

In another recent study regarding stage acoustics for symphony orchestras, Dammerud (2009, 2012) recognised that the direction and time of arrival of early reflections on stage could influence subjective aspects of the acoustic conditions for musicians. It was found that a narrow and high stage enclosure provided beneficial reflections to the orchestra to provide improved communication with other musicians and also acoustically coupled the main auditorium with the stage allowing musicians to hear the auditorium acoustics clearly. The proposed geometrical parameters are shown in Figure 2.23.

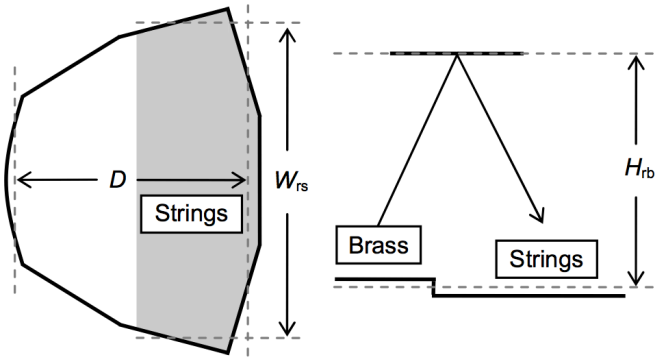


FIGURE 2.23: *Diagram showing geometrical parameters proposed by Dammerud. A stage is shown in plan (left) and section (right). W_{rs} is the average distance between surfaces occupied by string instruments. H_{rb} is the average height between the stage floor and ceiling between brass and string sections. D is the distance between the back edge of the stage and the average stage front. Image reproduced from (Dammerud, 2011)*

For orchestral music, reflections from the stage ceiling are of great importance to ensemble communication given that the direct sound path may be severely attenuated due to the presence of other musicians. A reflecting ceiling allows the other musicians to hear the rest of the ensemble. The height and angle of the ceiling (or ceiling reflectors) will determine how well distributed the sound is to the ensemble but also the time delay of the reflected sound. A very high ceiling may cause sound to be excessively delayed which may exacerbate timing errors for the musicians. This suggests that there is a basic ratio of stage dimensions required to ensure effective communication and ability to hear the auditorium acoustics.

Dammerud (2012) proposes the use of the following architectural parameters. W_{rs} is the average distance between lateral surfaces positioned at the front half of the stage (i.e. where the strings are normally seated). H_{rb} is the average height of the ceiling from the floor between the brass and string sections. D is the distance between the back of the stage and the average stage front. He also suggests the use of various ratios of these parameters.

The acoustic conditions and requirements for a soloist are likely to differ greatly from those of a symphony orchestra and so the measures proposed by Dammerud (2012) may not be directly applicable in the same manner. However, this work does highlight that the direction and time of arrival of early reflections are of importance to performing musicians. In the context of a soloist, there may be a basic ratio of dimensions which promotes a feeling of envelopment for the musician or is perceived to enhance the bloom of the sound into the auditorium. For a soloist, their position or orientation on stage may change the spatial and temporal distribution of early reflections. It is therefore speculated that the basic geometry of the space and the musician's position within it could change the perceived impression of acoustics.

2.7.8 Evaluation of stage acoustic parameters

The preceding section summarised the most recent set of acoustic parameters used to assess concert hall acoustics. The majority of these parameters follow a similar paradigm to auditorium acoustic parameters such as C_{80} by comparing the relative level of energy in discrete time windows. Typically, these parameters compare the relative level of early reflections with the direct sound. Objective support is the most widely used stage acoustic parameter in concert hall stage design, having been included in ISO 3382:1 (International Organisation of Standardisation, 2009). It is used in conjunction with reverberation time parameters to assess or design concert hall stages. From interviews with musicians undertaken in previous research (Dammerud, 2009, Gade, 1982), it is clear that a musician's impression of stage acoustics is multidimensional, where subjective aspects such as dynamics, projection, timbre, bloom etc. can influence a musician's preference towards a particular space. It is proposed that in order to design successful stage acoustics, additional parameters should be developed which take these additional subjective aspects into account.

As with other stage acoustic parameters described in this chapter, ST_{early} is measured with an omnidirectional microphone and loudspeaker. The parameter is measured such that the directivity or orientation of the sound source does not contribute to any variation observed between measurements in different venues. This is to ensure ease of standardisation between measurements made by different practitioners and is common practice for other known acoustic parameters such as C_{80} and T_{30}. When a stage enclosure is measured with an omnidirectional sound source, reflections are excited equally in all directions and at all frequencies. However, when the stage is excited with the sound from a musical instrument, which has a unique directional radiation pattern, reflections from particular directions will be excited more strongly than others, with different frequency-dependent magnitudes. It is proposed that an impulse response measured with an omnidirectional sound source does not sufficiently represent the acoustic response experienced by a musician.

Similarly, for reasons of standardisation, stage acoustic parameters are measured with an omnidirectional microphone. Measuring in this way ensures that the measurement is independent of any spatial variation of early reflections. As discussed previously, the spatial distribution of early reflections could determine how audible some reflections are to different musicians (due to factors such as spatial unmasking). In addition, the spatial distribution of early reflections may produce different subjective effects for the musician. Recent work by authors such as McCarthy et al. (2008) and Guthrie (2014) have developed spatial equivalents of stage acoustic parameters by using directional microphones. Dammerud (2009) has suggested a similar approach which makes use of various stage dimensions which may encompass the spatial and temporal distribution of early reflections.

It is also the case that the majority of current stage acoustic parameters are based upon assessing the overall energy in finite time-windows in relation to various reference energies (such as the direct sound). The main difference between the current proposed stage acoustic parameters is that they utilise different length time-windows. In doing so, these parameters are independent of any

temporal variation which causes a similar issue with the spatial independence of the measurement. Two halls with very different temporal reflection sequences, but with identical reflection energy, would produce the same value of objective early support. This is demonstrated in Figure 2.24 which shows two different reflection sequences with identical levels of ST_{early}. Similarly, assessing the early energy in this way does not distinguish between specular or diffuse reflections which are known to produce very different perceptual effects for both audience members and performers (Bermond and Davies, 2001).

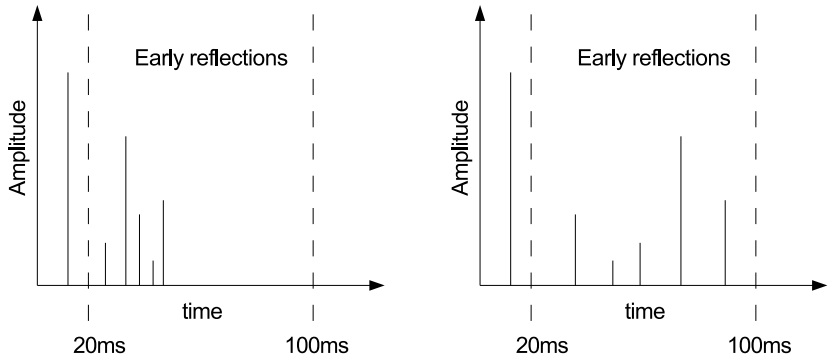


FIGURE 2.24: *Two impulse responses with significantly different early reflection patterns will result in the same level of Objective support due to the energy between 20ms and 100ms being equivalent.*

The upper integration time for ST_{early} (100ms) was originally chosen as it corresponds to the length of a short musical tone and also thought to equal to the integration time for the ear (Gade, 1982, 1989) although the appropriateness of this value has never been investigated fully. The integration times for ST_{early} might be considered as a way of segregating the effects of early reflections from the effects of diffuse reverberation. This approach reflects other standardised acoustic parameters such as C_{80} and D_{50} which separate early reflections from diffuse reverberation at 80ms and 50ms respectively. These integration times are generally dependent on the nature of the sound source i.e. music or speech. It has been demonstrated that for audience members, the perceived mixing time is affected by the physical volume of the auditorium and is a highly subjective value. It is possible that this is also the case from the performer's perspective and that a fixed integration time is an oversimplification.

The choice of 100ms ensures that reflections from surfaces within a radius of approximately 17m from the performer are included in the analysis. This guarantees that reflections from the stage enclosure are included in the analysis however reflections from surfaces such as the audience-rear wall may not be included depending on the size and shape of the venue. It is hypothesised that reflections from the audience-rear wall could influence a musician's impression of projecting into the audience (Kahle, 2013).

As the integration times determine the temporal limit of a rectangular assessment window, there is a distinct possibility that high amplitude reflections at, for example, 101ms will not be included in the measurement but could feasibly have an audible impact for the musician. It is however uncertain if reflections after this point influence the impression of support or if they influence some other subjective aspect of the acoustic response.

The lower integration time of 20ms determines a minimum distance from reflecting surfaces for measurement equipment. Reflecting surfaces that are within this distance will not be included in the measurement unless they are within the assessment window of the direct sound. It is possible, however, that musicians will perform at positions that are within this minimum distance (Lautenbach and Vercammen, 2013). A simple example would be when a musician is using a music stand or standing next to an accompanying piano. It is equally likely that for practical reasons a musician in an orchestra may be required to be located near a wall. The lower integration time has been questioned by authors such as Wenmaekers and Hak (2013) who proposed an adjusted version of ST_{early} which includes reflections occurring directly after the direct sound.

In summary, it is evident that there are many subjective dimensions that drive musician preference towards particular stage acoustic conditions. Yet, current stage acoustic parameters are independent of many physical variables that could influence a musician's impression of a venue. The spatial and temporal distribution of early reflections could vary with aspects related to the performer, for example, the location of the performer, the directivity of the instrument or the orientation of the performer. Similarly, variations could be caused by aspects of the venue, such as the geometry of the stage and the materials used in the stage enclosure construction.

While the overall level of early reflections is likely to be an important factor for feelings of support, it is feasible that the spatial and temporal distribution of early reflections influences, what Cabrera et al. (2010) refers to as, the *'quality of support'* . This may account for additional dimensions of musician preference towards a concert hall acoustic. Therefore, it is proposed that additional stage acoustic parameters are required to characterise this.

However, further research is required to determine if variations in the temporal or spatial distribution of early reflections are audible to a performing musician, given the dominance of the direct sound from their instrument. In addition, it is important to determine, for example, if a spatially diffuse impulse response improves the perceived control of dynamics for a musician. Furthermore, current literature does not appear to include how the stage acoustic conditions vary on stage due to factors such as instrument directivity or musician position. Therefore, it is proposed that further observations are required to characterise the acoustic conditions found on stage. As the ultimate objective of stage acoustic research is to create a successful stage acoustic environment for the musician, further work is required to understand if these factors are drivers of musician preference.

2.8 Summary and discussion

This chapter reviewed the elementary aspects of sound generation, propagation, sensation and perception; including basic approaches to auditorium design. It also reviewed current theories regarding the aspects of acoustic response perceived by performing musicians in addition to various preferences towards particular acoustic conditions. It is evident that, while the study of auditorium acoustics is highly developed, the effect of stage acoustic condition on a performer are comparatively less well explored. There is a growing body of evidence that suggests that

musicians are sensitive to the acoustic conditions they experience on stage and adjust their technique accordingly to fit their performance (Brereton, 2014, Kato et al., 2008). This includes aspects such as intonation, dynamics, note length and tempo (Kato et al., 2008).

Many musicians appear to adjust their performance technique subconsciously whereas for others this action is considered and deliberate, based on careful attention to the acoustic response of the space. It is common for musicians to perceive the venue as being an extension of their musical instrument (Ueno and Tachibana, 2005). This process can be influenced by the experience level of the musician, musical genre and instrument family. Given that the performance is influenced considerably by the acoustics heard on stage, it follows that in order for a musician to perform well for an audience, the acoustic conditions on stage must assist rather than hinder the performer.

Stage acoustic conditions are currently assessed using the objective support parameters, ST_{early} and ST_{late} which are measured with an omnidirectional microphone and loudspeaker. These parameters quantify the ratio of energy within specific time windows, to the energy in the direct sound and are related to the impression of how much effort the musician has to exert to get a sufficient response from the hall.

It was discussed that the early reflections reaching the musician on stage could vary in their spatial or temporal distribution, independently of the level of ST_{early}. This could occur due to musician location on stage, the directivity of the instrument and the geometry/construction of the stage enclosure. It is feasible that these variations could influence the musician's subjective impression of the venue acoustics (as has been demonstrated for audience members). However the close proximity of the musical instrument, in addition to the stressors of performance, may mask these effects. Therefore it it is unclear if the early reflections have a similar effect from a musician's perspective. If the temporal or spatial distribution of early reflections are found to have a significant influence on the musician's impression of the venue then additional acoustic parameters may be required to assess concert hall stages.

It is proposed that a detailed investigation of stage acoustics in performance spaces is required to determine how aspects of the acoustic response (in addition to ST_{early}) vary under different circumstances. For instance, this may include variables such as sound source directivity, measurement location and stage geometry. The survey should attempt to determine how spatial and temporal aspects of the acoustic response vary in relation to these variables. It is also proposed that a series of listening tests be performed with musician test subjects to determine if these physical variations are audible during a performance and what subjective effect they have on a musician's performance.

Chapter 3

Stage acoustic measurements

In Chapter 2 it was discussed how, for an audience member, the spatial and temporal distribution of early reflections could greatly influence the subjective impression of the sound of a performance; including aspects of timbre and spaciousness etc. Accordingly, there are numerous acoustic parameters which are used to predict the intensity of these subjective attributes.

It was speculated that the distribution of early reflections could potentially influence a musician's impression of the venue. However, it was found that current stage acoustic parameters are independent of the temporal or spatial distribution of early reflections, mainly due to the transducers used to excite and capture the stage acoustic response. Consequently, there is a possibility that two halls with the same level of ST_{early} could be perceived differently by musicians due to differences in spatial or temporal distributions of early reflections.

It was acknowledged that the proximity of the instrument and the pressures of performing could potentially reduce the magnitude of these effects, however there was some initial evidence to suggest these may be salient factors (Miranda Jofre et al., 2013). If shown to be the case, additional acoustic parameters may be required in order to correctly assess stage acoustic designs. These parameters would need to be sensitive to the temporal or spatial distribution of early reflections.

In order to determine the subjective impact of different acoustic conditions, it was considered that interactive listening tests with performing musicians should be conducted. The auralisation system would use impulse response data measured in existing concert halls to allow musicians to instantaneously compare different acoustic conditions. By allowing a musician to play in different virtual acoustic conditions, it would be possible to measure their subjective impression in a controlled laboratory environment.

To accommodate this, it was necessary to conduct acoustic surveys of different concert hall stages. Additionally, it was considered important to perform a detailed analysis of the measured impulse response data to determine how different physical factors could affect the temporal and spatial distribution of early reflections. A comparison of the subjective responses from musicians with the objective data measured in each hall would help to determine if the distribution of reflections influenced the musician's impression of the venue's acoustics.

This chapter will review the background theory related to the measurement of the acoustic conditions on concert hall stage enclosures. It will go on to propose a measurement procedure that allows the temporal and spatial distribution of early reflections to be analysed and effectively auralised to allow objective and subjective testing to take place. It will also describe how this procedure was applied in eight different performance spaces.

3.1 Objectives

The spatial and temporal distribution of early reflections on stage can vary in response to venue-related and musician-related variables. Specifically, the material construction and physical geometry of reflecting surfaces of the stage enclosure can vary the direction, time of arrival, frequency content and temporal diffuseness of each reflection. Additionally, the directivity of the instrument, its orientation and position on stage can also vary the distribution of early reflections.

The measurement technique should therefore be capable of capturing the temporal and spatial distribution of early reflections and be able to vary physical variables such as directivity, orientation or stage position. In addition, the measurement technique should allow the use of traditional (or slightly adapted) acoustic parameters to accurately assess the acoustic conditions in each venue.

The measurement system should be capable of capturing the acoustic response with sufficient detail so that early reflections can be localised spatially and temporally. The captured data should also be required to have minimal distortion or artefacts so that resulting auralisations will be as close as possible to the original soundfield. Finally, the measurement system should be highly portable, easily assembled and straightforward to implement so that surveys could be performed in a efficient manner.

From the review in Chapter 2, it is clear that existing measurement techniques cannot accommodate many of these objectives therefore a new approach is required. The remainder of this chapter will describe how high quality impulse responses were captured that met the above specifications.

3.2 Impulse response measurement

In Chapter 2, it was discussed how the direct sound, generated by the instrument, propagates around a room and is modified as it interacts with reflecting surfaces. Consider a recording of the direct sound made at a location in the room, the recording would consist of both the direct sound and numerous, attenuated and delayed, reflections of the sound. As the reflected sound has a significant impact on the subjective impression of the space, it is of interest to study these reflections directly.

The acoustic impulse response is considered one of the most effective methods of studying the acoustic conditions of a performance space. The majority of acoustic parameters are derived from impulse responses obtained via direct measurement in the venue or by using acoustic models.

The sound of a musician playing in a concert hall is widely considered to be the result of the direct sound being processed by a Linear, Time-Invariant (LTI) system; which in this case is the room itself. The impulse response describes the transfer function of the LTI system. The impulse response can therefore be used to predict the output of the system to a known input.

In systems theory, an LTI system can be broadly defined as one which does not change over time and fulfils the superposition, causality and linearity properties (Patynen, 2007). This type of system can be characterised by its reaction to an impulse, presented at its input (Kuttruff, 1979). An impulse in this case refers to a Dirac delta function ($\delta(t)$) which consists of an infinitely high amplitude, infinitely brief pulse centred around the time origin (Smith, accessed (17/11/14), as defined in equation (3.1) below:

$$\delta(t) = \begin{cases} +\infty & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases} \tag{3.1}$$

The impulse is presented to the input of the system $x(t)$ and the resulting system response is recorded at the output $y(t)$ as illustrated in Figure 3.1.

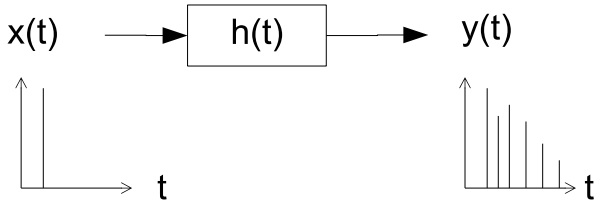


FIGURE 3.1: *Diagram showing the result of processing an impulsive signal, $x(t)$ with an LTI system, $h(t)$ to produce the impulse response $y(t)$*

Mathematically, the output of an LTI system, $y(t)$ can be considered as a convolution of the source signal $x(t)$ with the impulse response $h(t)$, as shown in equation (3.2), where '$*$' denotes convolution and $w(t)$ represents measurement noise.

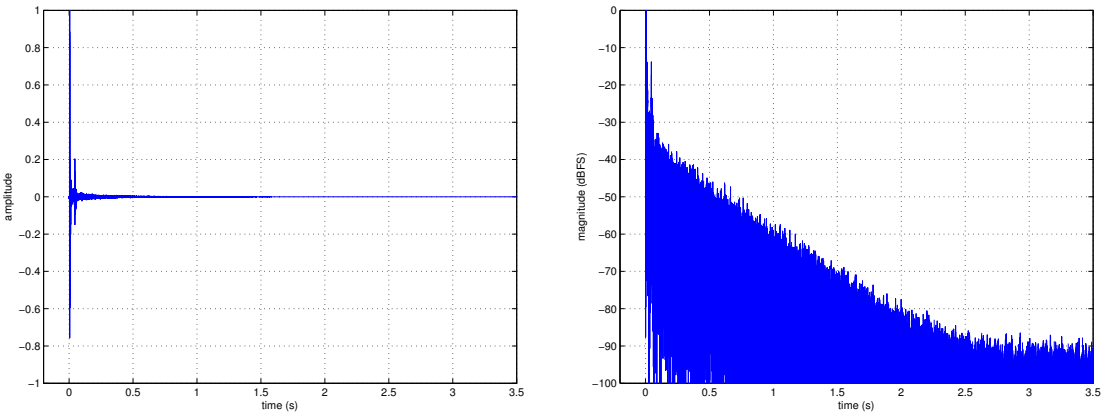
$$y(t) = x(t) * h(t) + w(t) \tag{3.2}$$

In theory, a room is widely considered to be an LTI system and therefore the effect of the concert hall on the direct sound can be evaluated by measuring with an appropriate sound source and receiver (Farina, 2000). However, in practice this is merely an accepted approximation as the acoustic conditions can vary over time due to aspects such as temperature, humidity and movement of air within the space. It does however mean that aspects of systems theory can be used to measure and assess the effect of the space on a sound source. In many cases, the audible effect of these environmental variations are minimal and are often considered acceptable.

In the context of room acoustics it is clearly not practical to produce or measure an infinitely loud and infinitely brief impulse as sound energy, therefore in practice an impulsive noise source, such as a starter pistol, balloon burst or electrical spark, is considered to generate a sufficient

approximation (Rossing, 2007). The room impulse response represents the transfer function between the source and receiver and as such is sensitive to the characteristics of the room (such as room volume, geometry and material construction), transducer attributes and relative positions of source and receiver.

To illustrate this, an example impulse response can be seen in Figure 3.2 in both linear and decibel scales. This impulse response was measured on a stage with the source and receiver positioned close together. In this example, it is possible to see the early reflections occurring shortly after the direct sound followed by the reverberant decay.



(a) Measured impulse response

(b) impulse response shown in dB scale (re: Full scale)

FIGURE 3.2: *Example of a measured impulse response shown in a linear (3.2(a)) and dB scale (3.2(b)). The direct sound, early reflections and reverberation can clearly be seen.*

Impulsive noise sources, however, can exhibit unpredictable frequency and directional responses (Abel et al., 2010) therefore in some instances, it is often necessary to average the response over a number of repetitions in order to account for this variability (International Organisation of Standardisation, 2009). Furthermore, the sound pressure level produced by these sound sources is often very high which can easily overload a microphone recording the impulse response. This overloading can produce distortion on the recorded signal rendering it unreliable for analysis. In order to record signals of this kind, the recording gain of the microphone is reduced. This, however, can cause the signal to noise ratio of the measurement to be reduced which can limit the quality of analysis (Abel et al., 2010). In stage acoustic research, the source and receiver are positioned at close proximity, exacerbating the issues of recording gain and signal to noise ratio.

It is now more common to measure a room impulse response by exciting the room with a known signal played through a loudspeaker and measuring the resulting signal at the desired location. The impulse response of the room can be recovered via deconvolution (correlation) of the input and output signals, as shown in Figure 3.3. Exciting the space with a loudspeaker ensures that the sound source has a known, fixed directional and frequency response which increases reliability. In addition, the use of steady state signals reduces the likelihood of transducer overloading, allowing microphone and sound source gains to be set much higher than with impulsive sources.

Consequently, this approach tends to result in an increased signal to noise ratio of the measured data.

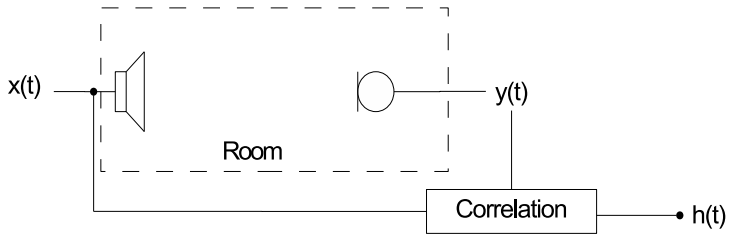


FIGURE 3.3: *Systems diagram showing a room excited with an input signal, $x(t)$, via a loudspeaker and the response recorded with a microphone $h(t)$ By correlating the input and output signals, the transfer function, $h(t)$, can be derived.*

The Maximum Length Sequence (MLS) is an example of an excitation signal frequently used in room acoustic measurements and loudspeaker testing (Farina, 2000). The signal is a pseudo-random binary sequence with white noise spectrum characteristics. These signals are generated using a linear feedback shift register (a specific circular arrangement of delays and eXclusive-OR gates) (Holter et al., 2009). An autocorrelation of this signal produces a periodic impulse occurring at intervals dependent on the length of the MLS signal. The impulse response of a room may be measured by energising the room with the MLS signal and recording at an appropriate location. The impulse response is then recovered by cross-correlation of the original MLS signal with the recorded signal. Due to the periodicity of the MLS signal, the length (dictated by the signal order) of the signal must be longer than the reverberation time of the room to avoid time-aliasing. The number of samples found in one period of the MLS signal can be calculated using equation (3.3) where m is the signal order which refers to the number of shift registers in the signal generator (Vikko and Tomi, 2008).

$$L = 2^m - 1 \tag{3.3}$$

Another commonly used excitation signal is a logarithmically swept sine wave of increasing frequency. This technique attempts to spread the broadband frequency content of an impulse over a longer time which further reduces the likelihood of the transducers overloading. This type of signal can be generated using equation (3.4) where ω_1 and ω_2 denote the start and end frequencies respectively and T represents the total length of the signal (Farina, 2000, Tervo, 2012).

$$s(t) = \sin\left(\frac{\omega_1 T}{\ln\{\omega_1/\omega_1\}}\left(e^{\left(\frac{t}{T}\ln\{\omega_2/\omega_1\}\right)}\right) - 1\right) \tag{3.4}$$

The impulse response, $h(t)$, can be de-convolved from the recorded signal, $s(t)$ via correlation with the original signal played into the room, $x(t)$. This can be achieved by division in the

frequency domain using equation (3.5).

$$h(t) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\{x(t)\}}{\mathcal{F}\{s(t)\}}\right) \tag{3.5}$$

where $x(t)$ denotes the recorded sine sweep, $s(t)$ denotes the original sweep signal, $\mathcal{F}$ denotes a Fourier Transform and $\mathcal{F}^{-1}$ denotes the Inverse Fourier Transform. It is also convenient to note that a convolution of two signals in the time domain is equivalent to a multiplication in the frequency domain.

Both of the aforementioned measurement techniques are sensitive to the characteristics of the source and receiver. For example, when the signal is played back at a high level, the loudspeaker can introduce non-linear harmonic distortion onto the signal which will eventually appear on the resultant impulse response as audible distortion (Farina, 2000). This harmonic distortion is largely due to the phase and frequency response of the loudspeaker.

This can be observed by measuring a swept sine wave, played at high level, from a loudspeaker as shown in Figure 3.4. This figure shows a time-frequency plot of a logarithmic sine sweep played through a loudspeaker in a performance space. The sweep signal can be seen as the highest amplitude curve (red) ending at $22050Hz$ at a time of $10s$. The non-linear harmonic distortion can be seen as lower amplitude sweep signals occurring alongside the main sweep signal. This distortion can cause issues for both auralisation and analysis of the acoustic impulse responses.

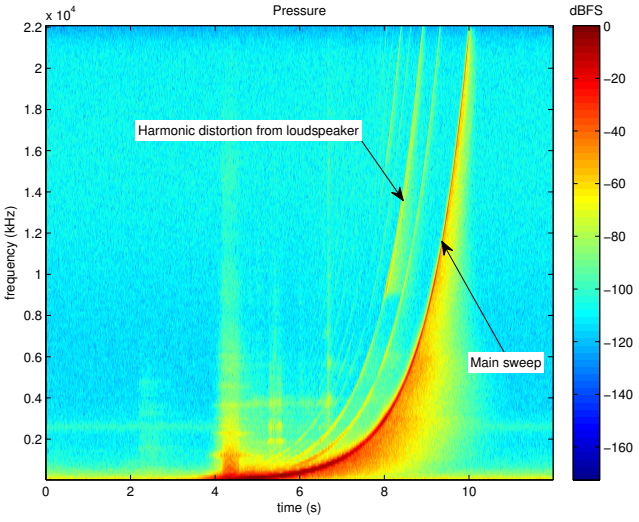


FIGURE 3.4: *Time-frequency plot of a measured sweep showing the increasing frequency of the sine sweep and the resultant harmonic distortion shown as higher frequency sweeps. The colour axis represents dB relative to Full scale (dBFS)*

A useful property of the swept-sine signal is that the phase difference between points on the sweep with integer multiple instantaneous frequency are constant. By inverse filtering the measured response, any non-linear contributions from the measurement system will be displaced backwards in time relative to the linear impulse response. As the inverse filter is constructed from a time reverse of the excitation signal, the impulse response itself is displaced forward in time to exactly

half way along the resultant signal. This allows the non-linear components to be easily discarded from the resultant impulse response (Farina, 2000).

The inverse filter can be generated in the time domain by time-reversal of the original sweep signal and applying a $-6dB$ per octave envelope. The room impulse response is then recovered via convolution of the inverse signal with the recorded sweep signal (Farina, 2000). Figure 3.5 shows the resultant impulse response when the recorded sweep is convolved with the inverse sweep. The room impulse response occurs exactly half way along the signal at $t = 0$s. The artefacts preceding the room impulse response are caused by the harmonic distortion introduced by the measurement system seen in Figure 3.4. It can be seen that, in Figure 3.5, these artefacts have been displaced in time relative to the main impulse response and so can easily be removed from the signal. Therefore, following this approach allows the signal to be played at a much higher level without the resulting harmonic distortion affecting the results. Ultimately, this results in an impulse response with a much larger signal to noise ratio.

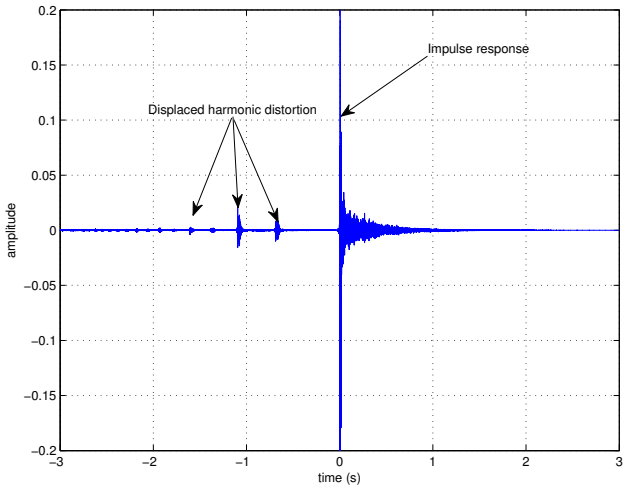


FIGURE 3.5: *Results of a convolution of a recorded sweep with an inverse sweep. The impulse response appears exactly half way along the resultant signal preceded by artefacts caused by non-linear harmonic distortion. These artefacts can be easily discarded from the signal by truncating the beginning of the signal*

There are numerous software packages specifically designed to simultaneously to playback and measure sine sweeps in concert halls including WinMLS (Morset Sound Development, 2012). It is also possible to achieve this by generating the excitation signals manually and using any Digital Audio Workstation (DAW) that is capable of simultaneous multichannel playback and recording. While specialist software such as WinMLS is capable of extracting the impulse response as well as the required acoustic parameters, it is also possible to achieve this using software packages such as MATLAB (Mathworks, 2013) which allows an increased flexibility for creating customised acoustic parameters. In this study, impulse responses were measured with a DAW (Reaper (Reaper, 2013)) and were analysed using MATLAB.

3.3 Spatial impulse response measurement

The room impulse response measurement has been hitherto considered only as a monaural signal, independent of any spatial information. However, as discussed in the measurement objectives, it is necessary for the impulse responses to include spatial data for both auralisation and analysis. Therefore, the following section will review the various available methods of capturing the spatial room impulse response and discuss the appropriateness of each method to research objectives.

3.3.1 Binaural

Binaural sound is based upon the principles of human auditory localisation which were discussed in detail in Chapter 2. It is a frequently used method of capturing spatial room impulse responses for acoustic analysis and auralisation.

For humans, sound is localised using cues such as ITD, ILD and filtering caused by the interaction of sound with the head, shoulders and pinnae. By placing microphone transducers in a listener's ear canals it is possible to encode these cues onto an audio signal. When played back to the listener over headphones, the brain interprets these cues as the sound arriving from a specific location. Measuring a room impulse response in this way ensures that the spatial characteristics of the acoustic response are encoded within the signals via the Head Related Transfer Function (HRTF). An obvious advantage to this approach is that the spatial information is encoded entirely using only two channels of audio data which can be easily stored and played back.

Binaural impulse responses are often measured in concert halls to analyse spatial aspects of the soundfield (such as spaciousness or envelopment or apparent source width) as these aspects are widely associated with the correlation of signals received at the ears. It is also possible to produce high-quality auralisations of concert hall acoustics by processing anechoically recorded sound from a musical instrument with the Binaural impulse response. By playing the resultant signals over headphones, the person listening can perceive highly plausible auralisations of particular scenarios.

High quality binaural recordings are often made using a dummy-head system which consists of the head and torso of a specially designed mannequin. Microphone capsules are placed inside the ear canals, which are usually surrounded by artificial pinnae, emulating a specific HRTF. An example of this type of system is shown in Figure 3.6.

FIGURE 3.6: *Image showing a Brüel Kjær 4128C - Head and Torso Simulator (HATS) (Brüel & Kjær, No date). The dummy head features microphone capsules arranged behind latex pinnae allowing binaural impulse responses to be recorded. This system also features a loudspeaker in the mouth of the dummy in order to measure Oral-Binaural Room Impulse Responses as demonstrated by Cabrera et al. (2013)*

However, if the HRTF is very different from the listener's own, the brain can have difficulty interpreting the frequency response which results in localisation difficulties (Rumsey, 2001). Therefore, binaural auralisations are often rendered with personalised HRTF data which must be measured prior to rendering. Another frequently encountered issue is that when the listener turns their head, the sound source tends to move sympathetically rather than remain fixed. For that reason, many binaural auralisation systems incorporate a head-tracking system which essentially crossfades between numerous binaural impulse responses. This allows the sound source to remain in a fixed position regardless of the orientation of the listener's head.

Binaural recording techniques have been used previously in the field of stage acoustic research to study the acoustic conditions experienced by vocalists. Specifically, a so-called Oral Binaural Impulse Response (OBRIR) is measured using a mannequin system that is equipped with a sound source positioned in the mouth of the dummy.

The OBRIR is representative of the effect of the room on the listener's own voice (Cabrera et al., 2013) and includes the spatial distribution of the acoustic conditions. In addition, these signals can be used to produce interactive auralisations for subjective testing. Cabrera et al. (2010) made use of this technique to determine the effect of orientation of a singer on the acoustic conditions they hear during a performance. In their auralisation system, the sound of a singer's voice was captured with a headset microphone and processed with the OBRIR signals. The processed signals were then rendered back to the singer in real-time using a pair of head-worn ear speakers.

While highly effective and efficient to implement, this approach to auralisation relies on the accuracy of the HRTF for each participant. To be as accurate as possible, individual HRTF will be required to be measured for each participant. In addition, musician test subjects would be required to wear headphones which may be disruptive in listening tests.

3.3.2 Multichannel recording method

An alternative approach to capturing the acoustic conditions of an auditorium is to use an array of directional microphones that are arranged to capture sound arriving from specific directions. By rendering the recorded signals over a corresponding loudspeaker array, it is possible to emulate the physical soundfield that was captured. In addition, it is possible to isolate sound which arrives from specific directions for analysis, a simple example of this being LEF which utilises monopole and dipole microphones to determine the dominance of lateral early reflections.

When used for authoring auralisations, an impulse response is measured for each loudspeaker in the auralisation array. Processing an anechoically recorded signal with these impulse responses and rendering over a loudspeaker array will produce a soundfield in the centre of array with similar spatial characteristics to that of the original soundfield. The arrangement and orientation of microphones in the array is entirely dependent on the geometry of the loudspeaker array.

For loudspeaker-based auralisation, there are two generally accepted measurement approaches. A *spaced* microphone technique consists of spatially separated directional microphones angled in directions that correspond with the position of loudspeakers in the array that is to be used for rendering the soundfield. The signals picked up by these microphones are then rendered directly through the corresponding loudspeaker. For example, the INA-5 (Ideal Cardioid) arrangement uses a spaced array of five cardioid microphones to sample the soundfield in the directions of a 5.1 loudspeaker array (Rumsey and McCormick, 2003), as defined in ITU-R BS.775-3 (International Telecommunications Union (ITU), 2012). The use of a spaced microphone array, ensures that the direction of a sound source is encoded both with amplitude and temporal differences.

A *coincident* microphone technique consists of closely spaced directional microphone capsules, similarly oriented in the direction of loudspeakers in an array. The proximity of these microphones ensures that the angle of arrival is encoded mainly by amplitude differences between each microphone channel (Laitinen, 2014). An example of this for two-channel reproduction is an XY arrangement of a pair of cardioid microphones.

Coincident microphone techniques produce signals with a high degree of coherence which can produce timbral artefacts (due to comb filtering) or spatial artefacts when used in multichannel loudspeaker arrays. In auralisations, this can result in the reverberant sound being perceived as sounding un-natural. Spaced microphone techniques do not experience the same degree of signal coherence (depending on the spacing of the microphones) and so the reproduced soundfield is often perceived as very natural. However, the microphone spacing must be considered carefully such that the temporal and amplitude differences do not create conflicting localisation cues (Laitinen, 2014).

Ueno and Tachibana (2003) previously used a multichannel microphone technique to measure existing performance spaces in order to create interactive auralisations for stage acoustic research. The auralisation system was used extensively to study the effect of various acoustic conditions on soloists and ensembles (Kato et al., 2007, 2008, Ueno and Tachibana, 2003, 2005, 2010, Ueno et al., 2005, 2007). The auralisation system consisted of six loudspeakers positioned along Cartesian axes mounted in an anechoic chamber. The direct sound from the musician was picked

up using a single directional microphone and processed in real-time with six impulse responses (one per loudspeaker channel), the resultant audio was then played back to the musician to generate the virtual acoustic conditions.

The impulse responses for each source position were measured using a single directional microphone and making repeated measurements with the microphone oriented in directions which corresponded to the loudspeaker array. The measurement system is shown in Figure 3.7. The acoustic response was excited using an omnidirectional loudspeaker positioned near the microphone. The microphone was repositioned after each measurement such that the orientation varied but the capsule remained at a fixed position. Therefore, this technique should be considered a coincident recording method.

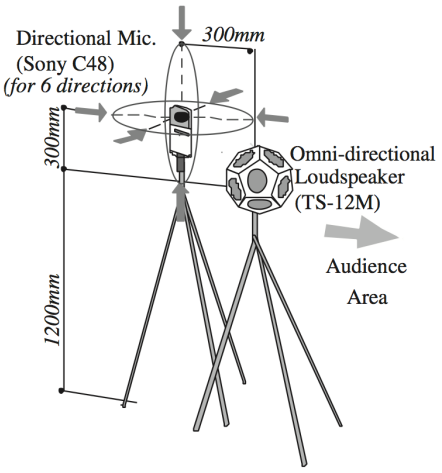


FIGURE 3.7: *Measurement system employed by Ueno and Tachibana (2003) to measure stage acoustic impulse responses for laboratory tests with musicians. Repeated measurements were made with the directional microphone pointing in the direction of each loudspeaker used in the auralisation system*

Woszczyk et al. (2012) produced a similar interactive auralisation system allowing a musician to play in specific halls. The system was developed as a way of providing virtual acoustic support for rehearsal and recording of performances. The auralisation system works in a similar manner to that described by Ueno and Tachibana (2003). The direct sound from the musician is captured and convolved with a multichannel impulse response and rendered in real-time over a 24 channel loudspeaker array which surrounds the performer. Woszczyk employs a multichannel method of recording room impulse responses by using an array of eight spaced microphones to record impulse responses at three separate heights thus producing an impulse response per loudspeaker channel. Four of the microphones are omnidirectional, arranged in a square with 2 metre spacing and the remaining microphones are figure of eight microphones arranged as orthogonal spaced-pairs crossing at 90° angles. An image of the microphone array is shown in Figure 3.8.

FIGURE 3.8: *Microphone array utilised by Woszczyk et al to record spatial impulse responses of different performance spaces. This microphone array consists of eight microphones arranged in identical directions to a loudspeaker array. In this particular measurement system, impulse responses were captured with the microphones at three different heights in order to feed a 24-channel loudspeaker array. Reproduced from (Woszczyk et al., 2012)*

The advantage of using a multichannel recording approach is that high quality directional microphones can be used to capture the 3D soundfield as required. However, in order to produce the correct acoustic cues, the arrangement of microphones must closely match the geometry of the intended loudspeaker array. If the loudspeaker array features a large number of loudspeakers then either a large microphone array or numerous repeated measurements with a single microphone are required.

Furthermore, new measurements are required if the loudspeaker array geometry changes. The use of a large array of microphones may be physically cumbersome to move efficiently during a survey and all microphone would be required to be carefully calibrated prior to recording. Repeated measurements with a single microphone may be feasible however there it is possible that the acoustic conditions could change between measurements, especially if there are a large number of loudspeaker channels in the auralisation system.

3.3.3 Ambisonics

Ambisonics is an approach to the capture, transformation and rendering of 3D soundfields which is based on the decomposition of a 3D soundfield into a finite number of spherical harmonics (Gerzon, 1992). Spherical harmonic decomposition attempts to approximate a complex spherical distribution of sound pressure using a series of spherical harmonic base functions. This is similar to how a complex signal can be represented as a weighted series of sinusoidal functions. A first-order Ambisonic recording encodes the soundfield into four spherical harmonics, which results in the soundfield being described by only four signals. These signals are often referred to as Ambisonic B-format. A soundfield encoded in B-format can be easily transformed (for example, rotated) using simple matrix relations.

Crucially, it is possible to use an array of loudspeakers to reconstruct the spherical harmonic representation of the soundfield by feeding each with a weighted sum of the B-format signals. This produces a faithful 3D reconstruction of the soundfield that is spatially isotropic, i.e. localisation quality is constant for all source directions. Furthermore, by analysing the B-format signals it is possible to evaluate the spatial distribution of sound at any given time instant.

For these reasons (and others), Ambisonics is a popular format for authoring auralisations and for spatial audio analysis and has been used extensively in auditorium acoustic research. The following sections will describe how Ambisonics can be used to capture or produce three dimensional soundfields. In Chapter 4, it will be demonstrated how impulse response captured using B-format can be analysed to determine the spatial distribution of an impulse response. In Chapter 6, the techniques required to reconstruct an Ambisonic soundfield over a loudspeaker array will be discussed.

Spherical harmonics

Spherical harmonics provide a method of approximating complex spherical distributions in a similar manner to how the Fourier Series can be used to approximate a complex audio signal. Spherical harmonics are commonly used in many different scientific fields (including quantum mechanics and gravitational fields) to represent spherical functions (Nachbar et al., 2011).

In the context of 3D soundfield reproduction, it is of interest to know the exact distribution of sound energy over a spherical boundary. In theory, if the exact boundary distribution is known and can be reproduced, the original soundfield can be reproduced. However, current microphone and loudspeaker technology is not capable of capturing or reproducing a continuous spherical distribution of sound.

In theory, it is possible to recreate a spherical distribution of sound by using a large number of loudspeakers, creating secondary wavefronts (Huygens principle) which combine to form the original wavefront. This principle is utilised in practice by a spatial audio technique called Wave Field Synthesis (WFS) (Ortolani, 2014). WFS tends to use a very large number of loudspeakers to ensure that the reconstructed soundfield is accurate at high frequencies. To encode the soundfield with sufficient resolution, a large number of microphone transducers is required pointing in a large number of different directions. This is very difficult to achieve due to the number and spacing of microphone capsules required. Ambisonics operates under similar principles, however it does not attempt to capture (or reconstruct) the exact spherical distribution of sound, rather it approximates this using spherical harmonic functions.

Mathematically, spherical harmonic decomposition of sound is an exterior problem where all the sound sources exist outside the microphone array. The acoustic wave equation can be developed into a Fourier-Bessel series (Ortolani, 2014), by expressing in spherical coordinates, consisting of a product of the spherical harmonic functions $Y_n^m(\theta, \phi)$ with the radial functions $j_n(\omega r)$ which are the spherical Bessel functions of the first kind for incoming energy. The term $A_n^m(\omega)$ represents the coefficients of each spherical harmonic in a similar manner to the Fourier series (Menzies and Al-Akaidi, 2007).

$$p(r, \theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m(\omega) j_n(\omega r) Y_n^m(\theta, \phi) \tag{3.6}$$

Figure 3.9 shows a limited set of spherical harmonics defined up to third order. The spherical harmonics are defined as:

$$Y^{\sigma}_{n,m}(\theta,\phi) = N_{n,m}P_{n,m}(\sin\phi)\cdot\begin{cases}\cos m\theta & \text{for } \phi = 1\\ \sin m\theta & \text{for } \phi = -1\end{cases}$$

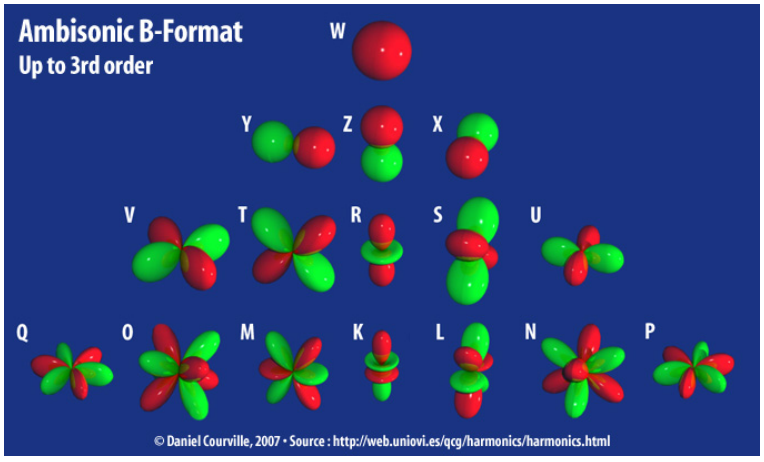


FIGURE 3.9: *Image showing spherical harmonics up to third order. Reproduced from (Courville, 2007)*

Where $P_{n,m}$ are the associated Legendre polynomials and $N_{n,m}$ describes a normalisation factor:

$$N^{\sigma}_{n,m} = \sqrt{2m+1}\sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}} \tag{3.8}$$

A common normalisation factor is the Schmidt Semi-Normalization scheme (SN3D) (Ortolani, 2014) which simplifies the spherical harmonic functions to:

$$Y^{\sigma}_{n,m}(\theta,\phi) = P_{n,m}(\sin\phi)\cdot\begin{cases}\cos m\theta & \text{for } \phi = 1\\ \sin m\theta & \text{for } \phi = -1\end{cases}$$

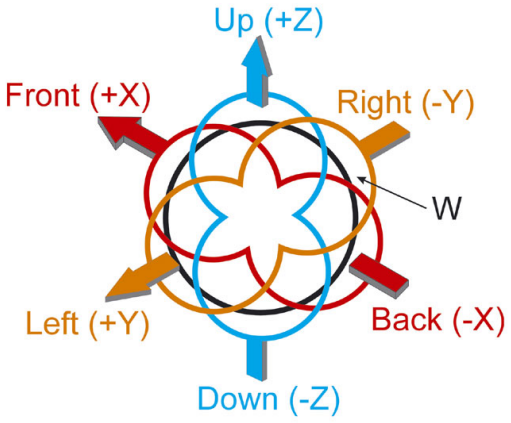
This then ensures the spherical harmonics form an orthonormal basis allowing them to be linearly summed together to describe, in this case, the sound pressure on the surface of a sphere. It is worth noting that there are other normalisation schemes in common usage in Ambisonics, Some of which are less mathematically correct but adhere to certain legacy schemes to improve the overall signal to noise ratio. Examples these are the Normalized set (N3D) and the Furse-Malham set (Fu-Ma). The differences in normalisation scheme often cause the mathematical notation (and hence the gain strategies for various transformations) to vary.

Ambisonic microphone

First-order Ambisonic soundfields are encoded using the zeroth and first-order spherical harmonics. These consist of a monopole and three mutually orthogonal dipole patterns. To encode the soundfield in this way, a coincident arrangement of a single omnidirectional microphone and three dipole microphones is required. In practice, it is not possible to co-locate these microphones precisely and so a near-coincident array of microphones is used.

A typical Ambisonic microphone consists of a tetrahedral arrangement of four, near-coincident sub-cardioid microphones (Rumsey and McCormick, 2003) as shown in Figure 3.10(b). The signals from the capsules are equalised to correct for microphone spacing. The tetrahedral arrangement ensures that the distribution of sound is captured equally in all directions and that the microphones are equally 'non-coincident' which spreads the angular error over the entire soundfield. The signals from these capsules are collectively known as A-format and are usually transformed into Ambisonic B-format by summation of specific combinations of these transducers (Ortolani, 2014), resulting in the polar patterns shown in Figure 3.10(a).

The B-format signals can be thought of as representation of the coefficients $A_n^m(\omega)$ shown in equation (3.10). The omnidirectional channel (commonly referred to as W) measures the pressure component of the soundfield while the dipole channels (referred to as X, Y and Z respectively) measure pressure-gradient signals in three dimensions.



(a) Arrangement of B-format polar patterns (Sound on Sound magazine, 2001)

(b) Tetrahedral arrangement of microphone capsules in an Ambisonic microphone (Studiocare, No date)

FIGURE 3.10: *Tetrahedral arrangement of subcardioid microphone capsules from which it is possible to obtain B-format signals by linear summation*

A scaling factor $\frac{\sqrt{2}}{2}$ (equivalent to a 3dB attenuation) was originally devised to improve the dynamic range of recordings based on the observation that signal levels in the W channel were often much higher than in the X,Y, Z channels (Benjamin et al., 2006). This formulation is included within the Fu-Ma standardisation and so recordings made with the soundfield microphone are compatible with this standard. Care must then be taken when using alternative formulations such as SN3D or N3D.

Virtual microphones

A weighted summation of the B-format signals can be used to derive an electronically steerable virtual microphone (or beam) with a controllable directivity pattern. This operation is shown in equation (3.10) where a virtual microphone pointing towards polar coordinates (θ, ϕ) can be obtained by deriving the weights for each B-format channel (Zolzer, 2011). The directivity of the virtual microphone is controlled using by k where $k = 0$ would produce an omnidirectional microphone and $k = 2$ would produce a dipole microphone polar pattern. In the context of this research, this technique is very useful as it allows the spatial distribution of energy to be analysed in particular directions without having to physically alter the orientation of the microphone.

$$S(t) = \frac{2-k}{2}w(t) + \frac{k}{2\sqrt{(2)}}[\cos(\theta)\cos(\phi)x(t) + \sin(\theta)\cos(\phi)y(t) + \sin(\phi)z(t)] \tag{3.10}$$

The use of virtual microphones is also a key component in rendering B-format over a loudspeaker array. By orienting the virtual microphones in the direction of each loudspeaker and carefully selecting the directivity it is possible to determine the correct signal weights for each loudspeaker feed necessary to reconstruct the recorded soundfield. This will be discussed in further detail in Chapter 6.

It is clear that Ambisonics represents a convenient and elegant approach to soundfield capture, processing, reproduction and analysis. Ambisonic microphones are widely available and the microphone capsules housed in a single piece of equipment making acoustic measurements quick, robust and efficient to conduct. B-format impulse responses can be decoded to many loudspeaker array arrangements providing a great deal of flexibility for subsequent auralisation. Furthermore, B-format signals can be easily analysed to provide detailed information regarding the spatial distribution of the recorded soundfield.

Ambisonics has been widely used to measure spatial room impulse responses for auralisation and detailed acoustic analysis (Farina and Ayalon, 2003). A further advantage of this technique is that the B-format signals can be used to measure existing monaural acoustic parameters (RT_{60}, C_{80} etc) which requires a measurement of sound pressure. In addition, spatial parameters such as LEF can be obtained which require coincident combinations of dipole and omnidirectional microphones. Ambisonics has also featured prominently in recent stage acoustic research with authors such as Brereton et al. (2012a) and Guthrie (2014) using measured Ambisonic room impulse responses to create interactive auralisations over loudspeaker arrays.

Higher Order Ambisonics

The concept of encoding a 3D soundfield into a series of spherical harmonics can be expanded further using Higher Order Ambisonics (HOA) which encodes the spatial attributes of a soundfield using a larger set of spherical harmonics. This allows the soundfield to be encoded at a much higher spatial resolution which is accurate to a much higher frequency than first-order Ambisonics.

The spatial resolution of the soundfield is recorded at the expense of a larger number of storage channels and often more complex processing requirements. The number of storage channels required for HOA can be calculated by $(n+1)^2$ where n is the required Ambisonic order. For

example a third order ambisonic signal set will encode a soundfield using 16 spherical harmonics stored as 16 audio channels. The zeroth and first-order ambisonic components (W, X, Y and Z) are accompanied by the second order factors (R, S, T, U and V) and the third order factors (K, L, M, N, O, P and Q) (Malham, 2003). Each of these channels corresponds to the spherical harmonics shown in Figure 3.9.

HOA soundfields can be manipulated in a similar manner to first-order ambisonics, including matrix transformations and beam forming. The additional channels allow the polar pattern of the virtual microphone to much more directional than in first-order Ambisonics. Additionally, HOA panning functions can be used to spatialise mono sound signals. Sound sources spatialised in this way are generally more easy to localise than with first order ambisonics.

HOA soundfields can be recorded directly using similar principles to the first-order Ambisonic microphone described above. However, in order to encode the sound using higher order spherical harmonics, a spherical arrangement of microphone transducers is required, often mounted on a solid structure. By filtering and summation of the signals from each microphone capsule, the more complex spherical harmonic patterns can be synthesised.

The number and arrangement of microphone transducers is determined by the required Ambisonic order of the microphone and the sampling scheme used. A second order HOA microphone may use up to 16 microphone channels (as shown in Figure 3.11), a third order HOA microphone will use an even greater number.

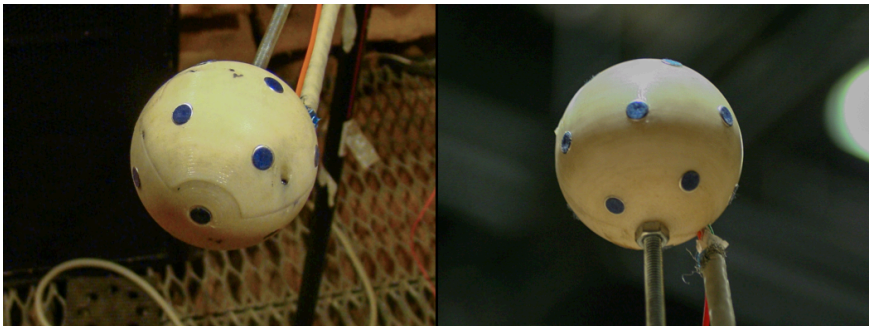


FIGURE 3.11: *Image showing a second order microphone built by Guthrie. An HOA microphone has an increased number of microphone capsules from which it is possible to synthesise the higher order spherical harmonics. Reproduced from (Guthrie, 2014)*

At the time of writing their exists one commercially available microphone array which can be used to record HOA, known as the EigenMike (MH Acoustics, 2014). Typically, HOA microphones are built for specific research activities by institutions. Consequently HOA microphones are not as widely available as first-order ambisonic microphones.

HOA has been used recently by Guthrie (2014) to analyse stage acoustic conditions in numerous concert halls. The data obtained in these surveys was also used to interactively auralise stage acoustic conditions using a loudspeaker array. This allowed musicians to experience different stage acoustic conditions rendered with a high spatial accuracy. The highly directional beam patterns which can be obtained with 2nd order Ambisonics allowed the spatial impulse response

to be explored in great detail with spatial versions of established stage acoustic parameters being implemented.

3.3.4 Intensity measurement

In Chapter 2, the concept of sound intensity was introduced which describes the mean energy flow of sound energy transported through an area. The intensity is a product of the sound pressure and particle velocity which are scalar and vector quantities respectively. The direction of arrival of sound energy can be estimated from the intensity as the opposite direction of the mean flow of energy.

This concept forms the basis of a recently developed family of spatial audio techniques used to analyse and auralise spatial soundfields. These methods are known as Spatial Impulse Response Rendering (SIRR) (Merimaa and Pulkki, 2004), Directional Audio Coding (DirAC) (Pulkki et al., 2009) and the Spatial Decomposition Method (SDM) (Tervo et al., 2013a). All of these methods rely on the direct measurement or estimation of intensity vectors using either an intensity probe, Ambisonic or HOA microphones. While SIRR and SDM are primarily used for auralisation of spatial impulse responses, DirAC is optimised to work on more general spatial audio recordings. In this thesis, the analysis and synthesis methods used by these spatial audio techniques are utilised regularly at different stages. Therefore, a more detailed discussion of these processes will be discussed later in this chapter as well as in Chapter 6.

A direct method of measuring the intensity of a soundfield is to use an intensity probe which consists of 6 omnidirectional microphone capsules, two microphones are arranged on each cartesian axis. This arrangements allow the sound intensity to be measured using a 3D coordinate system (Tervo, 2012). An example of this arrangement is shown in Figure 3.12 which shows a G.R.A.S. 50VI-1 Vector intensity probe.

FIGURE 3.12: *Image showing a G.R.A.S. 50VI-1 Vector intensity probe consisting of 6 omnidirectional microphone capsules. Reproduced from (G.R.A.S. Sound Vibration, 2013)*

The sound pressure can be estimated at the centre of the microphone array by calculating the mean sound pressure captured by all six microphones. The particle velocity is estimated in

each cartesian dimension by comparison of the sound pressure measured in each microphone pair taking into account the distance between the transducers and the characteristic acoustic impedance of air (Tervo, 2012). Typically the microphones are equally spaced at a radius of 10cm. The accuracy of the sound intensity estimation is limited by the distance, d, between the transducers. Frequencies above $f > c/d$ (where c is the speed of sound) experience spatial aliasing and so the correct direction of arrival cannot be estimated (Tervo, 2012). The microphone signals can also be processed to produce B-format signals.

Alternatively, sound intensity can be estimated from Ambisonic or HOA impulse responses as the W-channel is proportional to sound pressure and the X, Y and Z channels are proportional to particle velocity. The increased spatial resolution of HOA microphones can be used to obtain a more accurate estimate of the sound intensity.

When estimating the direction of arrival of reflections from B-format signals, the direction of arrival is obtained by observing the intensity vectors. If a single reflection is present, the vectors will provide a clear indication of the direction of arrival. When there are multiple coincident reflections however, the intensity vectors will point towards the vector sum of these reflections. An advantage of the intensity probe arrangement is that it is possible to localise reflections based on Time of Arrival (TOA) and Time Delay of Arrival (TDOA) methods. These methods exploit the time delays caused by the spatial separation of the microphone capsules and use cross-correlation methods to accurately localise reflections in time. This allows the direction of multiple coincident reflections to be resolved accurately, thus resulting in a more accurate representation of the soundfield (Tervo, 2012).

3.4 Sound source

So far, this discussion has focussed primarily on the recording apparatus, however, the room impulse response is widely known to be influenced by the directivity of the sound source (Lokki and Patynen, 2009). The directional radiation characteristics of a sound source causes variations in the frequency response measured at different angles around the source. Musical instrument directivity patterns are known to fluctuate due to numerous factors (including which note is played and the performance technique used) and vary significantly across different musical instruments (Otondo and Rindel, 2004). Careful consideration must therefore be given to the characteristics of the sound source used to measure room impulse responses so that the source directivity is representative of the desired sound source.

When measuring concert hall impulse responses for analysis, omnidirectional loudspeakers are often used in order to adhere with international standards (International Organisation of Standardisation, 2009). This ensures that variations in source directivity are reduced between measurements in different halls and measured data is easily comparable.

An omnidirectional loudspeaker typically consists of a number of matched loudspeaker drivers arranged on the faces of a platonic solid, often a dodecahedron, as shown in Figure 3.13. Due to the trade-offs necessary in arranging near coincident transducers, omnidirectional sound sources often feature a poor low frequency response and are prone to 'beaming' at higher frequencies.

It is often the case that results obtained from measurements with this type of sound source are averaged over a number of rotations (Hak et al., 2011).

FIGURE 3.13: *Image showing the relative positions of the dodecahedron loudspeaker and sound-field microphone.*

Due to the directivity of this type of loudspeaker it can be expected that early reflections from certain directions will have a different spectrum and different amplitude than if the sound source had the same directivity characteristics as a musical instrument. These differences may produce an audible effect during auralisation which could influence a musician's subjective impression of the space. It is also argued that the impulse response may not be representative of the acoustic conditions experienced by the musician, introducing an error into any objective acoustic parameters.

While it is very difficult to predict, measure or recreate the exact time-varying complex radiation patterns of musical instruments, it may be possible to gain an improvement in auralisation quality and in measurement accuracy by including broad approximations of source directivity in both the analysis and auralisation (Kearney, 2009). This idea has been applied previously to orchestral auralisation (Patynen and Lokki, 2010) by measuring impulse responses representative of each instrument in the orchestra separately. Each impulse response is measured using a directional loudspeaker ensuring that each instrument has a radiation characteristic that is more representative of a musical instrument. The loudspeakers are oriented in a similar manner to how instruments would be oriented in an orchestral configuration.

An ideal measurement system would allow the resultant measured impulse response to be post-processed so that arbitrary directivity patterns could be derived at will from a single acoustic survey. Kearney (2009) found that this was possible by combining impulse responses that had been measured with a directional loudspeaker oriented at different angles of azimuth. The basic premise is illustrated in equation (3.11) where h_{sn} denotes a measured impulse response from a directional sound source n oriented in a known direction. α_n denotes the gain coefficient applied to each impulse response. This formula shows a weighted summation of impulse responses measured at different angles of orientation. The aim is to derive the gains for each impulse response α_n to produced the desired directivity characteristics. The premise requires that the directional characteristics of the measurement loudspeaker are known.

$$h_i(t) = \alpha_1 h_{s1}(t) + \alpha_1 h_{s1}(t) + \ldots \alpha_n h_{sn}(t) \tag{3.11}$$

Pollow et al. (2013) and Kunkemoller et al. (2011) expanded on this concept by developing a method of measuring the room impulse response with a sound source capable of producing spherical harmonic directivity patterns. Using the same principles as applied in Ambisonics, the measurements can then be combined after the survey to produce arbitrary directivity patterns. The sound source used for these measurements consists of a spherical arrangement of different sized loudspeakers laid distributed in a specific manner.

In this research, one of the primary motivations for measuring the stage acoustic condition in concert halls is to extract meaningful data regarding the spatio-temporal distribution of early reflections. In similar work by Tervo (2012), it was found that near-coincident early reflections of similar energy were difficult to localise in space when measured with an omnidirectional source. This was due to the increased likelihood of high amplitude reflections overlapping in time, causing individual reflections to be harder to isolate. As a directional sound source emits more energy in particular directions, some reflections are excited with more energy than others. This has the effect of increasing the signal-to-interference ratio for near-coincident reflections allowing them to be more accurately localised. Tervo (2012) also introduced the notion of a compound sparse impulse response which combines numerous impulse responses measured with a sound source oriented in many angles. This allowed the room reflections to be analysed in greater detail and also to be auralised using parametric auralisation techniques.

In stage acoustics research, measurements are typically made using an omnidirectional sound source to excite the space, as recommended by ISO:3382-1 (International Organisation of Standardisation, 2009). The auralisation systems and analysis performed by Ueno and Tachibana (2003) and Guthrie (2014) were also based around the measurement of spaces using omnidirectional loudspeakers. However, more recently, Brereton et al. (2012b) made use of a directional sound source to represent the directivity of a singer for analysis and auralisation purposes. The acoustic survey of a recital space utilised a Genelec 8040 loudspeaker positioned directly below an Ambisonic microphone. The arrangement of transducers thus representing the source (voice) and receiver (ears) of the performer.

In reference to the survey objectives in this research, it was considered that the source directivity of, for example, a trumpet would have very different characteristics to a French Horn. It was further considered that a musician often changes orientation on stage, reorienting the directivity pattern of the instrument. Therefore, multiple impulse responses should be measured at each location on stage with the source oriented in different angles of orientation to provide suitable flexibility for auralisation.

It was further speculated that these variations in directivity could influence the musician's impression of the stage acoustic conditions as suggested by Cabrera et al. (2010). Therefore, the measurements should allow the temporal and spatial characteristics of the stage acoustic conditions to be studied in detail. The compound-sparse method (Tervo, 2012) of measuring impulse responses with a directional sound source oriented at regular angular increments was considered ideal for this purpose. Finally, as an additional aim was to synthesised arbitrary source directivity patterns by linear summation of impulse response measurements made at each location, as described by Kearney (2009).

In summary, it was considered appropriate in this research to use a directional loudspeaker to excite the stage acoustic response. At each measurement position, multiple measurements were captured where the loudspeaker was reoriented in regular angular increments after each measurement. This approach was found to be sufficiently flexible for auralising many different scenarios in addition to a detailed acoustic analysis. Consequently, the measurements made as part of this research are not directly compatible with those that adhere to international standards. However, by conducting measurements in a number of different concert halls, it was still possible to observe the influence of the concert hall geometry on the stage acoustic conditions.

3.5 Transducer arrangements

So far, the background theory of impulse response measurement has been reviewed. In addition, it was discussed how Ambisonic recording methods and the use of directional sound source could produce a very flexible data set which could be used for both detailed acoustic analysis and auralisation. As discussed previously, the position of the source and receiver can greatly influence the measured impulse response. Similarly, it must also be considered carefully for stage acoustic auralisation and analysis.

ISO:3382-1 (International Organisation of Standardisation, 2009) describes the source and receiver arrangements necessary for measuring objective support. It recommends the source and receiver should be positioned one metre apart, both at a height of either 1m or 1.5m. The heights reflect the difference between a seated musician and a standing musician and the distance represents a similar distance between the musician's instrument and their ears. Measurements made by Ueno and Tachibana (2003) reduce this distance to 0.3m where the source height is 1.2m and the receiver height is 1.5m. Similarly, room acoustic measurements made by Brereton et al. (2012b) have slightly altered the source and receiver positions by positioning the microphone directly above the loudspeaker so that they are nearly coincident.

As mentioned previously, measurements will be conducted using a directional loudspeaker that is rotated after each measurement in regular angular increments. It is of interest to observe how the acoustic conditions vary with source orientation, in isolation from the direct sound. Positioning the loudspeaker directly underneath the ambisonic microphone, as demonstrated by Brereton et al. (2012b), allows the loudspeaker to freely rotate underneath the microphone such that the direct sound and floor reflection remained unchanged but the early reflections would vary with source orientation. This arrangement of ambisonic microphone and loudspeaker was therefore considered appropriate for capturing the stage acoustic impulse responses for this research. A diagram of the arrangement of transducers is shown in Figure 3.14.

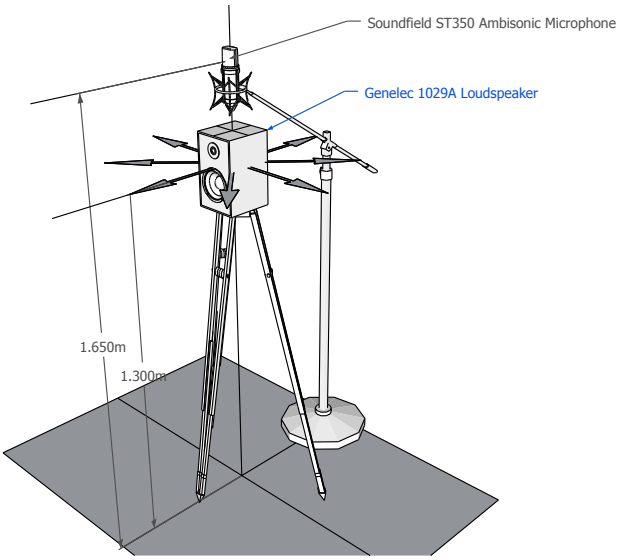


FIGURE 3.14: *Diagram of the relative positions of source and receiver used in performance space surveys. The directional loudspeaker is positioned at a height of 1.3m (to the top of the low frequency driver) and the Ambisonic microphone is positioned directly above at a height of 1.65m. Arrows show source orientations used for each measurement position (45° increments).*

3.6 Discussion

As outlined in the beginning of this chapter, acoustic surveys were conducted on a number of concert hall stages to observe how the spatial and temporal characteristics vary between different stages, at different positions on stage and at different sound source orientations. In addition, the acoustic surveys enabled the capture of spatial room impulse responses suitable for auralisation for stage acoustic laboratory testing.

The impulse responses were captured using a swept sine wave technique and extracting them by convolution of the recorded signal with the inverse sweep signal. This method removes non-linear distortion introduced by the measurement system, providing an enhanced signal to noise ratio. This method is highly repeatable and can be conducted easily and efficiently in the field.

By measuring the room impulse response with a directional sound source oriented in different directions, the effect of source orientation on the measured room impulse response could be observed. Measuring impulse responses in this way allows early reflections to be more easily localised in time and direction of arrival and possible combined as a compound-sparse representation as described by Tervo (2012).

It is proposed that impulse responses measured in this manner are more appropriate for auralisation than the use of an omnidirectional sound source as it provides a rudimentary similarity with the directional characteristics of a musical instrument which typically have a radiation pattern with a dominant direction.

Ambisonics is an ideal method for capturing room impulse responses primarily due to the ease by which spatial data is encoded onto a relatively small number of audio channels using a single compact microphone array. These data can be easily analysed to observe the spatial and

temporal characteristics of a room impulse response. In addition, Ambisonic impulse responses are highly flexible in that they can be easily transformed and manipulated and can be used to auralise virtual acoustic soundfields over any loudspeaker array layout. First-order Ambisonics was chosen primarily due to the widespread availability of this type of microphone despite the documented improvements that are possible by using HOA microphones or intensity probes.

It was determined that the sound source should be positioned directly beneath the Ambisonic microphone with the source and receiver positioned at heights representative of a musical instrument and the head of a standing performer. This would allow the sound source to freely rotate underneath the microphone without any overall increase in level from the loudspeaker producing a set of impulse responses where the level of the direct sound does not change significantly between measurements but the relative level of reflections changes in response to the source orientation.

The measurement technique used in these surveys departs from the methodology outlined in ISO:3382-1 (International Organisation of Standardisation, 2009). The consequence of this is that the parameters extracted from the measurements cannot be directly compared to those obtained in other studies. However, the measurement technique has been designed to explore variables that are not normally included in standard acoustic measurements and so the overall approach was deemed appropriate for this study. Acoustic surveys were performed in eight venues of different geometries and sizes and so will allow internal comparisons to be made between venues. A detailed description of the measurement procedure used in each venue is described in the following section. This will outline the specific apparatus used in each survey.

3.7 Venue measurement procedure

In order to capture the acoustic characteristics of the space, impulse response measurements were made at various source and receiver combinations around each venue. The loudspeaker was mounted on a tripod so that the height from the stage to the top of the low frequency driver was 130cm. An Ambisonic microphone (usually a Soundfield ST350) was positioned directly over the loudspeaker at a height of 165cm from the stage floor. A Genelec 1029A loudspeaker was used to excite each performance space. After each measurement, the loudspeaker was rotated laterally in 45° increments. This procedure was carried out at each stage position resulting in 8 measurements per source-receiver. Figure 3.15 shows the arrangement of equipment in one of the concert halls included in the acoustic surveys.

FIGURE 3.15: *Image showing the relative positions of the loudspeaker and microphone. The height from the stage floor to the top of the low frequency driver of the loudspeaker is approximately 130cm whereas the the height of the soundfield microphone above the stage is 165cm.*

A Genelec 1029A Active loudspeaker was chosen as a suitable sound source for measuring impulse responses. Figure 3.16 shows the directional characteristics of the loudspeaker at different angles of azimuth. The magnitude is shown in dBFS (relative to Full Scale) with the colour bar range shown adjacent. Details of the measurement procedure used to obtain these results are included in Appendix A.

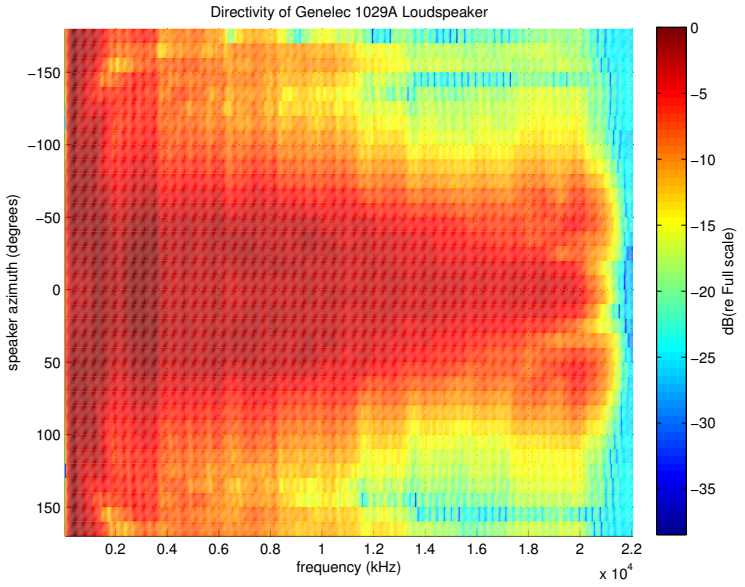


FIGURE 3.16: *Directivity characteristics of the Genelec 1029A loudspeaker measured in the transverse plane. Positive angles represent angles in the clockwise direction. As expected the loudspeaker contains more high frequency content at angles close to* 0° *(on-axis)*

It can be seen from Figure 3.16 that at lower frequencies (below $2kHz$) the energy from the loudspeaker does not vary much with speaker orientation. Between $2kHz$ and $10kHz$, the

Equipment
Macbook (OS X Snow Leopard 10.6.8) 2.1GHz Intel Core 2 Duo, 2GB SDRAM
M-Audio Profire Lightbridge Firewire Soundcard
Behringer ADA8000 A/DAC unit
Genelec 1029A Loudspeaker
Rion NA28 Precision integrating Sound Level Meter (AAcG Kit A)
1/2" diameter pre-polarised condenser microphone
Type 1 Sound Pressure Level calibrator
Soundfield ST350 (or ST450) microphone with pre-amplifer unit
Bosch Laser Range finder

TABLE 3.1: *Typical Equipment list for venue surveys*

amplitude is greater at angles between $\pm 50°$ where $0°$ denotes the on-axis orientation. As the frequency increases further, the amplitude is greatest between $\pm 10°$ and off axis energy is severely attenuated.

The loudspeaker was connected via balanced XLR leads to the line output of a Behringer ADA8000 A/DAC which was connected via ADAT to an M-Audio Profire Lightbridge soundcard. The soundcard was connected to a Macbook via Firewire cables. A diagram of this arrangement can be seen in Figure 3.18. An equipment list is shown in Table 3.1.

The measurement methodology was adapted slightly from the first survey in the Grand Hall at the Glasgow City Halls. In this particular survey the sound source and receiver were positioned at heights of $1.30m$ and $1.65m$ respectively, however the sound source was positioned at a distance of 20cm in front of the soundfield microphone. Furthermore, the sound source was rotated at this radius in $45°$ increments and the Soundfield microphone was rotated in tandem so that the front of the microphone pointed in the same direction as the main axis of radiation of the loudspeaker. This was to emulate a musician holding an instrument in front of them turning on the spot.

Additionally in the Grand Hall survey, it was also possible to measure impulse responses with a Brüel & Kjær 4296 dodecahedron loudspeaker. This loudspeaker was driven in the same manner as described previously with the exception of the loudspeaker being driven by a Funktion1 A4 amplifier. A picture of this arrangement can be seen in Figure 3.17. It can be seen that a number of the loudspeaker drivers directly face the microphone and are positioned at close proximity, therefore in order to avoid distortion the gain of the loudspeaker was adjusted such that clipping did not occur. This had the effect of reducing the signal-to-noise ratio of the recording. The dodecahedron loudspeaker was used without a subwoofer. In typical concert hall surveys a subwoofer is often used to increase the amplitude of the sine sweep at low frequency regions beyond the capabilities of the main loudspeaker.

FIGURE 3.17: *Image showing the relative positions of the dodecahedron loudspeaker and soundfield microphone.*

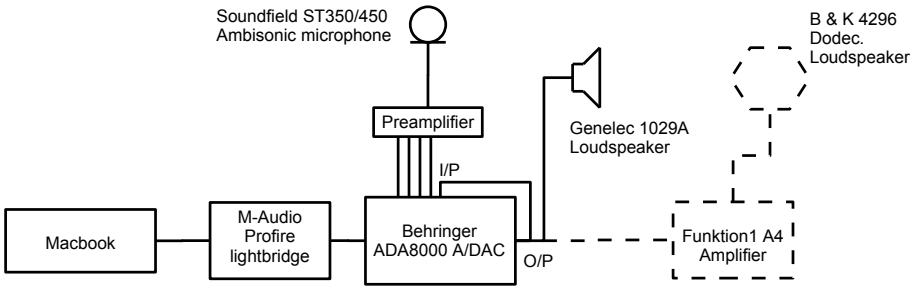


FIGURE 3.18: *System diagram of measurement system showing how the loudspeaker and microphone were connected to the laptop. The equipment marked in dashed lines was used during one particular survey.*

The impulse responses in each venue were measured using a 10-second, logarithmically swept sinusoidal signal which was generated using MATLAB code (Mathworks, 2013) developed by Wells (2012). This code was used to generate a sine sweep signal between 0Hz and 22050Hz with an 100ms amplitude ramp at the start and the end to avoid any unwanted transient signals. Furthermore, it generates an inverse sweep by time-reversing and applying a -6dB per octave envelope. The sine sweep was played back and measured simultaneously in Reaper (DAW) (Reaper, 2013) at a sampling frequency of 44.1kHz and a 32-bit floating-point bit depth.

The recorded sweeps were processed after the survey was concluded in order to extract the impulse response data. As described previously in Chapter 3, the impulse responses were recovered by convolution of the recorded sweep with the inverse signal. The impulse response was then truncated from the start by an amount equal to the measurement system latency. The latency of the measurement system was obtained every measurement by sending the sine sweep via short XLR cable from the output to the soundcard input and extracting the resultant impulse response. The impulse response featured a single peak which occurred at a delay time equal to

the system latency. For each measurement location the recorded latency remained unchanged. However, it was found that the latency varied slightly when the equipment was moved from one location to another. This was thought to be caused by powering down the sound card devices in order to move them.

Once the impulse responses were extracted, the signals were normalised to ensure that the direct sound as measured by the W channel was the same amplitude and also that no B-format channels exceeded an amplitude of ± 1.0 to avoid clipping when writing the data to audio files. This was achieved by finding the impulse response with the highest amplitude peaks and applying a gain such that the W-channel peak from the direct sound was an amplitude of ± 0.5. The gain required to do this was applied to the remaining impulse responses to ensure the relative level of each channel was maintained. For impulse responses where the source and receiver were positioned at separate locations, the normalisation scheme was adjusted so that the relative amplitude of the impulse response was maintained for changes in source directivity or source distance.

The background noise level was measured in each venue at different positions using a Rion NA28 (or B&K 2260) Integrating Sound Level Meter (SLM) positioned alongside the Ambisonic microphone. The relative position of the transducers is shown in Figure 3.19. The SLM calibration was verified prior to each measurement to a sound pressure level of $94.0dB$ by using the accompanying calibrator. The SLM and soundfield recording system were set to record simultaneously for a duration of 5 minutes.

FIGURE 3.19: *Image showing the relative positions of the Ambisonic microphone and sound level meter for background noise measurement in the Ledger Recital Room.*

The basic dimensions of the performance space and the measurement locations were also measured using a laser range finder.

3.8 Performance space descriptions

The performance spaces included in this research occupied a wide range of attributes and sizes, ranging from small informal recital spaces to large symphonic concert halls that could seat 2000 audience members. Some of the venues were designed for a specific genre of music whilst others were used as multi-purpose venues that could accommodate small recitals as well as wedding receptions or conferences. The acoustic response measured in each hall is often inextricably linked to a specific context, for example, the hall may designed purely for symphonic music or alternatively may be used for a wide range of performance types. Therefore a brief description of each of the venues is given below.

In this thesis, positions on the stage are referred to from the point of view of the performer, for example the up-stage left position would be situated at the rear of the stage with the nearest wall on the musician's left. Down-stage centre would be positioned at the front of the stage in the centre of the stage.

3.8.1 The Grand Hall, Glasgow City Halls

The Grand hall at Glasgow City Halls (Glasgow Life, 2013) is located in the Merchant City area of Glasgow and is one of the main performance venues in the city. The Grand Hall was originally designed by George Murray and was opened in 1841. The City Halls was recently renovated in 2006 by Arup Acoustics and the Grand Hall is now the home of the BBC Scottish Symphony Orchestra. The Grand Hall is a shoebox concert hall design seating around 1066 audience members. The hall is primarily used for orchestral performances (seats approximately 90 musicians) but however sees regular use for amplified contemporary music and broadcast (Arup, 2013). The stage features permanent risers with a rear wall that is slightly angled towards the centre of the hall on either side. The side walls throughout the venue feature a regular pattern of cylindrical protrusions which act as diffusers. There is a seating area arranged around the rear of the stage for choirs (approximately 110 seats). There are high arched windows on either side of the hall.

FIGURE 3.20: *Panoramic view of the Grand Hall at Glasgow City Halls taken from the stage*

(a) View of audience from stage

(b) View of stage from audience

FIGURE 3.21: *Views around the Grand Hall at Glasgow City Halls.*

The stage is raised above the front stalls by approximately 1.5m, this seating area follows a shallow rake towards the rear of the hall. There are also two rows of seating on each side of the hall with a height equal to that of the stage. The venue also has a balcony seating area at the sides and rear of the hall. At the rear of the hall there are steeply raked seats. Variable absorption is often applied in this hall for certain performances which can be deployed in the ceiling cavity and on the rear walls (stage front) in the form of heavy curtain material. The ceiling cavity is open to the venue and is also used as a place to position stage lamps. This was not deployed during this survey.

3.8.2 The Recital Room, Glasgow City Halls

The Recital Room is a small recital space also located at Glasgow City Halls (Glasgow City of Music, No date). This performance space was 'rediscovered' during refurbishments to the Glasgow City Halls in 2006. It is a small, rectangular shaped venue with a vaulted ceiling which seats approximately 100-150 audience members. The venue is suitable for chamber music and soloist recital however it is regularly used for jazz recitals, traditional music and also as a venue for weddings and other events. It is often used as a pre-concert venue during high profile concerts in the Grand Hall.

FIGURE 3.22: *Panoramic view of the Recital Room taken from the stage*

(a) View of stage from audience

(b) View of audience from stage

FIGURE 3.23: *Views around the Recital Room at Glasgow City Halls. Audience seating was removed prior to the survey.*

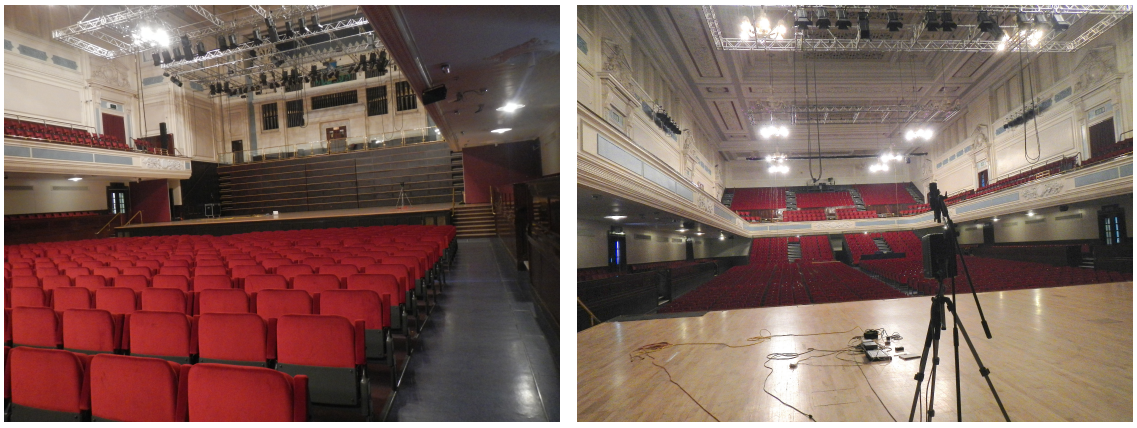
The space is approximately $865m^3$ in volume. It features large windows (two panes of single glazing separated by a 10cm, non-sealed air gap) along the west wall which look out onto flats and bars on Candleriggs (a street in Glasgow). There is no fixed seating and stage arrangement allowing the venue some flexibility in terms of usage. The floor is of timber construction and is flat in profile. The walls appear to be of plasterboard construction and do not feature any noticeable diffusing features. The ceiling appears to be the original timber construction which is vaulted. On one side of the ceiling three air handling units are visible but did not appear to be in operation at the time of the survey. There also appeared to be ventilation behind grilled panels at the stage front and stage rear walls.

3.8.3 The Caird Hall, Dundee

The Caird Hall (Wikipedia, 2014) was completed in Dundee in 1922 and was named after James Key Caird, a jute baron and major benefactor of the venue. It is located in Dundee city square and overlooks Dock Street and the riverside development (which was being constructed at the time of writing). The venue has a very wide programme hosting graduation events, exhibitions, conferences, stand up comedy, popular music, opera, symphonic and chamber music. It is also

used on occasion for recording. The hall regularly hosts The Royal Scottish National Orchestra (RSNO) and the Scottish Ensemble.

FIGURE 3.24: *Panoramic view of the Caird Hall taken from the audience*



(a) View of stage from audience

(b) View of audience from stage

FIGURE 3.25: *Views around the Caird Hall. The stage is set in its normal orchestral configuration*

The venue is a shoebox configuration and has a seated capacity of 2000 with stalls seating being removable for a standing audience (Caird Hall, No date). For some performances, drapes are flown on the stage and at the rear of the hall to form a 'black box' configuration. The hall features extensive plaster ornamentation. A balcony runs around the perimeter of the hall (with the exception of the stage area). The audience area features a shallow rake until the last few rows of seats which incline towards the height of the balcony. On either side of the stalls area there are wooden panels which act as the front of raised seating which runs the length of the venue on either side.

The stage is approximately $1.2m$ above the audience floor and has an area of approximately $214.2m^2$. The rear of the stage features a wooden bleacher style arrangement of seats, presumably to seat a choir. During the survey the bleachers were in their fully retracted position and formed a solid reflecting surface. Part of the reflecting side walls are of wooden construction. Above the rear of the stage is a pipe organ built by Harrison & Harrison (Wikipedia, 2014).

3.8.4 The Ledger Recital Room, Royal Conservatoire of Scotland

The Ledger Recital Room (formerly The Guinness Room) at The Royal Conservatoire of Scotland is one of five venues used for teaching, workshops and public performance. It is situated at the Renfrew Street Campus in Glasgow City Centre (Royal Conservatoire of Scotland, No datea). The Ledger Recital Room can seat approximately 108 audience members on retractable bleacher style seats which are upholstered.

FIGURE 3.26: *Panoramic view of the Ledger Recital Room taken from the stage*

(a) View of stage from audience

(b) View of audience from mezzanine

FIGURE 3.27: *Views around the Ledger Recital Room.*

The stage area comprises of a mezzanine which runs around the perimeter of the stage end of the venue which is accessible via staircases on either side. On top of the mezzanine there are a number of rare organs and harpsichords. Underneath the mezzanine forms a storage area and access route that is concealed by operable walls set on rails (without acoustic seals). At the stage front wall there is a control position with equipment for operating lighting and PA systems. The stage area is at the same level as the first row of seats and has a footprint of approximately $70m^2$.

The hard ceiling in this venue is mostly obscured by gantries, air handling units and lighting trusses. At the top of each side wall are windows which are covered with curtains. There are

also fixed hanging baffles near the windows on each side. The walls of the venue are a mixture of exposed brickwork on the upper portion of each wall and wooden panels.

3.8.5 The Stevenson Hall, Royal Conservatoire of Scotland

The Stevenson Hall is another venue in the Royal Conservatoire of Scotland also based at the Renfrew Street Campus. It is named after former Lord Provost of Glasgow, Sir Daniel Stevenson (Royal Conservatoire of Scotland, No dateb). The venue is large enough to accommodate symphonic orchestras but is also used for chamber music and master classes for music of many genres. It is regularly used for recording and broadcast for national radio. It can seat approximately 355 audience members on upholstered seating with a reasonably steep rake. There is also a single level balcony on either side of the main seating area with the edge of the balcony forming the side walls for the audience.

FIGURE 3.28: *Panoramic view of the Stevenson Hall taken from the stage*

(a) View of stage from audience

(b) View of audience from stage

FIGURE 3.29: *Views around the Stevenson Hall.*

The stage is level with the front row of audience seats. The stage walls are constructed of exposed masonry however the rear wall is completely covered with a heavy curtain and the majority of

the stage right wall is comprised of an organ. Two of the side access doors to the venue form large reflecting surfaces which are angled towards the stage. At the rear of the stage there are temporary stage risers (3 levels) which run much of the width of the stage. The stage footprint is approximately $220m^2$ not including the organ. At the stage front wall there is a control room with a glass window with equipment for operating lighting and PA systems.

The hard ceiling in this venue is mostly obscured by gantries, air handling units and lighting trusses. At the top of each side wall are windows which are covered with curtains. Over the stage, there are four rows of overhead reflectors (approximately 7.5m above the stage) angled presumably to assist communication between members of an orchestra and to reflect sound towards the audience. Over the audience there is a hung wooden lattice which may serve to assist diffusion of sound over the audience.

3.8.6 The Younger Hall, University of St Andrews

The Younger Hall is situated on North Street in St Andrews and is one of the main performance venues seating around 1000 audience members. It was built between 1923 and 1929 and designed by Paul Waterhouse as the graduation hall for St Andrews University and is now home to the University music department (Unversity of St Andrews, No date). The venue generally hosts a wide variety of performances including chamber music, orchestral, Scottish traditional and jazz music. It is often used as a conference venue in addition to graduations. It also known for hosting traditional Scottish dancing classes and Ceilidhs.

FIGURE 3.30: *Panoramic view of the Younger Hall taken from centre stage position*

(a) View of stage from audience

(b) View of audience from stage

FIGURE 3.31: *Views around the Younger Hall. The majority of front stalls seating was removed to accommodate a Scottish country dancing class prior to the survey.*

The hall is a shoebox shape which features a rear audience balcony and a wooden panelled stage enclosure which is raised approximately $1.15m$ above the front stalls. The stage enclosure is approximately $11.25m$ wide at the front and approximately $8.6m$ wide at the rear. The audience area on the ground floor is wider than the stage itself, approximately $20m$. The stage front wall is approximately 31m away from the organ on stage. The stage has angled side walls with operable steps at the rear leading to an organ designed by Harrison & Harrison. These steps are presumably able to accommodate a small choir. The seats are upholstered with vinyl and front stalls seats can be removed as required. A number of surfaces in the hall have acoustic absorption applied in the form of perforated panels. A number of these surfaces can be seen on the face of each balcony level. There appears to be absorptive panels fixed on the ceiling of the stage enclosure. On either side of the hall there are large single glazed windows which overlook one of the main entrances to St Andrews University and student halls. The windows can be covered with heavy curtains. There are numerous archways over each window in the space.

3.8.7 The University of Glasgow Concert Hall

The Glasgow University Concert Hall is a small concert venue in the main Glasgow University Campus (Gilmorehill) on University Avenue. The campus was constructed in 1870. The concert hall is generally used for small lunchtime recitals, performance practice and teaching. In addition, the venue is also used occasionally as a recording venue for the Music department and for electroacoustic music performance.

FIGURE 3.32: *Panoramic view of the Glasgow University Concert Hall taken from the front seats. Please note the panoramic image has failed on the left hand side showing the rear of the stage.*

(a) View of stage from audience

(b) View of audience from stage

FIGURE 3.33: *Views around the Glasgow University Concert Hall.*

As with other venues of this type, a number of rare instruments are housed here. This includes two Model D Steinway Grand Pianos, an 1840s Broadwood Grand and a Mozart-era fortepiano. There are also two chamber organs in this venue, one of which occupies a major section of the venue rear wall (Glasgow University, 2014). These instruments were stored in the cylindrical alcove and could not be moved prior to the survey.

The concert hall is of shoebox construction, approximately 25m in length and 15.5m in width at the widest point in the room. The ceiling height is constant throughout and is approximately 6.5m from floor to ceiling. The stage area is a cylindrical alcove at one end of the hall which is comprised mainly of two sets of glazing separated by an air-gap of approximately 10cm. The inner set of glazing can be opened and does not feature any airtight seals. The windows overlook the campus main entrance and also University Avenue.

This cylindrical stage area is approximately 9m in diameter and is the same height as the audience area. The audience area is considerably wider than this alcove (approximately 15.5m). All wall surfaces appear to be of plasterboard construction which presumably covers the building‘s

original stone work. Porous acoustic absorption has been applied to the upper half of the rear and side walls of the venue housed within a perforated steel enclosure.

The venue was carpeted throughout including the stage area with a thin pile carpet and all seating (which was upholstered) in the venue is temporary and was stored at the sides of the venue during the survey. The seating was stored alongside a range of musical equipment (i.e. percussion instruments, music stands etc). A lightweight curtain is positioned at the front of the stage alcove which was fully retracted during the survey.

3.8.8 The Reid Concert Hall, University of Edinburgh

The Reid Concert Hall is a venue in the School of Music at Edinburgh University (The University of Edinburgh, 2014). It is located at the main university campus at Bristo Square and is adjacent to the McEwan Hall and also the University Students Union. It is also situated next to the Musical Instrument Museum. It was completed in 1859 and designed by Professor John Donaldson with Reid Bequest funds (Edinburgh Guide, 2014). The venue is used regularly for teaching, recording and lectures as well as public concerts. Performances are typically small ensembles, solo recital or organ recital, however it is used as a venue for the Edinburgh Fringe Festival every summer.

FIGURE 3.34: *Panoramic view of the Reid Concert Hall taken from the front seats*

(a) View of audience from stage

(b) View of stage from audience

FIGURE 3.35: *Views around the Reid Concert Hall.*

The venue is a narrow shoebox configuration and has a seated capacity of 250 seats. Seating is raked with the height of each row of seats increasing on an 'exponential-like' curve. The seating area is split into two with a central stairway. At the rear of the audience area, two small stairways are used for egress. The seats are upholstered with a vinyl/leather cover with a fabric base.

The venue configuration can be likened to a 'live end-dead end' room where the stage side walls are mainly of plaster construction (with the bottom portion constructed of wood panelling) whereas the rear half of the audience side walls have porous acoustic absorption applied in the form of perforated panels. The audience rear wall also features similar acoustic treatment. There is minimal ornamentation on the walls aside from a number of hanging portraits. The venue features a vaulted ceiling with regular recessed areas. Along each side wall there are numerous tall windows which look north onto McEwan Hall or south onto other university buildings.

The main stage area is at the same level as the first row of seats aside from a two level riser at the back of the stage. The main stage floor is approximately $88m^2$ in area with an additional $36m^2$ when the risers are included. The rear stage wall protrudes into the hall at the centre of the stage and is recessed back on both sides in two discrete steps. Behind the rear stage wall is a storage space alongside access to the organ. The rear stage wall (on the upper half) features the Ahrend Organ which was built in 1978 by Jürgen Ahrend (The University of Edinburgh, 2014) and is the only remaining example in the UK. The organ forms an overhang over the highest riser. It is normal that a number of musical instruments, including a grand piano, harpsichord and percussion instruments (tympani, chimes etc) remain on the side of the stage when not in use.

The venue features a permanent sound reinforcement system which consists of two raised loudspeakers adjacent to the organ in addition to an array of small loudspeakers used for public address.

3.8.9 Physical dimensions

The venues measured represent a wide range of concert halls from small recital venues of around 100 seats to large symphonic concert halls of around 2000 seats. The basic dimensions of the venues are shown in Table 3.2. The table shows the stage width (W_s), depth (D_s), ceiling height (H_s), stage area (A_s) and number of audience seats (N).

Venue	W_s(m)	D_s(m)	H_s(m)	A_s (m^2)	N
LRR	10.87	5.58	8.8	60.7	108
RR	7.44	2.5	8.47	19.0	120
GUCH	9.0	7.5	6.47	64	140
RH	11	11.5	14.99	126.5	250
SH	17.55	10	7.45	175.5	355
YH	11.25	10	9.68	99.2	1000
GCH	18.42	11	10	203	1066
CH	18.89	12.55	13.11	237.1	2000

TABLE 3.2: *Basic dimensions of the stage in each venue including audience seating capacity*

The two largest venues, in terms of seating capacity are the Caird Hall and the Glasgow City Halls which are well known for programming regular orchestral performances. The smallest venues are the Recital Room (at Glasgow City Halls) and the Ledger Recital Room. The Caird Hall has the largest stage area whereas the Recital Room is much smaller; The Recital Room had a small temporary platform stage.

3.8.10 Measurement positions

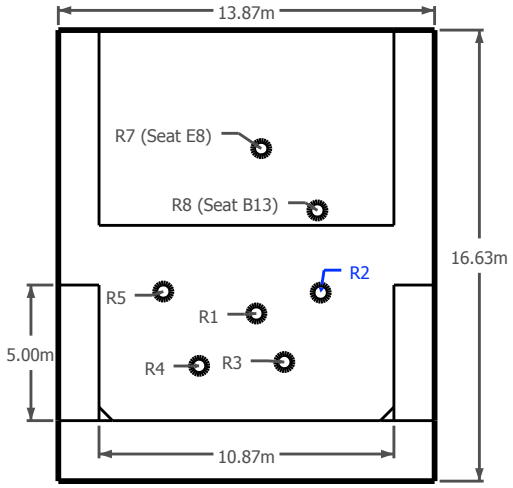
The measurement positions were chosen to reflect possible positions a soloist may perform from. Normally during the surveys, this would consist of three down stage positions, left, right and centre. It can be seen that the measurement locations are not consistent in each performance space with different measurement locations chosen due to physical dimensions of the venue. On some occasions, rare instruments were positioned on stage and could not be moved prior to the survey commencing.

Additional measurement positions were measured up-stage as time allowed to increase the coverage over the remainder of the stage. In addition to this, impulse responses were measured across the stage to obtain data suitable for assessing ensemble parameters. Finally, impulse responses were also measured with the source positioned on stage and receivers positioned in the audience area. This was to enable future auralisation and analysis from the audience position.

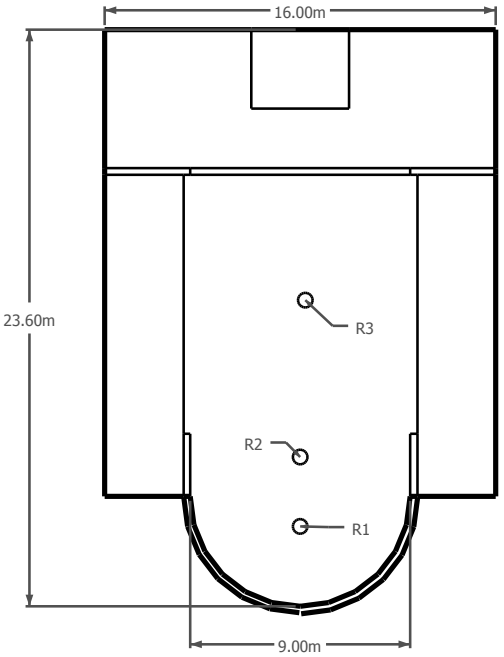
Sketches of the basic geometry are shown in Figures 3.36 and 3.37. Each sketch is shown in plan with the stage located at the bottom of each plot; measurement locations marked as circles. The plots are not scaled relative to each other.

The measurement locations show the positions chosen as receivers. The receiver positions located on stage also show the location of the sound source for soloist stage acoustic measurements.

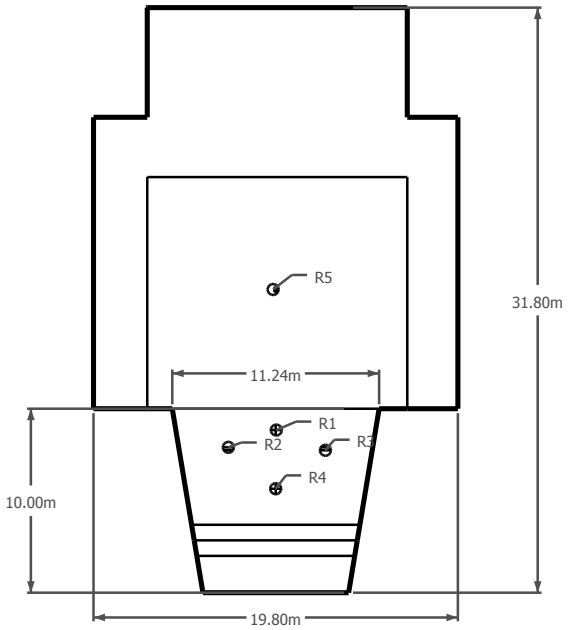
Specific measurement positions are referred to by the location of the source and receiver. i.e. S1R1 denotes both the source and receiver are located at position R1.



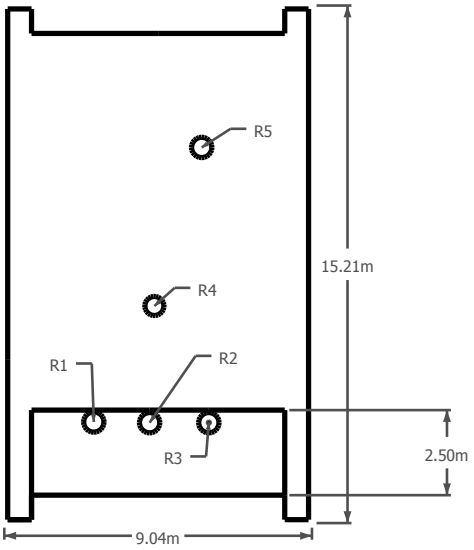
(a) The Ledger Recital Room,
Royal Conservatoire of Scotland, Glasgow (108 seats)



(b) Glasgow University Concert Hall,
Glasgow (approx. 140 seats)

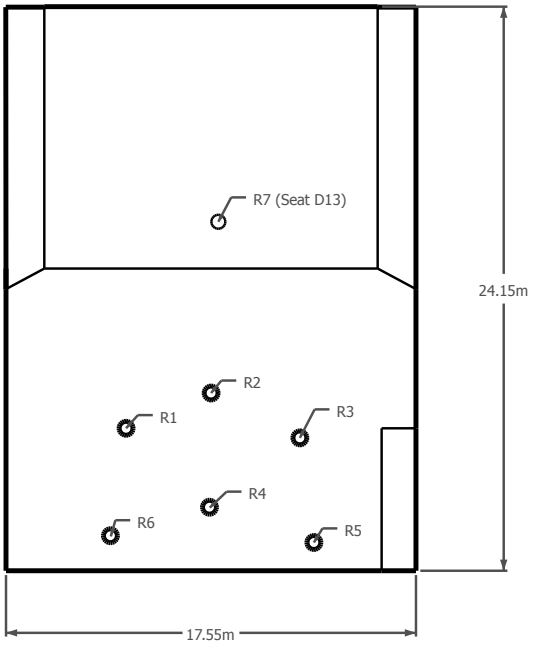


(c) The Younger Hall, St Andrews (1000 seats)

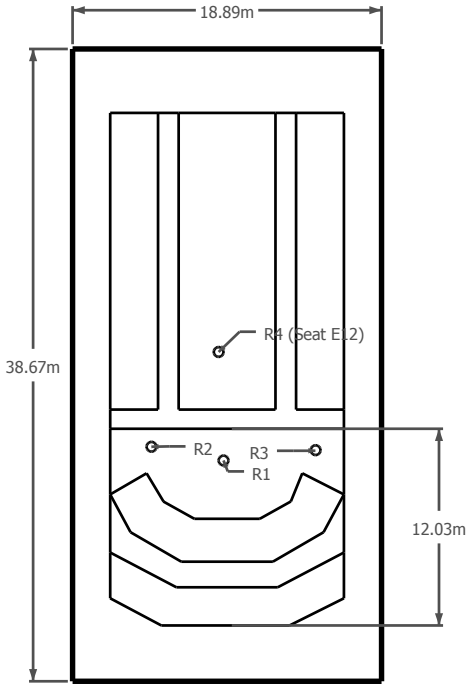


(d) The Recital Room,
Glasgow City Halls (approx. 120 seats)

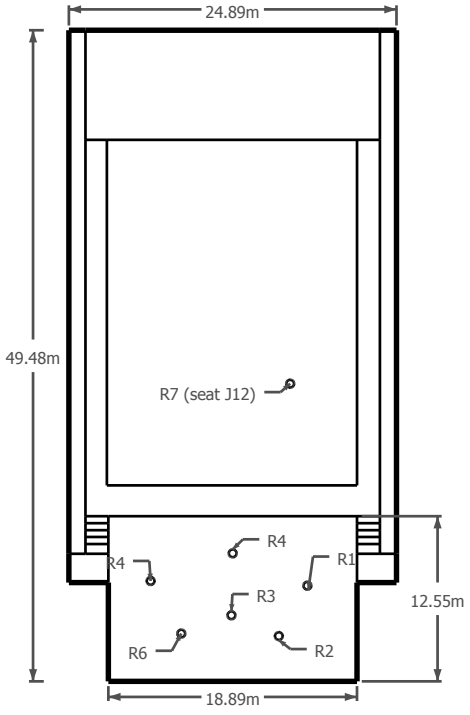
FIGURE 3.36: *Views of the concert halls surveyed as part of this research. Diagrams are not scaled relative to each other. The stage is situated at the lower portion of each diagram.*



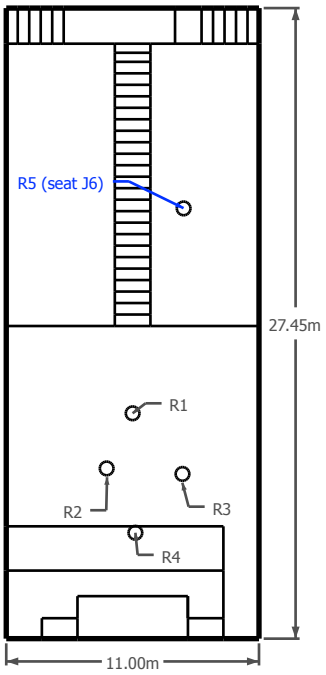
(a) The Stevenson Hall, Royal Conservatoire of Scotland, Glasgow (355 seats)



(b) The Grand Hall, Glasgow City Halls (1066 seats)



(c) Caird Hall, Dundee (2000 seats)



(d) Reid Hall, Edinburgh University (250 seats)

FIGURE 3.37: *Views of the concert halls surveyed as part of this research. Diagrams are not scaled relative to each other. The stage is situated at the lower portion of each diagram.*

3.9 Summary and discussion

This chapter has reviewed how room acoustic responses are measured and has proposed a new method of capturing stage acoustic conditions suitable for both auralisation and detailed analysis of the spatio-temporal distribution of early reflections. This method was applied in eight performance spaces in order to observe how the acoustic conditions varied due to hall-related (stage geometry) and performer-related variables (location on stage and source directivity/orientation)

The proposed measurement procedure differed significantly from the methods recommended by international standards (International Organisation of Standardisation, 2009). The deviations can be attributed to the use of a directional loudspeaker, the use of an Ambisonic microphone and the relative location of the transducers on stage. The relative position of source and receiver was considered as much more appropriate for auralisation, in addition the use of repeated measurements with a directional sound source allowed the source directivity to be adapted for auralisation which will be discussed in the following chapter. Consequently, these measurements might not be considered to be compatible with other stage acoustic data measured in the standardised manner. However, this procedure was considered appropriate as it allowed these variables to be explored in detail as well as being suitably efficient to allow a full analysis of each stage.

At the time of writing, it has been widely recognised that measurements made with omnidirectional transducers, while standardised, do not provide a complete representation of the acoustic conditions in a room mainly due to the lack of directional characteristics of the sound source (Tervo et al., 2013b). Currently, numerous loudspeaker arrangements are being explored that arbitrary, complex radiation patterns to be synthesised (Tervo et al., 2013b). This approach could improve the plausibility of auralisations but also provide a more relevant analysis of concert hall acoustic responses in relation to specific instruments.

This can be achieved using sound sources which are capable of producing spherical harmonic radiation patterns allowing the directivity of a sound source to be controlled arbitrarily (Pollow et al., 2013). A similar approach is to use an array of loudspeakers (or a single loudspeaker used to measure at different angles of orientation) to measure the acoustic response. The amplitude and phase of each of measured impulse response can then be altered and summed to produce an approximation of a particular instrument directivity (Waxman, 2005). In this study, measurements were made with a sound source rotated on its axis to different directions in order to emulate the effect of a directional musical instrument changing orientation or directional response.

The use of an Ambisonic microphone for measuring the stage acoustic impulse responses proved to be a highly effective method of measuring the spatial characteristics of the stage acoustic response. It is widely recognised that a first-order Ambisonic microphone can only provide a limited spatial resolution. Other researchers in stage acoustics have reported improved results when using a higher-order Ambisonic microphone to record the impulse response for auralisation or analysis (Guthrie, 2014). A further improvement may be obtained by using an intensity probe as proposed by Tervo (2012).

Overall, it was found that the measurement procedure was a highly efficient way of capturing the stage acoustic response for the objectives of this study. With the measurement procedure established, it is now possible to perform different analyses on the measured impulse responses to determine how the spatial and temporal aspects vary with the independent variables discussed. Chapter 4 will demonstrate how the spatial and temporal distribution of reflections can be extracted and characterised from the measurements. The results of the analysis will be discussed in Chapter 5.

Chapter 4

Stage acoustic analysis techniques

In Chapter 2, it was discussed how existing stage acoustic parameters only broadly consider the temporal structure of the impulse response and do not include the spatial distribution of reflections. As this research is concerned with the subjective impact of the spatial and temporal structure of early reflections, it is necessary to develop acoustic parameters which reveal the spatial and temporal structure of the acoustic response in detail.

Observing how these parameters vary on different concert hall stages will provide valuable insight into the acoustic conditions found on stage. In addition, a comparison of these parameters with subjective responses from musicians will reveal if these aspects of the acoustic response influence a musician's impression of the concert hall.

In order to examine the spatial and temporal properties of a stage acoustic response a measurement approach was presented in the previous chapter which utilised a directional loudspeaker and Ambisonic microphone to capture the acoustic response experienced by a musician on stage. This chapter will examine the available methods of extracting the required spatial and temporal information from the captured impulse responses and determine which is most applicable to the overall aims of this research.

4.1 Temporal analysis methods

The temporal structure of room impulse responses is of great interest to acousticians as it is widely accepted that it can has a strong influence on the perception of music played in a concert hall. Many acoustic parameters, for example Clarity (C_{80}), segregate the impulse response into early and late time regions and compare their relative energy. In the case of C_{80}, a concert hall is perceived as being all the more clear if the early sound (including the direct sound) contains more energy than the late time region.

A similar approach has been applied in the field of stage acoustics, for example, where a musician's impression of support is linked to the relative energy of the direct sound and early

reflections (arriving between 20-100ms after the direct sound). This is encompassed by the parameter, early objective support ST_{early}.

While these parameters recognise the importance of the relative energy in different time regions of the impulse response, they do not account for the detailed structure of reflections within these regions. It is probable that existing stage acoustic parameters are unable to distinguish between halls that have the same overall energy of early reflections but with different temporal distributions. If the temporal structure of early reflections is found to influence a musician's impression of the space, then it is possible that important subjective aspects of the acoustic response are being ignored.

The following sections will focus on different approaches to assessing the temporal distribution of early reflections and propose an appropriate method of analysis for this research.

4.1.1 Reflection density profile

As discussed in Chapter 2, the basic structure of an impulse response consists of the direct sound, early reflections and diffuse reverberation. After the direct sound, the density of reflection arrivals grows to a point where individual reflections are no longer discernible and the sound decay can be treated statistically. The transition time (often referred to as the mixing time) between the early reflections and reverberation is of interest to acousticians so that early reflections and reverberation can be considered separately in analysing the acoustic conditions.

Often, for practical purposes the boundary between early reflections and reverberation is assumed to be a fixed value. Commonly found values are $50ms$ for speech, $80ms$ for music and $100ms$ if analysing stage acoustic conditions. These values are commonly used as the upper integration time for room acoustic parameters where the time region assessed end abruptly at these points. This can be problematic if, for example, a strong reflection arrives just outside of that time window. The effect of this reflection may be audible but it might not be included in the objective parameter.

It is useful to determine the mixing time for stage acoustic impulse responses measured in the course of this research. This is primarily to isolate the early reflections so that they can be studied in more detail but also to observe if the mixing time corresponds with the value generally used for early objective support. One method of determining the mixing time is to observe the density of reflections over time. Once the reflection density increases beyond a set value, the corresponding delay time can be considered the mixing time. Furthermore, analysing the density of reflections over time provides an additional method of quantifying the temporal distribution of early reflections. Therefore, a method of quantifying the reflection density profile is discussed below.

It can be demonstrated that the rate of arrival of reflections depends on the geometry and size of the room and this can have a significant impact on the timbre of music playing in a room. It has been demonstrated that the amplitude of late reverberant sound exhibits a normal distribution (as described by the central limit theorem Tervo (2012)) around a mean however early reflections tend to occupy values that are less normal i.e. more kurtoic Usher (2010). The normalised echo

density profile, $\eta(t)$, was proposed by Abel and Huang (2006) and is defined as the fraction of impulse response taps which lie outside the standard deviation in a finite time window. This is obtained by using a sliding analysis window over an impulse response $h(t)$.

$$\eta(t) = \frac{1}{erfc(1/\sqrt{2})} \sum_{\tau=t-\delta}^{t+\delta} w(\tau) 1\{|h(\tau) > \sigma|\} \tag{4.1}$$

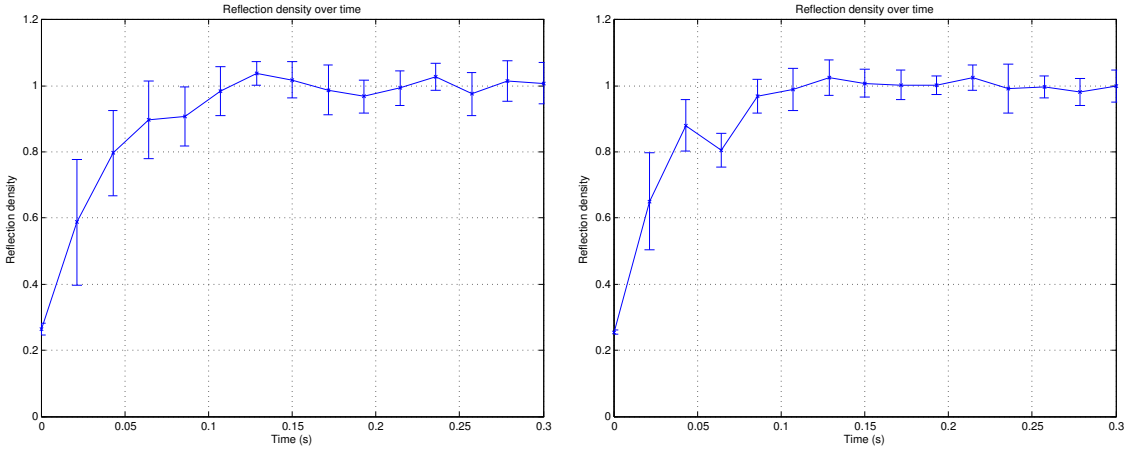
Where $w(\tau)$ is a weighting function i.e. a rectangular or Hanning shaped window of $2\delta + 1$ samples in length. σ is:

$$\sigma = \sqrt{\sum_{\tau=t-\delta}^{t+\delta} h^2(\tau))} \tag{4.2}$$

where $w(t)$ is normalised to have a unit sum $\sum_\tau w(\tau) = 1$. The indicator function$1\{|\cdot|\}$ returns a true or positive state when the current sample is outwith a standard deviation. The function is multiplied by $1/erfc(1/\sqrt{2}) = 0.3173$ where $erfc$ is the complementary error function which is the expected fraction of samples lying outside a standard deviation from the mean for a Gaussian distribution. This ensures that final function varies between 0 for a set of sparse reflections and close to 1 for fully diffuse reverberation.

The echo density profile can therefore be used to estimate the mixing time or the point at which reflections in the impulse response become sufficiently dense to be considered diffuse reverberation i.e. when the echo density profile first exceeds a value of 1. This is potentially very useful as in auditorium acoustics it is widely accepted that the mixing time occurs between 80ms and 100ms despite these numbers being known to be largely arbitrary.

Figure 4.1 below show the echo density profile for the first 0.3 seconds of impulse responses measured on a concert hall stage at stage left and stage centre positions. The impulse responses were measured as described in the previous section (using a rotating sound source) resulting in eight impulse responses per measurement location. The mean and standard deviation of these measurements are plotted for each location. In these cases, the plots show the reflection density is low at the beginning of the impulse response but quickly reaches a value of 1 shortly after a delay of $100ms$. After this point, the reflection density remains close to a value of 1 as the reverberant energy arrives at the microphone.



(a) Stage left

(b) Stage centre

FIGURE 4.1: *Plot showing Echo Density Profile over the first 0.3 seconds of impulse responses measured in the Ledger Recital Room. This parameter was obtained with a 20ms rectangular sliding window*

In both examples, the reflection density increases at different rates. The increase in arrival density at the stage centre position is slightly steeper than at stage left, implying that reflections are more clustered together at this position in contrast to the stage left position where they are increase more steadily over time. This is to be expected when comparing these two positions due to the increased likelihood of coincident reflections when the measurement is made equidistant from the side walls of the stage.

At the stage centre position, the mean reflection density decreases slightly before reaching a value of 1. A reduction in reflection density could be caused by a short period of silence before additional reflections arrived. Furthermore, in both examples, measurements made with different source orientations produce quite different reflection density profiles; as can be seen by observing the standard deviation. These differences provide a broad indication that the temporal distribution of early reflections varies at different positions on stage and with different source orientation.

It is of interest that in this example, both plots show the mean reflection density profile reaches a value of around 1 at about $100ms$ which corresponds with the upper integration limit of early objective support as discussed previously.

4.1.2 Reflection detection

In order to examine the temporal structure of early reflections, it is necessary to develop a reliable method of detecting the presence of early reflections in an impulse response. As shown in equation (4.3), the time of arrival $\hat{t_n}$ for each reflection can be estimated by applying a detection function, $D_n(t)$ to the measured impulse response. When $D_n(t)$ produces maxima, a reflection has been detected and therefore the time delay in relation to the beginning of the signal can be computed.

$$\hat{t_n} = t_{start} + \underset{t}{argmax}\{D_n(t)\} \tag{4.3}$$

The detection function $D_n(t)$ can be implemented in numerous ways. Correlation-based methods of reflection detection are typically based on the premise that reflected sound simply consists of delayed, filtered and attenuated versions of the direct sound. Therefore, it is possible to detect reflections by finding the maxima of correlation between the direct sound and the remainder of the impulse response. A measurement of the direct sound can be obtained either by measuring in free-field conditions or by windowing the direct sound from in-situ measurements. Similar approaches, namely Matching Pursuit (Defrance et al., 2008), perform an iterative cross correlation of the impulse response with a dictionary of atoms which represent the direct sound. After each iteration it subtracts the corresponding atom from the original impulse response and adds it to a sparse representation. It is normally set to repeat until the residual signal is low enough in amplitude.

Another estimation method, suggested by Tervo et al. (2010), is based on assessing the ratio of energy in sliding analysis windows of differing lengths. The analysis windows are shifted along the impulse response. When a reflection is encountered,a larger variation in energy will occur in the shorter analysis window than in the longer analysis window. When this occurs the ratio of energy in the analysis windows will increase, producing a peak in the detection function. The detection function is shown below in equation (4.4).

$$D_h^{Peak}(t) = \left[\frac{1}{2T_l}\int_{t-T_l}^{t+T_l} |h(\tau)|d\tau\right] / \left[\frac{1}{2T_g}\int_{t-T_g}^{t+T_g} |h(\tau)|d\tau\right] \tag{4.4}$$

where T_l and T_g are the local and global window lengths respectively, with window lengths of 1.3ms and 50ms. This function produces a peak when there there is a large difference between local and global time windows. Peaks which exceed a threshold value are identified as reflection arrivals. The threshold value can be obtained by the value of $D_h^{Peak}(t)$ during measurement of the noise floor.

An alternative method utilises the Hilbert transform which produces an analytical signal which preserves the relative amplitude of each reflection allowing the time of arrival and the amplitude envelope to be observed. The amplitude envelope is obtained from the impulse response as in equation (4.5):

$$e(t) = |h(t) + j\hat{h}(t)| \tag{4.5}$$

where $\hat{h}(t)$ is the Hilbert transform of the impulse response $h(t)$

$$\hat{h}(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{h(\tau)d\tau}{t-\tau} \tag{4.6}$$

Local maxima are then detected by a peak finding algorithm which identifies peaks if they are exceed surrounding data by a given amount (Yoder, 2014). Figure 4.2 shows an example of the results of this analysis. The figure shows the Hilbert transform of an impulse response after being normalised to a maximum value of 1. The red marks on the reflections identify reflections which have been identified by the peak finding algorithm.

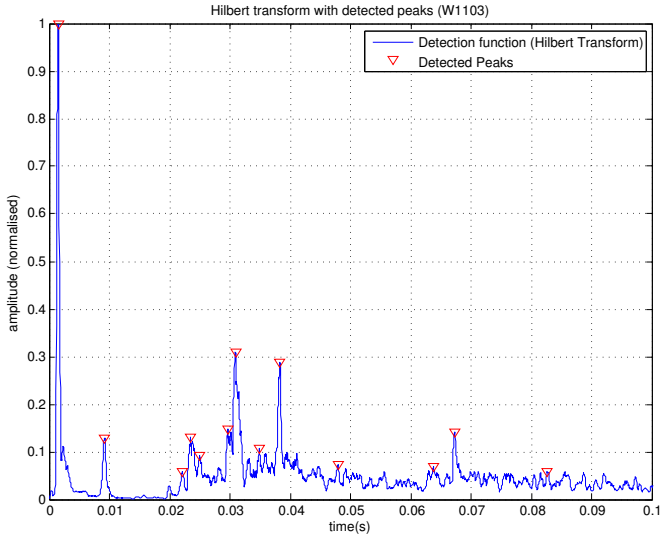


FIGURE 4.2: *Amplitude envelope of impulse responses measured at Front stage centre position in Ledger Recital Room. The amplitude envelope is obtained by applying the Hilbert transform to the W-channel of the impulse response. The detected reflections, denoted by red triangles, are detected using a peak detector.*

4.1.3 Temporal distribution of early reflections

A concert hall may be perceived differently if the musician hears reflections that arrive mainly within a short time frame; compared with the reflections being more spread out over time. Furthermore, on a small stage, a tightly clustered group of early reflections may arrive after a short time delay relative to the direct sound, whereas on a larger stage, the reflections may arrive clustered together but occur much later in time relative to the direct sound.

In order to characterise the temporal distribution of reflections, a set of descriptors are required which are able to quantify how clustered the reflections are and the time of arrival of this cluster relative to the direct sound.

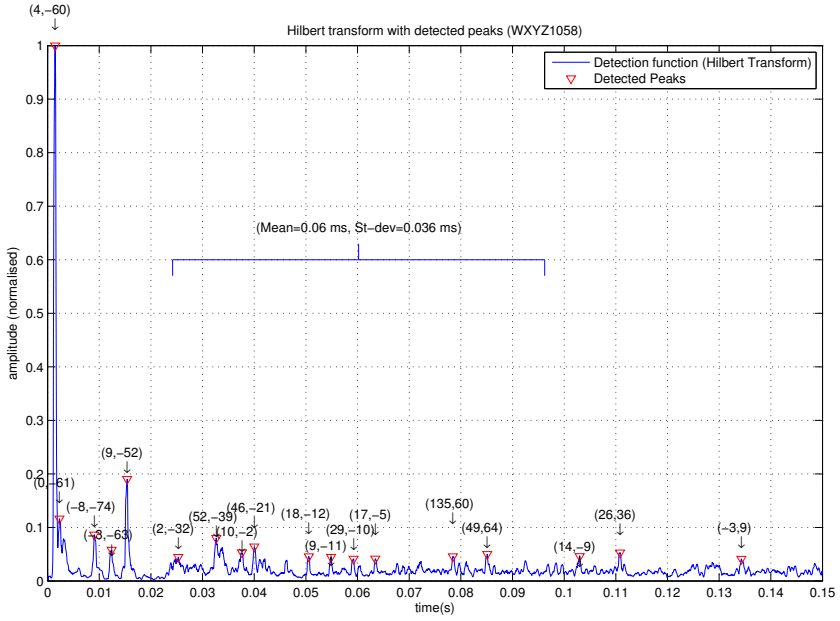
Previous research by Jeon et al. (2014) into preferred stage acoustic condition for soloists found there was a general preference for musicians playing from a centre stage position. It was speculated that this was due to reflections arriving from many directions, clustered close together temporally. Therefore, Jeon evaluated the temporal variation of reflections by deriving the standard deviation of reflection arrival times. The standard deviation provided a measure of how clustered together the reflections are in time.

Also in reference to stage acoustics, Miranda Jofre et al. (2013) performed a number of listening tests whereby a singer was asked to rate a number of interactively auralised soundfields, where

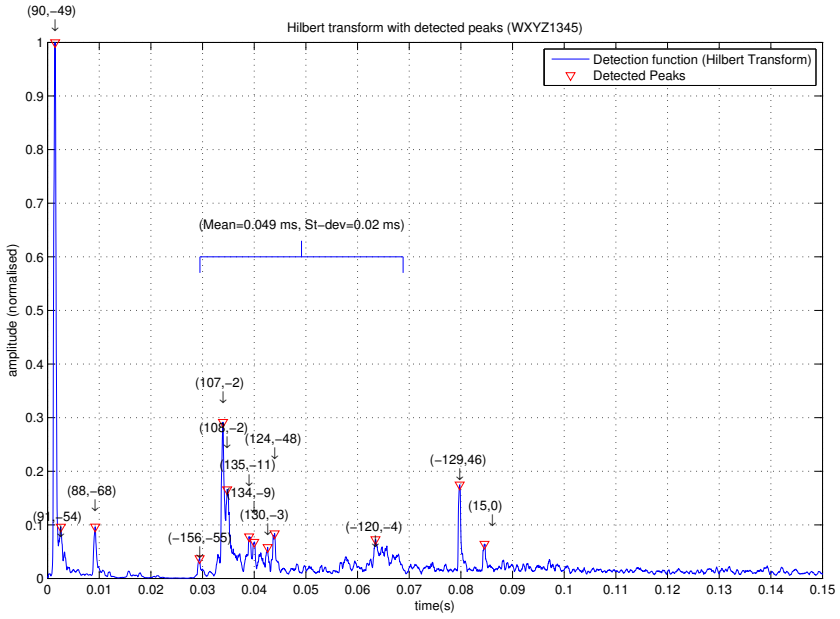
each soundfield was representative of on-stage concert hall acoustics. The soundfields were controlled parametrically by the spatial and temporal distribution of early reflections which were controlled using 'centroid' and 'diffusivity' parameters. The temporal centroid was a delay time around which the reflections could be spread out in time by varying their diffusivity. These parameters appear very similar to those proposed by Jeon et al. (2014).

As mentioned previously, it is feasible that a tight cluster of reflections could occur at any time delay after the direct sound which could also influence the perceived acoustic conditions on stage. Therefore, the method proposed by Jeon is expanded slightly in this research by including the mean time of arrival to determine when the majority of reflections have arrived.

An example of this is shown in Figure 4.3 which shows the reflections detected in two different concert halls. As before, detected reflections are indicated with a red arrow. In addition, the t_{mean} and t_{σ} is represented using the horizontal line in the centre of the plot. The centre of the line and the length of the line represents the mean and standard deviation respectively.



(a) Hall 1



(b) Hall 2

FIGURE 4.3: *Amplitude envelope of impulse responses measured in different halls. Detected reflections are denoted with a red triangle. The mean and standard deviation are represented by a brace above the amplitude data. Hall1 shows the reflections arriving spread over a longer time than in Hall 2 where the reflections are clustered together. The temporal structure is characterised by the mean and standard deviation of the times of arrival.*

In Hall 1 it can be seen from the amplitude envelope that reflections are spread out over time whereas in Hall 2 the reflections are localised in clusters. This is characterised by the mean (t_{mean}) and standard deviation (t_σ) of time of arrival. In Hall1, $t_{mean} = 60ms$ and $t_\sigma = 36ms$

which suggests reflections are spread out considerably over time. In Hall 2, $t_{mean} = 49ms$ and $t_{\sigma} = 20ms$, implying that reflections tend to occur much earlier in this hall and are clustered more closely together.

4.1.4 Discussion

This section discussed how the temporal distribution of early reflections could be extracted from a measured impulse response. There are numerous methods of detecting the presence of early reflections.

Stage acoustic measurements made as part of this study use a microphone positioned above a directional loudspeaker. This results in the direct sound measured by the Ambisonic microphone being markedly different to the direct sound emanating from the front of the loudspeaker. Correlation methods have been found not to work as well in this situation as the direct sound and reflections are sufficiently different for the correlation to remain low.

However, the Hilbert transform method was considered to be especially effective as it produces a visual representation of the impulse response amplitude envelope. In conjunction with a peak-finding algorithm, early reflections can be isolated accurately in time. Consequently, this technique was considered appropriate for use in analysing stage acoustic impulse responses for this research.

It was also discussed how the temporal distribution of early reflections could be characterised by a limited set of parameters. In agreement with previous research by Jeon et al. (2014) and Miranda Jofre et al. (2013) it was determined that the central tendency and temporal spread of early reflections were appropriate parameters to use for this research. The approach demonstrated by Jeon et al. (2014) was adopted and augmented to incorporate the mean time of arrival and standard deviation of time of arrival.

4.2 Spatial analysis methods

As discussed previously in this chapter, ST_{early} is independent of the spatial distribution of early reflections. Consequently, early reflections could exhibit different spatial distributions but still result in the same measured early objective support. Reflections arriving from a particular direction may be masked by the sound from the musician's instrument, which could influence their impression of the acoustic conditions.

As this research is concerned with the effect of the spatial distribution of early reflections, it is necessary to be able to examine the spatial distribution of early reflections both visually and described in terms of a set of objective parameters. This section will therefore review existing methods of analysing the spatial distribution of early reflections and propose an appropriate method based for this research.

4.2.1 Directional energy parameters

As discussed in Chapter 2, for audience members, a feeling of spaciousness adds an attractive quality to music played in a concert hall. An element of spaciousness is the feeling envelopment or being surrounded by the acoustics of the room. It is widely known that late lateral reflections arriving at the listener increase the perceived feeling of envelopment. This has led to the development of the Lateral Energy Fraction (LEF) which measures the energy ratio of reflections arriving between 5ms and 80ms from a lateral direction in relation to the energy received between 0ms and 80ms from all directions. To measure this parameter, a coincident pair of microphones are required where one is omnidirectional and the other has dipole characteristics and is oriented with its lobes pointing stage left and right.

Dammerud (2009) previously suggested the use of LEF for use in stage acoustic research relating to orchestras. It is suggested that a directionally dependent parameter could be used for assessing the relative level of early reflections in particular directions which may be relevant for feelings of *'projection'* on stage. Dammerud also comments on the lack of flexibility and angular resolution inherent in this approach. He suggests that Ambisonic microphones could be used to more effectively observe the spatial distribution of early energy. By processing the Ambisonic signals, a virtual microphone of specific directivity and orientation could be derived allowing directional equivalents of existing stage acoustic parameters to be derived. This use of virtual microphones is advantageous as many different directions can be sampled from only a single measurement.

For example, ST_{early} could be adjusted so that the numerator in equation (2.29) represents the response measured with a directional virtual microphone pointing in a particular direction. This was demonstrated previously by McCarthy et al. (2008) where a number of B-format impulse responses were measured on stage. From each impulse response, a directional response was obtained by deriving a virtual cardioid microphone pointed in six cartesian directions. By comparing the directional and omnidirectional responses, a directional early objective support, $ST_{early,dir}$, could be obtained using equation (4.7).

$$ST_{early,dir} = 10log_{10}\left(\frac{\int_{20ms}^{100ms} p_{dir}^2(t)dt}{\int_{0ms}^{10ms} p_{omni}^2(t)dt}\right) \tag{4.7}$$

Where p_{dir} is the sound pressure as measured with the directional microphone and p_{omni} is the sound pressure as measured by the omnidirectional microphone. It is noted that the signal measured with the directional microphone response is strictly a pressure-gradient signal as opposed to pressure (Cabrera et al., 2012).

It is noted (as it was by McCarthy et al. (2008)) that the diffuse energy response of a cardioid microphone is a 1/3 compared to the response of 1 for an omnidirectional microphone. This would result in a $-4.77dB$ difference between omnidirectional and cardioid measurements if the early energy was spatially diffuse. McCarthy recommends increasing the directional signal by $4.77dB$ to compensate for this and to allow the directional measurements to have a similar average value to the omnidirectional measurements.

The directional parameters can be studied quantitatively by selecting a small number of directions (as demonstrated by McCarthy et al. (2008)) making it possible to compare, for example, the level of early objective support arriving from the up stage direction at various locations around stage. It is also possible to derive a heat map of early objective support by assessing it at every direction of arrival and displaying as a heat map. This can be useful for visualising the directional distribution of the early energy.

An example of this type of visualisation is shown in Figure 4.4. The plot is derived by sequentially pointing a cardioid virtual microphone to each angle of azimuth and elevation and calculating $ST_{early,dir}$. The angular distribution of $ST_{early,dir}$ is represented as an unwrapped heat map where X and Y axes show azimuth and elevation respectively. In this case, the origin $(0°, 0°)$ is centred in the plot and is associated with the stage front direction. The example in Figure 4.4 is an impulse response measured at a stage centre location where the sound source is oriented at 90° (towards stage right). maximum and minimum levels of $ST_{early,dir}$ are indicated with green and blue points respectively. It can be seen in this case that the majority of energy arrives from an azimuth of 104°.

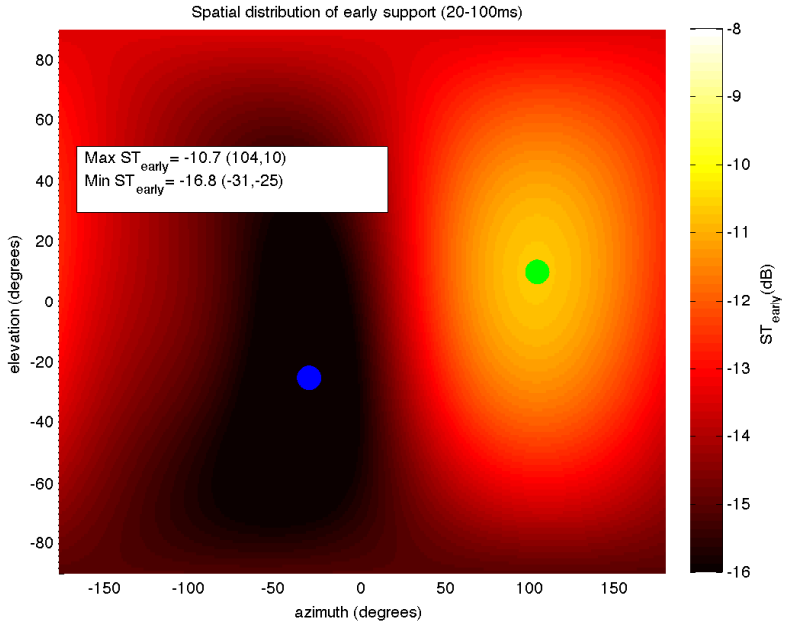


FIGURE 4.4: *Spatial distribution of ST_{early} (20-100ms) at stage centre in the Ledger Recital Room with source orientation. The brighter areas show areas of increased energy whereas the darker areas show reduced early energy*

Guthrie (2014) also adopted this approach for observing the directional distribution of early energy. Guthrie extended this concept to all stage acoustic parameters obtaining directional distributions in specific look directions. Furthermore, a number of the directional parameters were related to each other. For instance, one parameter used was $ST(Top/Sides)$ which related the level of early energy from above the performer to the energy received from the side walls. As mentioned previously, in Guthrie's research the acoustic response on stage was measured with an HOA microphone. By encoding the soundfield with higher order spherical harmonics, it is

possible to derive virtual microphones with a higher directivity and so the directional distribution of each acoustic parameter could be observed at a much finer resolution.

This technique is a straightforward method of adapting existing acoustic parameters to show their directional distribution. However, in the context of a directional variation of early objective support, it is not straightforward to see how the directional distribution develops over time. In the previous section, it was demonstrated how early reflections could be isolated in time. By assessing the B-format signals at the times of reflection arrival, it should be possible to estimate the angle of arrival of each reflection, thus allowing the spatial distribution of early reflections to be analysed.

4.2.2 Intensity vector analysis

A further set of methods involves assessing the B-format signals at the reflection time of arrival, it is possible to estimate it's direction of arrival. This can be estimated by evaluating the active sound intensity, $I(t)$ which is a product of the sound pressure $p(t)$ and particle velocity vector $u(t)$ as shown in equation (4.8). This describes the transfer of energy of the soundfield and therefore the opposing vector will describe the direction of arrival of the sound.

$$I(t) = p(t)u(t) \tag{4.8}$$

As discussed in Chapter 3, the four channels of Ambisonic B-format consist of the the omni-directional signal, $W(t)$, and the mutually orthogonal, figure-of-eight pressure-gradient signals $X(t)$, $Y(t)$ and $Z(t)$. While the pressure gradient channels do not measure particle velocity directly they are assumed to measure a property proportional to the components of the particle velocity $u(t)$ (Merimaa and Pulkki, 2004). Similarly, the pressure channel, $W(t)$ is assumed to be proportional to the sound pressure $p(t)$. The active intensity (and therefore the direction of arrival) can be estimated from a B-format recording. As discussed earlier in this chapter, the active intensity can also be measured using an intensity probe as demonstrated by Tervo (2012). In the Fourier domain the active intensity can be estimated using the equation (4.9) below:

$$I_\alpha(\omega) = \frac{\sqrt{2}}{Z_0}\Re\{W^*(\omega)\acute{X}(\omega)\} \tag{4.9}$$

Where $\acute{X}(\omega) = (X(\omega)e_x + Y(\omega)e_y + Z(\omega)e_z)$, '$*$' denotes complex conjugation and $Z_0 = \rho_0 c$ is the acoustic impedance of air.

By calculating the active intensity of each reflection, it is possible to localise it in space. In the context of this research, this will enable all the early reflections to be spatially localised. It is common for this type of analysis to be applied in frequency bands, which can be achieved using a filter bank model where the B-format channels are separated into different frequency bands prior to directional estimation. Alternatively, the analysis can take place in the time-frequency domain by making use of the Short Time Fourier Transform (STFT), producing an estimated direction of arrival for each time-frequency bin.

Once the active intensity has been estimated, the direction of arrival is derived using equation (4.10) and (4.11) giving azimuth and elevation respectively. It should be noted that the active intensity points in the direction of flow of sound energy and so direction of arrival is calculated from the opposing vector angles.

$$\theta(\omega) = \tan^{-1}\left[\frac{-I_y(\omega)}{-I_x(\omega)}\right] \tag{4.10}$$

$$\phi(\omega) = \tan^{-1}\left[\frac{-I_z(\omega)}{\sqrt{(I_x^2(\omega) + I_y^2(\omega))}}\right] \tag{4.11}$$

Where θ is the angle of azimuth and ϕ is the angle of elevation. I_x for instance refers to the vector component in the x direction.

Figure 4.5 shows an example of this analysis as applied to a measured stage acoustic impulse response. This plot shows a time-frequency representation (spectrogram) of the impulse response that has been limited between $0Hz$ and $5kHz$ and between $0ms$ and $100ms$. The pressure response is shown as a colour axis where darker colours represent greater signal amplitude. For clarity, time-frequency bins with a sound pressure level lower than $-30dBFS$ have been omitted from this plot. Vectors are shown overlaid onto the spectrogram showing the estimated direction of arrival for each time-frequency bin. In this plot the vectors show azimuth only. The vectors will be parallel to the increasing x-axis when the sound appears in front of the microphone. When sound arrives from the left of the microphone the vectors will be parallel to the increasing y-axis.

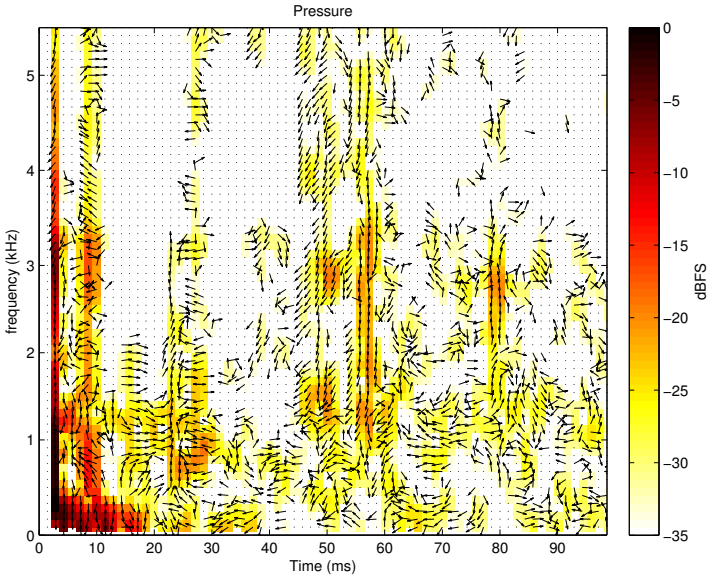


FIGURE 4.5: *Spectrogram of the first 0.1 seconds of measured impulse responses overlaid with a quiver plot showing the direction of arrival (azimuth only) of each time-frequency bin. For clarity, sounds of lower amplitude than $-30dBFS$ have been omitted from this plot*

It can be seen that the direct sound occurs at a time delay of 2ms followed by the floor reflection at a delay of 9ms. Further reflections occur at 25ms, 50ms, 57ms and 80ms. At these time delays, it can be seen that the intensity vectors tend to point in a similar direction. It is clear, for example, that the reflection occurring at 57ms arrived from the right of the microphone whereas the reflection arriving at 80ms arrived from the left of the microphone.

Due primarily to the capsule spacing of Ambisonic microphones, an angular error is introduced which increases as the wavelength approaches the dimensions of the microphone capsule. This results in a maximum frequency with which the direction of arrival can be accurately computed (Vilkamo, 2008). A similar study by Protheroe and Guillemin (2013) found that that the directional accuracy of energy vectors reduces significantly for sound above a frequency of 5kHz where microphone capsules can no longer be considered coincident. This cut off frequency was confirmed by experimentation as documented in Appendix B.

In addition, when using a short time window and low hop size settings for the STFT, a single reflection may be identified a number of times with slightly different results per iteration. This results in the direction of arrival of each reflection being represented by a distribution of estimates. Furthermore, it is highly likely that two or more reflections will arrive in a single time-window of the analysis, introducing estimation errors. This can cause a number of overlapping distributions describing the angle of arrival at a particular time estimate. This can make it difficult to obtain a reliable indication of the angle of arrival of each reflection.

By analysing the distribution of intensity vectors it is possible to obtain an estimate of the correct angle of arrival. This can be achieved by finding the maximum value of a Kernel Density Estimate (KDE), which is similar to observing the maximum value of a histogram. This is shown in Figure 4.6, where the KDE of estimated angles reveals the most common estimated angle of arrival. While out-with the scope of this discussion, it has been demonstrated (Tervo, 2012) that the histogram data can also be weighted to improve the accuracy of the estimation.

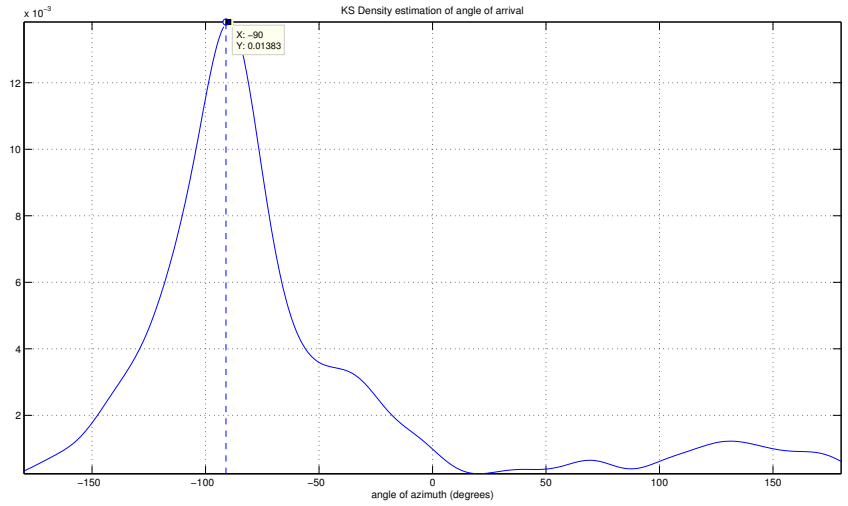


FIGURE 4.6: *Example of Kernel Smoothing Density Estimate being used to indicate the angle of arrival of a reflection which, in this case, arrived from 90°. The dashed line denotes the angle of arrival associated with the maximum value of the estimation.*

This technique is very powerful and can reveal the direction of arrival and spatial diffuseness of each reflection. The type of plots produced with this type of analysis (as shown in Figure 4.5) however can be difficult to interpret and can be difficult to characterise using objective acoustic parameters. Therefore the following discussion will focus on how these data can be organised such that a clear visualisation can be obtained and objective acoustic parameters extracted.

4.2.3 Directional energy analysis

A more recent method of analysing the spatial development of arriving energy can be achieved using directional energy histograms. The advantage of this technique is that it is capable of displaying the relative energy of incoming energy in arbitrary time windows as well as the variation over time. This technique, inspired by a similar technique proposed by Patynen et al. (2013), performs a directional analysis of the impulse response between $t_0 = 10ms$ and a delay variable, τ, which in this case is set at 20ms, 50ms, 100ms and 200ms. Currently, the estimated angle of arrival is obtained from the angle of the mean resultant vector.

The results of the directional analysis attribute a direction of arrival for every pressure value $h(t)$ allowing a directional energy histogram to be produced. The technique has been adapted for stage acoustics by relating the amplitude arriving from each angle to the direct sound and floor reflection so it resembles early objective support. A further adaption to the technique is that it performs the directional analysis over very short time windows. The published method uses a technique known as Spatial Decomposition Method (SDM) which is capable of obtaining a directional estimate for every sample.

$$(h^{'}(t|\hat{\theta}(t),\hat{\phi}(t))) \triangleq [h(t),\hat{\theta}(t),\hat{\phi}(t))] = DIR\{h(t)\} \tag{4.12}$$

$$ST_{0,\tau}^{DIR}(\phi) = 10log_{10}\left(\frac{\int_{t=t_0}^{\tau+t_0}\left[w_{med}(\hat{\theta(t)},\hat{\phi(t)})h^{'}(t|\hat{\theta}(t),\hat{\phi(t)}) = \phi\right]^2 dt}{\int_{t=0}^{10ms} h(t)^2 dt}\right) \tag{4.13}$$

$$ST_{0,\tau}^{DIR}(\theta) = 10log_{10}\left(\frac{\int_{t=t_0}^{\tau+t_0}\left[w_{lat}(\hat{\phi(t)})h^{'}(t|\hat{\theta}(t),= \theta,\hat{\phi}(t))\right]^2 dt}{\int_{t=0}^{10ms} h(t)^2 dt}\right) \tag{4.14}$$

where $h(t)^2$ refers to sound arriving before 10ms (i.e.the direct sound and floor reflection), $\hat{\theta}$ is the estimated angle of arrival (azimuth).

A toroidal weighting factor, $w(\hat{\theta(t)},\hat{\phi(t)})$, is applied to the energy analysis to suppress the effect of energy in the lateral or median plane as required. These weighting factors are shown in equations (4.15) and (4.16) when observing the median and lateral planes respectively.

$$w_{med}(\theta,\phi) = |1-\cos(\phi)||\sin(\theta-(\frac{\pi}{2}))| + |\sin(\phi+(\frac{\pi}{2}))| \tag{4.15}$$

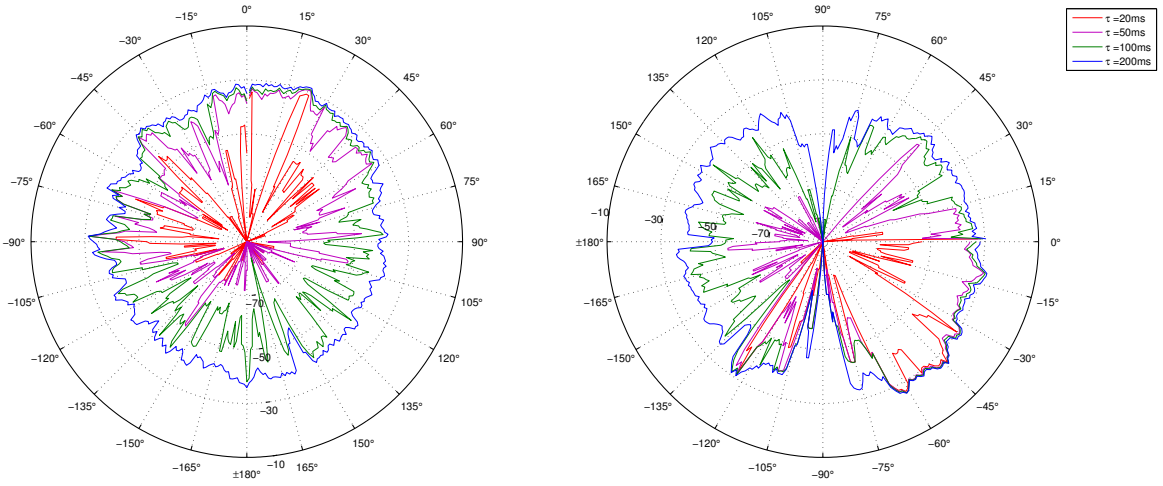
$$w_{lat}(\phi) = |\cos(\phi)| \tag{4.16}$$

For clarity, the analysis process is summarised in the steps below:

1. Directional analysis (SIRR) performed for longest region of interest of the impulse response. Estimated angle is obtained from mean resultant vector and broadband sound pressure is obtained from the average energy in each time window.
2. Attribute pressure value with estimated angle of arrival (rounded to nearest degree)
3. Isolate section of impulse response between $10-\tau$ ms
4. Sort pressure data into bins (angle of arrival over time)
5. Integrate each bin over window length τ
6. Express as dB related to pressure of direct sound and floor reflection
7. Smooth results with a 2-sample moving average filter
8. Plot as polar diagram

Figure 4.7 shows an example of the directional analysis method described above. These plots show the development of early energy over four time regions in relation to the direct sound. Figure 4.7(a) shows the distribution of early energy in the lateral plane (i.e. looking down on the musician), stage front is oriented to the top of this figure. Figure 4.7(b) shows the same distribution but in the median plane (i.e. looking through the musician from right to left), stage front is oriented to the right of this figure and the top of the figure points to the ceiling.

The energy in individual time windows can be seen with the different coloured traces. The red trace shows the energy between 10 and 20ms, the purple trace shows the energy between 10 and 50ms, the green trace between 20 and 100ms and the blue trace shows the distribution of energy between 20ms and 200ms.



(a) Lateral plane

(b) Median plane

FIGURE 4.7: *Example of Directional analysis results obtained at front stage centre of the Reid Concert Hall, Edinburgh. In this measurement, the sound source was oriented at* 0° *i.e. straight towards the audience rear wall. In figure 4.7(a)* 0° *points to stage front and* -90° *points to stage left. In figure 4.7(b)* 90° *points vertically while* -180° *points to stage rear.*

In this particular example it can be seen that between 10ms and 20ms, energy arrives mainly from in front and below the measurement position. By 50ms, the early energy begins to arrive also from above and in front of the measurement position. By 100ms the arriving energy is beginning to arrive almost equally from all directions and by 200ms the energy is largely distributed equally around all angles of arrival.

This method produces an intuitive plot showing how the directional distribution of reflections develops over set time intervals. However, it is difficult to attribute quantitative values to describe the development over time. Furthermore, a considerable amount of information is displayed making the plot challenging to interpret.

4.2.4 Image source plots

An alternative method of visualising the direction distribution of reflections have been demonstrated by Bassuet (2010) and Protheroe and Guillemin (2013), which displays each reflection as a vector on a 3D graphic producing a so called 'hedgehog plot', as shown in Figure 4.8.

In this example, the time of arrival of early reflections is coded by colour and vectors show the direction and broadband amplitude of sound energy. This results in a much clearer representation of the impulse response without significant loss of information and is particularly useful to observe how the impulse response evolves over time.

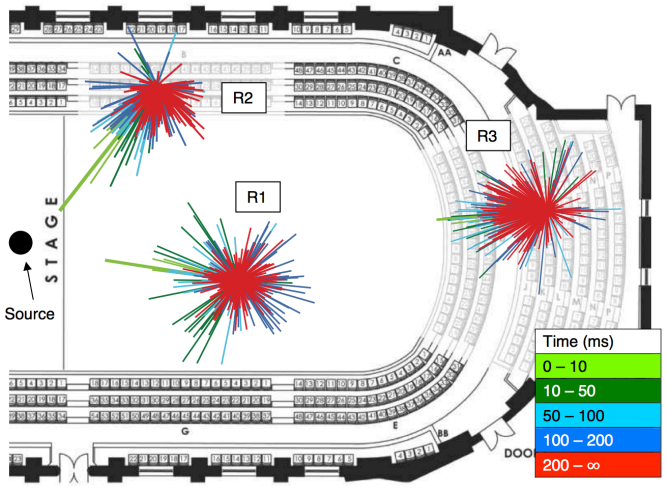


FIGURE 4.8: *Figure showing an example of a hedgehog plot as viewed from above. Spatial distribution of early reflections are displayed as coloured vectors where colour determines time of arrival and length determines amplitude. In this example, the difference in spatial distribution of reflections over time can be seen clearly. Reproduced from (Protheroe and Guillemin, 2013)*

This method produces similar issues to the directional energy analysis described earlier in that the temporal information of individual reflections is lost and the plots can, in some cases, be difficult to interpret.

As discussed previously, in this research it is necessary to extract the spatial and temporal distribution of early reflections in terms of an intuitive visualisation and in terms of a number of objective acoustic parameters. By combining the intensity vector analysis and the reflection detection method described earlier in this chapter. It is possible to derive the spatial and temporal information of each reflection. With these data it is possible to visualise both the spatial and temporal distribution of early reflections as an image source plot.

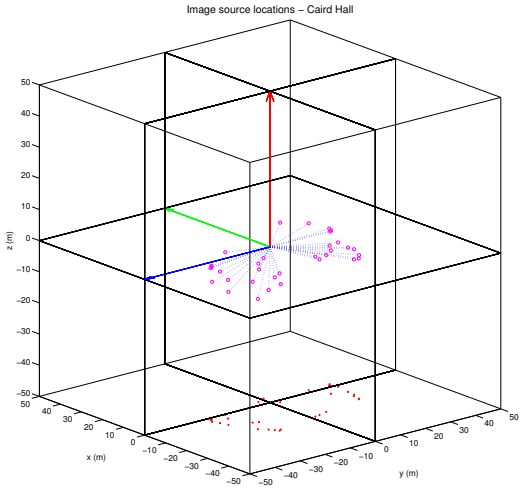
This is based around the concepts used in image source acoustic models which assume a geometric model of sound propagation. The image source model assumes each reflecting surface acts as a mirror and so determines, the position of mirror images of the sound source in order to determine reflection path lengths and angles of arrival to the receiver (Allen and Berkley, 1979). By using the reflection detection method and the intensity vector methods described previously, it is possible to determine the spatial location of image sources associated with each early reflection captured on stage. This information can then be used to display the reflections as a 3 dimensional point cloud. The time-of-flight of each reflection is represented as the distance of the each image source from the origin, r_n, which can be calculated using equation (4.17)

$$r_n = c \cdot \hat{t_n} \tag{4.17}$$

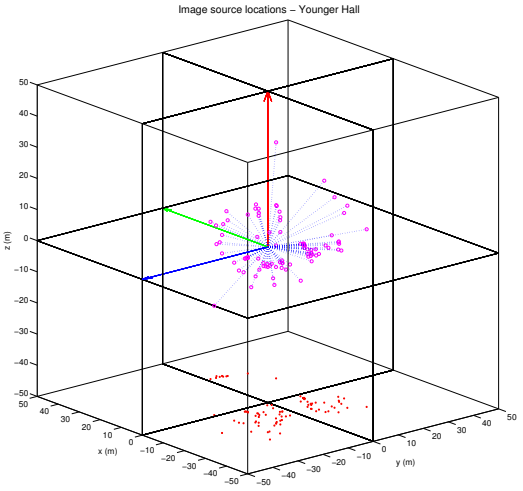
where c is the speed of sound in air and $\hat{t_n}$ is the time of flight of the n^{th} reflection

Examples of the image source plots are shown in Figure 4.9 which are produced from measurements made in the Younger Hall and Caird Hall, both from a front stage centre position. In each plot, the origin represents the location of the Ambisonic microphone.

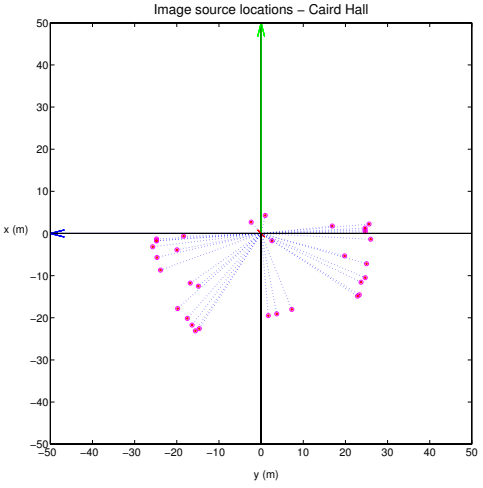
The green axis pointing in the positive X direction is pointing to the stage front, whereas the blue axis pointing in the positive Y direction is pointing to stage left. Each image source is shown in 3D space using a purple marker. For clarity the X-Y position is shown as red markers. These plots were constructed by combining the image source data from all measurements at each location; where multiple measurements were made with the sound source positioned at different orientations.



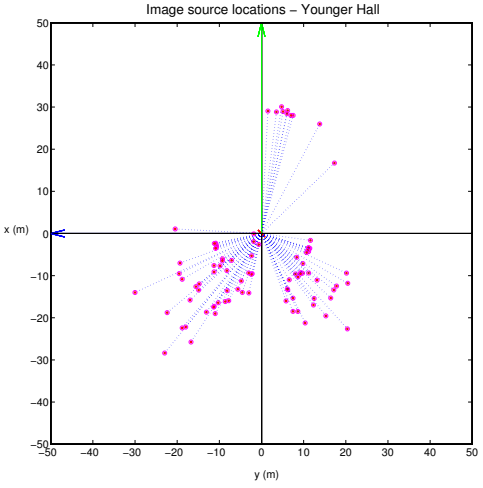
(a) Caird Hall - Profile



(b) Younger Hall - Profile



(c) Caird Hall - Plan



(d) Younger Hall - Plan

FIGURE 4.9: *Image source plots of the Caird Hall in profile (4.9(a)) and plan views 4.9(c) respectively. Also Image source plots of the Younger Hall in profile (4.9(b)) and plan views 4.9(d) respectively. Purple markers show position of image sources in 3D space. Red markers show their position in the X-Y plane. All plots were obtained by layering the results of multiple measurements as described previously*

The image source plots for the Caird Hall show reflections arriving only from behind and laterally. This position in the Caird Hall was 12.48m and 12.41m from the stage left and stage right walls

respectively. It was also 9.73m in front of the stage rear wall. The image sources appear at twice these distances, reflecting the total propagation distance of sound. Caird Hall is a large shoebox, symphonic concert hall seating a large number of people. The rear wall is therefore too far away to appear on this plot.

In the Younger Hall, the majority of image sources appear from behind the measurement positions, however there is also a cluster of image sources directly in front of the measurement position. The reflections from in front are caused by a balcony front which occurs at a similar height to the sound source on stage. The reflections arriving from the rear are clustered in lines angled away from the Y axis. This is caused by the stage enclosure in this hall featuring angled side walls. Two sets of these clusters can be seen on either side of the Y-axis indicative of 2nd order reflections (i.e. reflections that have encountered both side walls).

This method produces a simple visualisation of early reflections showing both the temporal and spatial information. Furthermore, this type of plot is commonly encountered by practitioners in the construction of acoustic models. This method could therefore be easily used in industry and allows for clear comparisons of different acoustic conditions.

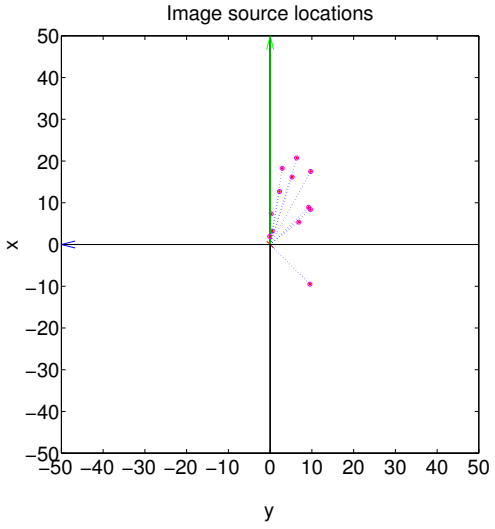
4.2.5 Spatial distribution of early reflections

In order to objectively compare the distribution of early reflections it is necessary to conceive a set of parameters which describe the spatial arrangement of detected image sources. The temporal analysis of early reflections was discussed earlier where the aim was to compute a mean and standard deviation of the time of arrival for detected reflections. It is possible to apply a similar approach to describe the central tendency and spread of early reflections in the spatial domain.

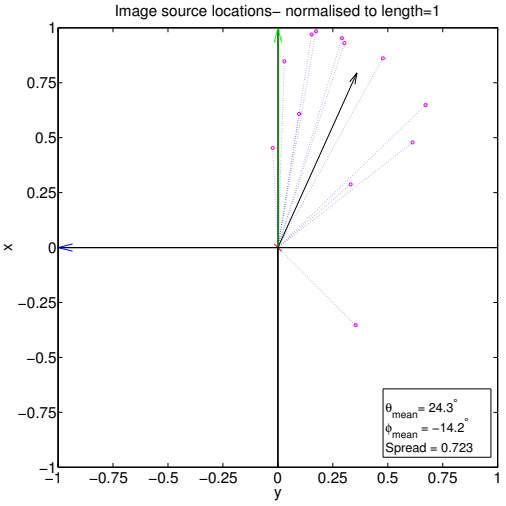
The spatial location of each image source is represented using spherical coordinates, (r, θ, ϕ). When examining their spatial distribution the temporal information, represented by the radius r, is discarded such that each point lies on a unit sphere. This allows the spatial distribution to be assessed in terms of a central tendency and spread of points. As the data is circular, i.e. azimuth is wrapped around a value of 0°, the traditional mean calculation would produce incorrect data. An angular mean can be computed instead by vector addition. In this case vectors are unit length and point in the direct of each detected reflection. The angular mean is computed by converting the spherical coordinates of the reflections into cartesian coordinates and summing the vector components to obtain a resultant vector which points in the mean direction of the data.

$$\begin{aligned} S_x &= \sum_i^n x_i \\ S_y &= \sum_i^n y_i \\ S_z &= \sum_i^n z_i \end{aligned} \tag{4.18}$$

Where x_i, y_i, z_i are the directional cosines of the spherical coordinates (θ, ϕ). Figure 4.10(a) demonstrates this by showing the image source plot measured at the front stage centre position of the Reid Hall with the source oriented at 0° azimuth. The image source plot is shown in plan with the axes showing the location of each image source in metres. In Figure 4.10(b), the temporal distribution of the image sources is disregarded and all image sources are represented by unit length vectors. The black arrow shows the mean resultant vector (normalised to unit length) with the mean azimuth θ_{mean}, elevation ϕ_{mean} and *spread* shown in the bottom right corner of the plot. It can be seen that this vector is oriented in the mean direction of the image sources. The unit vectors that appear in the graphic are distributed in elevation also and so appear to be different lengths when all have been normalised.



(a) Reid Hall - Front stage centre



(b) Normalised vectors and mean resultant vector

FIGURE 4.10: *Figure 4.10(a) shows the image source plot measured at the front stage centre position with the source oriented at 0° azimuth. The plot axes show the distance of each image source. The green arrow is pointing towards the rear of the auditorium. Figure 4.10(a) shows the same distribution of image sources with unit vectors. The normalised mean resultant vector is shown as a black arrow which points in the mean direction of image sources.*

The length of the vector R varies between 0 and n and can provide a measure of dispersion or spread of the reflection data, as shown in equation (4.19).

$$R = \sqrt{S_x + S_y + S_z} \tag{4.19}$$

The mean resultant vector length $\bar{R}$ is defined below (equation (4.20)):

$$\bar{R} = \frac{R}{n} \tag{4.20}$$

The value of $\bar{R}$ varies between 1 if all the angles are coincident; and a value of 0 if the angles are spread over the sphere. In the example shown in Figure 4.10 it can be seen that the image

sources appear to arrive from a similar direction but are spread out over a limited angle. This is reflected in the spread value of 0.723.

In conjunction with the temporal analysis proposed earlier, it can be seen that this method provides a concise description of the spatial distribution of early reflections.

4.2.6 Discussion

The preceding discussion reviewed a number of different methods of characterising the spatial distribution of early reflections detected in an impulse response. It was shown that, virtual microphones and estimation of intensity vectors could provide different representations of the spatial distribution of early reflections.

It was found that intensity vector estimation produced accurate estimates of the direction of arrival of individual early reflections. Estimation errors caused by the construction of the Ambisonic microphone could be averted by using Kernel Density Estimation to determine the most likely direction of arrival for each reflection. When combined with the temporal information, it was possible to construct a 3 dimensional point cloud showing the spatial distribution of early reflections. This resulted in an intuitive visualisation of early reflection distribution. It was also found that directional statistics could be applied to the spatial data from each reflection, allowing the spatial distribution to be characterised in terms of a mean direction of arrival (in azimuth and elevation) and spatial spread. As this approach allowed for a clear visualisation of the spatial and temporal distribution of early reflections and also allowed spatial and temporal acoustic parameters to be extracted, it was considered to be suitable for use in this research.

4.3 Summary and discussion

This chapter has reviewed currently available methods of assessing the temporal and spatial distribution of early reflections measured on stage. It was discussed how the temporal structure of a room impulse response could be broken down into early and late regions segregated by the mixing time. The mixing time could be estimated by assessing reflection density over time. It was further discussed how a more detailed analysis of the temporal distribution was possible by isolating individual reflections and calculating their time delay relative to the direct sound. This was shown to be possible by obtaining the amplitude envelope of the impulse response (using the Hilbert transform) and using a peak-finding algorithm to isolate the time of arrival of each reflection.

It was also discussed how the spatial structure of early reflections could be assessed. It was shown how the spatial distribution of early energy could be obtained by adapting traditional acoustic parameters (such as ST_{early}) to include spatial information. This approach has been used previously by Guthrie (2014) and McCarthy et al. (2008) and works especially well with impulse responses captured with Ambisonic impulse responses. By using virtual microphones, derived from the B-format impulse responses, it is possible to determine the proportion of reflected energy arriving from particular directions. This approach provides a straightforward

method of assessing the spatial distribution of early reflections and could easily be adapted to work with most other acoustic parameters. However, when used to assess the spatial distribution of ST_{early}, this approach continues to discard any temporal information from the analysis.

It was also demonstrated how the direction of arrival of individual reflections could be estimated from Ambisonic impulse responses by estimating the active intensity. The time and direction of arrival of early reflections could then be used to construct an image source plot showing the spatio-temporal distribution of early reflections. The spatio-temporal development of a measured impulse response was found to be well represented by using a directional energy histogram. These methods provided a useful method of qualitatively assessing the distribution of early reflections on different stages.

It was found that this analysis worked particularly well with the measurement technique described in Chapter 3. As previously reported by Tervo (2012), the directional sound source reduced the likelihood that multiple early reflections would arrive at a similar amplitude (as would be the case with an omnidirectional sound source). This reduced the potential for erroneous directional estimation.

By considering the spatial and temporal distributions separately, some simple measures were developed in order to make quantitative comparisons of different stage acoustic measurements. These measures assessed the central tendency and the temporal spread of measured early reflections both spatial and temporally. The temporal distribution was described using the mean time of arrival (t_{mean}) and standard deviation of detected reflections (t_{σ}). The spatial distribution utilised circular statistics to provide an equivalent mean direction of arrival (azimuth (θ_{mean}) and elevation (ϕ_{mean})) and spatial spread of early reflections. These spatial and temporal parameters allow these aspects of stage acoustic conditions to be studied in greater detail.

With the measurement and analysis methods established, the following sections will focus on the performance spaces that were measured as part of this study and a detailed analysis of the measured data.

Chapter 5

Objective analysis of stage acoustic conditions

There is currently little available data characterising how the spatial or temporal distribution of early reflections vary due to hall-related or performer-related variables. It is therefore difficult to predict if spatial or temporal variations of early reflections are audible for performing musicians. By applying the analysis techniques presented in Chapter 4 to the measurements described in Chapter 3, it will be possible to observe how hall-related and performer-related variables affect the acoustic conditions on stage.

The analysis will include both traditional stage acoustic parameters and the methods outlined in Chapter 4. The acoustic parameters measured over each stage will be collated and compared to give a broad summary of the acoustic conditions measured in each space. An analysis of the impulse responses measured at centre stage positions in each venue will follow to determine how the geometry of the stage enclosure affects the acoustic conditions. A further analysis will observe variables related to the musician (position, orientation/directivity) across a subset of the stages measured in this study.

5.1 Venue comparison

The performance spaces included in this research were chosen to represent a wide range of potential conditions that a musician might experience; from small, informal recital venues to large symphonic concert halls. A description of each hall and the techniques used to measure the impulse response data can be found in Chapter 3. This initial analysis will summarise the traditional acoustic parameters measured in each venue. This will include EDT, T_{30}, ST_{early} and ST_{late} as described in Chapter 2.

The venues included are listed below and have been abbreviated as follows:

1. GUCH = Glasgow University Concert Hall

2. RH = Reid Concert Hall
3. YH= Younger Hall
4. LRR = Ledger Recital Room
5. CH = Caird Hall
6. SH = Stevenson Hall
7. RR = Recital Room
8. GCH = Glasgow City Halls (Grand Hall)

Table 5.1 shows the number of measurements made on each stage, excluding measurements where the source and receiver were separated. As described in Chapter 3, eight measurements were made at each location on stage with the sound source pointing in different directions. This was repeated over a number of different locations on stage representative of different performer locations. Multiple locations were chosen to ensure a wide coverage of each stage. Due to differences in the size and layout of each stage, the total number of measurements made on each stage varies.

Hall	Number of measurements
GUCH	16
RH	32
YH	32
LRR	40
CH	48
SH	48
RR	24
GCH	24

TABLE 5.1: *Number of measurements made on each concert hall stage with close-proximity source and receiver. Additional cross-stage and auditorium measurements were made in each venue but not included in this analysis. Eight measurements were made at each location on stage.*

Single values of the objective support parameters were obtained by averaging octave bands between 250Hz and 2kHz as recommended by ISO 3382-1 (International Organisation of Standardisation, 2009). For reverberation time parameters, the single figure values were obtained by averaging third octave band measurements between 400Hz and 1.25kHz.

As described in Chapter 2, ST_{early} must be measured at a minimum of 3.4m from the nearest reflecting surface for reflections to occur after 20ms and be included in this parameter. In many cases this was not possible to achieve in some of the concert hall surveys due to the size of the stage or the presence of immovable stage furniture. Therefore, in this comparison ST_{early} is presented alongside $ST_{early,mod}$ (equation (5.1)) which uses a lower integration time of 10ms (following Wenmaekers et al. (2012)).

$$ST_{early,mod} = 10log_{10}\left(\frac{\int_{10ms}^{100ms} p^2(t)dt}{\int_{0ms}^{10ms} p^2(t)dt}\right) \tag{5.1}$$

5.1.1 Objective support

By comparing each venue in terms of ST_{early} and ST_{late}, it is possible to gain initial insight into the stage acoustic conditions measured in each space. In addition, it will be possible to see if two or more halls have similar levels of objective support. Figure 5.1 below shows a summary of ST_{early} measured in each hall. The plot shows the mean and standard deviation of all measurements made on each stage.

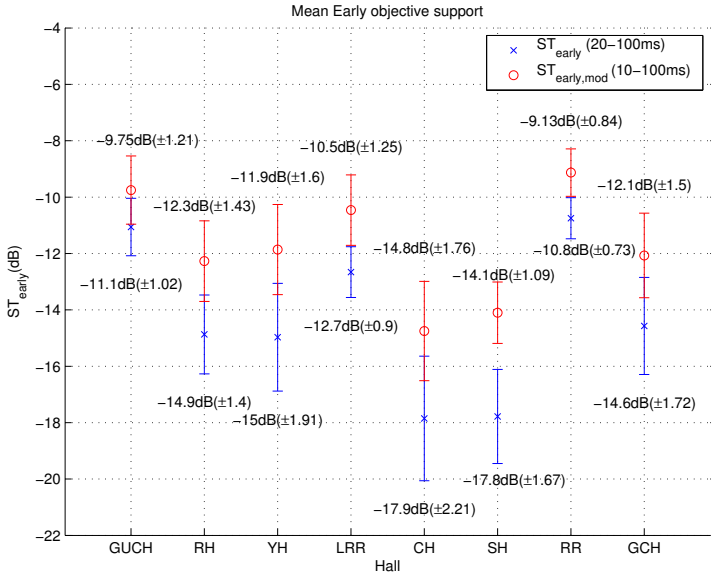


FIGURE 5.1: *Mean ST_{early} and $ST_{early,mod}$ measured on concert hall stages. Each parameter is shown with an associated standard deviation. The mean value for each measurement is obtained by the average ST_{early} between 250Hz and 2kHz octave bands. The plot shows these values averaged for all measurements made on each stage. The total number of measurements on each stage is shown in Table 5.1.*

It can be seen that on all stages, the mean level of $ST_{early,mod}$ is higher than the mean ST_{early}. This suggests that in the majority of cases there is additional energy arriving between 10ms and 20ms. The largest difference between ST_{early} and $ST_{early,mod}$ occurs in SH where the difference is 3.7dB. This may have been due to the presence of nearby risers on stage and the proximity of some measurement locations to the stage rear wall.

The highest mean ST_{early} tends to occur in venues with smaller stages (GUCH, LRR & RR) in a range of -10.8dB to -12.7dB. Larger stages (such as RH, YH, CH, SH and GCH) appear to exhibit a lower average value of ST_{early} in the range of -14.6dB to -17.9dB. RR provides the highest average ST_{early} whereas CH provides the lowest. The results also show that there is a lower spread of values in smaller venues (i.e. RR & LRR) than in larger halls (CH, YH & SH). This suggests that the acoustic conditions in smaller venues vary more due to musician or hall-related variables than larger venues.

Figure 5.2 shows a similar plot for ST_{late} for each hall. RR provides the highest mean ST_{late} ($ST_{late} = -9.9dB$) with the lowest average value occurring on the stage of the Stevenson Hall

($ST_{late} = -19.0dB$). Similarly, there appears to be a wider range of values of ST_{late} recorded in larger halls than in smaller halls.

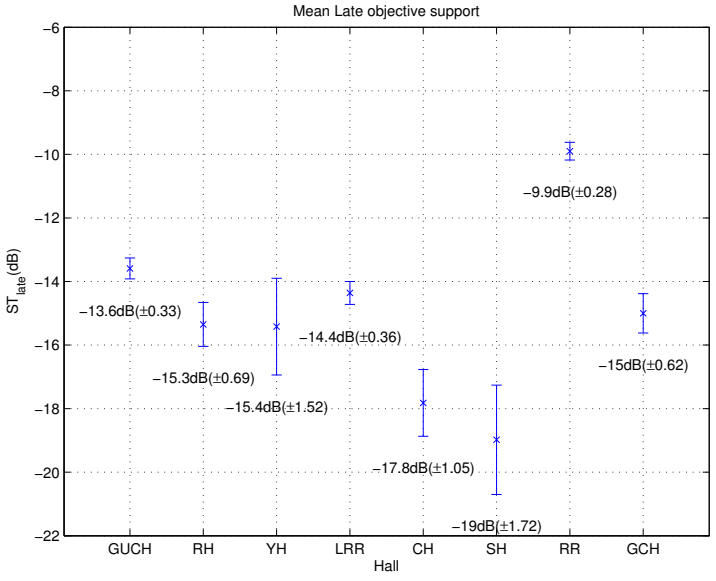


FIGURE 5.2: *Mean standard deviation of ST_{late} measured on concert hall stages.. The mean value for each measurement is obtained by the average ST_{late} between 250Hz and 2kHz octave bands. The plot shows these values averaged for all measurements made on each stage.*

In general, it can be seen that the level of early reflections is higher on smaller sized stages (such as RR and GUCH). In addition, it can be seen that many of the concert halls feature comparable levels of ST_{early} and ST_{late}, despite their stage dimensions being different (i.e. RH, YH and GCH). It is therefore feasible that these variations in geometry could cause these halls to be perceived differently, even with the halls returning comparable levels early reflected energy.

5.1.2 Reverberation

The measurements made on each stage can also be summarised in terms of their decay time. Figure 5.3 shows the mean EDT and T_{30} measured in each hall. The results show that the T_{30} varies between 1.23s (GUCH) to 2.49s (CH). As expected, the longest T_{30} was measured in the largest hall whereas the shortest was measured in one of the smallest venues (that also contained a large amount of acoustic absorption). The T_{30} measured in the RR is remarkably high given the relatively small size of the venue (when compared with the CH). It is highlighted that these measurements were made without the removable seating in the room, it is likely that the presence of an audience would significantly reduce the reverberation time in this space.

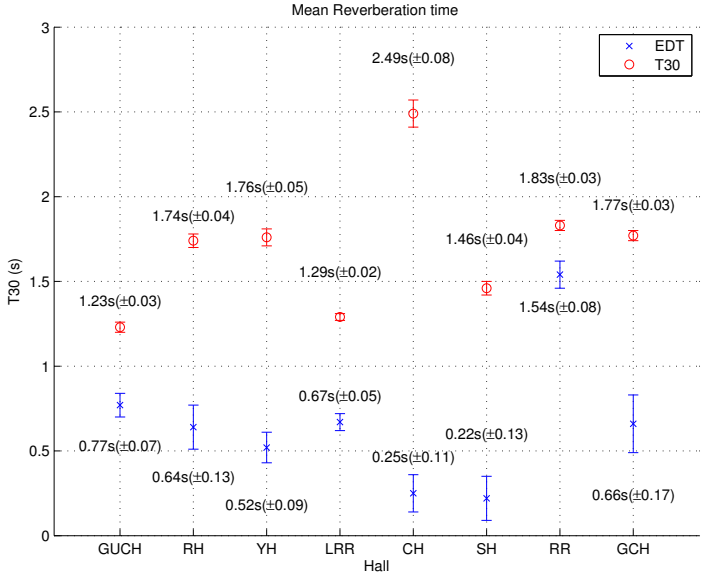


FIGURE 5.3: *Mean EDT and T_{30} measured on concert hall stages. Each parameter is shown with an associated standard deviation. The mean value for each measurement is obtained by the average values obtained between 400Hz and 1.25kHz octave bands. The plot shows these values averaged for all measurements made on each stage.*

The longest EDT values tend to occur in the smallest venues (i.e. RR/GUCH) and the shortest are found in the larger halls (i.e. CH/SH). In these cases the longer EDT values typically indicate the presence of very early, high amplitude reflections. Due to the large stage area in the larger concert halls, equivalent reflections occur much later and with less energy (due to propagation loss), therefore the early energy falls much faster thus truncating the reverberation time estimate. It can be seen from the standard deviation that overall EDT varied more than T_{30} in all of the measured halls. This suggests that the late arriving sound is affected much less by the orientation or position of the sound source on stage than early arriving sound.

5.1.3 Summary and discussion

Existing acoustic parameters were used to compare the stage acoustic conditions in a number of venues. These parameters were averaged over all measurement positions and source orientations to produce a single value for each stage. It was shown that ST_{early} tended to be higher in level in smaller venues. Similarly, ST_{late} appeared to be higher in smaller venues than in larger venues. These results are to be expected as the path length of each reflection is likely to be shorter in smaller venues and so reflections will have experienced less attenuation due to distance.

It was also observed that ST_{early} appeared to vary more than ST_{late} on each stage. This indicates that varying aspects such as source orientation and measurement location tends to affect early reflections, arriving between 20ms and 100ms, more than reflections arriving after 100ms. This is consistent with current understanding of auditorium acoustics that the diffuse reverberation varies relatively little when the source characteristics are varied.

It was also shown that concert halls could provide very similar mean levels of ST_{early} despite having very different dimensions, for example GCH, the YH and the RH. This provides initial support to previous assertions that two halls could produce the same ST_{early} but the distribution of early reflections could be very different. If the variation in distribution is audible to a performer then it provides further support to the notion that additional acoustic parameters are required to assess concert hall stages.

The mean values for ST_{early} and ST_{late} obtained appear to lie within a similar range observed in previous studies. The results of these previous studies are summarised in Table 2.1. ISO 3382-1 (International Organisation of Standardisation, 2009), for example, gives an expected range of values for ST_{early} of -24dB to -8dB and an ST_{late} range of -24dB to -10dB. This is a good initial indication that an average of numerous measurements made with a directional loudspeaker at different locations on stage is a robust approach for measuring stage acoustic conditions for this research. The main advantage of this approach is that the individual measurements can be observed in isolation which provide more detail regarding the distribution of early reflections.

It was found that the EDT tended to vary more than T_{30} in each hall which indicates that source orientation and measurement location produce more changes in early reflections than in the late reverberation. In most cases, there was a large difference between EDT and T_{30}. This difference was found to be increased in larger sized concert halls i.e. CH, SH and the GCH. This can be attributed to the temporal distribution of reflections. For instance, on smaller stages, early reflections will arrive with a higher amplitude ensuring that the early decay occurs over a longer period of time.

In general, these results provide an initial indication that stage acoustic conditions vary due to both hall related aspects (venue geometry) and musician related aspects (instrument orientation/directivity and position). These results also imply that these variations occur mainly in the early part of the acoustic response between 20ms and 100ms. It is feasible that these variations could account for different acoustic conditions experienced by musicians playing on different stages. The following sections will look at specific measurements made on stage to characterise how the distribution of early reflections vary in relation to hall related aspects and musician related aspects.

5.2 Hall-related aspects

As discussed earlier in this chapter, the acoustic conditions experienced on stage can vary due to aspects of the venue or aspects related to the musician. This section will observe how the stage acoustic conditions vary due to hall-related variables. By using traditional acoustic parameters and spatial/temporal parameters (developed in Chapter 4) it will be possible to provide a detailed analysis.

In Chapter 3, it was discussed how it was not always possible to measure at identical locations on each stage due to the differing geometries of each stage or the presence of rare pianos or stage furniture that could not be moved. Therefore, in order to make a direct comparison of the acoustic conditions measured in each hall, this section will focus specifically on the measurements

made at the down-stage centre position in each hall. Eight measurements were made at each measurement position where the source is rotated 45° in azimuth after each measurement is completed. Therefore, the results are summarised for each hall as mean and standard deviation of eight measurements. In every hall, the acoustic response was measured at this location which is situated approximately in the centre front of each stage. In the majority of performance scenarios, a soloist will perform from this location.

5.2.1 Objective support

In the previous section, it was demonstrated that ST_{early} varied significantly on different stages, generally varying more on larger stages (such as CH and SH). By comparing these parameters at a common position on stage it is possible to observe what extent this is caused by the geometry of each stage enclosure. Figure 5.4 below shows the mean and standard deviation of ST_{early} and $ST_{early,mod}$ measured at the down-stage centre position in each venue.

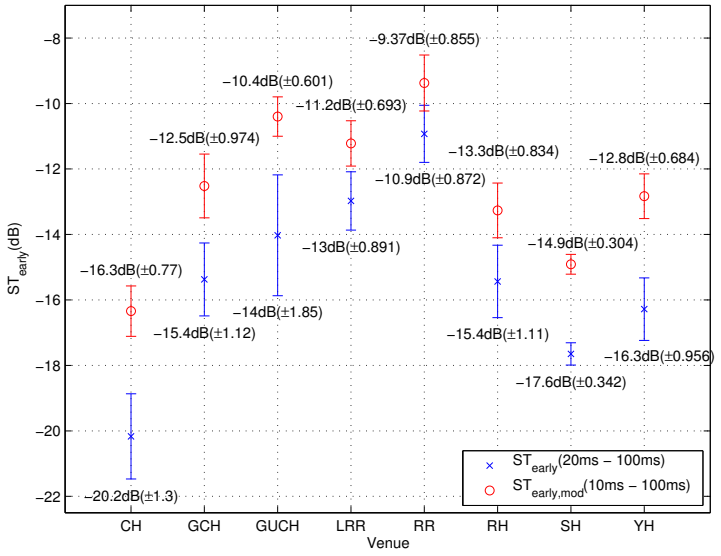


FIGURE 5.4: *Comparison of ST_{early} (blue) and $ST_{early,mod}$ (red) measured at the down-stage positions in each venue. Each plot summarises results (mean and standard deviation) obtained from eight measurements where the source orientation has been varied.*

It can be seen that mean ST_{early} varies significantly across different venues with measured values occupying a range of 9.5dB. The results indicate that the highest level ST_{early} is found in RR (mean ST_{early} = -10.7dB) whereas the lowest is found in the CH (mean ST_{early}=-20.2dB). Figure 5.4 also shows a similar pattern for $ST_{early,mod}$ where the highest level is found in RR (-8.93dB) and the lowest value is found in CH (-16.3dB). With the exception of RR, it can be seen that $ST_{early,mod}$ varies less than ST_{early}. In RR, this likely due to the proximity of the rear wall at the down stage centre position (2.13m). Due to the variations in source orientation, ST_{early} occupies a range of values in each hall. The largest variation occurs in CH ($\sigma = 1.3dB$) and GUCH ($\sigma = 1.85dB$) and least in venues such as SH ($\sigma = 0.34dB$).

As mentioned in Chapter 2, an informal JND for the ST_{early} measure is regarded as 2dB (Hak et al., 2012). Assuming this value is representative for the majority of musicians, it is possible

to group the concert halls in terms of similarity. Halls YH and SH can be grouped together as the difference in mean values (δ) is 1.3dB. RH, GCH and GUCH can also be grouped together ($\delta = 1.4dB$). LRR and RR are just outside the JND however showing 2.3dB difference in mean values. CH is closest to SH in terms of ST_{early} and is on average 2.6dB lower than SH.

Figure 5.5 shows the mean ST_{late} measured at the down stage centre position of each stage. It can be seen that, like in Figure 5.4, the highest mean ST_{late} occurs in RR (-9.95dB) whereas the lowest occurs in CH. It can also be seen that, in general, the standard deviation in each hall is lower than for ST_{early}. This suggests that, at the down-stage centre position, ST_{late} does not vary significantly with source orientation.

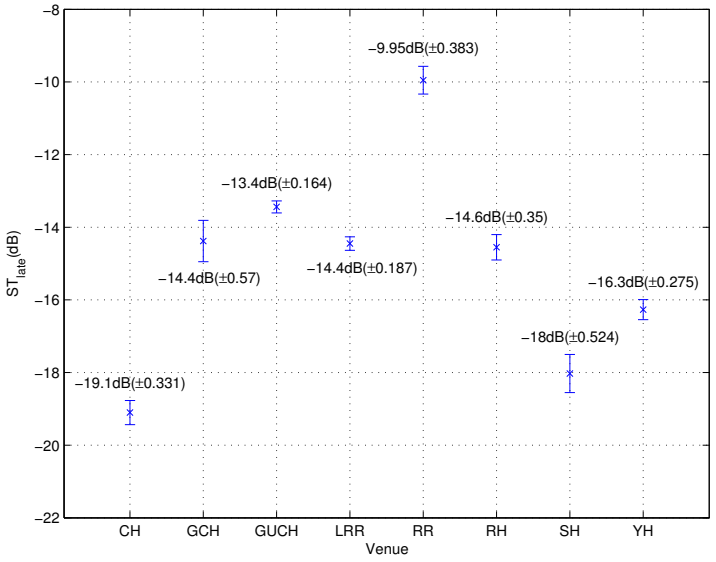


FIGURE 5.5: *Comparison of ST_{late} measured at the down-stage positions in each venue. Each plot summarises results (mean and standard deviation) obtained from eight measurements where the source orientation has been varied.*

It is useful to see if there is any relationship between the parameters ST_{early}, $ST_{early,mod}$, ST_{late} and the physical dimensions of the stage enclosure. As discussed in Chapter 2, Dammerud (2009) proposed using the physical dimensions of the stage in order to predict preference of orchestral musicians. Dammerud proposed using the basic dimensions of the stage (e.g. width, length, height) in conjunction with ratios of these dimensions. An adapted form of these dimensions is utilised in this analysis. Table 5.2 shows the correlation coefficient of each stage dimension with ST_{early}. Each correlation coefficient has been tested to verify it is significantly different from a correlation of zero. Statistical significance to pre-determined *p-values* are highlighted.

These parameters are the stage width (W_s), stage depth (D_s), the height from the stage floor to ceiling (H_s), the surface area of the stage (A_s) and the volume of the stage enclosure (V_s). In addition, following Dammerud (2009), correlation coefficients have been calculated for different ratios of stage dimensions. It can be seen that the highest correlation for ST_{early} occurs with stage surface area (A_s) at a value of -0.896. This is also the case for $ST_{early,mod}$ which has a correlation coefficient of -0.892. This implies that the level of early energy on stage is negatively correlated with A_s.

It can also be seen that the highest correlation coefficient for ST_{late} is with A_s at a value of -0.714. This implies that ST_{late} is negatively correlated with A_s. However, as ST_{late} assesses energy arriving between 100ms and 1000ms, it is more likely ST_{late} is correlated with the overall size of the venue rather than just the stage.

Parameter	W_s	D_s	H_s	A_s	V_s	$\frac{H_s}{W_s}$	$\frac{D_s}{W_s}$	$\frac{H_s D_s}{W_s}$
ST_{early}	-0.803*	-0.881**	-0.442	-0.896**	-0.848**	0.442	-0.354	-0.380
$ST_{early,mod}$	-0.825*	-0.847**	-0.511	-0.892**	-0.864**	0.361	-0.268	-0.410
ST_{late}	-0.764*	-0.793*	-0.290	-0.805*	-0.714*	0.522	-0.296	-0.278

TABLE 5.2: *Correlation coefficients of stage dimensions with mean ST_{early} measured at down-stage centre position in all venues. Correlation coefficients are significant to the following levels , $* = p \leq 0.05$, $** = p \leq 0.01$.*

Figure 5.6 presents the mean ST_{early} plotted against A_s. Each venue is represented as blue points. A simple linear regression has been applied which highlights the trend of linearly decreasing ST_{early} with increasing A_s ($R^2 = 0.803$). The largest residual (of magnitude 2.61dB) corresponds with GCH which has a higher ST_{early} than predicted by the linear regression. This may be due to the small differences in measurement procedure in this venue or the differences in the absorbing characteristics of each stage enclosure. It may also be possible that music stands and seating left on stage caused additional reflected energy to be received in this venue.

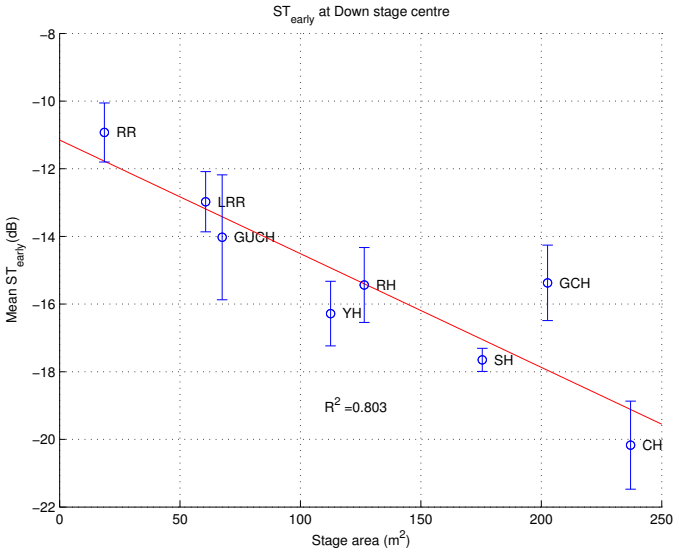


FIGURE 5.6: *Mean and standard deviation ST_{early} measured at the down-stage centre position plotted against stage surface area. Each venue is represented as a point and a linear regression has been applied ($R^2 = 0.803$). The results show that an increase in stage area will result in a decrease in ST_{early}*

Similarly, Figure 5.7 presents the mean $ST_{early,mod}$ versus A_s. A linear regression has also been applied to these results which shows a trend of decreasing $ST_{early,mod}$ with increasing A_s. It can be seen that the coefficient of determination is slightly lower ($R^2 = 0.796$) than for ST_{early}. It can also be seen that the gradient for $ST_{early,mod}$ is much shallower than for ST_{early}. It can be

seen that reducing the lower integration time to 10ms has a greater effect at larger halls than in smaller halls.

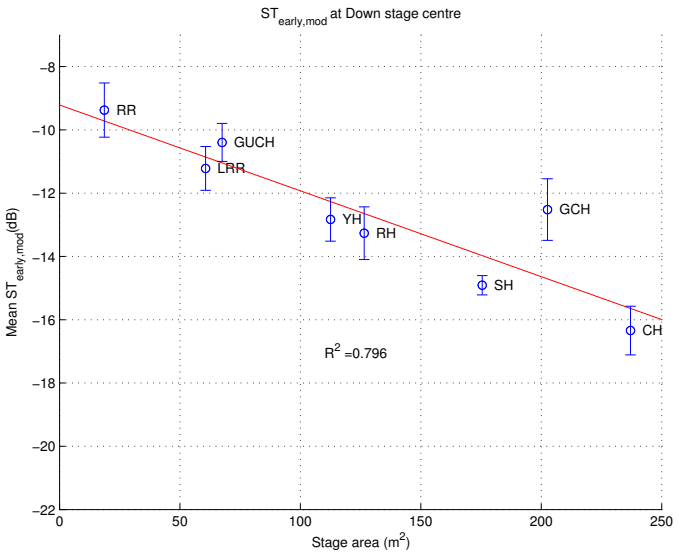


FIGURE 5.7: *Mean and standard deviation $ST_{early,mod}$ measured at the down-stage centre position plotted against stage surface area. Each venue is represented as a point and a linear regression has been applied ($R^2 = 0.796$). The results show that an increase in stage area will result in a decrease in $ST_{early,mod}$*

Figure 5.8 presents similar results for ST_{late} with a linear regression applied. These results show ST_{late} reducing linearly with A_s. However, the coefficient of determination is much lower than found for ST_{early}. As discussed earlier, it is more likely that ST_{late} is more closely related to the dimensions of the venue, as opposed to just the stage itself.

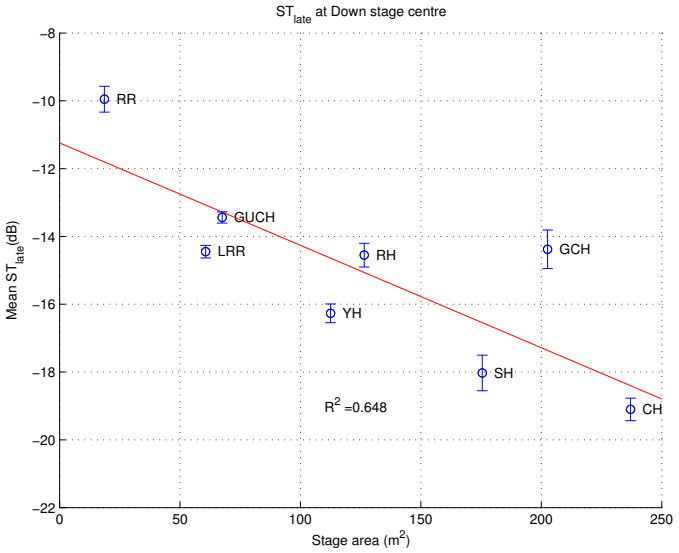


FIGURE 5.8: *Mean and standard deviation ST_{late} measured at the down-stage centre position plotted against stage surface area. Each venue is represented as a point and a linear regression has been applied ($R^2 = 0.648$). The results show that an increase in stage area will result in a decrease in ST_{early}*

From this analysis, it is suggested that ST_{early} varies significantly over the concert halls included in this study. When measured from the down-stage centre position, the level of ST_{early} tends to reduce linearly with increasing stage area (A_s). This is to be expected as reflected sounds will have travelled a further distance on larger stages before reaching the receiver and will therefore have significantly less energy. ST_{late} was found to be similarly related to A_s however it is more likely that the level of ST_{late} is more highly correlated with the dimensions of the venue.

The results also show that some of the concert halls exhibit very similar levels of ST_{early}, for example, RH and GCH. This is significant as stages featuring the same level of ST_{early} may be perceived differently by performers if there are large differences in the spatial or temporal distribution of early reflections. In the following section, the spatial and temporal analysis techniques, discussed in Chapter 4 will be applied to determine how the spatio-temporal distribution of reflections varies on each stage.

5.2.2 Temporal distribution

As discussed in Chapter 4, the temporal distribution of early reflections can be summarised by the mean and standard deviation of time of arrival. These parameters display when the majority of reflections arrive at the receiver and how spread out they are over time. These analyses have been performed for all measurements taken at the down-stage centre position in each venue.

Figure 5.9 summarises the values obtained for t_{mean} at the down-stage centre position in each concert hall. It can be seen that the average value of t_{mean} tends to appear between 47ms (GUCH) and 62ms (CH). CH varies more than other halls in terms of t_{mean}. In this particular case, when the source points into the audience, the reflections from the stage enclosure are very

low in amplitude and may not have been detected in the analysis. When the source points towards the stage enclosure, the reflecting surfaces are sufficiently far away to cause reflections to arrive much later. This explains the wide variation in the values obtained for this hall. In contrast, the LRR exhibits a very low spread of values ($\pm 1.7ms$). This is likely to be a consequence of the tiered seating arrangement in this venue and the proximity of many surfaces providing reflections regardless of source direction.

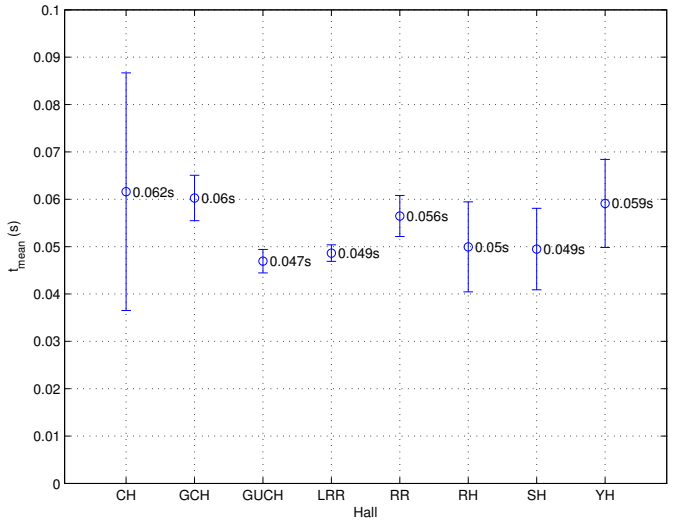


FIGURE 5.9: *Summary of t_{mean} values measured at the down-stage centre position in each concert hall. Values are summarised as mean and standard deviation.*

Figure 5.10 shows a similar plot for t_σ. In the halls measured it can be seen that the mean t_σ varies between 8.8ms (CH) and 23ms (LRR). The values obtained in larger concert halls tend to vary considerably with source orientation whereas smaller variations are produced in smaller venues (for example LRR and RR). These results imply that in smaller concert halls the temporal spread of early reflections does not change significantly with source orientation.

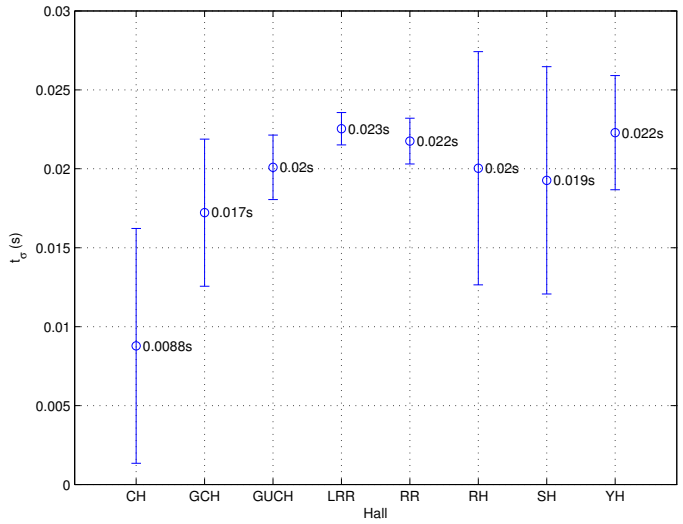


FIGURE 5.10: *Summary of t_σ values measured at the down-stage centre position in each concert hall. Values are summarised as mean and standard deviation.*

The temporal parameters have also been assessed in terms of their correlation with various stage geometrical parameters. The results are summarised in Table 5.3. It can be seen that the highest correlation occurs between t_σ and V_s (shown in bold). This suggests that stages of a smaller volume produce reflections which are more spread out in time. The highest correlation for t_{mean} is also with stage volume V_s albeit with a much lower coefficient than for t_σ. This implies that as the stage increases in volume, reflections tend to appear later in time.

Parameter	W_s	D_s	H_s	A_s	V_s	$\frac{H_s}{W_s}$	$\frac{D_s}{W_s}$	$\frac{H_s D_s}{W_s}$
t_{mean}	0.445	0.306	0.345	0.489	0.541	-0.108	-0.168	0.038
t_σ	-0.730*	-0.613	-0.454	-0.797*	-0.859**	0.323	-0.004	-0.189

TABLE 5.3: *Correlation coefficients of stage dimensions with temporal parameters measured at down-stage centre position in all venues. Correlation coefficients are significant to the following levels , $* = p \leq 0.05$, $** = p \leq 0.01$.*

Figure 5.11 shows how t_{mean} varies with stage volume. It can be seen that smaller venues tend to produce smaller average values of t_{mean}. It is also clear that the smaller stages (RR, LRR and GUCH) produce the lowest values of t_{mean}. This is to be expected as the closer proximity of reflecting surfaces will ensure that the average time of arrival is lower for early reflections. The distance travelled by early reflections is larger on stages with a higher volume. The coefficient of determination, $R^2 = 0.292$, however implies only a weak fit.

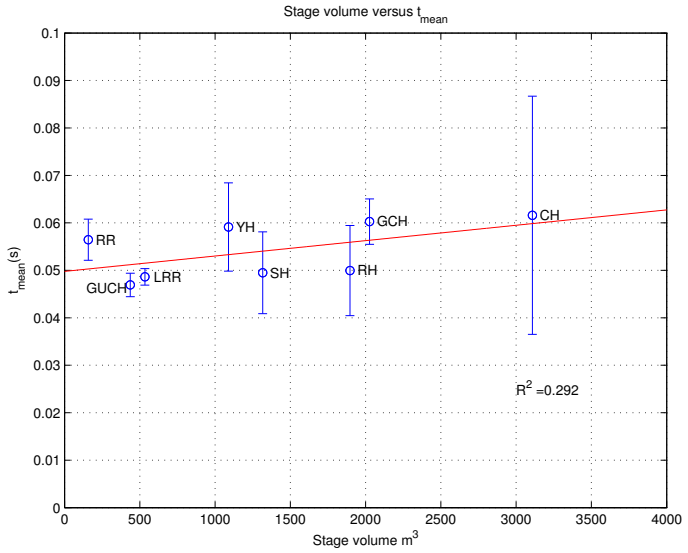


FIGURE 5.11: *Summary of t_{mean} values measured at the down-stage centre position in each concert hall. Values are summarised as mean and standard deviation.*

In Figure 5.12, the temporal spread is plotted against the stage volume. t_σ is plotted as mean and standard deviation for all measurements made at the down-stage centre position in each hall. It can be seen that larger halls produce less temporal spread than small halls. This is highlighted using linear regression where $R^2 = 0.738$. However, it should be noted that the range of mean t_σ is 13ms which is considered to be a relatively small range.

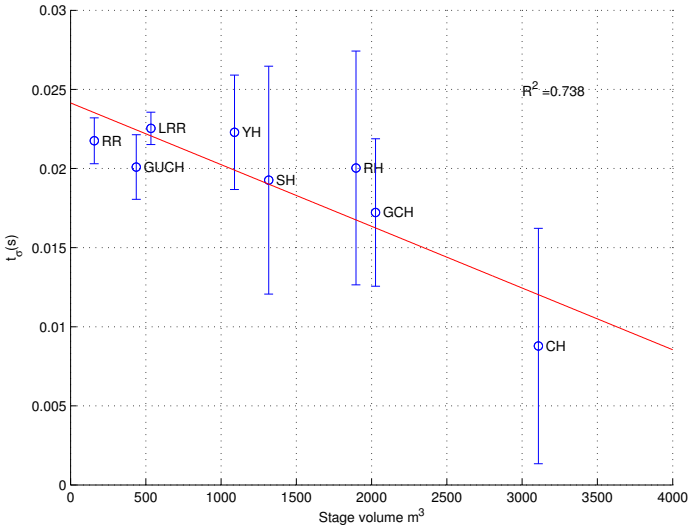


FIGURE 5.12: *Standard deviation of time of arrival plotted against stage volume for measurements made in each Hall. These results show that the reflections tend to become less spread out over time on larger stages.*

In summary, the results suggest that the mean time of arrival of early reflections increases with stage volume while the temporal spread decreases. This is to be expected as larger stages will result in longer path lengths for reflections to travel before encountering the receiver. This will

cause the reflections to arrive later but also increase the attenuation due to distance ensuring that higher order reflections, which contribute to a higher temporal spread have insufficient amplitude to be detected.

5.2.3 Spatial distribution

The spatial distribution of early reflections was also extracted from measurements made at the down-stage centre positions. As described in Chapter 4, the spatial distribution of early reflections was obtained from image source plots where the time of arrival was disregarded. By summing unit vectors pointing to each early reflection, the mean resultant vector was obtained. The azimuth and elevation of the mean resultant vector was considered to be representative of the mean direction of arrival of early reflections whereas the normalised vector length was deemed to be an appropriate measure of spatial spread.

In this research, it was necessary to characterise the spatial distribution when measured at the down-stage centre position of different concert halls. In Chapter 4, it was demonstrated that this could be achieved using the direction and length of the mean resultant vector. This denotes the dominant angle of arrival and the spread of early reflections.

Figure 5.13 shows the average azimuth of the mean resultant vector, θ_{mean} versus the angle of orientation of the loudspeaker. These results are summarised for each source angle and include measurements made in each venue at the down-stage centre position. It can be seen that θ_{mean} is oriented in a similar direction to the source orientation. This confirms that when a directional sound source is used to measure the stage acoustics, reflections are dominant in the direction the sound source is pointing in.

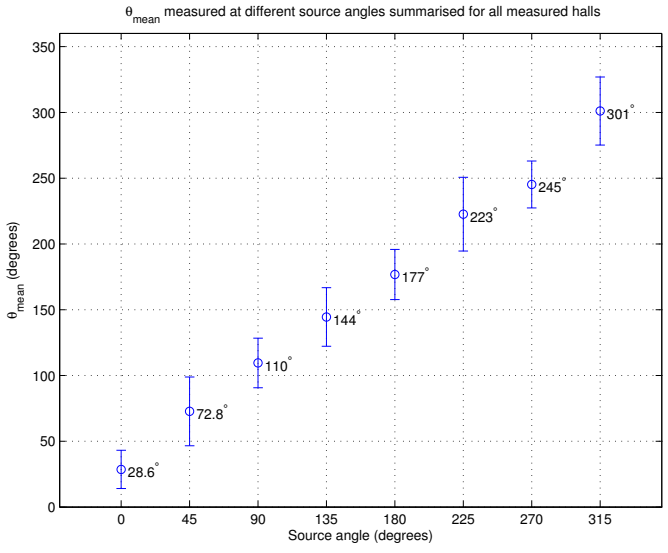


FIGURE 5.13: *Plot showing the source angle of orientation versus the angle of azimuth of the mean resultant vector. Values are summarised as mean and standard deviation. It can be seen that an increase in the source angle of orientation produces a sympathetic increase in* θ_{mean}

A similar plot is shown in Figure 5.14 which summarises the average angle of elevation, ϕ_{mean}, for each source angle. it can be seen that the average elevation for each source angle is very near the lateral plane. These results suggest that exciting the performance spaces in this way causes the majority of early reflections to arrive from the lateral plane. Conversely, exciting the space with an omnidirectional sound source will result in more reflections arriving from elevated angles.

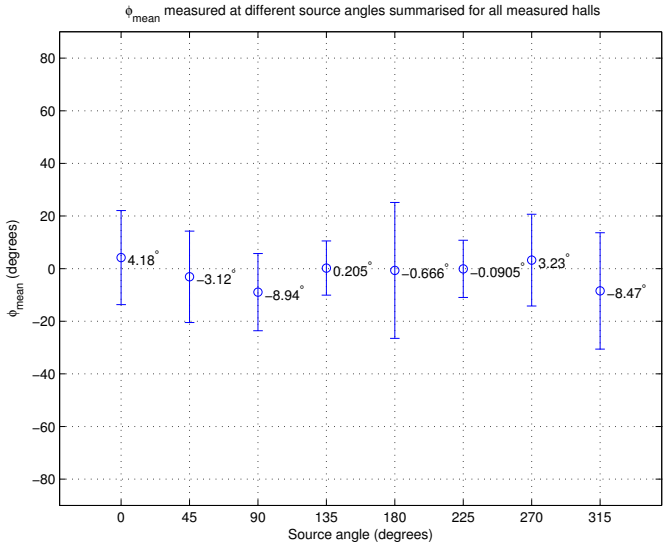


FIGURE 5.14: *Plot showing the source angle of orientation versus the angle of elevation of the mean resultant vector. Values are summarised as mean and standard deviation.*

Figure 5.15 shows the mean and standard deviation of spread values obtained in each hall. It can be seen that larger concert halls such as CH and SH tend to show reflections arriving clustered together in space (which results in a high value of spatial *spread*). Early reflections in smaller halls tend to arrive from a wider range of angles (as in RR).

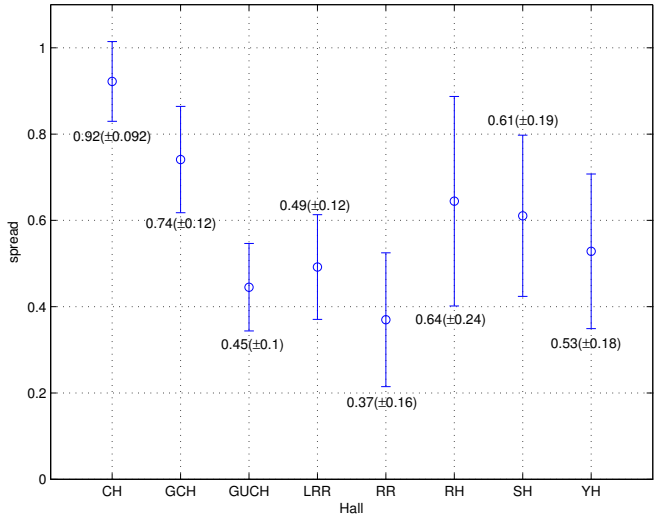


FIGURE 5.15: *Plot showing the source angle of orientation versus the spatial spread of the mean resultant vector. Values are summarised as mean and standard deviation.*

Table 5.4 shows the correlation coefficients of spatial spread against spatial parameters as described in the previous section. It can be seen that the highest correlation for spatial spread is with stage volume with a value of 0.986. This suggests that the spatial spread parameter increases with stage volume and stage area. This indicates that reflections are clustered more closely together on larger stages.

Parameter	W_s	D_s	H_s	A_s	V_s	$\frac{H_s}{W_s}$	$\frac{D_s}{W_s}$	$\frac{H_s D_s}{W_s}$
$spread_{mean}$	0.871**	0.847**	0.634	0.948**	0.986**	-0.314	0.201	0.422

TABLE 5.4: *Correlation coefficients of stage dimensions with spatial parameters measured at down-stage centre position in all venues. The spread parameter is the most highly correlated with stage dimensions, specifically the stage volume. Correlation coefficients are significant to the following levels ,* $* = p \leq 0.05$, $** = p \leq 0.01$.

Figure 5.16 confirms this by showing the mean and standard deviation of spatial spread measured in each hall plotted against stage volume. These data are labelled according to each hall. It can be seen that stages with a larger volume (m^3), such as CH, produce values closer to 1 than smaller stages (such as RR) which produce values as low as 0.33. This is highlighted by a linear regression which produces a coefficient of determination of $R^2 = 0.965$.

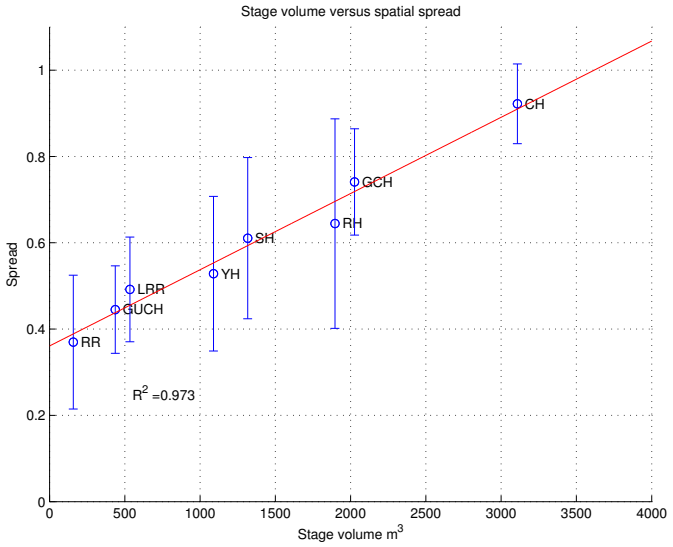


FIGURE 5.16: *Plot showing the mean and standard deviation of angular spread of early reflections measured in each concert hall. A spread value of 1 indicates that reflections are spatially coincident whereas a low value indicates reflections are spread over a larger area. It can be seen that an increase in stage volume produces a linear increase in mean spatial spread (which demonstrates reflections are clustered closer together)*

It can be seen that the spatial distribution of reflections appears to vary in different halls and is dependent on the source angle. Reflections tend to arrive predominantly from a similar angle to the sound source, for example, reflections will arrive mainly from 90° if the source is pointed in that direction. It was found that the angle of elevation varies much less. This is due to the sound source being rotated only in the lateral plane. It was also found that the spread of early reflections also varies on different stages and varies in response to the source angle. It was found that reflections tended to be clustered closer together on larger stages and more spread out on smaller stages. The spread value was found to decrease linearly with the physical volume of the stage (V_s). This is thought to be due to higher-order reflections having less energy on larger stages (due to the increased propagation distance). Overall, it can be seen that the spatial distribution of early reflections varies significantly on different stages.

5.2.4 Summary

This section utilised the analysis techniques presented in Chapter 4 to characterise the acoustic conditions, measured at the down-stage centre position, in relation to the physical dimensions of each stage. The acoustic conditions were assessed in terms of existing acoustic parameters in addition to temporal and spatial parameters developed as part of this research. The results presented were obtained from measurements at the down-stage centre position which were near-equidistant from side walls in all venues. It was shown that the mean level of ST_{early} decreased linearly with the surface area of the stage. High degrees of negative correlation were also observed for stage width, stage depth and stage volume.

In a number of cases, the level of ST_{early} measured in each hall was shown to be very similar i.e. within 2dB which corresponds to informal JND quoted for ST_{early}. It was therefore possible to group the concert halls together given that this parameter is the only parameter currently used to assess stage acoustic conditions. For example, it was shown that RH, GCH and the GUCH could be grouped together. The mean spatial spread measured in these halls at this position were found to range between 0.45 and 0.75 which imply the spatial distribution is very different.

However, it can be seen that reflection in GCH or RH are less spread out in space than GUCH. Similarly, SH and YH were grouped together using ST_{early}. The spatial spread of the reflections in YH vary less with angle of orientation than in the SH.

The results showed a weak, positive correlation between stage volume and t_{mean} which implied that larger stages tend to produce later arriving reflections which is to be expected. Furthermore, it was shown that t_{std} presented a strong negative correlation with stage volume which suggests that smaller stages tend to result in reflections being spread out more over time. This is more likely to be a consequence of higher order reflections containing higher energies on smaller stages (due to the shorter reflection path length).

It was also shown how the spatial distribution of early reflections varied. The spatial distribution was assessed in terms of the direction of the mean resultant vector, indicating the dominant angle of arrival, and the spatial spread which indicates how clustered the reflections are in space. It was shown that early reflections tend to be more clustered together in space on stages that had a larger volume. In a similar manner to the temporal spread of early reflections this is likely to be caused by reduced energy in higher order reflections on larger stages due to the increase in reflection path length. The lack of energy may have decreased the likelihood of detection.

In summary, the results presented above demonstrate that it is possible for concert halls with similar ST_{early} values to feature different spatial and temporal distributions of early reflections. In this study, the spatial and temporal distributions of early reflections were assessed by analysing image source plots as described in Chapter 5.

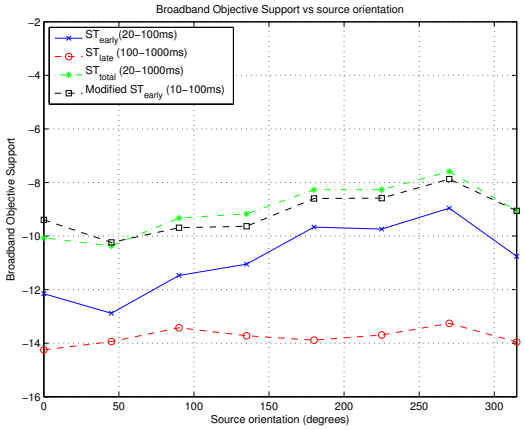
5.3 Performer-related aspects

The previous section discussed how the acoustic conditions on stage could vary due to architectural aspects including the geometry of each venue. The acoustic conditions can also change due to aspects related to the musician for example, their position on stage and their orientation (or instrument directivity). Measurements were made at other locations in each venue and at multiple angles of orientation to further explore the effect of these physical variables. In the following section the results of the down-stage measurements (left, centre and right) are compared. The effect of these physical variables on monaural acoustic parameters in addition to spatial and temporal analysis will be covered. For brevity, measurements on only a single stage will be considered in the analysis. LRR was chosen for this purpose given its frequent use as a recital space for soloists. In addition, the down-stage measurements were made at similar distances to the rear stage wall.

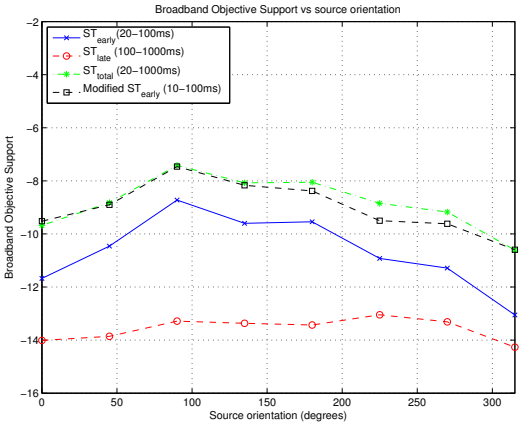
5.3.1 Objective support

All measurements were assessed initially by deriving monaural reverberation and objective support parameters. In this study it was of particular interest to determine if these parameters varied with source position and orientation on stage. Figure 5.17 shows the broadband objective support parameters plotted against the sound source orientation measured in LRR. In addition, the plots also show the level of a modified $ST_{early,mod}$. The plots show stage left, stage centre and stage right measurement positions. It can be seen that ST_{early} (solid blue trace) varies as the source orientation is changed.

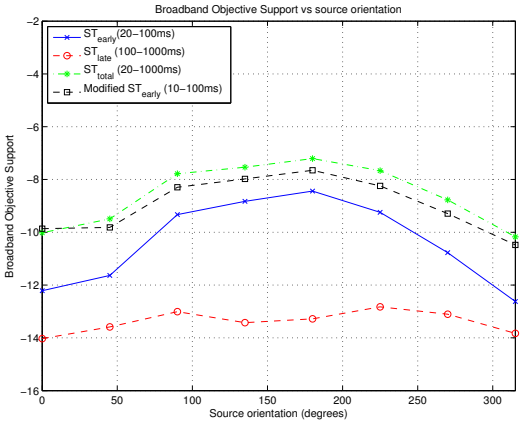
In each case a peak value is reached when the source is oriented towards the nearest reflecting surface. For example, at the stage left position, the peak value of ST_{early} occurs when the source is oriented to 270° i.e. towards the stage left wall. At the down-stage centre position, ST_{early} varies between -12.6dB and -8.8dB (range = $4.2dB$). At the down-stage right position, ST_{early} varies between -13.0dB and -8.7dB (range = $4.3dB$). At the down-stage left position, ST_{early} varies between -12.2dB and -9.0dB (range = $3.9dB$). These variations are considered significant given the estimated JND of 2dB for ST_{early} as discussed previously in this chapter.



(a) Down-stage left



(b) Down-stage right



(c) Down-stage centre

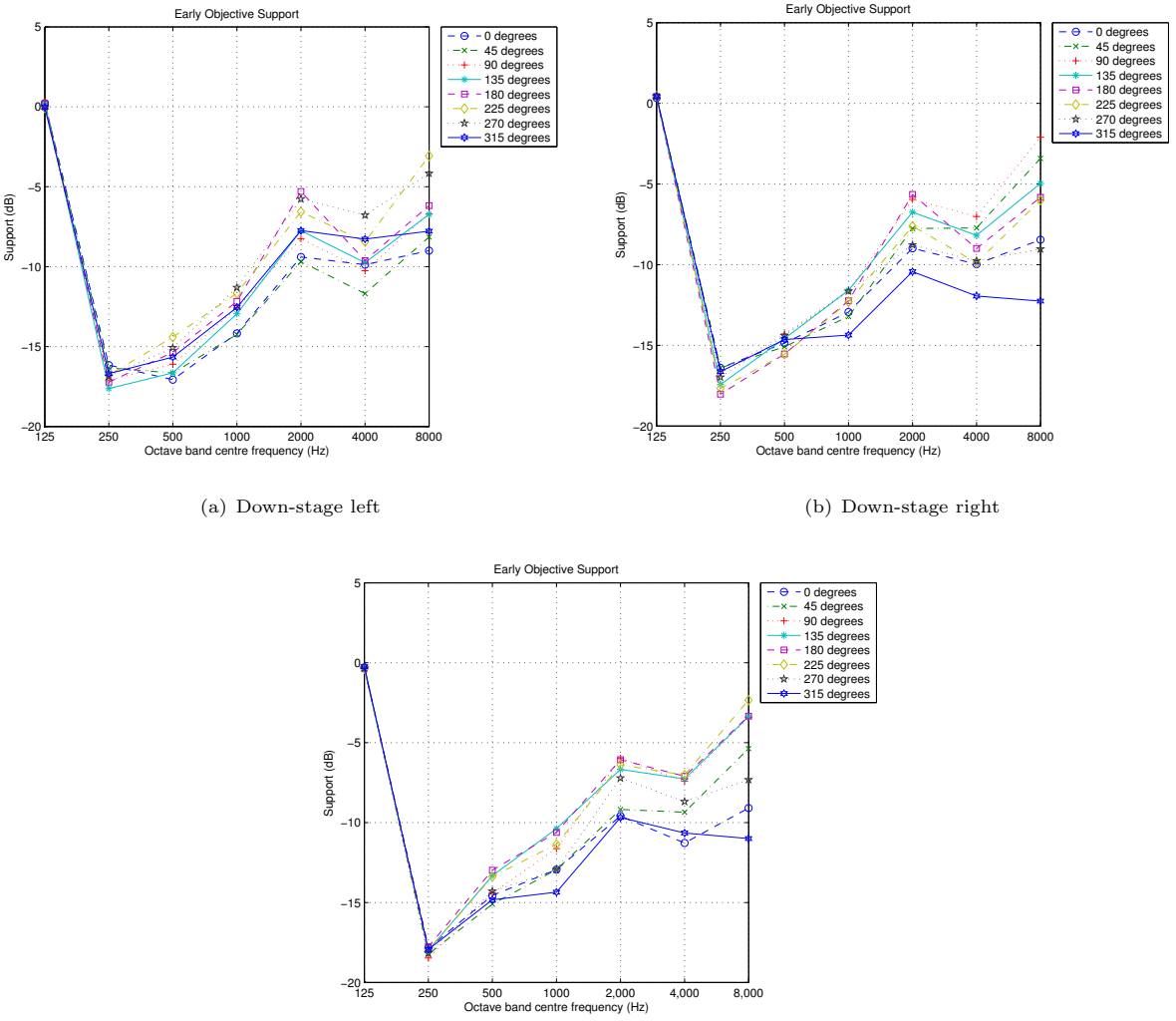
FIGURE 5.17: *Broadband objective support parameters (ST_{early}, ST_{late}, ST_{total} and $ST_{early,mod}$) measured at stage front positions in LRR.*

Figure 5.17 also shows how late objective support ST_{late} varies with source orientation (dashed red trace). It can be seen it varies much less than ST_{early} regardless of measurement position. At the down-stage centre position, the mean ST_{late} value is -13.4dB ($\sigma = 0.41dB$). At the down-stage right position, the mean ST_{late} value is -13.6dB ($\sigma = 0.42dB$). Finally, At the down-stage left position, the mean ST_{late} value is -13.8dB ($\sigma = 0.32dB$).

ST_{total} varies in a similar manner to ST_{early} in that the peak value always occurs in the direction of the nearest reflecting surface. However the results show that it is higher in value than ST_{early} typically by between 1.5dB and 2.5dB.

It can also be seen that $ST_{early,mod}$ parameter varies in a similar manner to the original parameter with peaks occurring in the same locations. In this case, it has a level very similar to ST_{total}. When the measurement has significant reflections occurring in the 10-20ms time period, the level of the modified and original ST_{early} parameters differ significantly.

Figure 5.18 shows ST_{early} varies with frequency for each sound source orientation and for each measurement position. It can be seen that ST_{early} generally increases with frequency. It can also be seen that ST_{early} appears to be high in the 125Hz band, it is likely that this is due to a lack of low frequency energy present in the direct sound due to the relative position of the source and receiver.



(a) Down-stage left

(b) Down-stage right

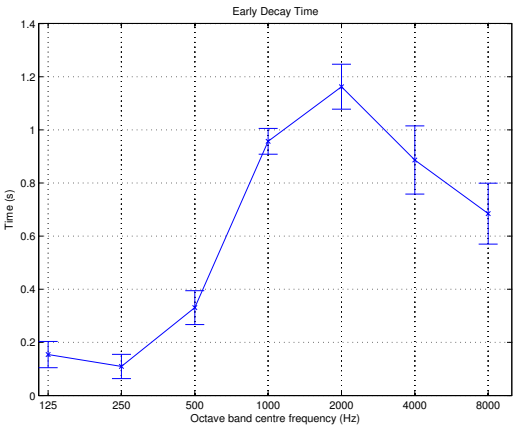
(c) Down-stage centre

FIGURE 5.18: *Each plot shows ST_{early} measured in octave band frequencies between 125Hz and 8kHz. Each plot shows how support varies in frequency with different source orientations. The results shown are for the three down-stage positions measured in LRR.*

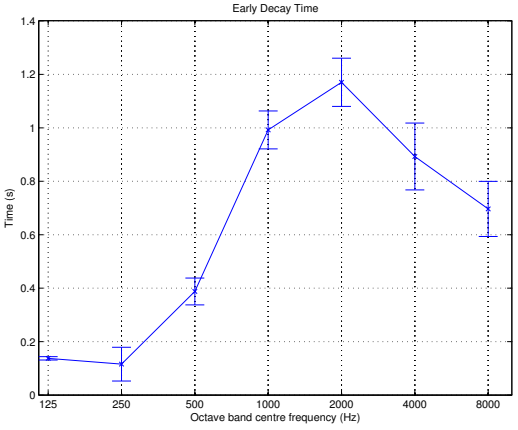
As the sound source is rotated it can be seen that ST_{early} varies more at high frequencies than at low frequencies. The high frequency ST_{early} tends to peak when the source is facing the nearest wall. For example, in Figure 5.18(b) high frequency ST_{early} is highest when the source is pointing towards the stage right wall (i.e. 90°).

5.3.2 Reverberation

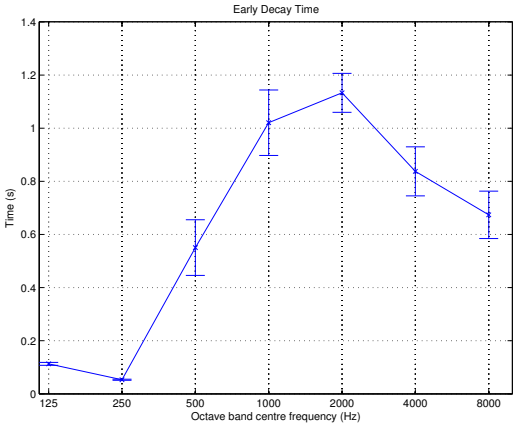
Figure 5.19 shows the mean and standard deviation of EDT in octave bands. The plots combine the results of all eight measurements made at each location. The results are very similar at each measurement position with a peak occurring at the 2kHz band. At the down-stage left position, the EDT shows a value of 1.2 seconds ($\sigma = 0.08s$). At the down-stage centre position the peak average EDT is 1.3 seconds ($\sigma = 0.07s$). At the down-stage right position the peak average EDT is 1.7 seconds ($\sigma = 0.09s$)



(a) Down-stage left



(b) Down-stage right

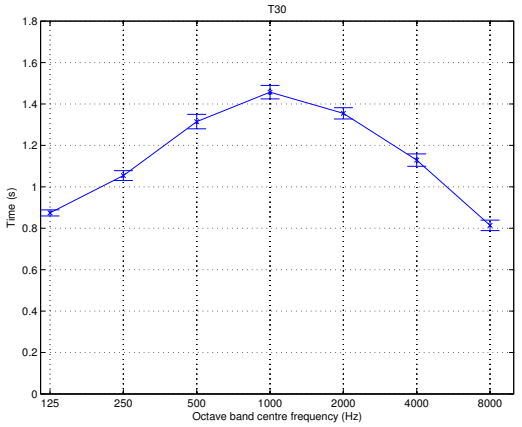


(c) Down-stage centre

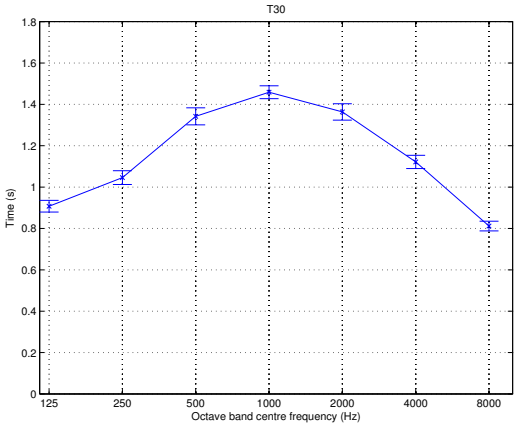
FIGURE 5.19: *Early Decay Time (EDT) measured at down-stage positions in LRR. EDT is displayed in octave bands and shows the mean and standard deviation of all measurements made at each measurement location.*

Figure 5.20 shows similar results for T_{30}. It can be seen that the peak value for T_{30} occurs at 1kHz and overall the variation in each octave band has reduced in comparison to EDT. Overall it can be seen that the low frequency reverberation time is significantly higher than when estimated with EDT. This includes the 500Hz octave band which is generally quoted when reporting reverberation times. At the down-stage left position, the peak average T_{30} is 1.46 seconds ($\sigma = 3ms$). At the down-stage centre position, the peak average T_{30} is 1.4 seconds

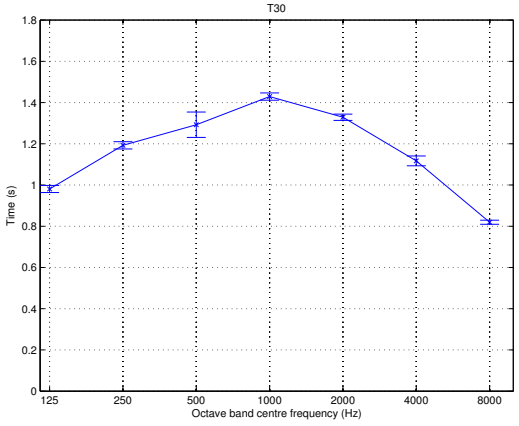
($\sigma = 1.7ms$). Finally, at the down-stage right position, the peak average T_{30} is 1.5 seconds ($\sigma = 3ms$).



(a) Down-stage left



(b) Down-stage right



(c) Down-stage centre

FIGURE 5.20: *Reverberation time (T_{30}) measured at down-stage positions in LRR. T_{30} is displayed in octave bands and shows the mean and standard deviation of all measurements made at each measurement location.*

5.3.3 Temporal distribution

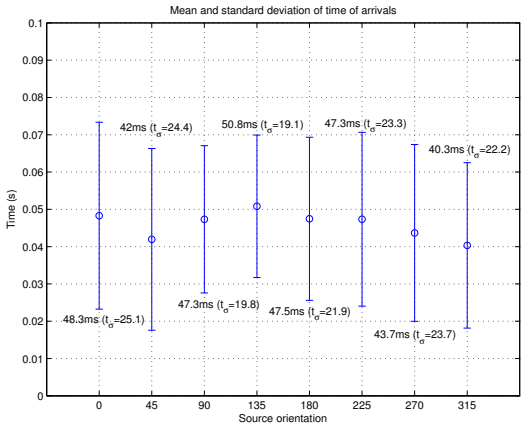
By utilising the temporal analysis method described in Chapter 4 it is possible to determine if the musician-related variables have any significant impact on the temporal distribution of early reflections.

Figure 5.21 shows the temporal distribution of early reflections for down-stage positions in LRR. Each plot shows the t_{mean} and t_σ for each source orientation. For each source orientation angle, t_{mean} is shown as a point and t_σ as the error bar.

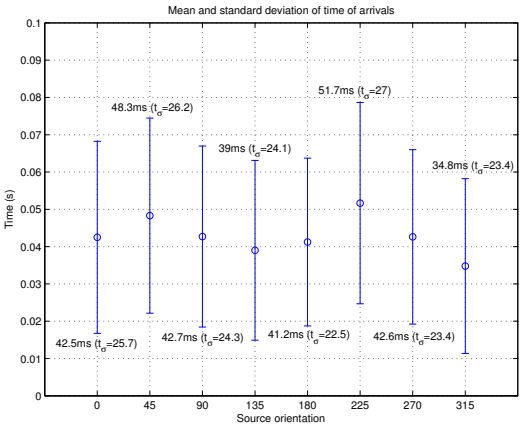
At down-stage left, t_{mean} varies between 40.3ms (at 315°) and 50.8ms (at 135°). t_σ varies between 19.1ms (at 135°) and 25.1ms (at 0°). At down-stage right, t_{mean} varies between 34.8ms (at 315°)

and 51.7ms (at 225°). t_σ varies between 22.5ms (at 180°) and 28.2ms (at 45°). At down-stage centre, t_{mean} varies between 42.4ms (at 315°) and 51.2ms (at 135°). t_σ varies between 21.4ms (at 180°) and 24.6ms (at 225°).

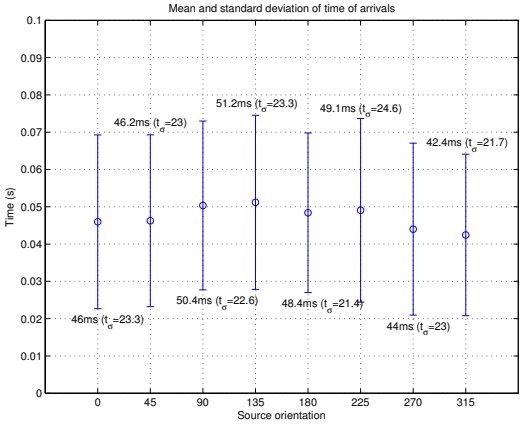
From this analysis it can be seen that in this hall, varying performer location and orientation has a small effect on the temporal distribution of early reflections. Varying these attributes appears to have a larger effect on t_{mean} than on t_σ which has a range of only 9.1ms throughout the analysis presented here.



(a) Down-stage left



(b) Down-stage right



(c) Down-stage centre

FIGURE 5.21: *Mean and standard deviation of time of arrival of early reflections at down-stage positions in LRR. Each plot shows the mean and standard deviation against source orientation angle.*

Overall it can be seen that the temporal distribution of early reflections does not change significantly in this hall with the mean time of arrival varying less than 10ms at any position. In the hall-related analysis performed earlier (Figure 5.11), it was shown that at the down-stage centre position, t_{mean} and t_σ varied comparatively less in LRR than in larger halls such as YH, RH and CH. Consequently, the temporal distribution of early reflections may be more sensitive to performer-related variables on larger stages

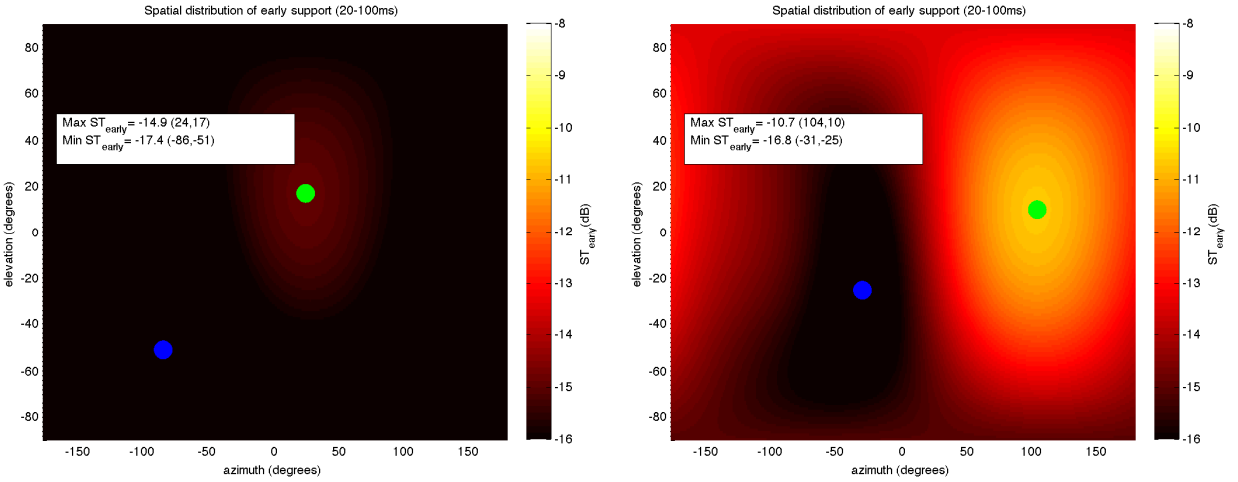
As the source orientation is varied at one particular position, the relative amplitude of early reflections will change due to the source directivity but the temporal distribution should not change significantly as the distance to each reflecting surface remains unchanged. In this case, the temporal distribution appears to vary as certain reflections drop below the detection threshold at certain source angles. Wider variations in temporal distribution are observable when the position on stage is varied.

5.3.4 Spatial distribution

By observing both broadband and octave band ST_{early} it was shown that the level of ST_{early} varies with source orientation and that change occurs mainly in the frequency bands where the sound source is more directional which in this case is in the higher frequency bands. It was shown that ST_{early} is highest when the source is oriented towards the nearest reflecting surface. It is of interest to observe the effect of source location and orientation on the spatial distribution of early reflections.

Figures 5.22, 5.23 and 5.24 show 3D distributions of ST_{early} measured at the down-stage centre position. Each plot shows the distribution for a single source orientation. ST_{early} is represented as a colour map with dark areas representing lower ST_{early} and brighter areas representing higher ST_{early}. A green marker is positioned at the maximum value of ST_{early} and a blue marker denotes the minimum.

It can be seen that the maximum value of ST_{early} tends to occur at a similar angle of azimuth to the sound source orientation. Furthermore, the results show very similar distributions of ST_{early} between measurement locations when comparing the same source orientation. For example, the distributions in Figures 5.22(b), 5.23(b) and 5.24(b) all show very similar distributions. This implies that the when a musician is performing on stage, the majority of early energy returning to the musician will be heard in the direction their instrument is pointing in, regardless of location on stage.



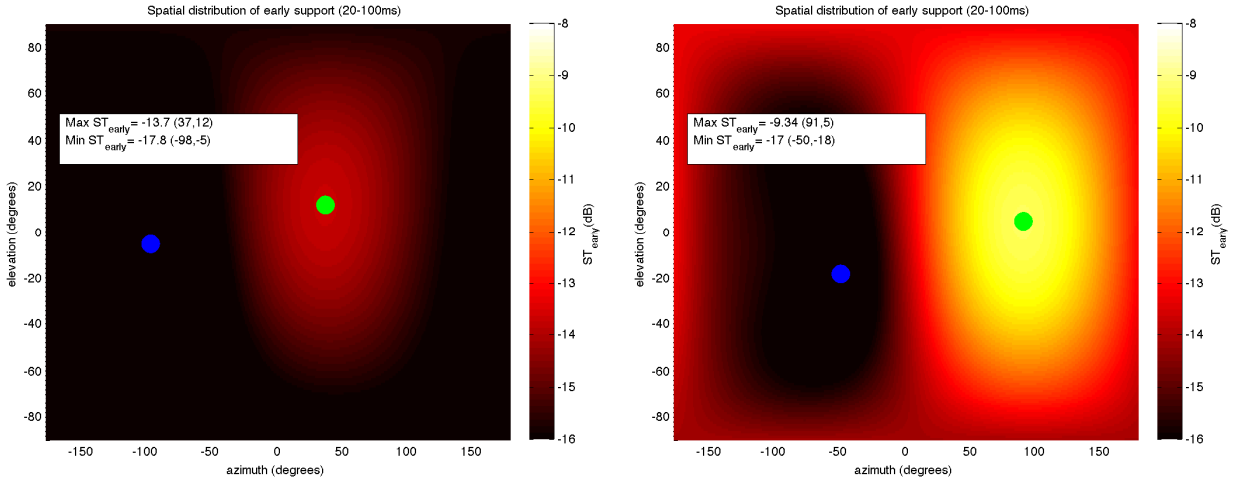
(a) source orientation = 0°

(b) source orientation = 90°

(c) source orientation = 180°

(d) source orientation = 270°

FIGURE 5.22: *3D distribution of ST_{early} measured at a stage centre position with source orientations of 0°, 90°, 180°, 270°. Green marker denotes the location of maximum ST_{early} and blue marker denotes the minimum. The maximum ST_{early} in each plot is very similar to the source orientation.*



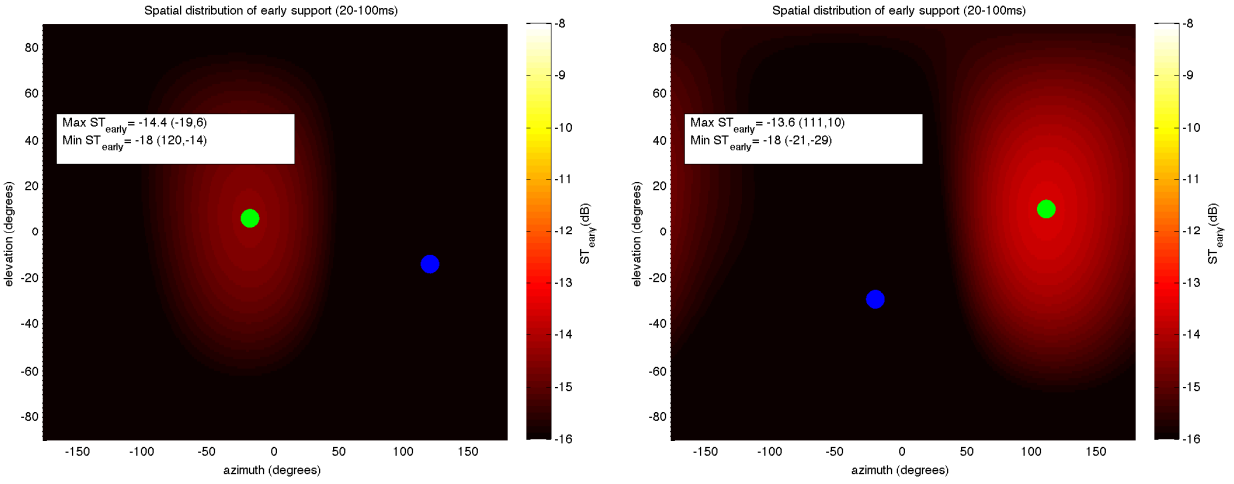
(a) source orientation = 0°

(b) source orientation = 90°

(c) source orientation = 180°

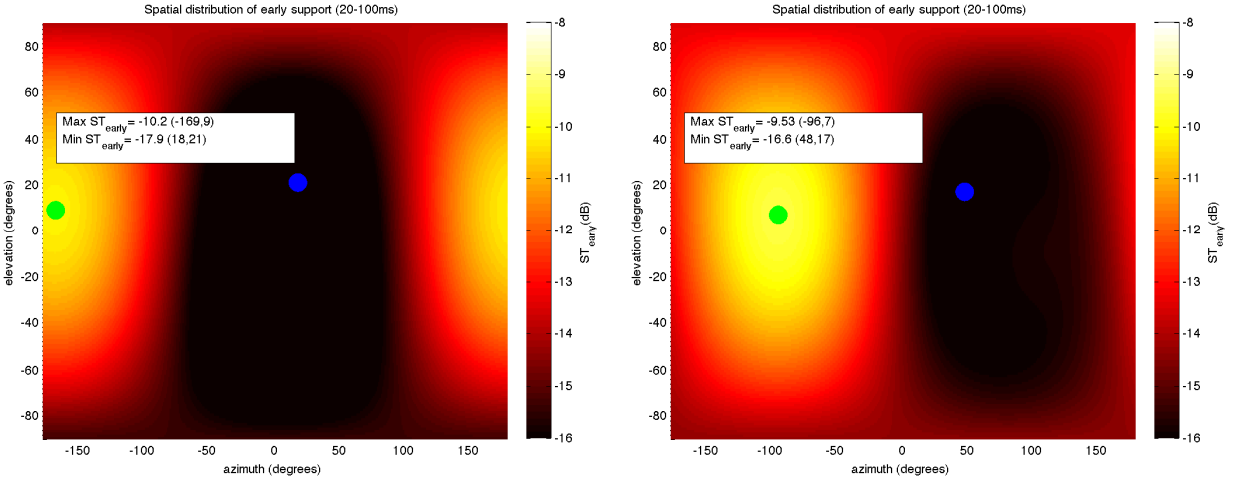
(d) source orientation = 270°

FIGURE 5.23: *3D distribution of ST_{early} measured at a stage right position with source orientations of 0°, 90°, 180°, 270°. Green marker denotes the location of maximum ST_{early} and blue marker denotes the minimum. The maximum ST_{early} in each plot is very similar to the source orientation.*



(a) source orientation = 0°

(b) source orientation = 90°



(c) source orientation = 180°

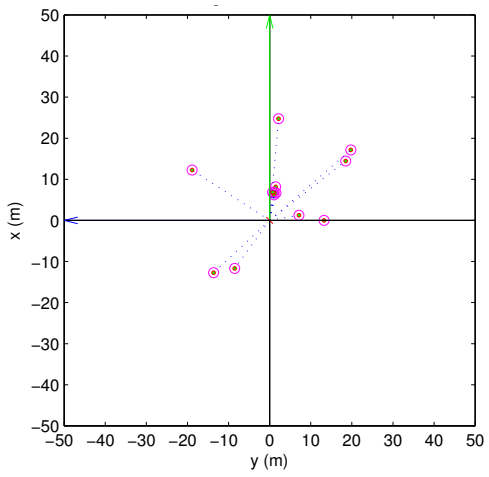
(d) source orientation = 270°

FIGURE 5.24: *3D distribution of ST_{early} measured at a stage left position with source orientations of 0°, 90°, 180°, 270°. Green marker denotes the location of maximum ST_{early} and blue marker denotes the minimum. The maximum ST_{early} in each plot is very similar to the source orientation.*

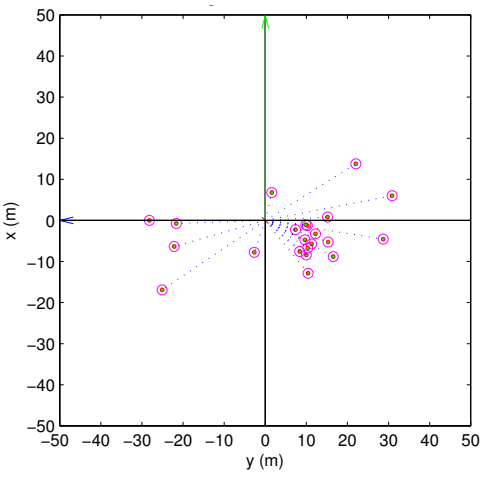
As different musical instruments have different radiation patterns, the distribution of early energy received will be different for performers of various instruments. Furthermore, performing musicians who change their orientation through gesture may cause the distribution of early energy to shift as they move or turn. These changes in spatial distribution may cause certain reflections to become more or less audible to the musician thus affecting their impression of the acoustic conditions on stage.

In order to observe the spatial distribution of reflections at these positions in more detail, image source plots are shown in Figures 5.25 (down stage centre), 5.26 (down stage right) and 5.27 (down stage left). Each figure consists of four image source plots for source orientations of 0°, 90°, 180°, 270°. The plots are shown in plan where stage front is oriented in the direction of

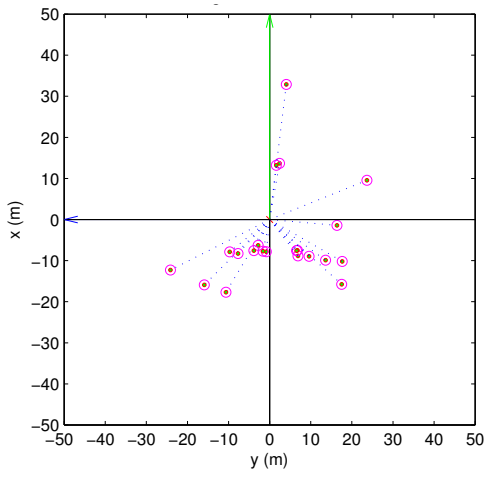
the green axis. The position of each image source is represented by a marker. In each plot, the azimuth and elevation of the mean resultant vector (MRV) is shown as well as the spatial spread.



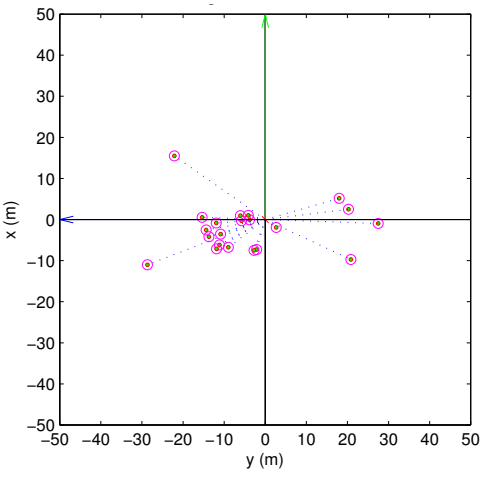
(a) source orientation = 0°,
MRV $(\theta, \phi) = (-18.7°, 10.7°)$, Spread = 0.45



(b) source orientation = 90°,
MRV $(\theta, \phi) = (117.4°, -18.0°)$, Spread = 0.54



(c) source orientation = 180°,
MRV $(\theta, \phi) = (-160.8°, -31.2°)$, Spread = 0.48



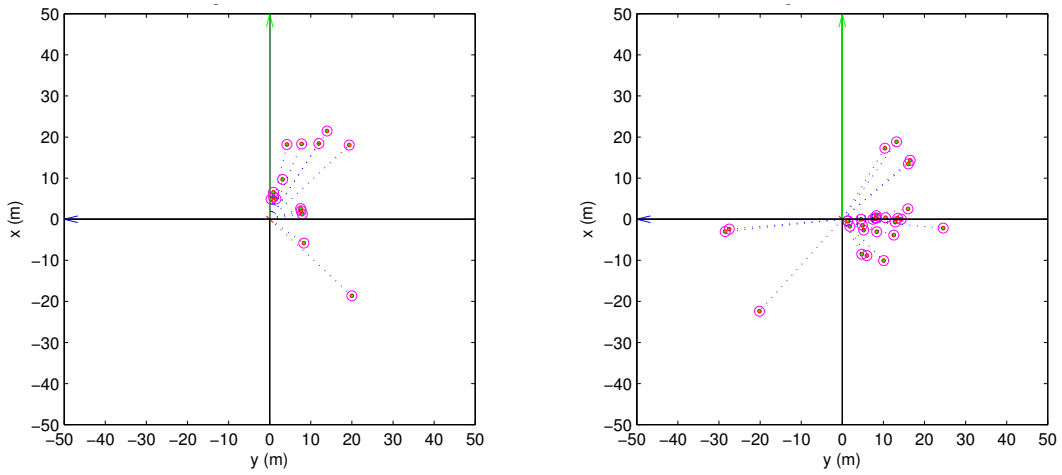
(d) source orientation = 270°,
MRV $(\theta, \phi) = (-114.1°, -24.0°)$, Spread = 0.51

FIGURE 5.25: *Image source plots of impulse responses measured at the down-stage centre position with the source oriented at 0°, 90°, 180°, 270°. Image source plots are shown in plan with stage front pointing to the top of the page. Each point represents the location of an image source. The azimuth and elevation of the mean resultant vector (MRV) is given for each plot in addition to the spatial spread.*

By observing the direction of the MRV and the position of the image sources it can be seen that the average angle of arrival varies with the sound source orientation. In general, the majority of image sources appear clustered around a similar angle to the source orientation with a small number of reflections appearing in other directions also. These image source are often much later in time (as represented by their larger distance from the origin) and have reflected off of more than one surface.

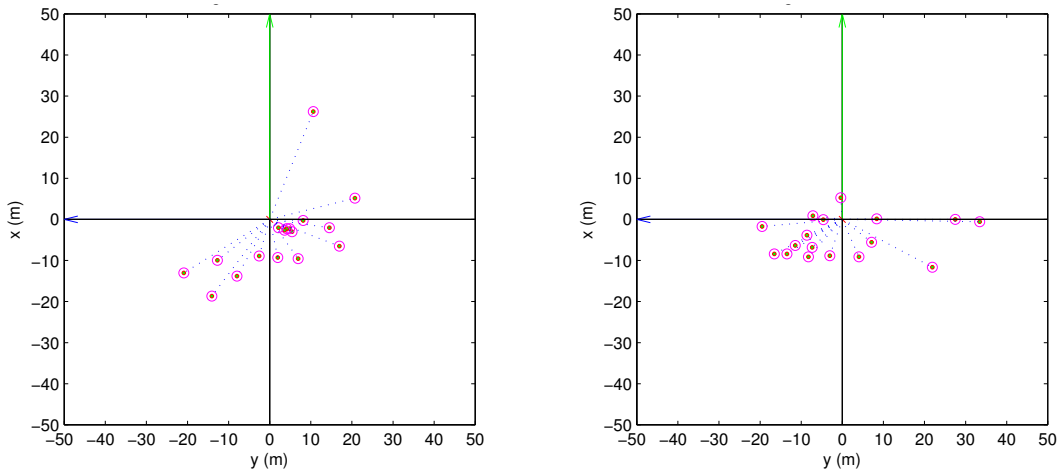
At the centre stage position, the distance to the left and right stage walls is 5.8m and 5.07m respectively and 3.95m from the stage rear wall. By observing Figures 5.25(b), 5.25(c) and

5.25(d), corresponding image sources can be seen to occur at positions twice these distances, incorporating the total time of flight of the reflection. Later reflections occur in the opposite direction to source orientation which correspond with 2nd order reflections that have encountered both stage left and stage right walls.



(a) source orientation = 0°,
MRV $(\theta, \phi) = (43.1°, -16.3°)$, Spread = 0.75

(b) source orientation = 90°,
MRV $(\theta, \phi) = (96.2°, -17.2°)$, Spread = 0.60



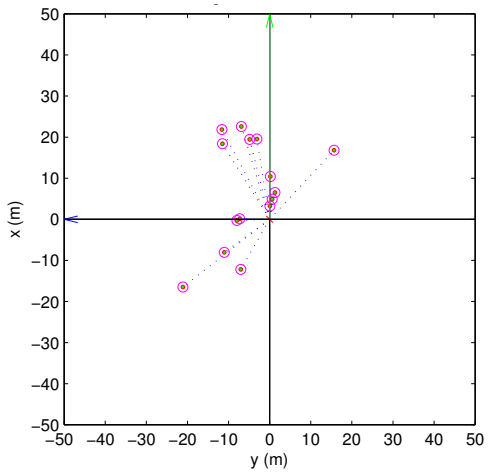
(c) source orientation = 180°,
MRV $(\theta, \phi) = (143.4°, -25.1°)$, Spread = 0.54

(d) source orientation = 270°,
MRV $(\theta, \phi) = (-153.4°, -34.0°)$, Spread = 0.42

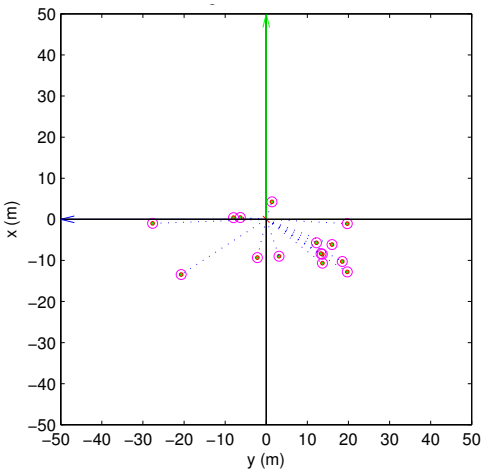
FIGURE 5.26: *Image source plots of impulse responses measured at the down-stage right position with the source oriented at 0°, 90°, 180°, 270°. Image source plots are shown in plan with stage front pointing to the top of the page. Each point represents the location of an image source. The azimuth and elevation of the mean resultant vector (MRV) is given for each plot in addition to the spatial spread.*

The image source plots obtained at stage left and right positions are shown in Figure 5.26 and 5.27 respectively. It can be seen that when the source is oriented towards the nearest wall the reflections are densely clustered together in that direction with a smaller number of later reflections arriving from other directions. In LRR, it can be seen that when the source is oriented towards the audience at 0° there are reflections which arrive from the audience area. These reflections are due to bleacher style seating in this venue. In larger halls, such as the

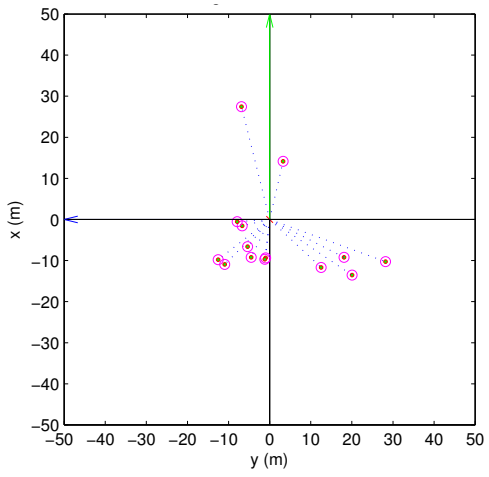
Caird Hall, there are no reflections detected from the audience area as the stage is raised and the distance to the nearest reflecting surface in that direction is very large.



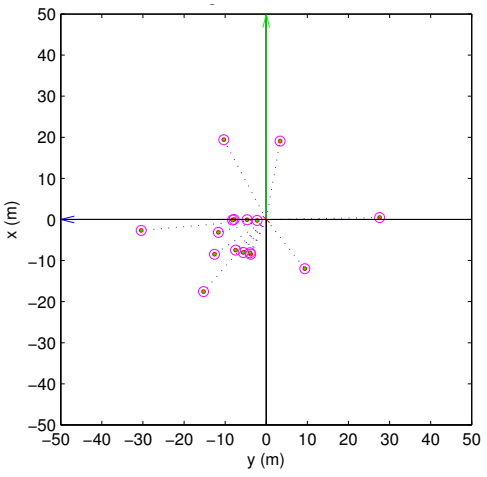
(a) source orientation = 0°,
MRV $(\theta, \phi) = (-35.1°, -9.9°)$, Spread = 0.54



(b) source orientation = 90°,
MRV $(\theta, \phi) = (144.0°, -25.7°)$, Spread = 0.45



(c) source orientation = 180°,
MRV $(\theta, \phi) = (-168.0°, -12.9°)$, Spread = 0.37



(d) source orientation = 270°,
MRV $(\theta, \phi) = (-115.9°, -18.3°)$, Spread = 0.54

FIGURE 5.27: *Image source plots of impulse responses measured at the down-stage left position with the source oriented at 0°, 90°, 180°, 270°. Image source plots are shown in plan with stage front pointing to the top of the page. Each point represents the location of an image source. The azimuth and elevation of the mean resultant vector (MRV) is given for each plot in addition to the spatial spread.*

5.3.5 Summary

The results presented in the preceding section focussed on variables relating to the musician (i.e. performer location and directivity/orientation). For brevity, the analysis focussed on the Ledger Recital Room (LRR) as a venue that more commonly hosts solo performance. The objective was to determine the effect performer-related variables on the acoustic conditions. Both established

acoustic parameter (such as ST_{early} and T_{30}) and also spatio-temporal analysis (as described in Chapter 4) were used in this analysis.

By observing broadband ST_{early}, it was found that the highest value would typically occur when the sound source was oriented towards the nearest reflecting surface. By observing the results in the frequency domain, it was shown that the change in ST_{early} was dominant in higher frequency bands with significantly more high frequency energy arriving when the sound source was pointed to a nearby reflector. It is likely that this was caused by the radiation pattern of the sound source which is significantly more directional at high frequencies. These results suggest that changes orientation may be more audible for musical instruments with a highly directional radiation pattern and that the early energy will vary mostly in the frequency bands that these radiation patterns occur in.

It was also observed that performer-related variables appeared to have a larger impact on parameters related to early reflections such as ST_{early} and EDT than ST_{late} and T_{30}. This suggests that the position and orientation of the performer mainly affects early reflections arriving from the stage enclosure.

The temporal distribution of early reflections was summarised using t_{mean} and t_{σ} as described in Chapter 4. Overall, it was found that these parameters only varied a small amount over all positions and orientations measured. As described in the hall-related analysis, LRR only showed a small variation in t_{mean} and t_{σ} when measured at the down-stage centre position. Larger halls such as Caird Hall and the Glasgow City Halls were found to vary much more. This suggests that performer-related variables affect the temporal distribution more on larger stages.

Two different techniques were used to demonstrate the variation in spatial distribution of early reflections. The first utilised a 3D distribution of ST_{early} which was obtained for each measurement. When measurements were compared across different stage locations with the sound source pointing in the same direction each time, it was found that the maximum level occurred in the same direction each time. While the overall level of ST_{early} changes with position, the directional distribution varies mostly with source orientation. This was confirmed using image source plots where it was evident that reflections tended to arrive from the direction the sound source was pointing in.

This has two important implications. Firstly, as a musician moves and gestures on stage, the dominant lobe of the instrument's directivity pattern will be changing direction, which will in turn vary the spatial distribution of early energy as well as the overall level of support. Secondly, as the directivity pattern of each musical instrument is unique, the spatial distribution of the early energy will vary with musicians of different instruments. This may cause certain early reflections to be masked (or un-masked) by the instrument contributing to musicians perceiving the same performance space differently.

The results obtained in this analysis suggest that the orientation/directivity of the musical instrument mainly affects the spatial distribution of reflections whereas the location on stage has a larger effect on the temporal distribution on stage. It is also clear that both of these effects are dependent on the geometry of the stage enclosure itself.

5.4 Summary and discussion

This chapter presented an objective analysis of measurements made on eight concert hall stages around Scotland. It was discussed how the early part of the acoustic response experienced by a performer could vary due to aspects related to the venue (i.e. stage geometry and construction) and also aspects related to the performer (i.e. musical instrument directivity, musician location on stage). The analysis techniques discussed in Chapter 4 were used to determine how the acoustic conditions vary in relation to these variables. Specifically, the analysis included traditional acoustic parameters, such as ST_{early} and T_{30}; and also spatial and temporal parameters (derived from image source plots derived from measurements) which were developed as part of this research.

As hypothesised, it was found that a number of halls which featured similar levels of ST_{early} could exhibit differing spatial and temporal distributions of reflections. The objective analysis implies that there are observable variations in spatial and temporal characteristics of early reflections as well as variations in overall energy (ST_{early}). It was found that the size of the stage affected the level of ST_{early} with higher levels captured on smaller stages. This follows, as reflections tend to have propagated a shorter path and have attenuated less through propagation losses. It was also shown that parameters related to the early part of the acoustic response, such as ST_{early} and EDT varied much more than those related to the later parts of the acoustic response i.e. ST_{late} and T_{30}. It was demonstrated that the early parts of the acoustic response varied according to source orientation and location on stage in addition to the physical dimensions of the stage itself.

In order to observe the impact of hall-related variables on the acoustic conditions, a comparison was made of the measurements made in the down-stage centre position in each hall. The results suggest that ST_{early} appears to reduce linearly with an increase in stage area. As the stage area increases so to does the average path length of early reflections which in turn increases the attenuation due to propagation, resulting in the reflections arriving later and at a lower overall amplitude. A linear increase in stage area was also found to result in an increase in ST_{late}, although to a lesser extent than ST_{early}. It was determined that this was more likely a consequence of the size of the venue as opposed to the size of the stage.

It was also shown that early reflections appeared more spatially clustered together when measured on larger stages. A linear increase in stage volume produced an increase in reflection clustering. The wider spatial spread of early reflections on smaller stages was attributed to off-axis reflections containing more energy on smaller stages (due to the lower distance attenuation) and the higher amplitude of higher order reflections that have encountered more than one reflecting surface. This was also the case for the temporal spread of reflections where reflections appeared to arrive spread over time in smaller halls than in larger halls. To a lesser extent, it was found that the mean time of arrival was later on stages with a higher physical volume.

An analysis was also conducted to determine the effect of varying performer-related variables such as source orientation and location on stage. This was achieved by focusing only on measurements made on one stage (LRR). It was shown that ST_{early} varies significantly with source orientation and also location on stage with the highest levels generally recorded when the sound source was oriented towards the nearest reflecting surface. The increase in level was found to be caused

mainly at higher frequencies, due to the directivity of the sound source used in the measurements. The variation in ST_{early} appeared to reduce when the measurement location was further from reflecting surfaces (such as stage-centre).

Three-dimensional distributions of ST_{early} suggested that regardless of measurement location, most early energy would arrive in a direction similar to the source orientation. This was confirmed using image source plots which showed clusters of early reflections occurring in these directions. The results also showed only small variations in temporal distribution with source position and orientation. However it was discussed that these effects may be amplified on larger stages (as suggested by the hall-related analysis). These results imply that the variations in performer directivity/orientation mainly affect the spatial distribution of early reflections while the position on stage has a larger effect on the temporal distribution.

It was also shown that the temporal distribution of reflections, as assessed using t_{mean} and t_{σ} did not vary significantly with performer orientation or location on stage. However, it was discussed that wider variations in temporal distribution appeared to occur on larger halls such as CH. In general, it was found that the temporal distribution varied more due to location on stage but was affected less by performer orientation.

In general, the results show that the temporal distribution of early reflections is affected mainly by the stage geometry and the performers location on stage. The spatial distribution of early reflections was found to vary mainly due to performer orientation. These variables also affect aspects of early arriving reflections such as ST_{early} and EDT but had little effect on the later, diffuse part of the acoustic response.

From these results, it can be seen that the spatial and temporal distributions of early reflections vary significantly on stage, even when the halls feature similar levels of ST_{early}. As discussed in Chapter 2, it is uncertain if these variations are audible to a performer, mainly due to the masking properties of the musical instrument and of the cognitive load required for a performance. Furthermore, it is uncertain what the subjective effect of these variations in reflection distribution is for a performing musician.

Therefore, it is proposed that a number of listening tests are required in order to determine if the spatio-temporal distribution of early reflections is audible for a performer and if there is an effect on any perceptual attributes of the space (timbre, support etc). Further investigation is proposed to determine if these variations result in a change in preference towards playing in a particular venue. The following chapter will describe how the survey data was manipulated and auralised interactively to allow perceptual testing to take place.

Chapter 6

Auralisation of Stage Acoustic Conditions

In the previous chapter, it was demonstrated that the spatial and temporal distribution of early reflections varied significantly with aspects such as source directivity, the musician's position on stage and the stage geometry. It was also found that it was possible for halls with similar levels of ST_{early} to have different distributions of early reflections. As the spatial and temporal distribution of early reflections are known to influence how sound is perceived in enclosed spaces, it follows that additional acoustic parameters may be required to completely describe the acoustic conditions on stage. However, it is currently unclear if variations in the spatial or temporal distribution of early reflections are audible to a performer, or if there is an associated subjective effect on the musician's impression of the space. Therefore, a series of listening tests are proposed in order to observe the subjective effect of varying the distribution of early reflections.

Previous research in stage acoustics approached this by interviewing musicians that had played in specific venues and relating their responses to measurements made later on stage (Gade, 1982). This revealed important clues as to what were the salient aspects of the acoustic response for various musician groups and what aspects assisted the performer's efforts. Gade (1982, 1989) realised the benefit of introducing musicians into laboratory conditions to investigate these effects and developed an interactive auralisation system capable of presenting the musician with specific acoustic conditions in response to the sound of their instrument. This approach allows the researcher to have increased control over the acoustic conditions experienced by the test participants and provides increased repeatability and convenience into the research. It also allows the musician to be introduced into abstract acoustic environments which allow researchers to determine, for example, the threshold of perception of a single reflection.

Since Gade's pioneering research, auralisation systems have developed significantly with a drive to understand how acoustic designs influence the audience's impression of a performance. A corresponding development in similar interactive auralisation systems have allowed researchers (Brereton et al., 2012a, Guthrie, 2014, Ueno and Tachibana, 2003, Yadav et al., 2013a) to introduce performers into increasingly realistic acoustic environments, providing further insight

into successful acoustic designs and appropriate acoustic parameters. Due to the requirement of auralisation interactivity, there are a number of technical challenges encountered when building interactive auralisation systems. Due to these technical limitations, stage acoustic research continues to utilise both in-situ and laboratory testing paradigms to test hypotheses.

In the context of this research, an interactive auralisation system could allow the temporal or spatial distribution of early reflections to be varied in isolation from ST_{early}. In addition, it is possible to instantaneously compare different concert hall stages by performing on them which will reduce the influence of external variables. Therefore, it is an attractive option for conducting listening tests with musicians. This chapter will compare and contrast previous approaches to interactive auralisation for stage acoustic research and go on to demonstrate a new approach based on recent developments in the field of auralisation. It will discuss how real spaces, such as those measured in Chapter 3, can be effectively emulated using this technique.

6.1 Interactive auralisation

Auralisation systems enable the acoustic response of a virtual or existing space to be applied to anechoic recordings, thus allowing the listener to hear the effect of the environment on the sound source. This can demonstrate the efficacy of different acoustic designs, especially when used in conjunction with advanced acoustic modelling techniques. Auralisations impart a measured or modelled acoustic response to an audio signal (that has been recorded in anechoic conditions) via a mathematical procedure known as convolution. The resultant signals are then either rendered over headphones or over a loudspeaker array to a listener.

As discussed in Chapter 2, the perceived impression of a space can be significantly influenced by the spatial distribution of early reflections as well as the temporal distribution. Therefore, it is common for the impulse response to be captured in such a way that the spatial information of each reflection is encoded into the impulse response. This ensures the resulting auralisation is as plausible as possible. At the time of writing there are numerous methods of spatialising audio signals over loudspeaker arrays or headphones, each of which has advantages and disadvantages. The technique used is often dependent on the resources available and the requirements of the individual auralisation.

More recently, auralisations have started to include dynamically varying elements allowing, for instance, the listener to feel as if they are moving around a space or allowing them to experiment with different acoustic designs (Aspock et al., 2014). Such auralisations are said to operate in real-time as the impulse response is dynamically updating or interactive in the sense that the listener can make changes to their acoustic environment with an imperceptible amount of delay.

However, the auralisation is not interactive in the sense that in a real-life situation, a person could produce sounds which are then modified by the environment and then heard by that person. In this thesis, an auralisation system is considered *passive* if the listener is only required to listen to an auralisation. By contrast an auralisation is considered *interactive* if the listener also acts as the sound source. An interactive auralisation system could allow a listener to excite the space,

for example, by clapping their hands or playing their musical instrument and listening back to the resulting response.

An interactive auralisation system continuously captures the direct sound created by a musician which is then processed by a system capable of rendering the desired acoustic conditions found on-stage in a particular venue. A musician is then able to play their instrument in a virtual representation of the target space and provide immediate comparisons with other stages. A system like this could allow the musician to practice in a space prior to a public performance in order to refine certain aspects of their technique. This research uses such a system to perform listening tests with musician test participants to ascertain what aspects of the acoustic response are audible to the performer and what is the preferred delivery of the acoustic response.

The requirement of interactivity presents a number of additional technical challenges in comparison to passive auralisation systems. Like passive systems, it must be capable of reproducing the acoustic response of a target space such that reflections arrive at the musician at the correct time, at the correct amplitude (with no timbral artefacts) and from the correct angle of arrival. In interactive systems, the timing of reflections is relative to the direct sound produced by the musician and so must be rendered without a noticeable delay.

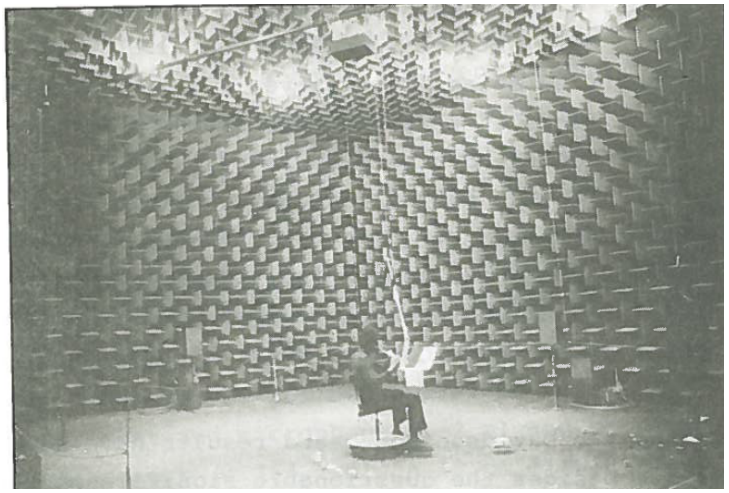
As with passive auralisation systems, the system should not introduce any further room response onto the auralisation as this may produce conflicting cues. Often, auralisations are rendered within an acoustically treated laboratory that does not contribute significantly to the virtual acoustic response. The audio equipment should allow for accurate capture of the direct sound of the musician but not be overly invasive or restrict movement of the musician. Furthermore, if the system is required to recreate the acoustic conditions of a real space, the spatial attributes of the soundfield must be accurately captured so that they can be reproduced in the laboratory.

The introduction of a live microphone into the auralisation system presents a number of potential issues as it is possible that positive feedback could produce audible ringing through the audio system. It is also required that the microphone be capable of capturing a fair representation of the tone of the musical instrument which may not be possible with a single transducer.

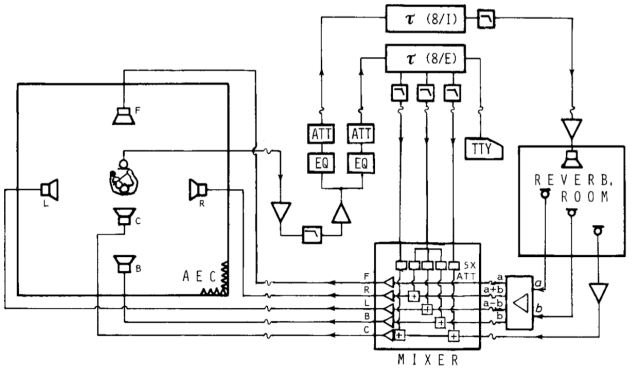
6.1.1 Existing interactive auralisation systems

Interactive auralisation systems have been used in the past for stage acoustic research and have generally kept pace with developments in the field of spatial audio and auralisation. One of the first examples of such a system was devised by Gade (1982, 1989). Figure 6.1(a) shows a musician performing in Gade's auralisation system. The musician was positioned inside an anechoic chamber at the centre of an array of five loudspeakers. The direct sound from the musician was captured by a highly directional microphone which was then electronically delayed, attenuated and filtered to simulate the effect of a limited number of early reflections (as shown in Figure 6.1(b)). The direct sound was also played through transducers set up in a reverberation chamber to simulate a diffuse reverberation. The resultant signals were summed and played back over the surrounding loudspeaker array presenting the virtual soundfield to the musician. The system allowed for basic control of a limited number of early reflections including the time and

angle of arrival, amplitude and frequency content. This system allowed Gade to study specific aspects of stage acoustics including the threshold of perception of single, or groups of, reflections and ultimately led to the development of ST_{early} discussed previously in Chapter 2.



(a)



(b)

FIGURE 6.1: *6.1(a) shows image of musician in interactive auralisation system constructed by Gade. 6.1(b) shows systems diagram of system. Early reflections were reproduced with delays and equalised to provide control over frequency content. A reverberation room was used to produce the reverberant decay. Both images from Gade (1989)*

In this pioneering study, Gade discussed in detail some of the limitations of the auralisation system in comparison to performing tests in real auditoria. It was shown that the fidelity of the auralisation could be influenced by many factors including microphone placement and type, loudspeaker quality and accuracy of acoustic response simulation.

As low-latency convolution engines were developed, Ueno and Tachibana (2003) devised a similar system which allowed the acoustic response of real stages to be more accurately measured and recreated in real-time over a loudspeaker array. This system, shown in Figure 6.2, recreated the acoustic response of measured spaces directly by convolving the direct sound from an instrument with six measured room impulse response simultaneously in real-time. The room impulse response in the auditoria was measured repeatedly using a directional microphone oriented along Cartesian axes (6 impulse responses per position). The loudspeakers recreating the acoustic

response were positioned in the corresponding directions relative to the musician where each loudspeaker recreated the associated acoustic response in that direction. As this system relied upon digital signal processing (as opposed to analogue delays and reverberation chambers) it was possible to reproduce highly plausible stage acoustic responses based on real environments. In addition, by editing the measured impulse responses prior to auralisation, it was possible to have an increased level of control over the virtual stage acoustic response. For example, by applying a gain envelope to parts of the impulse response, it was possible to emphasise reflections from a particular direction (Ueno and Tachibana, 2003).

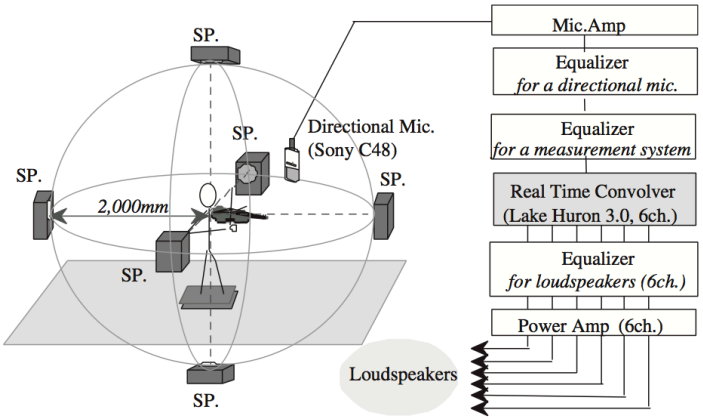


FIGURE 6.2: *System diagram of the interactive auralisation system developed by Ueno. The sound from the musician is picked up by a directional microphone and convolved with a 6-channel room impulse response. Each impulse response was measured using a directional microphone pointing in the direction of each loudspeaker. Image from Ueno and Tachibana (2003)*

This method was later expanded by Woszczyk (2006) who measured the acoustic response of a room, using a much larger array of microphones. The array consisted of a spaced arrangement of four omnidirectional microphones and four figure-of-eight microphones. Impulse responses were recorded with the array set at three different heights. These impulse responses were used to reproduce the acoustic response of the room in real-time for a performer over an array of 24 ribbon loudspeakers. Figure 6.3 shows a performer playing a harpsichord in a virtual space. This system has been used previously to record performances from a musician who is experiencing the acoustics of a specific space. This can help capture an improved performance from the musician.

The use of a spaced microphone array to record the acoustic response will ensure that the direction of arrival of reflections are encoded by inter-channel temporal and amplitude differences. This is in contrast to coincident microphone techniques which utilise only amplitude differences between microphone capsules (Laitinen, 2014). When rendered over a loudspeaker array, coincident microphone recordings cause each loudspeaker signal to be highly correlated which can cause timbral or phasing artefacts at the sweetspot. The inter-channel delays from a spaced microphone recording are much less correlated producing a much more natural reproduction of the soundfield. However, the delays between loudspeaker signals can produce conflicting localisation cues (time-intensity trading (Howard and Angus, 2001)) depending on the size of the microphone array which can lead to ambiguous localisation of sound sources (Laitinen, 2014). However, by

using an increased number of loudspeakers to recreate the virtual acoustic response it is possible that reflections are spatialised to a high degree of accuracy.

FIGURE 6.3: *Image of the interactive auralisation system developed by Woszczyk. The sound from the musician is picked up by a directional microphone and convolved with a 24-channel room impulse response. The room response is rendered over a 3D loudspeaker array which is visible behind and to the left of the musician. Impulse responses were measured using an array of 8 directional microphones positioned at 3 different heights. Image reproduced from Woszczyk and Martens (2008)*

An alternative approach involves the use of Binaural room impulse responses, where the direct sound from the instrument is processed with a two channel impulse response with spatial information encoded using HRTF. A binaural approach offers a number of important advantages in that the spatialisation of binaural auralisations is renowned for being highly accurate and does not require an extensive loudspeaker array to operate. This approach was previously used by Yadav et al. (2013a) to perform listening tests with vocalists. Figure 6.4 shows the system in more detail. The headphones used were open and positioned near the musician's ears so that the headphones would not interfere with the direct sound of the musician's voice. However, binaural systems work best when the binaural recording has been made with the same pinnae as the test subject which is often a significant challenge. A further aspect to note is that the system was head-tracked allowing the test participant to move their head freely in any direction with the rendered impulse responses updating accordingly to the resulting acoustic changes.

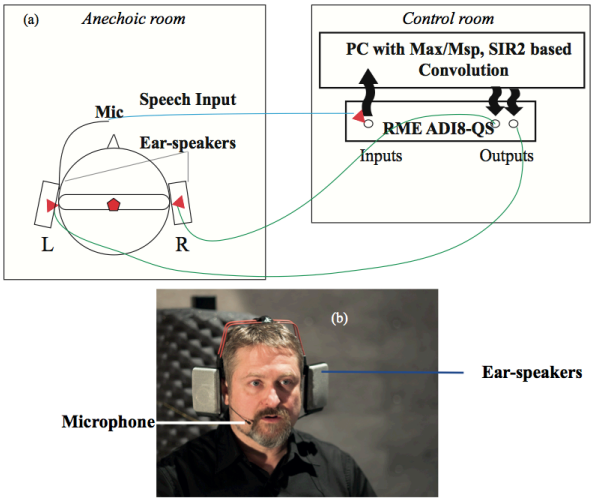


FIGURE 6.4: *System diagram of the interactive auralisation system developed by Yadav. The sound from a vocalist was picked up by a head-mounted microphone and convolved with an Oral Binaural Room Impulse Response (OBRIR). The resultant acoustic feedback was rendered over head-mounted loudspeakers positioned close to the ears. Image reproduced from Yadav et al. (2013b)*

More recently, authors such as Watson and Clark (2010), Guthrie et al. (2013) and Brereton et al. (2012b) have made use of Ambisonic techniques and real-time software convolution engines to render highly realistic stage acoustic conditions over a loudspeaker array. As discussed previously in Chapter 3, Ambisonics is an ideal format for measuring and transforming 3D soundfields. Ambisonic methods (in particular Higher Order Ambisonics) have also been shown to be highly effective at rendering accurate soundfields over loudspeaker arrays and can allow a high degree of flexibility in terms of the layout of the loudspeaker array (Kearney, 2009). Furthermore, many commercial acoustic modelling packages are capable of rendering synthesised impulse responses in Ambisonic B-format allowing the musician to play in spaces which have yet to be built or are in the process of being designed.

Similar interactive auralisation systems have also been used to augment the acoustic conditions of existing stages for example *The Virtual Acoustics Technology rendering system* at McGill University. Systems such as these can improve aspects of an existing stage by, for example, increasing the level of ST_{early} or rendering additional reflections from specific directions. These systems are analogous to assistive reverberation systems which use a series of microphones, signal processors loudspeakers to control the reverberation time of a performance space. Ko et al. (2013) have utilised such a system to investigate the subjective effects of stage support on ensembles.

In general, it can be seen that there are a number of processing stages common to all such interactive auralisation systems for stage acoustic laboratory experiments, all of which must be considered carefully in the overall system design. The *microphone system* (used to measure the direct sound from the musician), *a signal processing stage* (used to apply the acoustic response to the direct sound) and a *loudspeaker array or headphone system* (used to render the acoustic response back to the musician). A basic system diagram is shown in Figure 6.5.

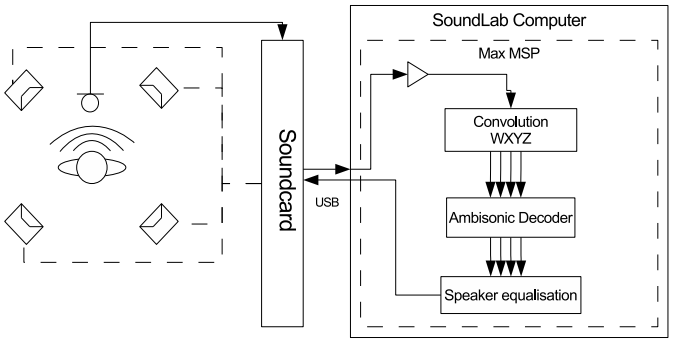


FIGURE 6.5: *System diagram of a typical FOA-based interactive auralisation system (only 4 loudspeakers are shown for clarity)*

It is clear that there are numerous methods of implementing an interactive auralisation system. Each system will be optimised for a specific study and will often be based on the available equipment. The design of the auralisation system and its intended use should be considered in tandem with the proposed measurement procedure as an alteration in one aspect can have consequences for other parts of the system. The fundamental aspects of the interactive auralisation system will be discussed throughout the remainder of this chapter.

6.2 Auralisation quality

Like many other sensory evaluation studies which use auralisation systems, the lack of plausibility and context can limit the reliability of the results in comparison to in-situ tests. A laboratory environment is a notably unnatural space for a musician to perform in. In general, these often feature heavy acoustic treatment and a large amount of audio equipment which may limit the movements of the performers. Auralisations may, out of necessity, focus only on simulating the acoustic response of the space rather than attempting to create a multi-modal sensory stimulus, i.e. including visual stimuli. For any sensory evaluation laboratory test, it is generally accepted that there are certain practical limitations to such a system and there may be a resulting affect on the responses of the musicians. Therefore, laboratory tests should aim to complement results from in-situ tests.

As part of this research, informal listening tests with musicians found that a poorly calibrated auralisation could induce what is described as a *'PA effect'* (Laird et al., 2011). This term refers to musicians feeling as if they were not playing naturally in the space, rather it felt as if they were playing through a sound reinforcement system in that space. This effect detracts from the auralisation being representative of a typical performance situation and so this effect should be minimised as far as possible. The PA effect can be caused by numerous contributing factors including, but not limited to:

- Proximity effect of a directional microphone
- Timbral distortion from sampling of the instrument's directivity pattern

- Dynamic range compression from proximity of microphone
- Unwanted acoustic feedback in the auralisation system
- Frequency response of loudspeaker and microphone
- Capture of impulse response (i.e. distortion, signal-to-noise ratio etc)
- Spatial audio rendering of impulse response (comb-filtering, lack of spatial resolution etc)
- Soundfield rendered at an incorrect level
- Contribution of acoustic response of laboratory

The following sections will discuss how the effect of these artefacts can be reduced by careful design of the auralisation system.

6.2.1 Direct sound capture

In order to accurately capture the direct sound from a musical instrument, careful consideration must be given to the microphone or microphone array and the position of transducer(s) around the instrument. Fundamentally, for interactive auralisation, the direct sound of the instrument should be captured so that the signal does not contain an audible room presence, contains similar spectral content to the original and does not introduce any unwanted positive feedback into the auralisation system (characterised by tonal ringing). Furthermore, the number and position of microphones should not restrict the movements of the musician or be invasive during the auralisation.

In previous stage acoustic auralisation systems (Gade, 1982, Ueno and Tachibana, 2003), a single directional microphone was positioned at close proximity to the musical instrument. The captured signal was then auralised with the appropriate stage acoustic impulse response. When testing vocalists, the microphone is often head-mounted and positioned very close to the musician's mouth (Brereton, 2014, Yadav et al., 2013b). A similar arrangement could be used for instrumentalists by physically attaching the microphone to their instrument. However, some musicians may be wary of attaching a microphone to their instrument for fear of damage. Furthermore, careful thought would have to be given regarding the appropriate position for the microphone on every instrument tested.

Positioning a microphone at close proximity to the musical instrument has numerous advantages including minimising latency introduced into the auralisation from propagation delay. It also has the effect of reducing the required microphone gain to capture a signal with sufficient signal-to-noise ratio; reducing the likelihood of unwanted acoustic feedback loops being created in the auralisation system. This effect is enhanced by using a directional microphone which will reject sound arriving from other directions.

However, close-proximity microphone techniques can introduce unwanted artefacts into the captured signal, the most well known effect being an un-natural emphasis of low frequencies (known as the proximity effect), which is inherent in all directional microphones, due to the design that

produces their directional response. The low-frequency boost can be generally described using equation (6.1) (Clifford and Reiss, 2011):

$$\beta_{dB} = \frac{\sqrt{\frac{1+(2\pi r}{\lambda}^2)}}{\frac{2\pi r}{\lambda}} \tag{6.1}$$

where β_{dB} describes the low frequency boost in decibels due to the distance r between source and microphone (in centimetres) and the wavelength, λ. The effect will differ with the construction of individual microphones but with this knowledge it is possible to produce an inverse filter to correct the low-frequency boost. This approach has been used with some success by Kearney (2009) in order to reduce the effect of the recording and reproduction chain on auralised sound-fields. However, in this case, the inverse filter will only be valid for a single source-receiver radius and so if the musician moves relative to the microphone, audible artefacts may be introduced.

The close proximity of the microphone also captures ancillary sounds from the instrument such as key clicks and breath sounds etc. These sounds are then auralised with the same gain as the intended sound from the instrument. This can sound very similar to the dynamic range compression often featured in PA systems when a microphone is positioned very close to an instrument.

A further concern is that performers are rarely static when they perform in public. They move and gesture for both expressive and functional reasons. These movements relative to a static microphone may cause unintended modulations in the frequency content of the measured signal. It is possible to alleviate this problem by attaching the microphone to the instrument, however this may not be possible for all instruments. Similarly, requesting a musician to restrict their movements can be problematic as they may not feel comfortable.

In addition to the proximity of the microphone, the angular location of the microphone relative to the musician can also affect aspects of the measured direct sound. As musical instruments radiate sound in a non-uniform, frequency-dependent manner, a static microphone positioned near the instrument is effectively sampling this radiation pattern. This sound will have a given frequency content which will differ from that measured at other angles, neither of which may be a true representation of the overall timbre of the instrument. This could influence the perceived impression of an auralisation created with these signals as recognised previously by Rindel and Otondo (2005).

Using a surrounding array of microphones to capture the direct sound from the musician would allow the radiation pattern of the instrument to be included in the auralisation. This technique has been used previously for study of the directional characteristics of musical instruments but not (at the time of writing) for use in interactive auralisation (Lokki et al., 2008, Nachbar et al., 2010). Figure 6.6 shows a basic microphone and DSP system that could be used to include dynamic variations in directivity as demonstrated by Menzies (2010). The sound source has two directional microphones positioned in front with an angular separation of 90°. The signal from each microphone is fed into a reverberation processor with different settings and summed together. As the sound source orientation varies, the frequency content and amplitude from each

reverberator varies. This causes the auralisation to vary dynamically with source orientation. The reverberation processors could utilise impulse responses measured with directional sound sources at various source orientations which would result in a more lifelike auralisation where the acoustic characteristics varied with source directivity.

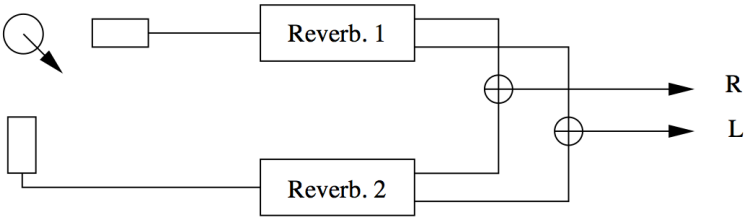


FIGURE 6.6: *Diagram showing two cardioid microphones positioned in front of a directional sound source with an angular separation of* 90°. *Each signal is fed into a separate reverberation processor, each with different settings. As the source changes orientation, the characteristics of the auralisation change also. If each reverb consisted of an impulse measured with a directional sound source at the same orientation then it would be possible to include some dynamic variations in directivity. Image extracted from Menzies (2010).*

It is possible to extend this idea even further by using additional microphones placed around the musician as described by Skavelik (2006) and Otondo et al. (2002). Extending this even further would involve the use of a spherical array of microphones positioned around the musician. The signals from these could then be decomposed into spherical harmonics and auralised (Pollow and Gottfried Behler, 2009). A technique such as this would have the effect of improving the overall timbre of the auralisation but potentially could also allow the dynamic directional characteristics of any instrument to be included in an auralisation without the need for different microphone arrangements. Introducing a large number of microphones as well as the existing loudspeaker array, could create a visual impact which leads to the musician being completely surrounded by transducers which may contribute to their responses when playing in the interactive auralisation. It has been demonstrated that a compressive sampling technique, using fewer microphones randomly positioned on the unit sphere, can be used to determine the radiation pattern of a sound source (Masiero and Pollow, 2010).

It should be noted however, that the use of more complex microphone arrays may introduce large processing demands on the auralisation system which may affect its ability to render the acoustic conditions in real-time. Therefore, future work in this area should investigate how the radiation characteristics of the instrument could be captured efficiently in addition and if there is an audible benefit in terms of auralisation plausibility.

6.3 Acoustic impulse response measurement

Chapter 3 discussed the different approaches to measuring stage acoustic impulse responses for the purpose of analysing the spatial and temporal distribution of early reflections. It was discussed how aspects of the sound source, receiver and their relative positions could affect what

information could be extracted from the impulse response. These aspects can also influence how the impulse response is captured for auralisation purposes and can have audible consequences when a sound is auralised. Therefore, the measurement technique should be developed so that the captured signals can be used for both objective analysis and auralisation

The proposed measurement technique was considered to be highly effective for obtaining appropriate impulse responses for objective analysis however it was acknowledged that the spatial resolution of the measured impulse responses was likely to be limited. The first-order Ambisonic microphone was capable of encoding the spatial characteristics of the soundfield onto a single impulse response measurement which ensured the venue surveys were as efficient as possible. As these microphones are relatively cheap and readily available the measurement procedure could be easily reproduced by other researchers. The use of first-order Ambisonics was considered appropriate for use in the laboratory set up and would also allow for complex manipulation of the spatial characteristics of the impulse response.

As discussed previously in Chapter 3, it is common for room impulse responses to be measured with omnidirectional sound sources so the measurements adhere to international standards. By doing so, the measured impulse response is independent of source orientation or directivity. One of the aims of this research is to determine if the spatial distribution of early reflections influences the musician's impression of the venue. The use of an omnidirectional loudspeaker may cause reflections from the rear of the stage enclosure to be un-naturally emphasised and thus exaggerate the differences between venues. Therefore, an approximation of an instrument's directivity pattern was considered to be sufficient. In Chapter 3, a technique was described where eight impulse responses were measured with a directional sound source oriented at set angular orientations. The use of the directional sound source ensured that the stage acoustic impulse response included some directional characteristics. By making repeated measurements with the sound source pointing in specific directions it was possible to auralise the effect of a musician pointing in different directions. It was also considered possible that more complex directivity patterns could be synthesised by a weighted, linear summation of these measurements.

It is clear that the measurement technique used for measuring the acoustic response should be considered carefully such that the results are compatible with available analysis tools, auralisation equipment and research objectives. It is currently unclear exactly how sensitive musicians are to aspects such as the spatial distribution of early reflections. Consequently, it is uncertain what degree of spatial resolution is required for suitable auralisation of stage acoustic conditions. Future stage acoustic research should aim to determine these aspects so that a suitable measurement technique can be agreed upon.

6.4 Convolution

A fundamental element of any auralisation system is the process by which the acoustic response of the space is imparted to the direct sound of the musician's instrument. This can be achieved using the mathematical process of convolution, which can be expressed in discrete-time notation

as in equation (6.2) (Battenberg and Avizienis, 2011) where '$*$' denotes the convolution operation, $x(n)$ is the input signal and $h(n)$ is the impulse response.

$$y(n) = x(n)^* h(n) = \sum_{k=0}^{N-1} x(k)h(n-k) \tag{6.2}$$

This operation filters the source signal $x(n)$ with the impulse response of the system, $h(n)$, which in this case is the impulse response measured on stage. As shown in Chapter 3, the impulse response measured on stage with an Ambisonic microphone contains the spatial attributes of reflections in addition to the amplitude response. Therefore, if a sound, recorded in free-field conditions, is convolved with a spatial room impulse response, that sound will inherit the spatial and temporal characteristics measured in that room. Consequently, when the sound is played back over a suitable audio system, the direct sound appears as if it was playing in the room (Farina and Tronchin, 2005).

For audience auralisation, the programme material is often convolved with the impulse response 'offline' so that the system playing back the auralisation is only required to play back preprocessed audio signals rather than convolving the audio in real-time. In the case of stage acoustic auralisation, the input signal is being generated live by the musician and so this operation must occur in real-time with minimal processing delay (latency).

Equation (6.2) refers to the operation performed in the discrete time domain. While this method has no inherent latency it requires significant processor power. The amount of processor power required increases linearly with the size of the impulse response. As was discussed in the previous chapter, a stage acoustic impulse response can be as long as 2.5 seconds resulting in a significant amount of required processing power. When there is insufficient computational resources, the resultant audio contains processing artefacts and distortion.

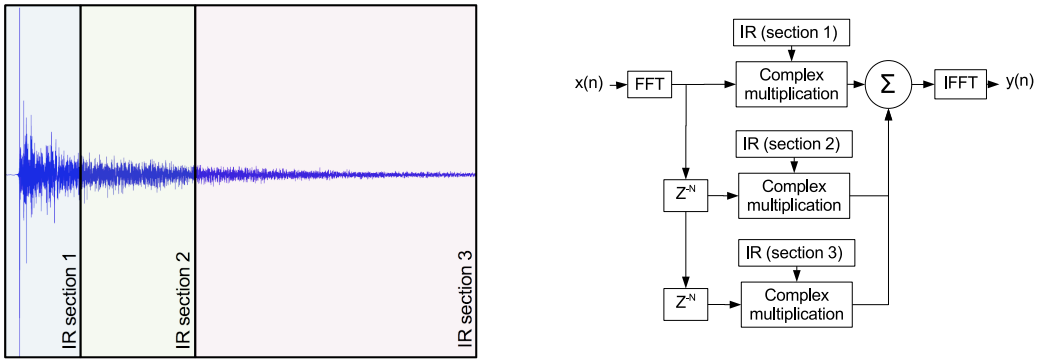
A more efficient method of real-time convolution utilises FFT methods which transform the impulse response and a buffered portion of the input signal into the frequency domain where they are multiplied together and then transformed back into the time domain using the IFFT. This operation is shown below in equation (6.3) which requires significantly less processing power at the expense of some latency due to buffering (Battenberg and Avizienis, 2011).

$$\{x(n) * h(n)\} = \mathcal{F}^{-1}(\mathcal{F}\{x(n)\} \cdot \mathcal{F}\{h(n)\}) \tag{6.3}$$

A method of reducing this latency without any additional computational cost is to partition the impulse response into non-overlapping time regions of length N. Each section of the impulse response is processed (in the frequency domain) simultaneously after an appropriate delay and summed together. The resultant response is then transformed back into the time domain. The result of this is that the latency has been reduced to the length of each section N.

Further optimisation is possible by rearranging the processing stages such that the input signal is transformed into the frequency domain only once, then performing the complex multiplication with each part of the impulse response, summing together and finally transforming the signal

back into the frequency domain. Such arrangements are known as Frequency-domain Delay Line-partitioned (FDL) convolvers (Battenberg and Avizienis, 2011). This approach can be further optimised by using non-uniform partitioning of the impulse response, separating the impulse response into very small regions at the beginning which get progressively longer towards the end . This allows the beginning of the impulse response to be processed at lower latency while the later parts are processed at an increased efficiency (Battenberg and Avizienis, 2011). An example of a non-uniform FDL convolver is shown in Figure 6.7 where an impulse response has been partitioned into three sections and are processed separately in the frequency domain.



(a) Partitioned impulse response

(b) FDL convolver structure

FIGURE 6.7: *Diagram showing how an impulse response can be partitioned and processed using a Frequency-domain Delay Line (FDL) Convolver structure.* Z^{-N} *denotes a delay of length N samples. In this case the impulse response is divided into three shorter sections for clarity.*

Convolution is now routinely implemented as a digital audio effect in many Digital Audio Workstations (DAW), many of which operate with very low latency. Some examples that were explored during this research are the SIR2 VST plug-in (Knufinke, 2010) and the Reaverb plug-in which is part of the Reaper DAW (Reaper, 2013). Previously, Watson and Clark (2010) demonstrated how an interactive auralisation could be implemented using Reaper and measured Ambisonic impulse responses. This provides a very accessible way of allowing musicians to practice in virtual environments. While the DAW implementations worked very well for experimentation it was found to be quite difficult to quickly switch between impulse responses which is important for the listening tests in this research. Therefore, Harker's HISS Tools external library (Harker and Tremblay, 2012) (which operate with Max MSP (Cycling74, 2013)) were used so that appropriate controls could be more easily programmed.

6.5 Spatial audio rendering

In Chapter 2 it was discussed how, in the context of audience auralisation, the spatial distribution of early reflections had a significant influence on the perception of sound in a venue. Therefore, in order to ensure the auralisation is as plausible as possible, the spatial attributes of the rendered soundfield must be preserved.

Most spatial audio systems require that the spatial attributes are encoded in a specific way, meaning the method used for measuring and rendering the soundfield are interrelated. Some methods, based around the principle of *'holophony'*, aim to reproduce a physical equivalent of the target soundfield using a dense array of loudspeakers. Other approaches attempt instead to create a perceptually equivalent soundfield. Both approaches have been used extensively in the context of audience auralisation and have been shown to produce highly plausible virtual environments.

The following section will describe a number of candidate spatial audio techniques that could be used for stage acoustic laboratory auralisations and will discuss which approach is appropriate for this research.

6.5.1 Holophony

Holophony is the main principle governing how the wavefront created by a virtual sound source can be reproduced using an array of loudspeakers. For concert hall auralisation, it is desired that a loudspeaker array reproduces the wavefronts of early reflections arriving from any given direction such that the listener experiences an equivalent soundfield as if they were inhabiting the real space (Zotter, 2010). As was introduced in Chapter 3, holophony is based around the Huygens principle which states that a wavefront can be regarded as a superposition of a number of secondary sources (Kuttruff, 2007). This is illustrated in Figure 6.8 which shows a wavefront created by a sound source. The original wavefront can be recreated using by using an array of secondary sources (loudspeakers).

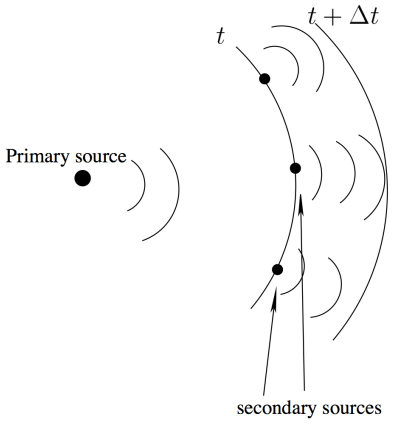


FIGURE 6.8: *Illustration of the Huygens principle where the wave front of a sound source is recreated using the superposition of a number of secondary wavefronts. In this case the secondary wavefronts are created using an array of loudspeakers. Image taken from Bourdillat (2001)*

This can be expressed mathematically by the Kirchhoff-Helmholtz integral (equation (6.4)) which states that if the sound pressure and velocity on the surface of an arbitrary, source-free, volume are known, then there is complete knowledge of the soundfield in the interior of that volume (Ortolani, 2014). Therefore, the entire soundfield inside the source free volume can be controlled by varying the pressure and velocity distribution of sound on the unit sphere (Zotter et al.,

2010). This can be achieved using a spherical arrangement of loudspeakers that are driven with the appropriate signals. Equation (6.4) shows the Kirchoff-Helmholtz integral which determines the driver signal, $P(r, \omega)$ for each loudspeaker on a radius r.

$$P(r,\omega) = \frac{1}{4\pi} \int \int_s \left[P(rs,\omega) \frac{\partial}{\partial n} \left(\frac{e^{-jk|r-r_s|}}{|r-r_s|} \right) - \frac{\partial P(r_s,\omega)}{\partial n} \frac{e^{-jk|r-r_s|}}{|r-r_s|} \right] dS \tag{6.4}$$

where $P(r, \omega)$ is the sound pressure in the Fourier Domain, k is the wavenumber ω/c, S is the surface of the volume, r is the coordinate vector of an observation point and r_s is the coordinate vector of the integrand function on S. The Kirchhoff-Helmholtz integral assumes that the boundary pressure and velocity can be described as a continuous function which implies the need for a spherically continuous driving function to control the soundfield. Current microphone and loudspeaker technology make both aspects impossible therefore a different approach is required.

Instead, a direct implementation of this, Wave Field Synthesis (WFS), uses a large number of loudspeaker channels to recreate specific wavefronts. This approach is considered highly effective at reproducing soundfields over a large area, however in order to be accurate over the full audio bandwidth, a very large number of closely spaced loudspeaker transducers are required in order to avoid spatial aliasing artefacts (Ortolani, 2014). Spatial aliasing occurs when the soundfield is either spatially under sampled by the microphone array or is recreated with an insufficient number of loudspeakers. The consequence of this is an upper frequency limit for spatial accuracy. The upper frequency limit is known as the spatial aliasing frequency (Daniel et al., 2003) calculated by (6.5) where c is the speed of sound and Δ represents the physical transducer spacing. When spatial aliasing occurs localisation and timbral errors can occur.

$$f_{sp} = \frac{c}{2\Delta_{transducer}} \tag{6.5}$$

As is shown by equation (6.5), a very large number of loudspeakers would be required to accurately reconstruct the soundfield at high frequencies. The number of loudspeakers often precludes the use of WFS for practical reasons such as cost or arrangement of each transducer.

6.5.2 Ambisonics

In Chapter 3, Ambisonics was introduced primarily as a method of capturing the spatial impulse response of a space via spherical harmonic decomposition. In reality, Ambisonics is a complete method of capturing, synthesising, storing and rendering 3D soundfields. Ambisonics (including Higher Order Ambisonics) is now commonly used in many auralisation systems. It is a highly flexible production format which can accurately render 3D soundfields over loudspeaker arrays or over headphones. This section will expand upon the description given in Chapter 3 to describe how an Ambisonic recording can be rendered for the purposes of auralisation.

Ambisonics uses spherical harmonics to approximate the spherical distribution of pressure and velocity. Decomposing a 3D soundfield into a series of spherical harmonics is analogous to how a complex one-dimensional signal can be decomposed into a weighted sum of sinusoidal

components. Ambisonics (Gerzon, 1972) achieves this by measuring the amplitude and phase of a set number of spherical harmonics. An array of loudspeakers is then used to recreate the spherical distribution of sound, encoded as spherical harmonics, using a number of discrete drivers arranged on a sphere. The main advantage of this approach is that far fewer channels of data are required to store or render the spatial distribution of the soundfield resulting in a much lower number of loudspeakers required.

There are numerous advantages to authoring auralisations using Ambisonics. One of the most pertinent, in relation to auralisation, is that Ambisonic soundfields (when rendered correctly) can be consider isotropic i.e. a sound is spatialised equally from any direction even if positioned away from loudspeakers (Benjamin et al., 2006). Therefore, in principle, it is possible to accurately reproduce the angle of arrival of reflections over a limited number of loudspeakers without significant timbral distortion when reflections arrive from directions other than the loudspeaker. Furthermore, as the spatial attributes of the soundfield are encoded using spherical harmonics, a limited number of convolvers (one per B-format channel) are required to impart the spatial impulse response onto anechoically recorded audio signals. This ensures a highly efficient rendering system that can feasibly operate at real-time.

Once in the spherical harmonic domain, numerous transformations of the soundfield are made possible by manipulating the gains of each B-format channel. This includes for example, rotation of the soundfield in any axis, mirroring the soundfield in any plane and zooming onto specific regions of the soundfield. Furthermore, the Ambisonic encoding and decoding process have been designed to be separate which allows an encoded recording to be rendered accurately over any loudspeaker array geometry (assuming an associated decoder can be created). This provides a great deal of potential for stage acoustic laboratory experiments where the same soundfield can be easily adapted to suit a particular experiment (Wiggins, 2004).

Rendering B-format audio over a loudspeaker system is achieved by driving each loudspeaker with a signal which is the weighted sum of each B-format channel. The weights of each channel are computed based on the location of each loudspeaker. As B-format encodes the soundfield into a set of spherical harmonics, the driving signals of each loudspeaker attempt to collectively reproduce the spherical harmonics. The weights form a gain matrix for each B-format channel to each loudspeaker and is generally referred to as an *Ambisonic decoder*. The correct design and implementation of an Ambisonic decoder is not trivial and there exist various constraints and trade-offs inherent in the array layout (and hence the decoder design) which will be discussed below.

Decoding through projection

When attempting to reproduce the spatial characteristics of a soundfield it is often required that a loudspeaker is used to playback sound arriving from a particular direction. Therefore, a suitable microphone technique is designed to capture sound in the direction of each loudspeaker in a particular system. When the signals from these microphones are played back over the corresponding loudspeaker array, they should each reproduce the spatial characteristics of the soundfield. This approach has been used for stage acoustic auralisation previously by Ueno and Tachibana (2003) to capture and reproduce the acoustic conditions on stage using a 6-channel loudspeaker array.

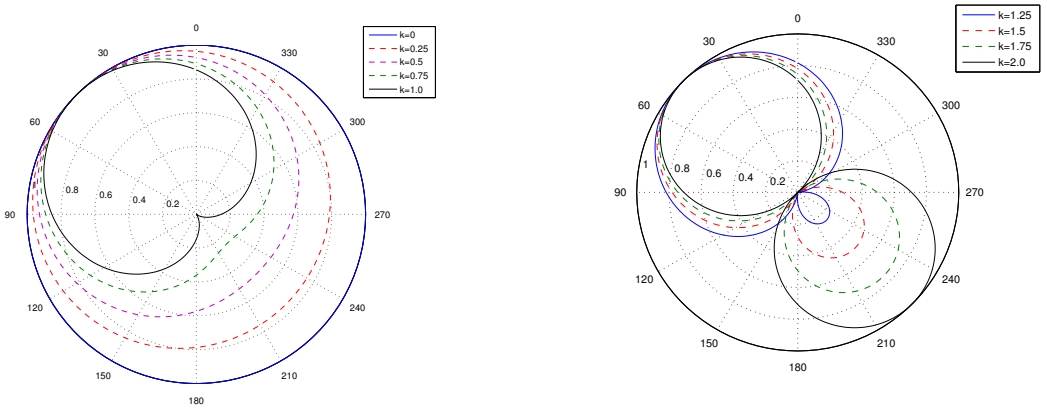
An equivalent arrangement could also be achieved using an Ambisonic microphone array. By carefully controlling the gains of each B-format signal a set of virtual microphones of specific directivity could be created and electronically steered in the direction of loudspeakers for any array geometry. The signals obtained from these virtual microphones are then played over the loudspeakers to recreate the soundfield. The set of coefficients used to obtain the appropriate array of virtual microphones is the same as the Ambisonic decoder matrix described previously. To ensure an accurate reproduction of the soundfield, the Ambisonic decoder matrix, governing the orientation and directivity pattern of the virtual microphones (and hence loudspeaker geometry), must be considered carefully.

Virtual microphones can be created by a weighted sum of signals from microphones with different polar patterns. An early example of this was developed by Alan Blumlein who, via the use of a coincident and orthogonal pair of figure-of-eight microphones, could derive an electronically steerable figure-of-eight microphone. It was also found that different virtual microphone polar patterns could be derived by a weighted summation of coincident omnidirectional and figure-of-eight polar patterns. By using three coincident figure-of-eight microphones and an omnidirectional microphone, a virtual microphone can be steered to any direction and made to have a different polar pattern (Wiggins, 2004).

Figure 6.9 show the virtual microphone polar patterns resulting from the weighted summation of spherical harmonics. These microphone polar patterns are derived using equation (6.6) where $w(t)$ refers to the omnidirectional signal from the Ambisonic microphone; $x(t)$, $y(t)$ and $z(t)$ refer to the dipole signals from the Ambisonic microphone, and k refers to the directivity factor which varies between 0 and 2.

$$S_n(t) = \frac{2-k}{2} w(t) + \frac{k}{2\sqrt{(2)}} [\cos(\theta_n)\cos(\phi_n)x(t) + \sin(\theta_n)\cos(\phi_n)y(t) + \sin(\phi_n)z(t)] \qquad (6.6)$$

The directivity factor, k, controls the ratio of the W-channel to the X, Y and Z channels. In Figure 6.9, it can be seen that a directivity factor of 0 results in an omnidirectional polar pattern, a directivity factor of 1 results in a cardioid polar pattern and a directivity factor of 2 results in a dipole (figure-of-eight) polar pattern.



(a) (b)

FIGURE 6.9: *Virtual microphone polar patterns where directivity factor k varies between 0 and 2 and the microphone is oriented towards 45°. For clarity 6.9(a) shows virtual mic patterns between $k = 0$ and $k = 1$, 6.9(a) shows $k = 1.25$ and $k = 2$*

The coefficients of a *'basic'* Ambisonic decoder are calculated by sampling the value of each spherical harmonic at each loudspeaker position on the unit sphere. This is known as decoding through projection and assumes that the loudspeaker array is regular in the Ambisonic sense, i.e. loudspeakers are arranged on the surfaces or vertices of the regular polyhedra (i.e. tetrahedron, cube, dodecahedron etc). If the array is not regular the decoder will continue to project the Ambisonic signals as if that was still the case (Kearney, 2009). This causes equal energy to be presented to each loudspeaker resulting in a non-uniform soundfield. This can affect the localisation accuracy of the auralisation.

Equation (6.6) calculates the driving signals required for each loudspeaker $S_n(t)$ in a loudspeaker array of n loudspeakers and can be viewed as a weighted sum of the B-format signals W, X, Y, Z. These weights are dependent on the locations of the loudspeakers in the array. A further weighting is applied via the coefficient k which controls what type of Ambisonic decoder is produced depending on the requirements of the system. The different types of Ambisonic decoder will be discussed later in this chapter.

Decoding through Pseudoinverse

An alternative method of deriving a basic Ambisonic decoder matrix is by using least-squares optimisation (or Pseudoinverse decoding). This method uses matrix inversion to derive a set of loudspeaker gains which, when multiplied with the encoding matrix, yield the correct source direction (Kearney, 2009). For example, consider the Ambisonic signals $[WXYZ]$ expressed as a matrix $\boldsymbol{B}$, where T denotes matrix transposition, s is equal to the source signal while $\boldsymbol{b}$ is equal to the B-format direction.

$$\begin{aligned} \boldsymbol{B} &= [WXYZ]^T \\ &= s\,[b_W b_X b_Y b_Z]^T \\ &= s\boldsymbol{b}^T \end{aligned} \tag{6.7}$$

If $\boldsymbol{p}$ is the column vector of loudspeaker gains and $\boldsymbol{C}$ is the value of each spherical harmonic sampled at each loudspeaker position (referred to as the re-encoding matrix) then the correct sound source direction, $\boldsymbol{b}$, is reproduced by multiplying the re-encoding matrix by p.

$$\boldsymbol{b} = \boldsymbol{C} \cdot \boldsymbol{p} \tag{6.8}$$

It is of interest to determine the gain values needed to reproduce the correct B-format direction which can be obtained by matrix inversion (6.9).

$$\boldsymbol{p} = \boldsymbol{C}^{-1} \cdot \boldsymbol{b} \tag{6.9}$$

It is often the case that the loudspeaker array features more loudspeakers than B-format channels, resulting in a non-square decoder matrix. As there are more unknowns than knowns the matrix is overdetermined and so cannot be inverted. Therefore, the Moore-Penrose pseudoinverse ($pinv$) (From MathWorld–A Wolfram Web Resource, 2012) is used as shown below in equation (6.10):

The gains can be solved using linear algebra and solving for $\boldsymbol{p}$ thus:

$$\begin{aligned} pinv(C) &= C^T(CC^T)^{-1} \\ p &= pinv(C) \cdot \boldsymbol{b} \\ &= C^T(CC^T)^{-1} \cdot \boldsymbol{b} \end{aligned} \tag{6.10}$$

If the loudspeaker array has a regular geometry, then decoders computed by pseudoinverse and projection will be equivalent (assuming that both are set to be 'basic'). In the case of the pseudoinverse decoder, if the array is non-regular the decoder matrix can contain numerical instabilities, resulting in poorly reconstructed spherical harmonics.

Localisation vectors

Gerzon (1992) formulated simple models to assess the quality of localisation exhibited by an audio system. These are the velocity and energy localisation vectors which predict the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) methods of auditory localisation. The direction of these vectors indicate the localisation perception whereas their magnitude indicates the quality of localisation. Gerzon recognised that the human auditory system localises low frequency sound predominantly using ITD and high frequency sound using ILD therefore a spatial sound system should aim to produce optimal cues at these frequencies to allow natural localisation. Gerzon developed a set of metrics to assist in quantifying the degree of localisation accuracy, namely the velocity and energy vectors.

The velocity vector (also referred to as the Makita localisation vector (Gerzon, 1992)) and the energy vector determine the direction a listener's head must be pointing towards to ensure the ITD and ILD are both zero. These situations generally occur only when a listener is directly facing a natural sound source. The vectors will have a magnitude of 1 when the phase or level differences are zero, indicating perfect localisation. Lower magnitudes of each vector indicate more ambiguous localisation (Wiggins, 2004).

The 2D velocity vector components $[v_x, v_y]$ can be calculated from the gain of each loudspeaker in an array where g_n and θ_i is the gain of the n^{th} loudspeaker and the angular location respectively.

$$\begin{aligned} v_x &= \frac{\sum_{n=1}^{N} g_n \cos(\theta_n)}{\sum_{n=1}^{N} g_n} \\ v_y &= \frac{\sum_{n=1}^{N} g_n \sin(\theta_n)}{\sum_{n=1}^{N} g_n} \end{aligned} \tag{6.11}$$

The 2D energy vector components $[e_x, e_y]$ can be calculated in a similar way using the square of the loudspeaker gains, g_n:

$$\begin{aligned} e_x &= \frac{\sum_{n=1}^{N} g_n^2 \cos(\theta_n)}{\sum_{n=1}^{N} g_n^2} \\ e_y &= \frac{\sum_{n=1}^{N} g_n^2 \sin(\theta_n)}{\sum_{n=1}^{N} g_n^2} \end{aligned} \tag{6.12}$$

Calculating the magnitude of these vectors will predict the localisation accuracy of a particular spatial audio technique over a given loudspeaker array. By observing the magnitude at different angles around the loudspeaker array it is possible to determine how the localisation accuracy varies if a source is panned to different locations around the array. Furthermore, by observing the angle of each vector it is possible to determine if the expected angle of arrival matches the intended source angle of arrival.

Ambisonic reproduction systems are defined by a set of criteria based on these localisation vectors. The following criteria, summarised by Heller et al. (2012), are deemed necessary for a reproduction system to be considered Ambisonic:

1. Constant amplitude gain for all source directions.
2. Constant energy gain for all source directions.
3. At low frequencies, correct wavefront direction and velocity
4. At high frequencies, maximum concentration of energy in the source direction.
5. Matching of high and low frequency perceived directions.

In practice this requires the magnitude of both velocity and energy vectors should be maximised as far as possible and be constant regardless of source angle. It also requires the vectors to be pointing in the same direction for any given source direction. These conditions ensure the system is truly isotropic.

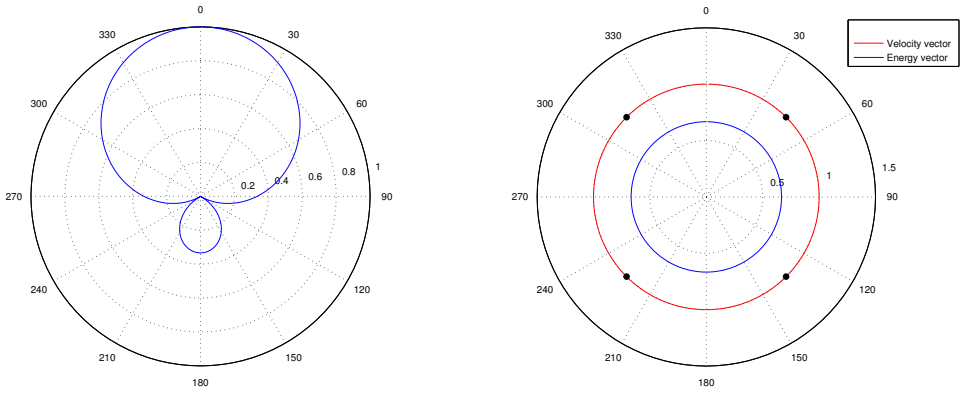
The 'basic' decoder described so far only partially fulfils these criteria. Consequently, a number of variations upon this decoder have been developed to ensure the remaining criteria are satisfied. These designs are described in the following sections.

Velocity decoder

A velocity decoder (termed also in some cases *'basic'*, *'exact'*, *'mode matching'*) applies gains to each B-format signal to ensure that the velocity localisation vectors are optimal at all directions. This type of decoder is obtained directly using the pseudoinverse method or using equation (6.6)

with the directivity factor $k = 1.33$ (Wiggins, 2004). Figure 6.10 shows the polar patterns and localisation vector plot for this type of decoder. In Figure 6.10(a) the polar pattern features a prominent rear lobe which is oriented in the opposing direction to the main lobe. Using this type of decoder causes an anti-phase signal from the opposing loudspeaker to correctly spatialise the sound. In simple terms, one loudspeaker 'pulls' while the opposing loudspeaker 'pushes'.

Figure 6.10(b) shows the magnitude of the localisation vector for signals decoded to a square shaped loudspeaker array, where loudspeakers are represented by black dots. It can be seen that the velocity vector and energy vector have good agreement in terms of angle and the magnitude of both vectors remains constant for all source angles (indicating isotropy). The velocity vector (shown in red) is a value of 1 for all source directions indicating that localisation is excellent for localisation in terms of ITD. However, the energy vector (shown in blue) is a value of 0.67 for all source directions indicating poor localisation in terms of ILD. This implies that the velocity decoder works very well for low frequency localisation but not high frequencies.

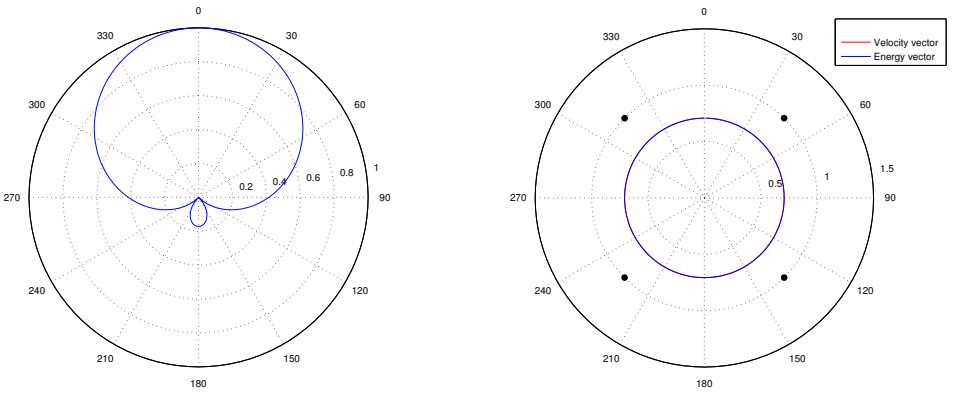


(a) (b)

FIGURE 6.10: *Figure 6.10(a) shows the polar pattern of a virtual microphone associated with a velocity decoder pointing to an angle of* 0°. *Figure 6.10(b) shows the associated localisation vector plot for a square loudspeaker array. Loudspeakers are shown as black dots. In this case, the velocity vector* = 1 *for all source angles and the energy vector is 0.67 for all source angles.*

Energy decoder

The energy decoder (termed also *'Max-Re'*) attempts to maximise the amplitude of the energy vectors and can be derived by modifying the relative gains of the omnidirectional and directional B-format signals. In equation (6.6) the required directivity factor $k = 1.15$ (Wiggins, 2004). This has the effect of altering the polar pattern of the virtual microphone, shown in Figure 6.11(a), such that it has a reduced rear-lobe but a wider frontal lobe. Figure 6.11(b) shows the energy and velocity vectors have the same value (0.707) for all source angles decoded over a square shaped loudspeaker array. As there is always more than one loudspeaker generating sound to produce a virtual sound source at any given angle, it is not possible for the energy vector to have a higher value than 0.707. This implies that the energy decoder has an improved localisation in terms of ILD in comparison to the velocity decoder but poorer localisation in terms of ITD.



(a) (b)

FIGURE 6.11: *Figure 6.11(a) shows the polar pattern of a virtual microphone associated with a velocity decoder pointing to an angle of* 0°. *Figure 6.11(b) shows the associated localisation vector plot for a square loudspeaker array. Loudspeakers are shown as black dots. In this case, both the velocity and energy vectors have a magnitude of 0.707 for all source angles.*

Dual-band decoder

A dual-band decoder is a combination of the velocity and energy decoders where each decoder operates over its optimal frequency range. This decoder will ensure a velocity vector magnitude of 1 for low frequencies and an energy vector magnitude of 0.707 for high frequencies (assuming the array is regular).

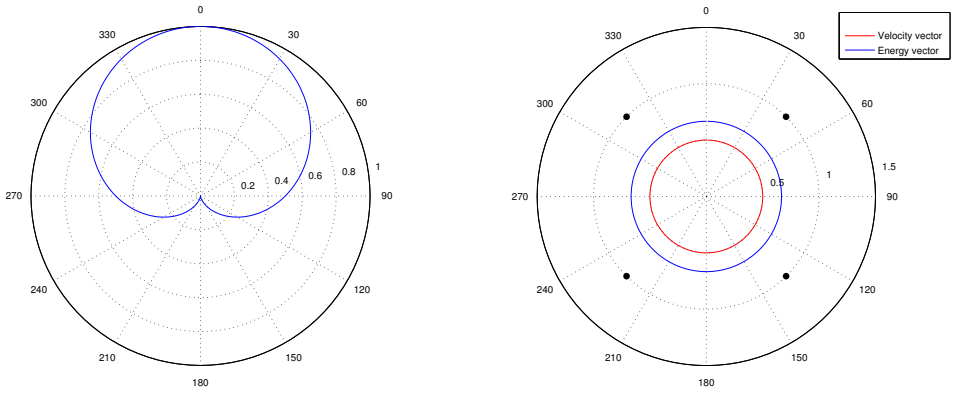
This is normally achieved by filtering the B-format with a crossover matrix or shelf-filter arrangement to ensure that each part of the decoder has the correct gains for each channel. The former is used if the two decoders are implemented separately and combined, whereas the latter is also used to adjust the gains of the B-format signals so that only one decoder is required. A typical crossover frequency is often quoted as $700Hz$ which is commonly quoted as the transition between ITD and ILD localisation (Wiggins, 2004). In reference to the aforementioned criteria for an Ambisonic decoder, this arrangement is the only one that fulfils all criteria.

Cardioid decoder

For completeness, the cardioid decoder (or *'controlled opposites'*) is an alternative decoding scheme which has been specifically designed for when rendering Ambisonic sound over large loudspeaker arrays for large audience performances. In equation (6.6) the required directivity factor for this decoder is $k = 1$. This results in a cardioid shaped virtual microphone response pointing to each loudspeaker, shown in Figure 6.12(a). This decoder is designed to reduce the possibility of an opposing speaker being active when a sound is panned to a particular location.

This ensures that listeners positioned well outside the sweetspot will not receive a contradicting, anti-phase contribution if they are closer to the opposing loudspeaker for a given sound source location. In order to achieve this, the width of the virtual microphone frontal lobe has been increased reducing the overall accuracy of the reproduction. Figure 6.12(b) shows the localisation vectors for this type of decoder. It can be seen that this decoder performs poorly for low

frequency localisation (velocity vector = 0.5) and performs as good as the energy decoder at high frequencies (energy vector = 0.67).



(a) (b)

FIGURE 6.12: *Figure 6.12(a) shows the polar pattern of a virtual microphone associated with a velocity decoder pointing to an angle of* 0°. *Figure 6.12(b) shows the associated localisation vector plot for a square loudspeaker array. Loudspeakers are shown as black dots. In this case, the velocity vector* = 0.5 *for all source angles and the energy vector is 0.67 for all source angles*

Array geometry

When B-format is played back over a regular array, it is very simple to compute a successful Ambisonic decoder where the reproduced angle of arrival would always be the same as the encoded angle for a particular sound source. In addition the energy and pressure vectors would remain unchanged with source angle. However, in practice, it is often the case that a regular loudspeaker array would require loudspeakers to be positioned in the floor, or alternatively requiring the listener to be elevated off the floor (which may become uncomfortable over long listening periods). Therefore, it is more often the case that an irregular loudspeaker array is used. Figure 6.13 shows the effect of decoding B-format over an irregular array such as an ITU 5.1 array using a Max-Re decode. This loudspeaker array shape was developed for movie surround-sound and features a stereo arrangement of loudspeakers (±30°) with a central loudspeaker in between. Two rear loudspeakers are located at ±110°. It can be seen that, due to the relative sparseness of the loudspeaker array towards the rear, the high frequency localisation suffers when a source is located at $\theta = 180°$. It can also be seen that the velocity and energy vectors no longer point in the same direction and the magnitude of both vectors varies with source angle. This demonstrates the system is no longer isotropic.

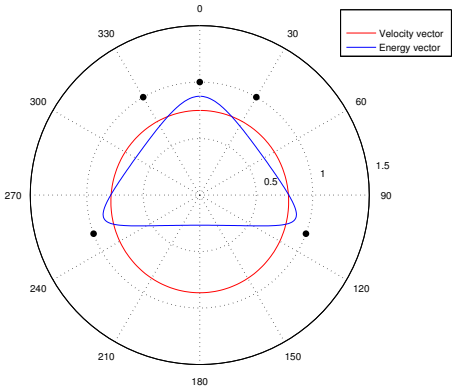


FIGURE 6.13: *Localisation vector plot for a Max-Re Ambisonic decoder computed for an ITU 5.1 loudspeaker layout. Loudspeakers are shown as black dots. In this decoder, the velocity vector (red) and Energy vector (blue) point in different directions and vary in magnitude significantly with source angle.*

As irregular shaped arrays are often easier to implement and already form popular loudspeaker arrangements in cinemas and home layouts, a number of optimisation schemes (Heller et al., 2010, Wiggins, 2004) have been developed to compensate for the poor localisation quality. These optimisation techniques use various genetic search algorithms to iteratively adjust the gains of each B-format channel such that the localisation vectors are improved in relation to various given criteria. In order to achieve an optimal decoder matrix the criteria may need to sacrifice one aspect of the decoder design in favour of another (Heller et al., 2010). For example, the magnitude of the localisation vectors may be sacrificed in order to ensure that the energy and velocity vectors point in the same direction.

6.5.3 Higher Order Ambisonics

The discussion of Ambisonics so far has focussed on the use of only the first four spherical harmonics. Higher Order Ambisonics (HOA) is an extension of this approach which encodes the soundfield using many more spherical harmonics allowing the soundfield to be captured and rendered at a higher spatial resolution. By encoding the soundfield using higher order spherical harmonics, the resultant auralisation is spatially accurate to a higher frequency. Furthermore, HOA systems tend to produce a larger accurate sweetspot than first-order Ambisonics. As the Ambisonic order increases, the spherical harmonic approximation of the soundfield becomes more and more accurate until it approaches the performance of WFS (Kearney, 2009).

The use of additional spherical harmonics requires more channels to store the encoded soundfield but also requires additional loudspeakers to ensure the higher order spherical harmonics are supported. For periphonic (i.e. full sphere with height) reproduction the minimum number of channels, N, used for encoding and also the minimum number of loudspeakers is determined for a given order, m, using equation (6.13) below (Hollerweger, 2008):

$$N = (m+1)^2 \tag{6.13}$$

Order	Channel Name	SN3D definition	FuMa weight
0	W	1	$\frac{1}{\sqrt{2}}$
1	X	$\cos(\theta)\sin(\phi)$	1
1	Y	$\sin(\theta)\cos(\phi)$	1
1	Z	$\sin(\phi)$	1
2	R	$(3\sin^2(\phi)-1)/2$	1
2	S	$(\frac{\sqrt{3}}{2})\cos(\theta)\sin(2\phi)$	$\frac{2}{\sqrt{3}}$
2	T	$(\frac{\sqrt{3}}{2})\sin(\theta)\sin(2\phi)$	$\frac{2}{\sqrt{3}}$
2	U	$(\frac{\sqrt{3}}{2})\cos(2\theta)\cos^2(\phi)$	$\frac{2}{\sqrt{3}}$
2	V	$(\frac{\sqrt{3}}{2})\sin(2\theta)\cos^2(\phi)$	$\frac{2}{\sqrt{3}}$

TABLE 6.1: *Definitions of spherical harmonic functions up to second order with accompanying FuMa weights. Replicated from Malham (2003)*

To illustrate this further, the spherical harmonic functions up to 2nd order have been defined in Table 6.1. From equation (6.13) it can be seen there are 9 channels required for 2nd order Ambisonics. These spherical harmonics are defined in their Semi-Normalized (SN3D) form alongside their accompanying Furse-Malham (FuMa) weight (Malham, 2003).

Like first-order Ambisonics, HOA can be used to spatialise mono sound sources using a set of panning functions. This is commonly utilised in acoustic modelling software which is capable of synthesising a HOA impulse response of a modelled space. However, capturing a measured impulse response is much more difficult. This is partially due to the shape of the spherical harmonics as there are currently no microphones with such directivity patterns. Instead, an array of microphones is used to capture HOA material. The signals from each capsule are weighted such that they produce the necessary spherical harmonic pattern. This can often require a large number of microphone transducers, and, unless extremely high-quality capsules are used, this can result in a reduced signal-to-noise ratio.

An HOA decoder is derived in a similar manner to first-order decoders with the exception that the decoder matrix contains many more weights in order to correctly render the higher order spherical harmonics. Similarly, the different 'flavours' of Ambisonic decoders (i.e. velocity, Max-Re, Cardioid) can be derived in the same manner as first-order Ambisonic decoders.

Figure 6.14 shows how the virtual microphone polar patterns (for the three main types of decoder) vary with Ambisonic order. In all cases, the main lobe of the virtual microphone becomes much more localised and the influence of rearward lobes (in particular with the velocity decoder) reduces.

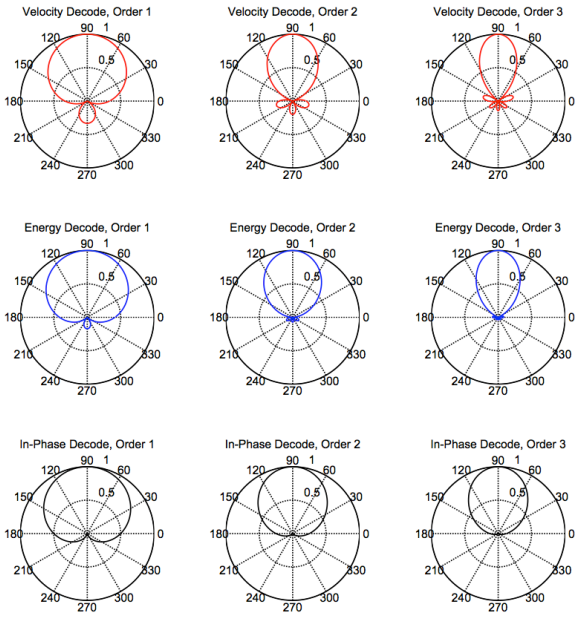


FIGURE 6.14: *Polar pattern of virtual microphone associated with a velocity, Max-RE and Cardioid decoder for 1st, 2nd and 3rd order Ambisonics. Image reproduced from Kearney (2009)*

In summary, Ambisonics is a highly flexible system for capturing, processing and reproducing 3D soundfields over a loudspeaker array. However, due to the high coherence of the signals from each loudspeaker, comb-filtering effects are often audible when listening at off-centre positions or when head movements are made around the sweetspot which can significantly degrade the fidelity of auralisations (Pulkki, 2007). Higher Order Ambisonic reproduction has been demonstrated to be of much higher spatial fidelity (Bertet et al., 2007) but requires many more loudspeakers and B-format channels to operate correctly and (if measured soundfields are to be recreated) requires the use of a highly specialised microphone array that is not yet commercially available.

6.5.4 Amplitude panning

Another approach for spatialising sounds using a loudspeaker array is amplitude panning. One of the most common, Vector Base Amplitude Panning (VBAP), was demonstrated by Pulkki (1997) who showed that sounds could be panned around a periphonic loudspeaker array using vector calculation to determine the gain of a triplet of local loudspeakers.

Consider three loudspeakers positioned on a 3D unit sphere all of which are equidistant from the listener who is positioned in the middle of the sphere. This is shown in Figure 6.15 below. A 3D unit vector, $l_1 = [l_{11}, l_{12}, l_{13}]^T$ points from the origin (listener position) to the position of each loudspeaker. Similarly, a 3D unit vector points to the desired position of the sound object to be spatialised $p = [p_{11}, p_{12}, p_{13}]^T$.

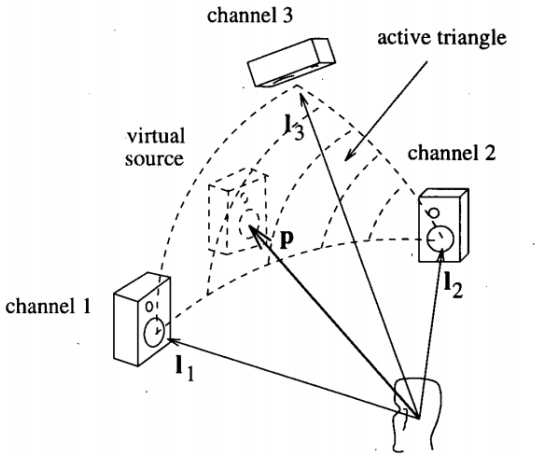


FIGURE 6.15: *Triangular arrangement of loudspeakers used to create a virtual sound source p in three dimensional space by calculating gains from vectors. Image reproduced from Pulkki (1997)*

g is a vector describing the gains of each of the loudspeakers and can be obtained via equation (6.14)

$$g = p^T L_{123}^{-1} = [p_1, p_2, p_3] \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \tag{6.14}$$

The gain factors are then normalised to give equation (6.15) where the factor $\sqrt{C}$ is a constant, representing the overall amplitude of the sound object being spatialised.

$$g^{scaled} = \frac{\sqrt{C}g}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \tag{6.15}$$

Like Ambisonic sound systems, the localisation accuracy of an amplitude panning system can be characterised by analysing the velocity and energy vectors. Figure 6.16 shows localisation plots for different 2D loudspeaker array layouts. In all plots it can be seen that the energy and velocity vectors are equal at all source angles. In all cases, the energy and velocity vectors reach a value of 1 when the sound source position coincides with the loudspeaker location. This increase in localisation quality occurs as only a single speaker will be active in this case, meaning that localisation will be as sharp as a natural sound source. However, unlike Ambisonics, the localisation quality is not constant with angle and reduces as the sound is panned between two loudspeakers.

For a quad layout, as shown in Figure 6.16(a), the localisation quality reduces to a minimum of 0.707 when the source is panned between two loudspeakers. As loudspeakers are added, as in 6.16(b) and 6.16(c), the localisation quality between loudspeakers improves monotonically (the minimum values are 0.866 and 0.925 respectively). This variation in localisation quality

between loudspeakers can cause the location of loudspeakers to be audible. It is also the case that the angle of the localisation vectors is only equal to the source angle when panned to the loudspeaker positions. When a sound source is panned around an array, the perceived location of a sound source tends to pull towards the nearest the loudspeaker. As more loudspeakers are added however, the variation in localisation quality reduces, meaning the localisation quality increases towards the localisation quality characteristics of Ambisonics.

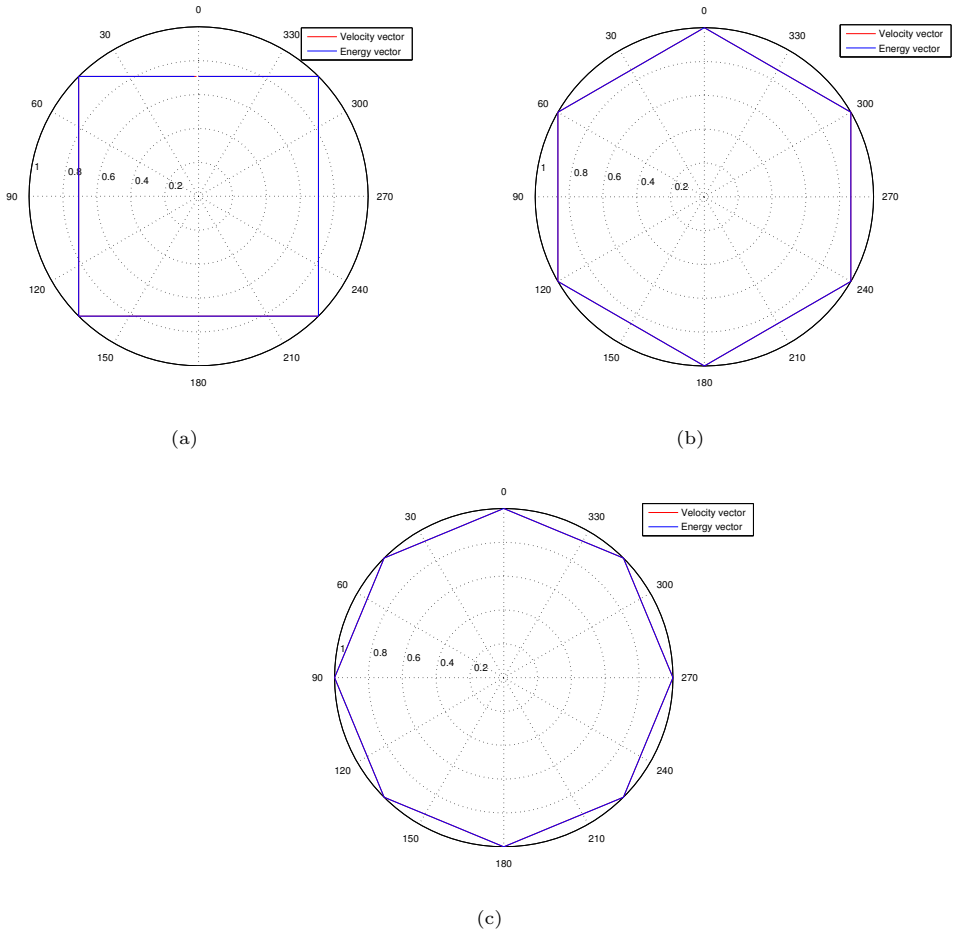


FIGURE 6.16: *Localisation vector plot for horizontal amplitude panning systems implemented on 6.16(a) a quad loudspeaker layout, 6.16(b) a hexagonal loudspeaker layout and 6.16(c) an octagonal loudspeaker layout. In this plot, both Energy and Velocity vectors are the same magnitude for all angles. It can be seen that the localisation accuracy varies with source angle however by adding more loudspeakers, the localisation accuracy increases monotonically.*

The variation in localisation vector magnitude can also be controlled by spreading the sound source over a range of similar directions. This, however, will reduce the maximum localisation quality (as the source is more spread out). This can be achieved using Multiple-Direction Amplitude Panning (MDAP) which calculates gain factors by summing the loudspeaker gains of numerous panning directions. By spreading these directions out, the sound source becomes more spread out.

The primary advantage of this technique is that for any given sound source there are only ever a maximum of three loudspeakers active at any one time. This has the effect of sounds being easily

localised even with the listener positioned out from the sweetspot. This becomes very useful in the context of interactive auralisation experiments with musician test subjects who may move in and out of the loudspeaker array sweetspot.

Amplitude panning techniques cannot be considered isotropic and so sounds spatialised to angles between loudspeakers can exhibit timbral colouration in contrast to when the sounds are spatialised from the angle of a loudspeaker. This means that sounds produced using VBAP can no longer be considered virtual (Kearney, 2009). The use of a sufficiently dense loudspeaker array can reduce this effect. Furthermore, unlike Ambisonics, amplitude panning techniques are exclusively used for panning sounds around a loudspeaker array and do not have a native microphone technique. Therefore, in order to auralise a room impulse response using this technique, knowledge of the time and direction of arrival of each reflection is required in order to synthesise an appropriate impulse response for each loudspeaker. An interactive auralisation system based on amplitude panning will require a convolution engine for each loudspeaker channel, in contrast to a more efficient Ambisonics based system which requires a fixed number regardless of the number of loudspeakers available.

6.5.5 Parametric decoding schemes

A more recent approach to soundfield reproduction takes advantage of both Ambisonic recording techniques and amplitude panning to produce a perceptually enhanced rendering of a 3D soundfield over a loudspeaker array or over headphones. Parametric decoding schemes refer to the extraction of directional parameters from a spatial audio recording which are then used to spatialise audio using any chosen 3D panning technique. It is stressed that it is not possible to physically increase the spatial resolution of an audio recording however this approach attempts to create the perceptual impression of a soundfield recorded at a higher spatial resolution.

As shown in Figure 6.17, parametric decoding schemes typically consist of an analysis phase, where spatial parameters are extracted from a spatial audio recording, and a synthesis phase, where signals are generated using the parametric data that aim to elicit a specific psychoacoustic response when rendered over a loudspeaker array or headphones. The separation of these processes provides an opportunity to alter specific aspects of the soundfield before the loudspeaker signals are generated. This can allow complex modification of spatial audio recordings to applied, such as spatial filtering or re-spatialisation.

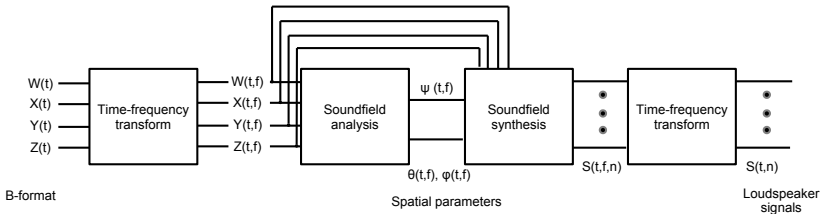


FIGURE 6.17: *Block diagram showing how B-format is decoded to loudspeaker signals via extraction of spatial parameters from the B-format signals. Speaker signals, $S(n)$, are synthesised by combining the spatial audio recording with parametric data. In this example, the B-format signals are used in the synthesis phase, however this is not always the case. DirAC and SIRR tend to extract diffuseness (ψ) and direction of arrival parameters (θ,ϕ). t, f and n denote time, frequency and speaker number respectively.*

Parametric decoding schemes allow a great deal of flexibility regarding how the soundfield is rendered, for example a first-order Ambisonic recording could be rendered using amplitude panning or Higher order Ambisonics. This can be very useful if, for example, it is not possible to use a regular geometry for a loudspeaker array or if a HOA microphone is unavailable.

These aspects make parametric decoding schemes an attractive option for auralisation experiments. There are currently a number of different implementations of this type of decoding scheme which are described in this section.

HARPEX

High Angular Resolution Planewave EXpansion (HARPEX) (Berge and Barrett, 2010a) is an example of a parametric decoding scheme which analyses Ambisonic recordings in the time-frequency domain and reproduces loudspeaker (or headphone) feeds based on this analysis. HARPEX can be used for any type of recorded soundfield and uses plane-wave decomposition to estimate the direction of sound sources in a recording. It is then possible to use this parametric data to create a physically correct reconstruction of the soundfield using the available loudspeakers. The synthesis of the plane waves can be achieved using any spatial audio panning technique such as HOA panning functions or binaural panning. This can produce a perceived increase spatial resolution of the recording (Berge and Barrett, 2010b). HARPEX processing is often applied to recorded soundfields but could feasibly be applied to room impulse responses used in auralisation.

HARPEX typically uses first-order Ambisonic B-format as input. B-format signals are transformed into the time-frequency domain in order to determine the direction of arrival over time. This is achieved by evaluating the phase and magnitude of the B-format signals. The original signal is then spatialised using a chosen spatialisation technique. In a typical spatial audio recording, it is possible that the directional estimates changes very quickly over time which can cause time domain artefacts. HARPEX applies temporal smoothing to the directional data in order to reduce this.

Listening tests show favourable results for HARPEX-rendered audio material in terms of perceived direction and spatial sharpness of sounds (Berge and Barrett, 2010b). Similar patterns

were observed for a 3rd order Ambisonic recording and first-order Ambisonic recording that had been decoded to the same loudspeaker array using pairwise panning (both over an octagon layout loudspeaker array).

Spatial Impulse Response Rendering

Spatial Impulse Response Rendering (SIRR) is an example of a parametric decoding scheme optimised for rendering room impulse responses over a loudspeaker array (Merimaa and Pulkki, 2004) as a convolving reverberator. The premise behind SIRR is that at any particular time instant and for each individual perceptual frequency band, the human hearing system is capable of processing a single directional cue and another cue for inter aural coherence. SIRR attempts to recreate these perceptual cues by synthesising loudspeaker signals based on an analysis of the soundfield. The loudspeaker signals therefore do not attempt to recreate a direct physical reproduction of the soundfield (like Ambisonics does) rather they aim to produce a perceptually equivalent soundfield.

The analysis technique therefore aims to extract direction of arrival and diffuseness parameters of each reflection from the spatial room impulse response. This is made possible by estimating the intensity of the room impulse response in the time-frequency domain. This technique was described in detail in Chapter 4 when it was used to determine the spatial distribution of early reflections. In this case, the parametric data (direction of arrival and diffuseness) is used to synthesise the appropriate loudspeaker signals.

SIRR generates an impulse response for each loudspeaker that will be convolved with an anechoically recorded signal for rendering over a loudspeaker array. The impulse responses are generated by spatialising the original impulse response (W-channel only) according to the parametric data obtained in the analysis. This process is shown in Figure 6.18. The most common implementations of SIRR use amplitude panning techniques, such as VBAP, to spatialise non-diffuse reflections. Spatially diffuse sound is generated by decorrelating the impulse response and distributing the result equally to all loudspeakers. The relative contribution of these synthesis techniques is controlled using the diffuseness parameter obtained in the analysis. This parameter varies between 0 and 1 where 1 is a fully diffuse soundfield and 0 denotes that the sound has arrived from a single direction only. The parameter is used as a crossfade to ensure the reproduced soundfield has the correct ratio of diffuse to non-diffuse sound.

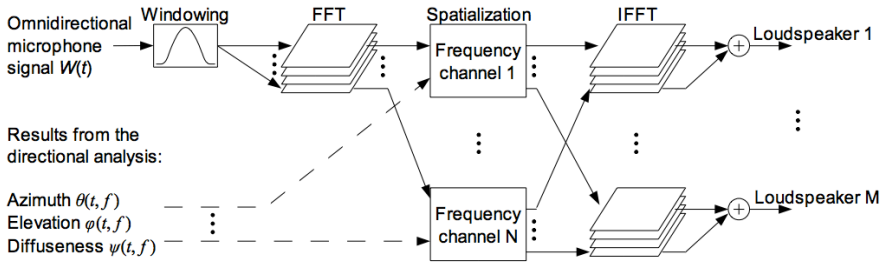


FIGURE 6.18: *Diagram of SIRR synthesis showing how the monoaural signal is respatialised using accompanying parametric data. Image reproduced from Merimaa and Pulkki (2004)*

There are no restrictions regarding the spatial audio technique used to spatialise the non-diffuse sound. It is largely decided by the individual application. For example, Kearney (2009) made use

of HOA panning techniques to spatialise reflections as point sources for auralisation. Kearney found that the subjective localisation accuracy of sound sources was equivalent for a soundfield rendered with 3rd order Ambisonics and VBAP. The relative trade-offs between HOA and VBAP have been discussed in the previous sections. It appears that the choice in panning technique is driven largely by practical constraints of the system, for example, the number and arrangement of available loudspeakers and the arrangement of listeners in the loudspeaker array.

Directional Audio Coding

Directional Audio Coding (DirAC) (Pulkki et al., 2009) is a generalised implementation of SIRR that can perform a similar function but for arbitrary soundfields (as opposed to impulse responses only). It is based on the same psychoacoustic premise as SIRR, that the human hearing system is capable of resolving a direction of arrival cue for a plane wave and a coherence cue for diffuse sound at any particular time instant and for each perceptual band. In its basic form, the general operation of DirAC is very similar to SIRR with the exception of processes which are optimised to suit the content of the spatial audio recording. For example, SIRR utilises an STFT that prioritises accurate transient reproduction over frequency resolution so that the temporal envelope of reflections is maintained. DirAC however must also ensure a high frequency resolution and so spatial audio recordings are commonly analysed and re-synthesised in perceptual audio bands to ensure the signals are reproduced with a high degree of perceptual accuracy. More recently, Multi-resolution STFT processing has been used to enhance the quality of DirAC reproductions (Vilkamo, 2008).

Specific implementations of SIRR and DirAC aim to reconstruct the soundfield by spatialising the omnidirectional pressure channel with the analysed parametric data. These implementations provide a highly efficient method of transmitting a 3D soundfield which requires a single channel of audio data and a single channel of accompanying data which contains the parametric representation of the soundfield's spatial attributes. By doing so, however, the pressure signal is replicated in all loudspeakers which will cause a high degree of correlation between the loudspeaker signals at the sweet spot, increasing the likelihood of comb-filtering around the sweetspot.

As an alternative, a so-called 'high quality' version of DirAC has been developed which assumes the original B-format recording is available at both analysis and synthesis phases. While this is less suitable for efficient transmission (due to the additional audio channels) it can significantly improve the quality of the synthesised soundfield. A diagram of the high quality implantation of DirAC is shown in Figure 6.19.

Prior to soundfield re-synthesis, in this example, the B-format audio is decoded using an Ambisonic decoding matrix to produce a signal for each loudspeaker channel. The parametric analysis is used to compute a set of frequency-dependent gains for each loudspeaker channel which are applied as a filter. These gains are determined by the VBAP panning gains and the ratio of diffuse to non-diffuse sound. The additional decoding stage reduces the likelihood of high-correlation of audio signals from each loudspeaker channel, thus in turn reducing the likelihood of audible comb filtering around the sweetspot.

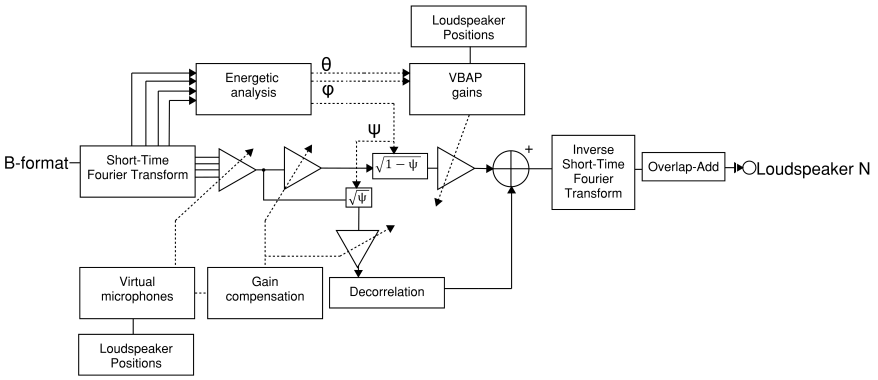


FIGURE 6.19: *System diagram of analysis and synthesis technique shown for a single loudspeaker channel. A B-format impulse response is analysed to determine the angle of arrival and diffuseness of each time-frequency bin. Audio signals are synthesised where non-diffuse sounds are spatialised using VBAP and diffuse sounds are reproduced as decorrelated signals. The relative gain of diffuse to non-diffuse signals is determined by this analysis.*

The parametric data for non-diffuse sound is applied to each loudspeaker as a filter with a response shown in equation (6.16). g_{vbap} applies the gain values for each frequency band and loudspeaker channel such that sound is panned to the correct direction using equations (6.14) and (6.19). These gains are multiplied by a factor $\sqrt{1-\psi}$ which determines the mixing coefficient of diffuse to non-diffuse sound.

$$g_{non-diffuse}(k,n) = \frac{g_{vbap}[n,\theta(k),\phi(k)]\sqrt{1-\psi(k)}}{\sqrt{[1-\psi]+\psi(k)g_{comp}^2}} \tag{6.16}$$

If the direction of arrival of sound is off-axis with respect to the virtual microphone, the energy of the sound is attenuated causing the sound to be reproduced with the incorrect amplitude. Therefore, a further modifier $\frac{1}{\sqrt{[1-\psi]+\psi(k)g_{comp}^2}}$ is applied which compensates for the energy lost via the use of virtual microphones (Vilkamo, 2008) where k denotes the directivity factor of the virtual microphone.

$$g_{comp} = \sqrt{1-k+\frac{k^2}{3}} \tag{6.17}$$

While it is possible for any virtual microphone polar pattern to be used, it is common to utilise a figure-of-eight polar pattern as it has the narrowest frontal lobe of those possible with a first-order Ambisonic microphone. The effect of the rear-lobe of the virtual microphone is considered less significant as for a non-diffuse sound, the loudspeaker gains in the opposing direction will be approaching zero. The diffuse soundfield is also not affected significantly due to this polar pattern choice as the sounds from the loudspeakers are being decorrelated (Vilkamo, 2008).

When DirAC is used for processing general soundfields, the intensity vectors can vary very quickly. This can cause noticeable distortion on the rendered soundfield. It is therefore common to apply temporal averaging to the intensity vectors in order to slow the rate of change of the vectors from one frame to another. The temporal averaging must slow the change in intensity vectors sufficiently so that distortion is reduced but not to the extent that spatialisation becomes sluggish. A common implementation is to average the intensity vectors within a Hanning shaped window (Vilkamo, 2008).

When DirAC is used as a means to transmit spatial audio data, an efficient data reduction technique is to perform the analysis and synthesis in perceptual bands. This can be implemented in the time domain by using a filter bank or by creating an equivalent set of filters in the frequency domain. However, for high-quality implementations of DirAC, there is no need to reduce the amount of data and so the processing stages are often applied on every time-frequency bin (Pihlajamaki, 2009).

Diffuse sound is re-synthesised by generating decorrelated audio signals distributed equally to all loudspeakers under the restriction that the timbre of the signal should not change (Pihlajamaki, 2009). This can be achieved by maintaining the magnitude response of the signal and randomising the phase of the signal, essentially applying an all-pass filter to the signal. There are numerous ways of achieving this and there are a number of trade-offs in each method. Often, there is a most appropriate type of decorrelation dependent on the type of signal being synthesised.

One method of decorrelating the loudspeaker signals involves introducing randomised time shifts into the signal. This is typically performed in perceptual frequency bands. The extent of the delay is set to reduce coherence between the audio signals but must be kept small enough (under 20ms) in order that the timbre of the signal is not changed significantly (Pihlajamaki, 2009).

Another proposed method is to convolve the diffuse signals with a noise burst. This noise burst can generate noise-like artefacts on the synthesised signals, therefore in some implementations an exponentially decaying envelope is applied to the noise burst prior to convolution. This method is more often associated with SIRR where a more aggressive decorrelation technique will not introduce significant artefacts. This method can introduce audible time-smearing onto dry or impulsive signals.

Another method of achieving this is to directly modify the phase response of the signal in the frequency domain. For each time-frequency block, the phase of the signal (denoted by the imaginary component) is multiplied by a random noise signal, $w(n)$, which takes values between 0 and 2π (6.18) . This process is similar to another audio effect known as *whisperization* (Zolzer, 2011).

$$S(\omega) = R \cdot e^{(i(2\pi w(n)))} \tag{6.18}$$

Similarly, this decorrelation method is often used in SIRR for decorrelation of reverberant sound. For DirAC, the extent of the randomisation is often reduced or the phase is modulated more subtly using continuous waveforms.

Once, the signals have been decorrelated, the diffuse signals are also weighted by a factor which compensates for the energy losses observed when using a virtual microphone. A further factor $\sqrt{\frac{1}{N}}$ weights the diffuse sound according to the number of loudspeakers N in the loudspeaker array (Pulkki, 2007).

$$g_{diffuse}(k,n) = \frac{1}{g_{comp}}\sqrt{\frac{1}{N}} \tag{6.19}$$

Due to the more widespread application of DirAC, a great deal of development has taken place in order to improve various aspects of the audio rendering. Tervo et al. (2013a) suggests some of the improvements could be applied to SIRR in order to improve the accuracy of auralisations.

It was demonstrated by Vilkamo (2008) (and confirmed in Appendix B) that an ST350 Ambisonic microphone, commonly used for measuring spatial soundfields, can accurately localise sounds up to a frequency of approximately 5kHz. Above this frequency, the capsule spacing and low number of transducers cause spatial aliasing. In order to avoid these errors being reproduced on the re-synthesised signal, the direction of arrival should be estimated from a low-pass filtered version of the signal. The higher frequency components can then be re-synthesised in this extrapolated direction.

SIRR and DirAC are often implemented assuming a source signal recorded in first-order B-format, however other microphone arrays can be used as a front end to this system. For instance, a higher order Ambisonic microphone or vector intensity probe. These microphone systems are capable of recording the soundfield at a higher resolution at higher frequencies, increasing the accuracy of the re-synthesised audio (Ahonen, 2013).

Spatial Decomposition Method

The Spatial Decomposition Method (SDM) (Tervo et al., 2013a) is a system optimised for analysing and auralising spatial room impulse responses. The premise of SDM is to decompose a spatial room impulse response into a set of image-sources. The image sources can be analysed (as demonstrated in Chapter 5) or can be used to synthesise audio signals which can be used for room acoustic auralisation.

In SIRR, the direction of arrival of a reflection is estimated using very small time-windows in order to reduce the influence of coincident reflections arriving at a similar time and amplitude. If two or more coincident reflections occur within one analysis window, the algorithm will localise that reflection as the vector sum of these reflections, which may not correspond to any of the arriving reflections.

SDM typically utilises an intensity probe to record the room impulse response. This features an array of six, spaced omnidirectional microphone capsules. The early reflections are localised by analysing the Time Difference of Arrival (TDOA) of the reflection at each microphone. The TDOA can be determined via cross correlation of reflections at each microphone. Due to the spaced array of microphones it is possible to isolate many near-coincident reflections and localise them without introducing significant angular errors.

Once the reflections have been localised temporally and spatially, an omnidirectional signal from one of the transducers is then spatialised to a loudspeaker array according to the spatial data. The analysis and synthesis utilises very short time window lengths allowing the spatial data for every sample to be determined. As SDM spatialises the pressure signal from one of the microphone transducers, the frequency response reproduced at the sweetspot is very accurate (as it is not the sum of numerous signals).

The spatial distribution of the diffuse sound part of the impulse response is reproduced naturally as the analysis technique results in a random direction for each audio sample providing a natural decorrelation. This is advantageous as there is no need to carefully design a decorrelation stage for soundfield synthesis. As discussed earlier, many decorrelation techniques can introduce audible artefacts into the auralisation. SDM has been shown to produce very successful results in comparison to SIRR (Tervo et al., 2013a) however this comparison was made with the

original version of SIRR as opposed to more recent approaches which use some of the recent advancements of DirAC.

6.6 Summary and discussion

This chapter has described how stage acoustic conditions can be interactively auralised in response to the sound of a musician's instrument. The purpose behind developing such systems is to introduce musicians into controlled virtual acoustic environments, ultimately allowing researchers to study the sensory and subjective experience of musicians performing on different stages. Different approaches to stage acoustic auralisation have developed since the first system developed by Gade. Each approach required specific measurement and signal processing techniques to suit the requirements of each particular study. These systems were discussed in detail. As shown by Gade (2010), the results obtained from stage acoustic laboratory tests are sometimes limited due to the lack of realism in the rendering of the acoustic response. This implies that further development is required to ensure the acoustic response is rendered as accurately as possible.

It was also discussed how various contributions of each processing stage could contribute to a feeling that the musician was not playing naturally in an environment, rather as if they were playing over a PA system in a venue, termed the *"PA effect"*. This was found to be caused by aspects such as low-frequency emphasis caused by the proximity of a directional microphone to the musical instrument, colouration due to spatial sampling of the musical instrument's radiation pattern and a reduced dynamic range as the proximity of the microphone also picks up and auralises ancillary sounds made by the musician (breath sounds, key clicks). It was also discussed how other aspects of the auralisation could contribute to a lack of naturalness from the performer's perspective. For instance, system latency, the presence of representative background noise in the auralisation or provision of a representative source directivity patterns.

More recently, the focus of auditorium acoustics research has centred around the spatial and temporal distribution of early reflections dictating the need for auralisation systems which are capable of rendering the acoustic conditions at a high spatial and temporal resolution. These more recent techniques were discussed in relation to their possible use in stage acoustic laboratory tests. Specifically, Ambisonic and parametric rendering techniques were explored in detail.

It was shown how an Ambisonic impulse responses (as shown in Chapter 3) could be auralised in real-time using recently developed convolution processors. It was then demonstrated how the resultant signals could be rendered over a loudspeaker array, emulating the musician's perspective, the experience of playing on a concert hall stage. It was shown that Ambisonic auralisation techniques provide a fully isotropic representation of the soundfield, where the accurate rendering of reflections was not determined by their angle of arrival with respect to the loudspeaker array. It was also shown that, where available, HOA rendering techniques could provide a marked improvement in auralisation accuracy. However, It was discussed that the high degree of correlation between loudspeaker signals often produced comb-filtering (and hence timbral colouration) around an often limited sweet spot. This was considered to be crucial for developing a plausible

virtual stage environment as a musician is likely to move around while they play their instrument and so would be aware of any comb-filtering artefacts.

It was also shown how Ambisonic impulse responses could be rendered using parametric decoding techniques (such as HARPEX, SIRR/DirAC and SDM), where the spatial characteristics of the impulse response are analysed and then reproduced using any spatial audio technique over a loudspeaker array. It was discussed how this approach could render an impulse response using amplitude panning techniques and the reverberant response using decorrelated loudspeaker signals. It was demonstrated how the use of VBAP to spatialise sounds over a loudspeaker array provided excellent localisation accuracy at the loudspeaker locations but reduced accuracy between loudspeakers. Consequently, a sufficiently dense loudspeaker array is required for accurate spatialisation of reflections. This technique however is regarded as robust method of enhancing this perceived spatial accuracy from first-order Ambisonic auralisations. As the correlation between each speaker channel is reduced, there is a reduced likelihood of perceived colouration around the sweet spot. Parametric decoding techniques can assist in improving the quality of auralisations where the impulse response was captured using a first-order Ambisonic microphone. Moreover, the technique is compatible with many different microphone arrangements and so is a viable technique for future stage acoustic research.

In general it can be seen that there are numerous trade-offs inherent in each spatialisation technique and often the decision to use one particular technique is driven largely by the physical resources available and the nature of the auralisation. As discussed in Chapter 3, stage acoustic impulse responses were measured using a first-order Ambisonic microphone. Given the objectives of this research, it was deemed important for the spatial characteristics of early reflections to be rendered as accurately as possible. In addition, it is likely that the musician may move around the sweet spot more than in a passive auralisation due to the physical act of operating their instrument. It is for these reasons that a parametric decoding approach was favoured for stage acoustic laboratory tests conducted as part of this study. Prior to constructing the interactive auralisation system, two pilot tests were carried out to evaluate the auralisation technique and to gain initial support for the research hypothesis. These pilot tests are described in the following chapter.

Chapter 7

Auralisation development and pilot tests

In Chapter 6, a number of approaches to interactive auralisation were discussed and it was considered that parametric decoding methods had a number of potential advantages over first-order Ambisonic methods. Parametric decoding methods such as SIRR, attempt to render the soundfield in such a way that colouration artefacts, produced by the listener moving around the sweetspot, are minimised. This technique was found to be compatible with many different microphone arrangements and could be used in the future to perform complex spatial audio effects on the impulse response. In addition, SIRR has been previously reported as providing a perceptually enhanced render of first-order Ambisonic soundfields. The interactive auralisation system was therefore designed using this technique.

During development of the interactive auralisation system, two pilot tests were conducted which aimed to assess the performance of the parametric decoding technique and to gather initial justification for conducting the main study. While these initial tests had differing objectives, the apparatus and general approach was very similar and so these aspects will be discussed in tandem. The testing method and results of each test will be discussed separately.

7.1 Objectives

In Chapter 6, a number of spatial audio techniques were reviewed as potential methods for interactive auralisation of concert hall stages. It was discussed that parametric decoding techniques could potentially improve an auralisation that was based on first-order Ambisonic impulse responses. This was mainly due to the techniques used to render early reflections (VBAP) and reverberation (decorrelation). The processing requirements of this approach are much more complex than simply using first-order Ambisonics and it was considered uncertain if there would be an audible difference between the two methods.

Therefore, the first listening test aims to evaluate the parametric decoding system in comparison to available first-order Ambisonic decoding. The objective was to determine if parametric decoding and Ambisonic decoding produced noticeably different sounding auralisations. As discussed in Chapter 2, the characteristics of the musical instrument may cause early reflections from certain angles of arrival to be masked more than others. A further objective of this test was to determine if the angle of arrival of a single early reflection had an audible impact on the resulting auralisation.

Based on the findings of this test, a second pilot test was developed which aimed to determine if listeners could distinguish between different stage acoustic environments that featured identical levels of ST_{early} but differing spatial or temporal distributions of early reflections. This test was considered to be a trial run of the main listening test which was conducted with musician test participants. In this test, the source signals were auralised using impulse responses measured on different concert hall stages as described in Chapter 3.

In these initial tests, it was considered necessary that the auralised sound source was kept constant so that aspects of the auralisation system could be evaluated in isolation. In previous research conducted by Gade (1982), musicians found it very difficult to repeat phrases exactly when comparing concert halls. This increased the difficulty of the test as variations in what the participant has heard may be due to their own technique rather than the physical difference in acoustic conditions. In addition, it was necessary to recruit participants internally from the department, many of whom had a range of musical experience and listening ability. Many of the participants were musicians of instruments commonly found in popular music which were considered inappropriate for this research.

Therefore, these initial listening tests were designed as passive auralisations where short musical samples were played through a directional loudspeaker positioned in front of the participant, imitating the direct sound from a musical instrument. This signal was simultaneously auralised and the results rendered over the surrounding loudspeaker array. It was considered that this arrangement of apparatus provided a sufficiently plausible facsimile of the acoustic conditions experienced by a performing musician, for a passive listener. The specific arrangement of sound sources will be discussed in further detail in the following section.

It is acknowledged that this arrangement removes the musician from the physical act of playing their instrument, resulting in a lower cognitive load when compared with a genuine performance situation. Therefore, it was considered likely that the participants may have been able to focus their attention more directly on the auralisation. In addition, the relative position of the participant and loudspeaker were much further apart than is typically seen for most handheld musical instruments. The participant was never in direct contact with the sound source which reduces the potential impact of physical masking through vibrations conducted through the tissue. These aspects may have allowed participants to detect finer details of the auralisation than a performer operating a real instrument, therefore the results of the pilot tests should be interpreted with care and can not be used to infer the sensory ability of the wider musician population.

7.2 Apparatus

Both pilot tests were conducted in the Arup-DDS SoundLab located at the Digital Design Studio in Glasgow. The SoundLab is an acoustically controlled auralisation space which features a 16 channel 3D loudspeaker array. The SoundLab and equipment set up is described in detail in Appendix C.

The SoundLab acoustic conditions are not anechoic and there are two large TV screens which are used for presentations and visualisation. In order to reduce the impact of reflections from these screens, additional absorption was applied to the TV screen. In addition, it was not feasible to auralise the floor reflection for this experiment and so this was reproduced naturally by placing a removable wooden floor directly on top of the SoundLab floor. The acoustic conditions of the SoundLab with and without these treatments are described in further detail in Appendix D.

As shown in Figure 7.1 the participant was seated in the sweetspot of the loudspeaker array with an additional loudspeaker positioned in front of them which was used to emulate the direct sound from a musical instrument. This loudspeaker was a Genelec 1029A Active loudspeaker which was mounted on a tripod at a height of 100cm above the floor (height to the top of the low frequency driver) and a radial distance of approximately 50cm from the sweetspot to the centre of the loudspeaker.

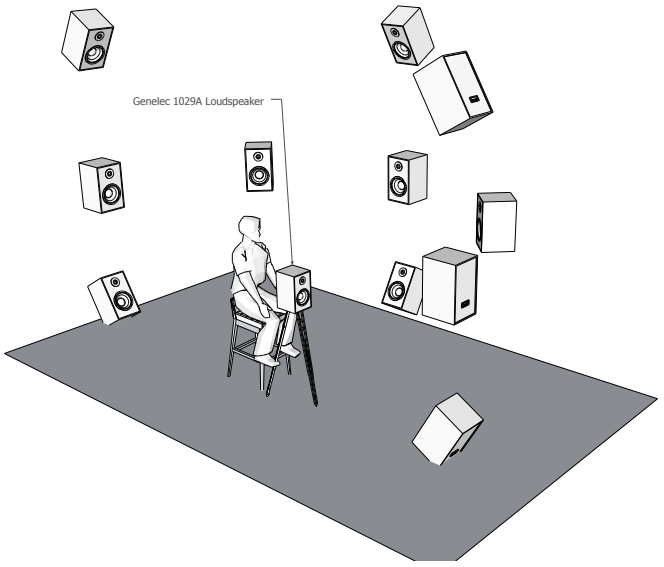


FIGURE 7.1: *Diagram showing the listener positioned in the sweetspot of the loudspeaker array. The loudspeaker in front of the musician was used to emulate the musical instrument and the surrounding loudspeaker array was used to emulate the concert hall acoustics. The loudspeakers on the right hand side of the sweetspot have been omitted for clarity.*

Short samples of music were rendered through this loudspeaker to emulate the direct sound. Simultaneously, the source signal was convolved in real-time with the required impulse response.

The results were rendered over the surrounding loudspeaker array to emulate the acoustic conditions of each space. For each trial, the sample of music would be auralised in different halls and the listener required to respond according to the specific question.

The digital signal processing was implemented in Max. Max is a graphical programming environment commonly used for audio and video programming (Cycling74, 2013). In the first test, real-time convolution was performed using SIR2 VST plug-ins (Knufinke, 2010) whereas in the second test this was implemented using the HISSTOOLS external objects (Harker and Tremblay, 2012). Additionally, the first pilot test used a different model of computer and soundcard to the second.

When auralising Ambisonics, the loudspeaker signals were processed using a Gerzonic Decopro Ambisonic decoder VST plug-in (Gerzonic, No date) set to a 'Max-RE' type for all frequencies. This decoder does not provide any dual-band decoding or optimisation routines to ensure accurate localisation.

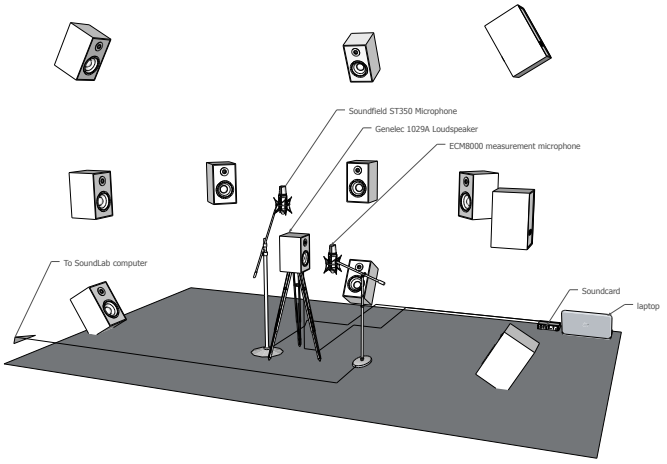
The following sections will describe the specific experimental methodology and results of each experiment.

7.3 Auralisation calibration

Prior to performing any listening tests, it is necessary to calibrate each auralisation to confirm the stimuli are presented correctly. This calibration includes minimising the effect of system latency, to ensure reflections are arriving at the correct time, and setting the level of reflections to an appropriate magnitude.

The calibration procedure utilised a measurement system shown in Figure 7.2. The measurement system is very similar to that used to measure the stage acoustic responses in venues as described in Chapter 3. An Ambisonic microphone is positioned in the sweetspot of the array with the loudspeaker positioned directly beneath. Sound from the loudspeaker is captured with a measurement microphone positioned in front of it. The signal from this microphone is convolved with the required impulse response and the resulting signals rendered over the loudspeaker array. By playing a swept sine wave signal through the loudspeaker and recording with the Ambisonic microphone, it is possible to capture the impulse response of the virtual space. By analysing the acoustic response it is possible to make appropriate adjustments so that the acoustic conditions can be verified against the target acoustic response.

(a) Image of experimental set up



(b) Diagram of experimental set up

FIGURE 7.2: *Photograph of the test apparatus arranged in the SoundLab. The microphone is positioned in the sweetspot of the loudspeaker array while the measurement loudspeaker is positioned directly below. An additional microphone is visible in front of the loudspeaker which is used to pick up the direct sound which will then be processed by the auralisation system. For clarity, a number of loudspeakers in Figure 7.2(b) have been omitted.*

7.3.1 System latency

Given the focus of this research, it is important to ensure that reflections rendered by the spatial audio system arrive at the sweetspot at the correct time. The processing equipment and the position of the musician relative to both the microphone and loudspeaker array introduce a delay between the expected and measured arrival time of a particular reflection. If the latency is large enough, the musician may become aware of reflections that were otherwise inaudible due to masking or be disturbed by late arriving reflections. In previous research, a latency of less than $10ms$ has been reported to be acceptable for virtual performance (Woszczyk et al., 2012).

As mentioned previously, the auralisation system is not required to reproduce the direct sound or floor reflection of the impulse response. These elements are replaced by a period of silence of specific length at the beginning of the impulse response. This period of silence can then be

truncated by an amount equal to the measured overall system latency to ensure that the early reflections arrive at the intended time of arrival. This is demonstrated in Figure 7.3. The upper diagram shows the effect of latency on the auralised impulse response. The auralised reflections (shown as dashed lines) occur after a delay equal to the system latency, t_{lat}. The lower diagram shows how the impulse response can be truncated by the same duration in order to minimise time-of-arrival errors.

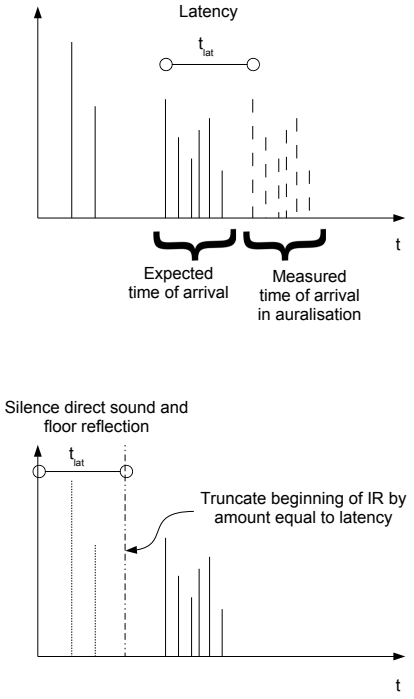


FIGURE 7.3: *Upper diagram showing the effect of system latency on an auralised impulse response in an interactive auralisation system. Lower diagram demonstrates how latency can be minimised by truncating the impulse response by a duration equal to the system latency. This is acceptable as the direct sound and floor reflection are recreated naturally in the SoundLab.*

The overall system latency can be decomposed into a number of main contributing factors as described in equation (7.1) below where $L_{src-mic}$ is the delay caused by acoustic propagation from the sound source to the microphone, $L_{I/O}$ is the input-output (I/O) buffer size of the soundcard, $L_{process}$ is the delay caused by the signal processing and $L_{spk-musician}$ is the delay caused by acoustic propagation between the loudspeaker array and the musician.

$$t_{lat} = L_{src-mic} + 2L_{I/O} + L_{process} + L_{spk-musician} \tag{7.1}$$

As the microphone is positioned 30cm away from the sound source and the loudspeaker array is on average $1.65m$ distance away from the sweetspot, the total delay due to acoustic propagation can be determined by dividing the total distance travelled by the speed of sound i.e. $(1.65m + 0.3m)/344 = 5.6ms$ which equates to 247 samples at a sample frequency of $44.1kHz$. The delay due to acoustic propagation will vary depending on the temperature of the SoundLab however this variation is likely to be small in magnitude.

A further delay occurs as the analogue signal from the microphone is converted to a digital signal in the soundcard. The soundcard periodically captures the analogue signal and stores it in an input buffer of a chosen size. Similarly, when the digital signal is converted back to an analogue signal a Digital to Analogue Converter (DAC) periodically stores digital signals in an output buffer before conversion. The size of these buffers is usually kept identical and is set in terms of samples. The length of the I/O buffer is normally a power of 2. Smaller I/O buffers require faster processing in order to accurately read or render the analogue signal. However, small I/O buffer sizes can result in processing errors which introduce noise-like artefacts into the rendered signals. Longer I/O buffers introduce a longer latency but are often more stable given the processes can occur over a longer period of time. Therefore, there is a trade-off between available processing power and I/O buffer size. If the sound card is operating at a sampling frequency of 44.1kHz and an I/O buffer size of 1024 samples, the delay between the signal input and output would be $46ms$.

Additional delay is introduced by the various DSP operations including the convolution of the input signal with the venue impulse response and the subsequent filtering of the signal to ensure the contribution of each loudspeaker is equal. Often, a longer impulse response requires a larger I/O buffer to render accurately. Furthermore, DSP processes can have their own buffer systems to increase efficiency but at the expense of a slight delay.

Figure 7.4 demonstrates the effect of truncating the beginning of the impulse response. A test signal (shown here in red - W channel only) was created which consisted of fixed amplitude impulses which occur ever 0.15 seconds after an initial delay of 0.1 seconds. The test signal was imported into the interactive auralisation system and the impulse response of the virtual soundfield was measured as described previously. The blue trace shows the time of arrival of the impulse responses with no latency correction applied where the I/O buffer was set to a value of 16 samples. The measured latency in this case was 11.5ms or 508 samples. By truncating the beginning of the impulse response by this amount, the first reflection (and therefore the rest of the impulse response) appear at the correct time delay shown by the green trace.

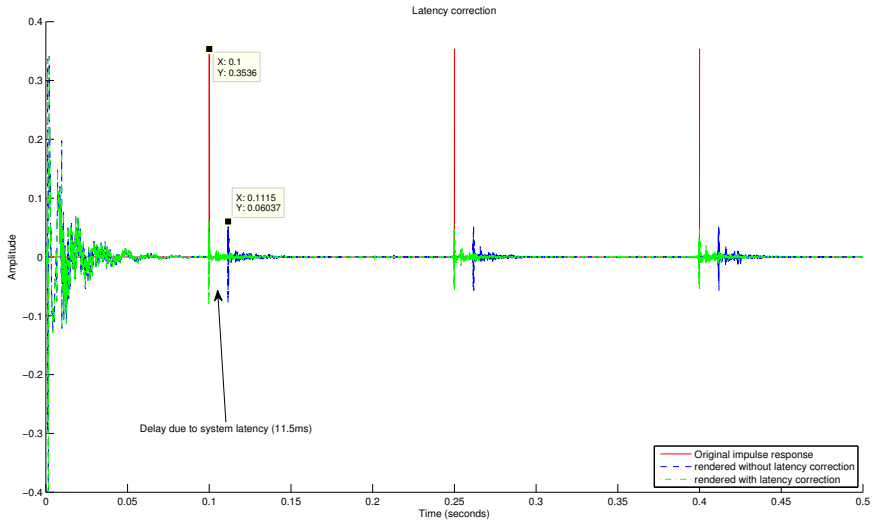


FIGURE 7.4: *An impulse response with three reflections (shown in red) is auralised and measured in the sweetspot of the loudspeaker array producing the impulse response shown in blue. The auralised reflections arrive 11.5ms after the expected time of arrival. By truncating the silence at the beginning of the impulse response by this amount, the reflections can be made to arrive at the correct time (shown by the green trace).*

The pilot tests are implemented as passive auralisations where the direct sound is rendered through a single loudspeaker, representing the musical instrument, and auralised through the loudspeaker array, rendering the acoustic environment. It is clear that this set up has slightly different latency requirements to an interactive system in that there is no delay caused by $L_{src-mic}$ and half of the $L_{I/O}$ delay. The latency for the pilot tests is still accounted for in the same way.

7.3.2 Level of acoustic response

As discussed throughout this thesis, it is clear that the overall energy of the impulse response relative to the direct sound is an important aspect of how performance spaces are perceived by musicians. Therefore, any stage acoustic laboratory experiments must present the acoustic response at the correct target level. Furthermore, the level of the virtual acoustic response must be set carefully to avoid unwanted acoustic feedback.

In passive auralisations, the gain of the auralisation can be adjusted such that it matches predicted levels. However, in interactive auralisations, the audio signal being auralised is unknown, therefore the gain of the auralisation must be calibrated via direct excitement and measurement of the virtual space. Adjusting the gain of the auralisation system so that the virtual space has the same properties as the measured space ensures that the auralisation is being rendered at the correct level. In principle, once the gain has been set correctly, the relative level of the acoustic response will be maintained regardless of the sound pressure level produced by the musician.

The use of ST_{early} is an obvious choice for setting the overall gain of the auralisation and has been suggested by previous authors such as Watson and Clark (2010). In this case, the virtual space is measured using a loudspeaker and Ambisonic microphone (as described in Chapter 3),

positioned in the sweetspot of the SoundLab. An additional microphone is positioned in front of the loudspeaker which will capture the excitation signal and auralise it as described previously. Prior to level calibration, the system latency is measured and corrected. This ensures that the measurement in the virtual space is as close as possible to that of the real space.

In perfectly anechoic listening rooms, reproducing the ST_{early} would be straightforward, however, in non-anechoic (but acoustically treated) listening rooms, the acoustic response may contribute to the level of ST_{early} measured at the sweetspot if the response of the listening room has a significant level between 20ms and 100ms. As the SoundLab is a non-anechoic space it is necessary to determine its acoustic contribution and to minimise it as far as possible. This is discussed in detail in Appendix D.

Figure D.4 and D.5 show ST_{early} and ST_{late} due to the acoustic response respectively of the SoundLab only. It can be seen that the level of ST_{late} is very low at $-52.84dB$ relative to the direct sound which is considered extremely low in relation to the direct sound. However, the ST_{early} of the SoundLab is $-14.36dB$ relative to the direct sound which is comparable to the level of ST_{early} measured in some concert halls. For example, the mean ST_{early} of the Reid Hall (RH) was found to be $-14.87dB$ as reported in Chapter 5. Therefore, it will not be possible for the SoundLab to reliably reproduce the acoustic response of venues that feature a lower value of ST_{early}. Typically, values lower than the SoundLab ST_{early} are observed on large stages used for orchestral performance. An example of this is the Caird Hall (CH), as discussed in Chapter 5 that featured a mean ST_{early} of $-17.85dB$.

When setting the overall gain of the virtual acoustic response, it is often the case that the target value of ST_{early} can be reproduced or the value of ST_{late} can be reproduced. However, it is less common for both values to be correct. Consequently, the gains of the early and late parts of the impulse response must be set separately. This is achieved by applying a linear cross-fade between the early and late part of the impulse response. The cross-fade is applied at a time delay of $t = 100ms$ and has a duration of $5ms$. In addition, a $1ms$ fade-in is applied at $t = 20ms$. This is to avoid introducing audible artefacts due to the impulse response beginning at a 'non-zero' crossing. This is illustrated in Figure 7.5. The red and blue coloured regions of the impulse response denote the early and late parts of the impulse response respectively after the amplitude envelopes have been applied. The amplitude envelopes are shown in green and magenta. The impulse response has been normalised to unity gains for clarity.

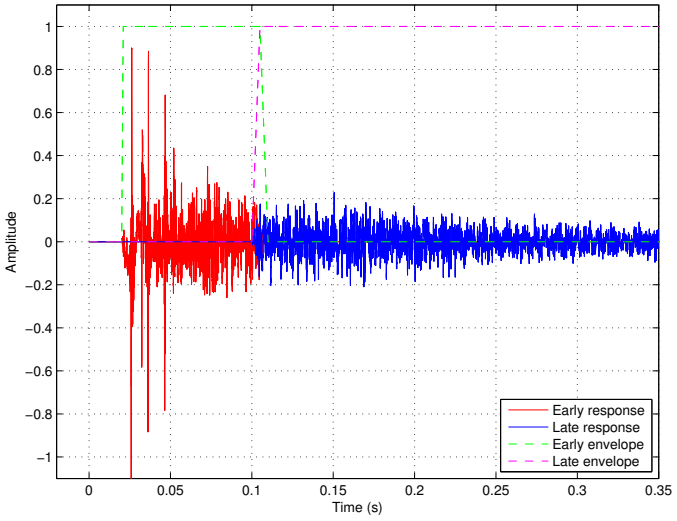


FIGURE 7.5: *This plot demonstrates how the impulse response is split into early (red) and late (blue regions). Overlaid is the amplitude envelope applied to the impulse response. The green trace shows the early amplitude envelope, whereas the magenta trace shows the late amplitude envelope. A linear cross fade of 5ms duration occurs at a time delay of 100ms.*

With the auralisation system and calibration procedure described, the following sections will report the methodology and results of each pilot test.

7.4 Pilot test 1 - Spatial audio technique

The first listening test aimed to determine if a listener, exposed to similar acoustic conditions to a performing musician, could perceive a difference between auralisations rendered with either first-order Ambisonics or SIRR. This was to help justify the use of parametric decoding techniques described in Chapter 6. It was expected that due to the different rendering methods used, participants would consistently identify one method over another due to the differences in how each approach renders the impulse response.

In addition, the experiment aimed to determine if the audibility of an early reflection was affected by the angle of arrival and the type of musical phrasing. It was expected that early reflections arriving from the front would be more easily masked by the sound source. This would provide an initial indication that the spatial distribution of early reflections influences how musicians perceive their acoustic environment. In addition, it was expected that early reflections would be easier to hear when the musical sample featured separated notes with staccato articulation.

7.4.1 Participants

Six volunteers were recruited for this test, all between the ages of 24 and 32 (4 male, 2 female). The participants were students from the Digital Design Studio, most of whom had a background

in audio engineering or acoustics, all of the volunteers had some prior experience of music performance. All test subjects reported no significant hearing loss. Prior to the beginning of the test, participants were asked to read the test instructions and provide informed consent. At the end of the experiment, participants were rewarded with a small amount of confectionary. The test took, on average, 45 minutes to complete however participants were free to work through the questions at their own pace.

As discussed earlier, it was considered appropriate in this case to use test participants that were not musicians. This was primarily due to the auralisation system being the main subject of the test, as opposed to the specific acoustic conditions. Using participants from the department allowed the initial testing to be conducted without revealing the nature of the research to potential candidates who may have taken part in the final listening tests. Furthermore, the passive auralisation did not require the participants to have any specific abilities or training ensuring that more participants could be recruited at this early stage.

7.4.2 Procedure

The listening test was conducted as an A/B (hidden reference) similarity test where listeners compared pairs of musical samples. Each sample was auralised with an impulse response and rendered using either FOA or SIRR. Participants were asked to listen to each sample and record how similar or different they thought the two sounds were on a five-point Likert scale. Participants recorded their responses on a continuous scale by marking an X on a paper scale. A value of 1 indicated that the samples were very similar, whereas a value of 5 indicated the samples were very dissimilar.

The responses from each participant were assigned a numeric value based on the position of the 'X' on the continuous scale. This value was obtained by measuring the distance of each mark from the nearest anchor point with a ruler. The recorded value would be between 1 and 5 and was recorded to a resolution of 1 decimal place. These data were of ordinal type, therefore responses for each question were summarised by calculating the median response and the 25th and 75th percentile ranges.

The stimuli were administered by the researcher using a Max patch. The researcher announced which sample was playing by saying either *"Sound A"* or *"Sound B"* and then activating the appropriate sample. The participant could request to hear samples again by saying which one they would like to hear. After they had recorded their response, the participant would say *"Next"* in order to move on to the next question. During the test there was no other communication between the participant and researcher. The researcher was positioned behind a screen throughout.

7.4.3 Stimuli

The stimuli consisted of anechoically recorded audio signals that were auralised with different impulse responses, rendered either with SIRR or first-order Ambisonics. The impulse responses

consisted of either:

- A synthesised specular reflection
- A measured stage impulse response
- No synthesised acoustic response

The specular reflection was rendered at a time delay of $60ms$ after the direct sound at an amplitude of $-6dB$ relative to the amplitude of the direct sound. This reflection was spatialised in the horizontal plane only at an azimuth of either 0°, ±60°, ±90° or ±110°. Reflections rendered at 0° and ±90° arrived from the direction of loudspeakers whereas reflections from ±60° and ±110° arrived from between loudspeakers. Reflections were spatialised to either side of the participant at random to give a symmetrical distribution of reflections. This reduced the possibility of the participant habituating to reflections arriving from a particular direction.

The stage impulse response used in this listening test was measured in the Grand Hall of the Glasgow City Halls as described in Chapter 3. The source and receiver were positioned at the stage right position (S3R3) and the sound source oriented to 0° i.e. towards the front of the stage. This impulse response was captured using an Ambisonic microphone and then processed using SIRR to obtain impulse responses for each loudspeaker channel.

The impulse responses consisting of single reflections were synthesised in MATLAB by rendering an impulse at the appropriate delay time and with an amplitude of -6dB relative to the direct sound. This impulse emulated a single specular reflection and was subsequently spatialised using FOA panning techniques to generate the B-format audio. This impulse response could then be processed using SIRR to generate the appropriate loudspeaker feeds. However in practice, when these impulse responses are processed using SIRR, the reflections are rendered only using VBAP due to the reflections being completely non-diffuse. Therefore, it was considered more efficient to render the single reflections using either a VBAP or FOA panner. The time of arrival and amplitude were controlled using a variable delay line and gain respectively. In this experiment, the reflection was rendered without any frequency domain filtering applied to emulate frequency dependent energy loss.

Prior to conducting this test, a series of measurements were made in order to calibrate the auralisation. To do this, the measurement system described in Chapter 3 was positioned in the centre of the loudspeaker array. The measurement system was used to measure the impulse response of the virtual space and allowed adjustments to be made in order to calibrate each auralisation. The calibration process was very similar to that described in Section 7.3.1 where adjustments were made to ensure that latency effects were minimised as far as possible and the gains adjusted so that reflections were presented at the correct level, or where appropriate, at a desired level of ST_{early}.

7.4.4 Programme material

The sound samples used in this experiment consisted of anechoic recordings of different musical instruments. The samples were chosen so that they were suitably brief and exhibited a range of musical articulation such as staccato and legato phrasing. The samples used were:

1. A short cello sample playing legato (Source 1)
2. A short clarinet sample playing staccato (Source 2)
3. A sustained long note from a clarinet (Source 3)

These signals are shown in Figure 7.6. The staccato clarinet sample features a wide range of phrasing and dynamics including a number of short separated notes at the beginning of the phrase. The legato cello sample is approximately twice the length of the clarinet sample and features long sustained notes with vibrato. The sustained clarinet note features a long crescendo followed by a diminuendo of similar length. These phrases were considered to have a suitable range of articulation for this experiment.

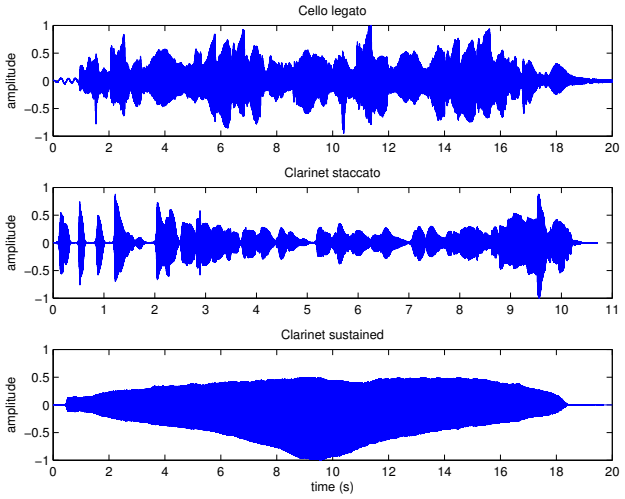


FIGURE 7.6: *Plot showing time-domain representation of signals used as stimuli for the listening test. The top plot shows the staccato clarinet, the middle plot shows a cello playing legato phrasing, the lower plot shows a clarinet playing a sustained note.*

There were 47 randomised combinations of stimuli in total. The order of stimuli was randomised to avoid expectation bias in addition, the hidden reference was randomly assigned to sound A or sound B. The test subjects were permitted to listen to each excerpt as many times as they liked before recording their answer. The test subjects were not given any visual references and were asked to face forward at all times but were not physically restrained in anyway. A number of null tests were introduced where both samples were the same in each pair and a single example test was presented to the listener at the beginning which was not included in the results.

7.4.5 Results

The results in Figure 7.7 summarise how similar or dissimilar the participants thought the musical pairs were when one of the samples was auralised with a single reflection (using FOA or SIRR) and the other sample was played without a simulated reflection (i.e. direct sound only).

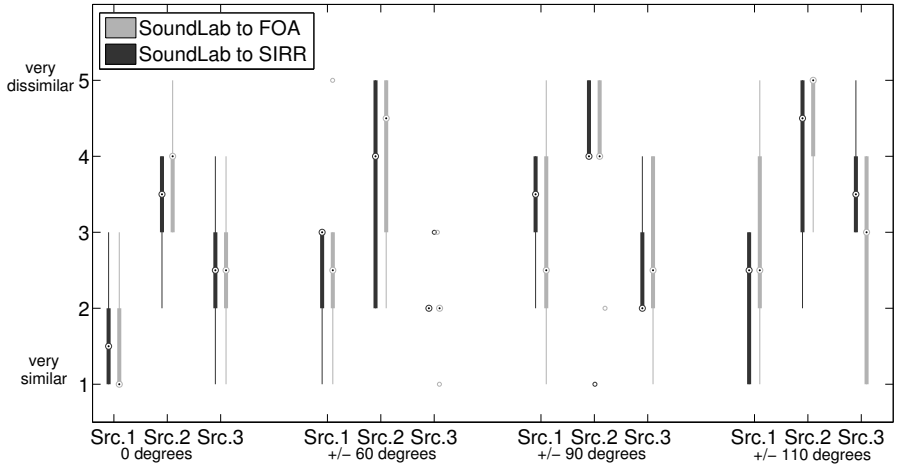


FIGURE 7.7: *Listening test results for sounds auralised with a single reflection rendered with FOA or SIRR compared to no reflection. Thick lines indicate 25th and 75th percentiles, thinner lines show the extremities of the data points, dots within boxes indicate the median while circles indicate outliers*

It can be seen that in most cases, auralising with either technique produces similar median scores throughout. This implies that the spatialisation technique did not have a strong impact on how the auralisations were perceived. The Wilcoxon rank-sum test (equivalent to the Mann-Whitney U test) can be applied to determine, for each test case, whether the results for FOA and SIRR are from distributions with equal medians. The results are shown in Table 7.1 where W is the sum of ranks and the p-value is computed from the z-statistic, Z (standard score). The p-values are much higher than a significance value of $p = 0.05$ which suggests the null-hypothesis of equal medians should not be rejected, implying that a single reflection is perceived similarly when rendered with SIRR or FOA.

Source	W	Z	p
Source angle = 0°			
Source 1	35.5	-0.49	0.62
Source 2	46	1.06	0.29
Source 3	36	-0.41	0.68
Source angle = 60°			
Source 1	37.5	-0.16	0.87
Source 2	39.5	0.00	1.00
Source 3	31	-1.21	0.23
Source angle = 90°			
Source 1	34	-0.73	0.47
Source 2	44	0.73	0.47
Source 3	39	0	1.00
Source angle = 110°			
Source 1	42.5	0.48	0.63
Source 2	43	0.58	0.56
Source 3	34	-0.72	0.47

TABLE 7.1: *Results of Wilcoxon rank-sum test to determine if scores for SIRR and FOA are from the same distribution. W denotes the rank sum, Z denotes the z-statistic and p is the p-value associated with a null hypothesis test that data in each comparison are from distributions with equal medians.*

It is also of interest to determine if the angle of arrival of the early reflection had an impact on how the auralisations were perceived. Therefore, a Kruskal-Wallis test was performed on the responses for different angles for each source type. This test is a non-parametric equivalent of ANOVA and can be used to determine if there is a statistical significant difference between two or more groups of ordinal data. The results are shown in Table 7.2 where the p-value, $(p > \chi^2)$, is approximated using the χ^2 value. It can be seen that for Source 1, the p-value is lower than the significance value of $p \leqslant 0.05$ suggesting the angle of arrival of a single reflection affects how similar the auralisations sound. However, the results for Source 2 and Source 3 suggest the angle of arrival did not have a significant impact.

Source	χ^2	$p > \chi^2$
Source 1	10.30	0.02
Source 2	5.58	0.13
Source 3	7.26	0.06

TABLE 7.2: *Results of Kruskal-Wallis test to determine if the angle of arrival of a single reflection can influence how it is perceived. The test has been repeated for each reflection angle. Where χ^2 is the test statistic and $p > \chi^2$ indicates the p-value associated with the null-hypothesis that samples are drawn from the same population*

For each reflection angle, it can be seen from the responses that the staccato clarinet sound source (source 2) resulted in reflections being identified more easily than the sustained clarinet tone or legato cello. To test this, the Kruskal-Wallis test can be applied in order to test the null hypothesis that responses are drawn from the same population. Table 7.3 shows the results of

this test. For all source angles, the test results cast doubt on the null hypothesis at a significance level of $p \leqslant 0.05$. This suggests that the musical sample can influence how the comparisons are perceived.

Source angle	χ^2	$p > \chi^2$
0°	19.48	$5.8 \cdot 10^{-5}$
60°	12.54	0.002
90°	9.05	0.011
110°	10.98	0.004

TABLE 7.3: *Results of Kruskal-Wallis test to determine if the musical sample has an impact on perception of single reflections. Where χ^2 is the test statistic and $p > \chi^2$ indicates the p-value associated with the null-hypothesis that samples are drawn from the same population*

The results in Figure 7.8 show how similar or dissimilar the participants thought the musical pairs were when musical samples were auralised with a stage acoustic impulse response rendered with FOA or SIRR. A Kruskal-Wallis test reveals that there are no statistically significant differences between responses recorded for each musical sample ($\chi^2 = 0.88$, $(p > \chi^2) = 0.65$). Given, the location of the median values, this suggests that participants could discern a slight difference between the musical samples and that this difference was consistent even if the source type was altered.

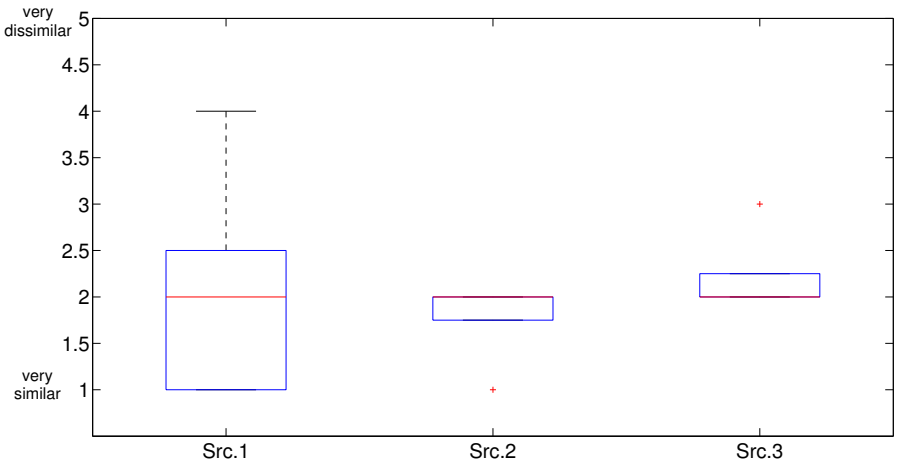


FIGURE 7.8: *Listening test results for sounds auralised with a measured impulse response rendered with FOA compared with SIRR. Central lines indicate the median response while box edges indicate 25th and 75th percentiles, outliers are indicated by crosses*

7.4.6 Discussion and remarks

The results of this pilot test suggest that the rendering technique (either first-order Ambisonics or SIRR) did not have a significant effect on how a single reflection was perceived. Furthermore, when directly comparing concert hall auralisations, it was found that SIRR and first-order Ambisonics were perceived to be similar by test participants. This is an initial indication that the

implementation of SIRR is performing similarly to the implementation of first-order Ambisonics. Given the potential benefits of using parametric techniques, as described in Chapter 6, this suggests it is a viable auralisation method for stage acoustic laboratory experiments.

The results also provide some initial indication that the audible effect of a single reflection was most apparent when source 2, the staccato clarinet, was playing. It is possible that the gaps between each note provided an opportunity for the reflection to be heard, in contrast to the other musical samples where the effect of the reflection may have been masked. It was also suggested that the angle of arrival of the single reflection produced an audible impact for one source type (Source 1). It is feasible that the masking effect of the sound source was lessened as the early reflection arrived from an angle with greater separation from the direction of the sound source (i.e. further away from 0°). This effect was not observed at a statistically significant level for the other two sources, however a similar effect can be seen to occur.

The test methodology was considered appropriate as it did not require any prior experience and allowed multiple variables to be explored in a relatively short period. However, as the participants were unlikely to be familiar with the samples they were being asked to rate and the variables under investigation, it is possible that the use of the scale varied between participants.

It is acknowledged the use of single reflections and the use of a speaker in place of a musical instrument are highly artificial in comparison to the acoustic conditions experienced by a musician. This, in combination with the relatively low sample size ($N = 6$), means the results of this pilot experiment can only be evaluated tentatively. The results, however, do provide a positive indication that parametric decoding techniques were operating as expected and some initial suggestion that the type of musical phrase and the angle of arrival of a reflection can influence how acoustic conditions are perceived by a listener.

In order to gain further insight, it was considered that a further pilot listening test should be conducted which compares concert hall impulse responses with different spatial or temporal distributions of early reflections. This experiment is described in the following section.

7.5 Pilot test 2 - Spatio-temporal distribution of early reflections

The second pilot test aimed to obtain an initial indication as to whether musicians could distinguish different stage acoustic conditions that featured identical levels of ST_{early} but differing spatio-temporal distributions of early reflections. A listening test was devised where sounds would be auralised with different room impulse responses that featured varying spatio-temporal distributions of early reflections. The listener was required to report if they hear a difference between these auralisations.

It was expected that differences would be detected due to differences in perceived timbre, or some other mechanism related to the presence of early reflections i.e. spaciousness, perceived size of space. It was also expected that the differences would be more easily detected when the level of ST_{early} was high, overcoming the masking properties of the instrument.

As with the first pilot listening test, it was considered appropriate to use non-musician test participants that were exposed to similar acoustic conditions to a musician. Therefore, volunteers for the experiment were recruited from the department. Like the previous experiment, In order to produce similar masking characteristics experienced from a musical instrument, a loudspeaker was used as a proxy for a musical instrument. Due to the specific arrangement of apparatus and the background of the participants, the hypothesis below was worded to take this into account.

Can a listener discern differences between different shaped concert halls that exhibit identical levels of ST_{early} in the presence of a distractor?

- H_A: Listeners are able to discern differences between auralisations of concert halls of differing geometry but identical ST_{early}.
- H_0: Listeners can not consistently discern differences between the auralisations of concert halls of differing geometry but identical ST_{early}.

If the results of the pilot experiment were found to support the alternative hypothesis then it would provide an initial indication that the impression of stage acoustics is only partially related to early objective Support, ST_{early}. This would imply that additional approaches to stage acoustic measurement are required which take into account the temporal and spatial characteristics of the impulse response.

The test (described in detail below) is a Bernoulli trial meaning there is a probability of 0.5 that the test participant reached the correct answer by guessing. It is of interest to prove within a 95% confidence interval that the result was not achieved by guessing. Thirteen participants were recruited for this test which will result in 13 independent trials for each comparison. By using a binomial distribution look up table, the number of successes required such that the probability of not guessing is over 95% is 9.

Therefore for each trial:

- *H_A is true if the number of correct responses is greater than or equal to 9.*

7.5.1 Procedure

A discrimination test was deemed to be the most appropriate test to gain initial support for the alternative hypothesis (Bech and Zacharov, 2006). The test utilised the ABX paradigm which is similar to a two Alternative Forced Choice test (2-AFC). Participants were presented with a test stimulus A, a hidden reference stimulus B and an open reference stimulus X. Participants were asked to identify which stimulus (A or B) matched the reference X. In this case, the stimuli were sounds that were auralised in different concert hall spaces that featured different spatio-temporal reflection sequences. The stimuli in each comparison were rendered at identical levels of ST_{early}. For each trial, ST_{early} was set to either a *high* or *low* level.

The test conditions were similar to the previous experiment described in Section 7.4 where participants were seated in front of a forward-facing 'source' loudspeaker which emulated the direct sound of the musical instrument. In this case, the acoustic response was rendered using SIRR only. The experimental arrangement can be seen in Figure 7.1.

The acoustic response was rendered 'offline' i.e. the programme material was played through the loudspeaker and array (as opposed to the direct sound being measured using a microphone, convolved and rendered over the loudspeaker array). The listening test took place in the SoundLab and was supervised by the researcher who was seated behind a screen.

The stimuli consisted of concert hall impulse responses convolved with anechoic recordings of a musical instrument. The concert hall impulse responses were obtained using both geometrical acoustic models and measurements from real performance spaces (as described in 3). In all cases the source-receiver position was down-stage centre. Comparisons of the modelled and measured acoustic responses were kept 'like for like' (i.e. measured against measured etc). As the test was primarily concerned with early refections, the late acoustic response was kept constant for all comparisons while the early refections were variable.

Participants recorded their responses using a handheld touch screen interface (iPod touch) which relayed signals to the Max patch over a WIFI connection. The signals were programmed to trigger specific events in the patch which allowed the participant to skip between each question, activate each sample and record their responses. The participants responses were recorded by the patch and later imported into Excel for statistical analysis. The user interface (shown in Figure 7.9) was programmed such that the participant could only trigger a sample after the current sample had finished playing. This was to avoid rapid switching between samples which may have allowed participants to focus on specific sections of the musical phrase potentially exaggerating any differences between auralisations.

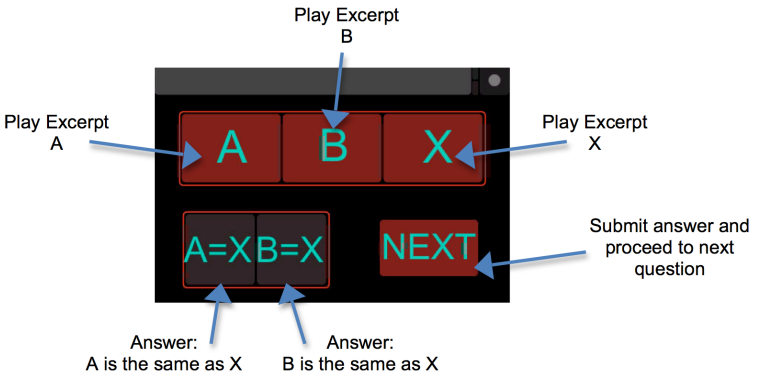


FIGURE 7.9: *Screenshot of user interface implemented on an iPod Touch using TouchOSC*

Figure 7.10 shows a system diagram which describes how the interface interacts with the Max patch. It can be seen that there are four convolver sections in the patch, where each section represents 16 individual convolvers (one per loudspeaker channel). One convolver section is responsible for rendering the late reverberation part of the impulse response and is controlled with a single fixed gain. The remaining convolver sections are responsible for rendering the

early reflections and run in parallel. The early convolvers are fed from a single audio player via a switch. Each convolver is controlled by some logic sections which load the correct impulse response into each convolver set and adjust the gain of each according to a prior calibration. This architecture was considered appropriate to allow fast switching between impulse responses with reduced likelihood of switching artefacts. The processed early and late sound are summed together and sent to the souncard via a loudspeaker equalisation phase (not shown).

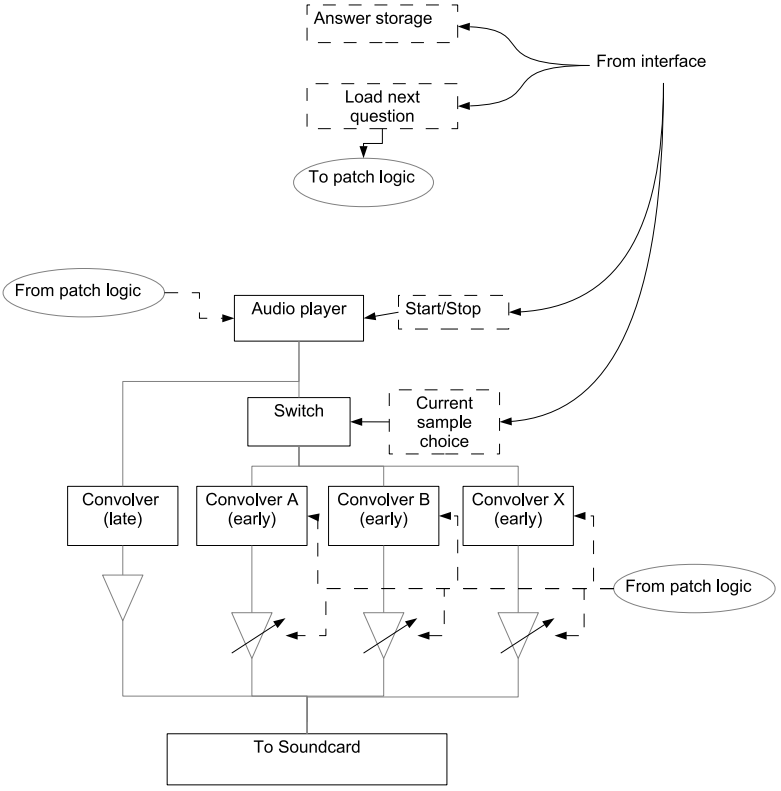


FIGURE 7.10: *Systems diagram of Max patch. The user interface was able to trigger the audio sample, switch between scenarios A, B or X and move onto the next question. When the next question was triggered, the patch would load the correct impulse responses, set the correct gains and audio samples and store the participants response.*

The dependent variable in this test was the correct or incorrect answer therefore classifying this approach as a Bernoulli trial. For this type of experiment the Binomial distribution was an appropriate method of testing response significance and calculating the number of required responses to support the alternative hypothesis. The independent variables in this test were the different stage enclosures, programme material and the level of ST_{early}.

Each participant compared six different concert hall geometries (3 modelled and 3 measured) at two levels of ST_{early} (low and high) and used two different musical samples (one featuring short phrasing and another featuring legato phrasing). Each hall was used as the open reference stimulus X for comparison with the other halls, resulting in 6 trials for a single sample of music and for a single level of ST_{early}. This resulted in a total of 48 trials meaning the average length of the test was approximately 40-45 minutes although participants were welcome to take as long

as they wished. It was not possible during this experiment to add a progress bar to the interface display, therefore the test was interrupted twice for every participant to inform them when they had completed half of the questions and when they had reached the final question. For each trial, the participant was allowed to listen to each sample as many times as they wished and they were free to proceed with the test at their own pace.

This combination of source attributes avoided any programme bias while ensuring the total number of questions does not result in listener fatigue. The test stimulus was randomly assigned to A or B to reduce any expectation bias and to ensure it is double-blind. During the test the participant was asked to face forward at all times but was not restrained in anyway allowing some minor head movements.

7.5.2 Participants

Thirteen participants were recruited for this test from the Digital Design Studio (6 female, 7 male). Participants were either staff or students with the majority of volunteers having at least 5 years of performance experience of some kind. For some participants, their involvement with music performance was extensive and ongoing whereas others had not performed for a number of years. The average ages of the participants was 31.25 years ($\sigma = 9.3$). Some of the volunteers were older than this and so expressed some uncertainty regarding their hearing ability at high frequencies. Otherwise all participants reported no hearing difficulties. The test subjects were asked to complete a single example question at the beginning of the test to familiarise themselves with the interface. Participants were rewarded for their time with a small amount of confectionary after the test had finished.

7.5.3 Stimuli

The acoustic responses required for the test were sourced from measured B-format impulse responses captured in three different concert halls; also the acoustic response of virtual spaces were generated using CATT Acoustic modelling software (Dalenback, No date). To reduce the degrees of freedom in this test it was required that the reverberant response (where $t > 100ms$) was kept constant for all comparisons and only the early reflections (where $20ms < t < 100ms$) were allowed to vary. This allowed the early reflections to be studied in isolation. The response used as the reverberant decay in all cases was measured in the Younger Hall (as discussed in Chapter 3). This concert hall has a T_{30} of 1.7 seconds at 1kHz.

The only differences between the acoustic models was the geometry of the stage enclosure, meaning that reflections will vary only in terms of their spatial and temporal distribution. Conversely, reflections in the measured impulse responses may feature different absorption or scattering characteristics owing to differences in their construction. Reflections in different halls are likely to comprise of differing frequency-content or temporal response (owing to differing amounts of scattering from each surface).

7.5.4 Programme material

The two musical phrases used in the test are shown below in Figure 7.11. The Bassoon (upper trace) has clear gaps between each note while the Flugelhorn (lower trace) plays with legato phrasing. These phrases were chosen as they represented two clearly different styles of phrasing (staccato and legato), were sufficiently different in timbre and were comfortable to listen to over a large number of repetitions. Both samples of music were recorded in free-field conditions.

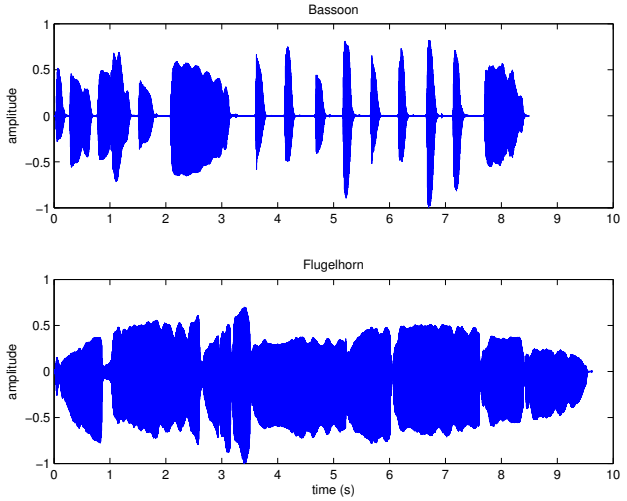


FIGURE 7.11: *Plot showing time-domain representation of signals used as stimuli for the listening test. The upper trace shows the Bassoon sample while the lower trace shows the Flugelhorn sample. Both samples are less than 10 seconds long.*

7.5.5 Measured impulse responses

The measured impulse responses were obtained during performance space surveys at the Ledger Recital Room (LRR), The Reid Concert (RH) and Younger Hall (YH) which have been reported previously in Chapter 4. The impulse responses were measured with an ST350 sound field microphone mounted directly over a Genelec 1029A loudspeaker. In all cases, the directional loudspeaker was facing directly out into the audience and was located down stage centre. Impulse responses were obtained by measuring a 10-second swept sine wave and convolving with the inverse sweep to ensure harmonic distortion was reduced and signal to noise ratio was maximised.

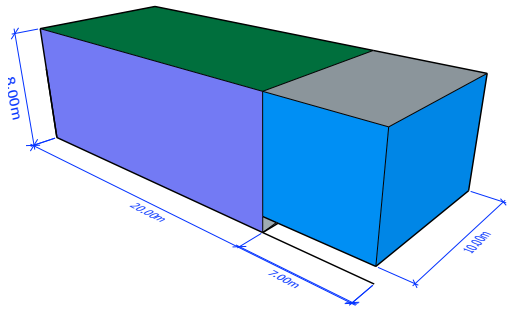
7.5.6 Modelled impulse responses

The modelled impulse responses were obtained from virtual spaces modelled in the acoustic modelling software CATT (Dalenback, No date). The initial geometry was created using the Rhino 3D-modelling software which was then imported into CATT using the DXF2GEO tool. Each model was run using 80,000 rays per octave with a truncation time of 3 seconds. In each model, the source was modelled as a Genelec 1029A loudspeaker with a coincident receiver where

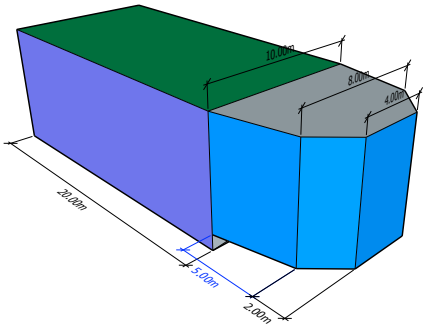
both source and receiver are facing the audience rear wall. The position of the receiver in all models was 2 metres from the very front of the stage and 1 metre off the centre line towards the stage right wall. The receiver height was 1.65m above the stage floor.

Figure 7.12 show standard views of each modelled concert hall. All concert halls were modelled as 27 metres in length, an audience area width of 10 metres and 8 metres in height. The audience area is rectangular in shape with a flat floor.

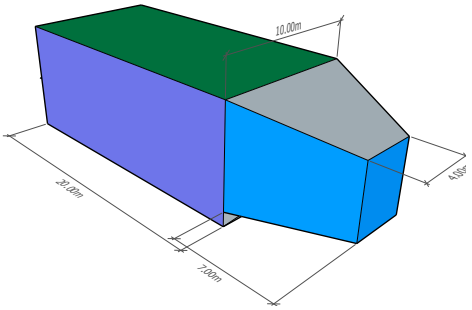
The stage in every modelled hall was elevated above the audience floor by a height of 1 metre and all stages had a maximum depth of 7 metres. The screenshots below show the varying geometries used in each hall. In Hall 1, the stage walls are completely parallel and are the same width and height as the audience area. The stage has an area of $70m^2$ and a constant height of 7 metres. In Hall 2, the side walls are angled inward in two steps where the very rear parts of the hall have a relatively steep angle to the frontal parts of the side walls. The concert hall stage has an area of $57m^2$ and the ceiling is at a constant height of 7 metres above the stage. In Hall 3, the side walls are angled more steeply in a single phase on each side. In addition, the ceiling is angled towards the audience as is sometimes observed in concert halls to direct sound towards the audience. At the front of the stage the ceiling is 7 metres above the stage whereas at the rear it is 5 metres above the stage. The stage area in this hall is $49m^2$.



(a) Hall 1 geometry



(b) Hall 2 geometry



(c) Hall 3 geometry

FIGURE 7.12: *Simple models of the concert halls used in this listening test. Models were created in Rhino and imported into CATT to perform acoustic modelling. Colours represent surfaces with common absorption characteristics.*

The surfaces in each model are displayed in different colours representing surfaces with common absorption characteristics. The absorption coefficients used in each model are shown in Table

7.4. The stage floor was modelled as a wooden floor mounted on joists. The stage walls and ceiling were modelled as 1-inch wooden panelling with an air space behind. The audience walls and ceiling were modelled as 9-mm plasterboard panels on battens mounted on top of an 18mm airspace packed with glass fibre. The audience floor was modelled with absorption characteristics approximating a seated audience. All surfaces were modelled with 10% diffusion characteristics to ensure that the modelled impulse response contained mainly specular reflections.

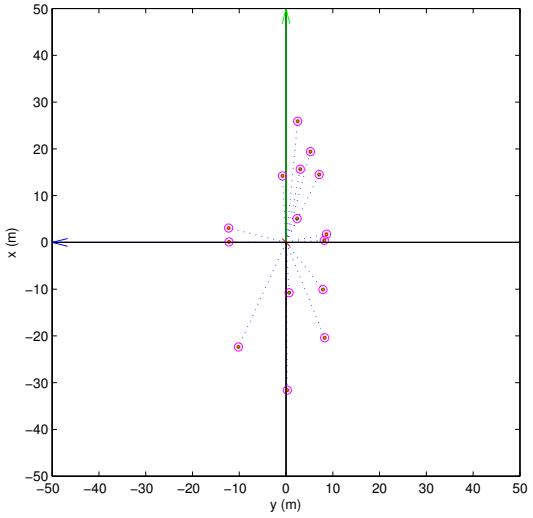
Surface	Octave band centre frequency (Hz)					
	125	250	500	1000	2000	4000
Stage Floor	0.15	0.11	0.1	0.07	0.06	0.07
Audience area	0.39	0.52	0.8	0.94	0.92	0.85
Stage walls and ceiling	0.19	0.14	0.09	0.06	0.06	0.05
Audience walls and ceiling	0.3	0.2	0.15	0.05	0.05	0.05

TABLE 7.4: *Absorption coefficients of surfaces used in the modelled concert halls*

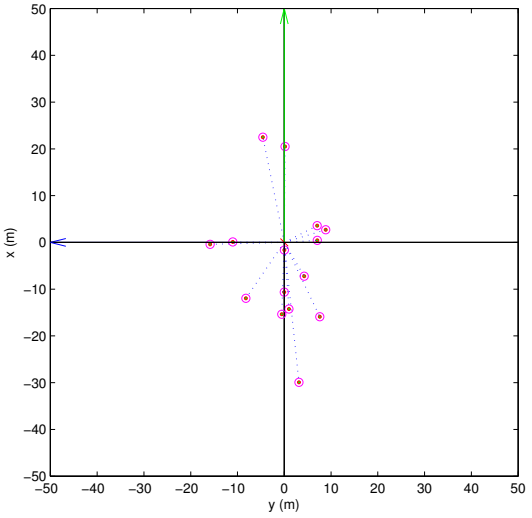
7.5.7 Objective analysis of impulse responses

The impulse responses used in this experiment were analysed to determine the extent to which they differed. The analysis extracted temporal and spatial parameters from an image source plot which was obtained as described in Chapter 4.

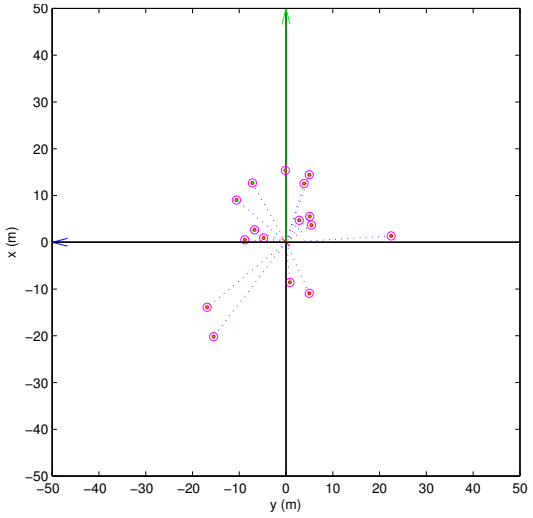
The image source plots for each impulse response are shown in Figure 7.13. All plots are viewed in plan, where the x-axis increases towards the audience rear wall (as shown by the blue arrow) and the y-axis decreases towards stage left (as shown by the green arrow). Each image source is displayed as a point with a line connecting it to the origin. The distance from the origin represents the time of arrival and the position relative to the origin indicates angle of arrival. Any reflections that arrive before 20ms have been omitted from the analysis as, in these cases, the reflections arriving in this time frame are from the floor.



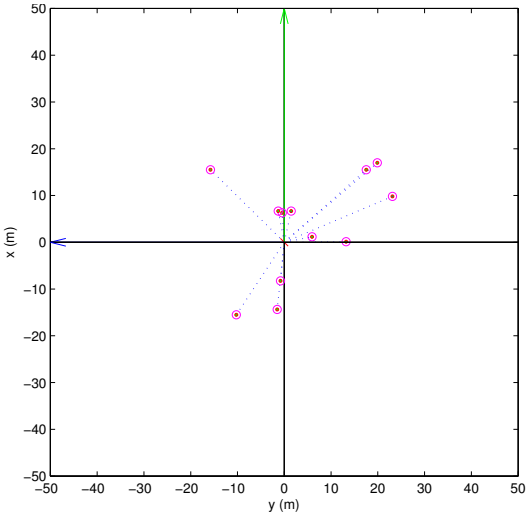
(a) Hall 1



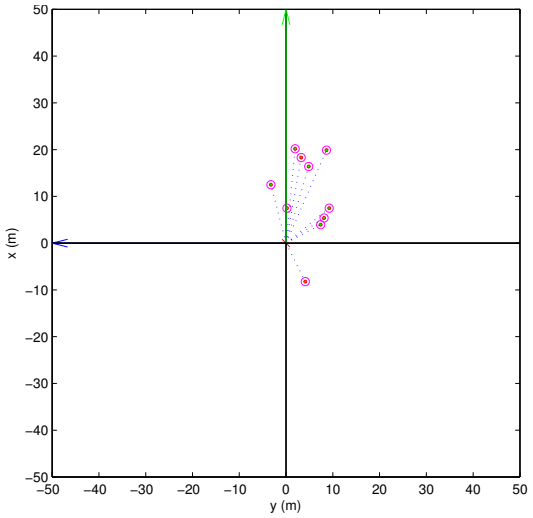
(b) Hall 2



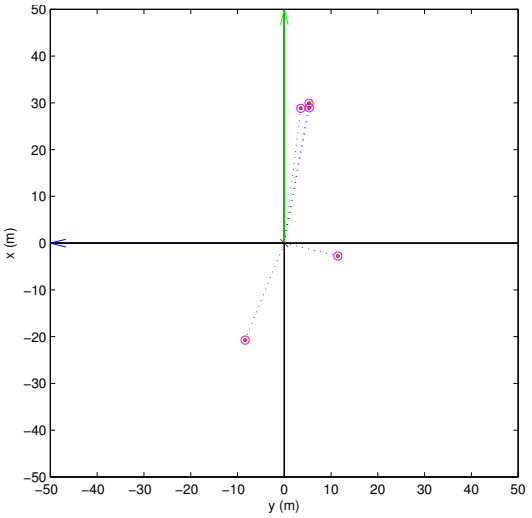
(c) Hall 3



(d) Hall 4 (LRR)



(e) Hall 5 (RH)



(f) Hall 6 (YH)

FIGURE 7.13: *Image source plots of each tested impulse response obtained with SIRR analysis techniques. Image source plots are viewed in plan with the top of the plot (green arrow) pointing to the audience rear wall*

The temporal and spatial parameters for each impulse response are shown in Table 7.5. In reference to the temporal parameters it can be seen that the average time of arrival (t_{mean}) is very similar for all halls (they fall within a range of 8ms) with the exception of Hall 6 (Younger Hall). As can be seen in Figure 7.13 the detected reflections arrive comparatively late in the impulse response from in front of the measurement position. Similarly, the standard deviation of time of arrival (t_σ) shows only small differences in the overall temporal spread of early reflections (they all fall within a range of 4ms). It can be seen however that the spatial parameters vary far more considerably with a wide range of spread values and mean resultant vector directions.

Hall	t_{mean} (s)	t_σ(s)	θ_{mean} (°)	ϕ_{mean}(°)	*Spread*
Hall 1	0.048	0.021	52.84	-1.03	0.156
Hall 2	0.046	0.021	158.82	-20.45	0.269
Hall 3	0.045	0.021	-14.26	-54.36	0.349
Hall 4 (LH)	0.048	0.023	52.8	26.72	0.280
Hall 5 (RH)	0.053	0.020	21.11	-2.14	0.690
Hall 6 (YH)	0.076	0.024	30.61	0.91	0.446

TABLE 7.5: *Table containing temporal and spatial parameters of the impulse response used in the pilot test*

To illustrate the objective differences between these impulse responses, the results from Table 7.5 are shown in Figures 7.14, 7.15 and 7.16.

Figure 7.14 shows the azimuth and elevation of the mean resultant vector for each impulse response. It can be seen that Hall 1, the Reid Hall (RH) and the Younger Hall(YH) are clustered together where the mean direction of arrival of early reflections appears between 21° and 53° azimuth, close to the lateral plane. The remaining halls feature reflections with more elevated reflections. It can be seen that in Hall 3, many of the reflections occur at a low angle of elevation whereas in the Ledger Hall many of the reflections are elevated towards 27°. Hall 2 appears to feature more reflections occurring from behind the measurement position.

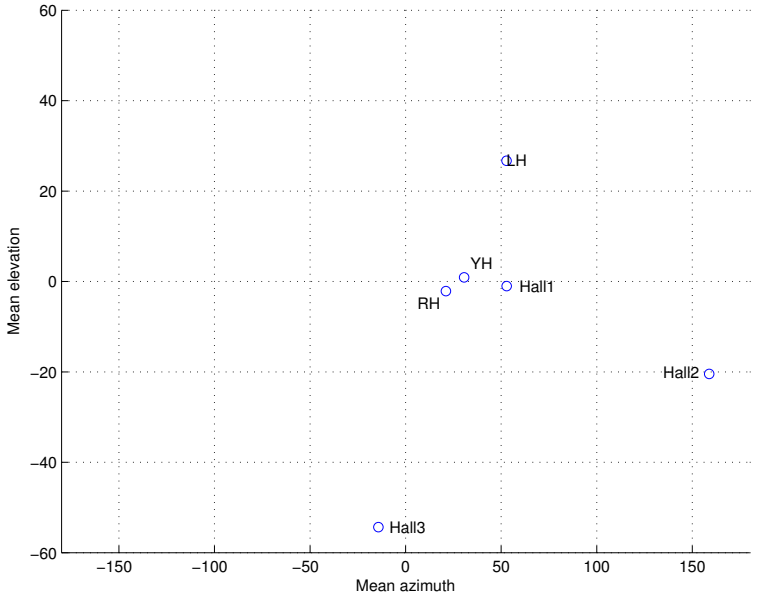


FIGURE 7.14: *Plot of the azimuth and elevation of the mean resultant vector obtained for each impulse response. Each impulse response is represented as a dot with an accompanying label.*

Figure 7.15 shows the average time of arrival versus the spatial spread of early reflections. It can be seen that there is a wide range of spatial spread values in the selected stimuli varying between 0.156 and 0.690. The spatial spread values indicate how spatially clustered together the reflections are where a value of 1 indicates all reflections arrive from the same direction and 0 indicates they are spread equally. Most of the impulse responses appear to have similar average time of arrivals varying between 45ms and 53ms. However, the Younger Hall shows a much later average time of arrival at $76ms$.

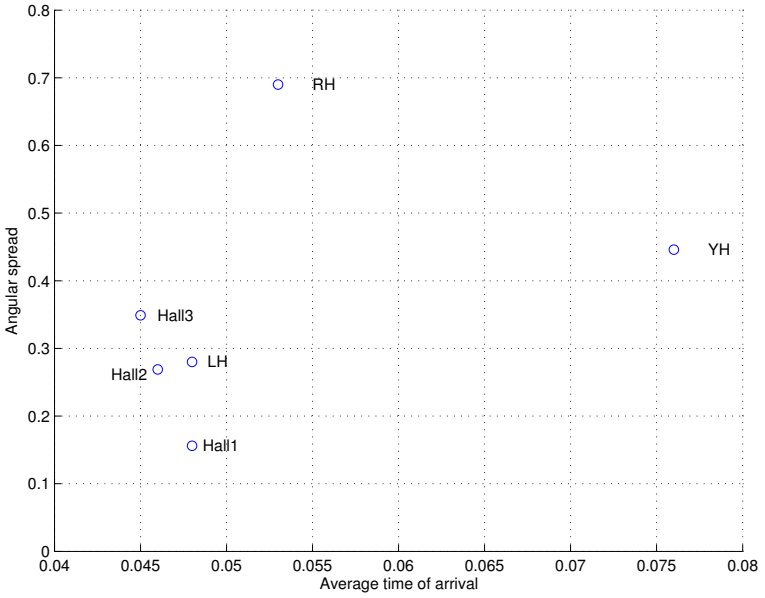


FIGURE 7.15: *Plot of the average time of arrival and spatial spread of early reflections in each impulse response. Each impulse response is represented as a dot with an accompanying label.*

Figure 7.16 shows the same halls in relation to the temporal parameters i.e. the average and standard deviation time of arrival. It can be seen that most of the impulse responses feature very similar temporal properties. As discussed previously, the Younger Hall (YH) is the exception where the average time of arrival is much later than the other impulse responses. This can be observed visually in Figure 7.13.

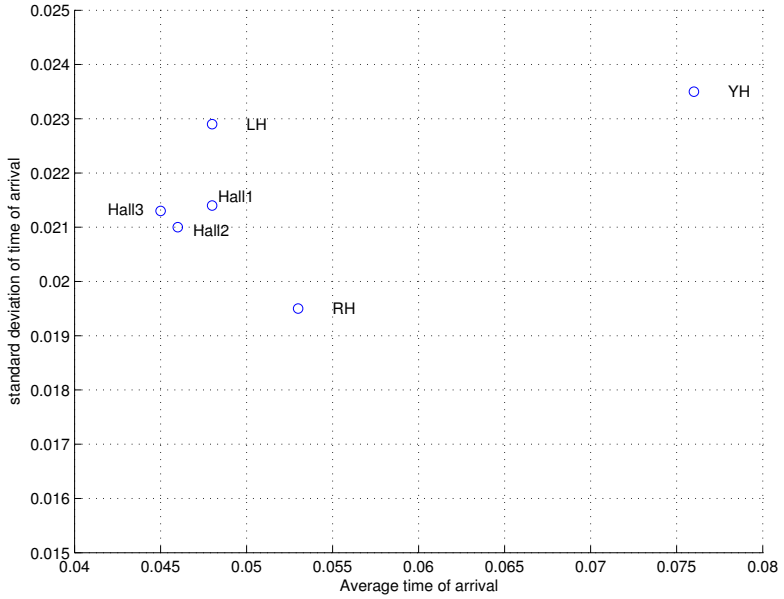


FIGURE 7.16: *Plot of the average time of arrival and standard deviation time of arrival for early reflections in each impulse response. Each impulse response is represented as a dot with an accompanying label.*

This analysis demonstrates that the impulse responses vary most in terms of their spatial properties, specifically spatial spread and mean direction of arrival. There is less variation in the temporal parameters with the exception of the Younger Hall where the reflections occur much later than in other impulse responses.

7.5.8 Stimulus calibration

Each auralisation was calibrated in terms of latency and the level of each part of the impulse response. This was achieved by measuring impulse responses in the sweetspot of the loudspeaker array with an Ambisonic microphone.

The Max patch was set to operate with an i/o buffer of 256 samples long which resulted in a 6ms delay between the expected time of arrival of a single reflection and the measured time of arrival. The latency was significantly less than measured in Chapter 7 as the signal is being rendered directly from the patch as a passive auralisation. This means the delay caused by acoustic propagation between source and microphone, and also that introduced by the input buffer, are eliminated.

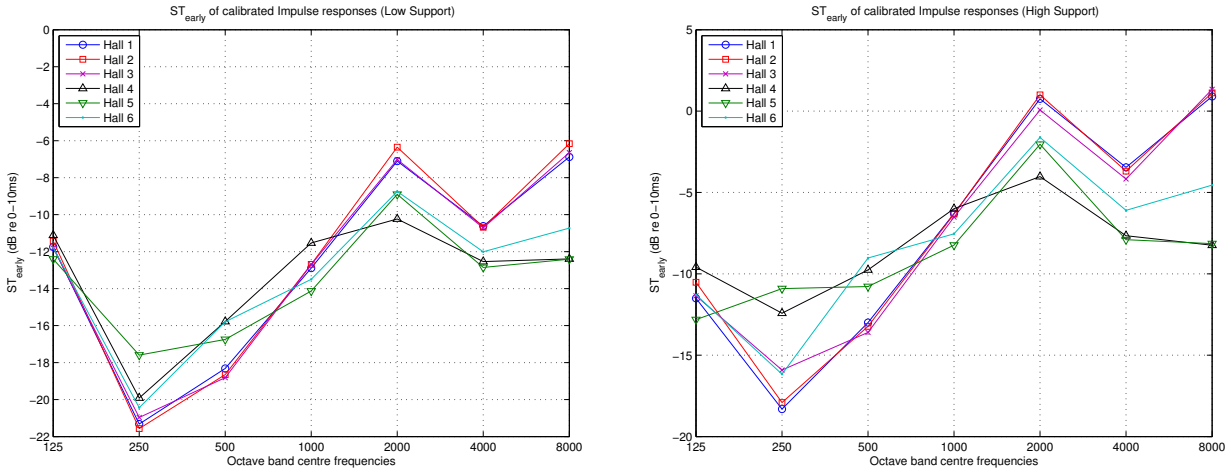
To reduce this measured latency to acceptable magnitudes, the buffer objects were set so that the initial pointer was set to a time delay equal to the measured latency, ensuring that early reflections would arrive at the correct time. The i/o buffer size was chosen so that the system could process the impulse responses reliably and without artefacts but efficiently enough so that the reflections could feasibly arrive at the correct time.

The impulse response level was calibrated using ST_{early} measurements where an average value between 250Hz and 2kHz was set to equal a value of either -8dB (for a high support value) and -14dB (for a low support value). These values were deemed appropriate as an early support value of -8dB appears to be the highest available in the literature (as summarised in Table 2.1). These values also represented a difference of approximately 3 JNDs based on informal feedback from the parameters creator (1 x JND is thought to be approximately equal to 2dB (Hak et al., 2012)). The late reverberation was calibrated in a similar manner to a ST_{late} value of approximately -14dB. The measured ST_{early} for each hall and for low and high support settings is shown in Table 7.6. The level of each impulse response was adjusted so that the measured ST_{early} was within 0.1dB of the target values. The calibration process resulted in a gain value for each impulse response that would ensure it rendered the sound field at the correct level. These gains were saved and recalled by the patch when required by each question.

Hall	High - ST_{early} (dB) (250Hz-2kHz)	Low - ST_{early} (dB) (250Hz-2kHz)
Hall 1	-8.06	-14.03
Hall 2	-8.03	-13.99
Hall 3	-8.02	-14.04
Hall 4	-7.97	-13.99
Hall 5	-7.97	-14.04
Hall 6	-8.09	-14.09

TABLE 7.6: *ST_{early} measured in each auralised concert hall after calibration. Values show the mean ST_{early} measured in octave bands between 250Hz and 2kHz.*

Figure 7.17 shows the ST_{early} of each set of early reflections in octave bands after they had been calibrated. It can be seen that the modelled halls (Halls 1, 2 and 3) appear to have more high frequency energy than the measured halls (Halls 4, 5 and 6). This is likely to be caused by the high frequency emphasis often observed in geometrically modelled impulse responses.



(a) Low ST_{early}

(b) High ST_{early}

FIGURE 7.17: *ST_{early} viewed in octave frequency bands for each concert hall after calibration had been implemented. Solid lines show the measured concert halls while dashed lines show modelled concert halls. The upper plot shows the low ST_{early} while the lower plot shows the high value of ST_{early}.*

7.5.9 Results

The responses from test participants were collected as nominal data i.e. number of correct answers. Therefore it was most appropriate to summarise the test results by the number of participants that correctly identified the test stimulus in each question.

In each trial there was an equal probability that the responses were based on guessing or by correct identification. Therefore, in order to support the alternative hypothesis it was of interest to demonstrate that the results were not as a result of guessing. The cumulative binomial distribution was used to calculate exact probabilities for observed detection rates for a known number of trials. It was therefore possible to determine the number of correct responses required out of a total number of trials in order that the result be considered statistically significant and not likely to be as a result of guessing. This is described by equation (7.2) which gives the probability of obtaining at most x successes in n independent trials where each trial has a probability, p, of success (Vetrivel, No date), where $r = 0, 1, 2..., n$.

$$P(X \leqslant x) = \sum_{r=0}^{x} C_r^n p^r (1-p)^{n-r} \tag{7.2}$$

Where C_r^n is the binomial coefficient:

$$C_r^n = \frac{n!}{r!(n-r)!} \tag{7.3}$$

For a Bernoulli trial, the probability that x number of a total of n trials was obtained successfully (rather than by guessing) is 0.5. As there were 13 participants who took part in the test, there

must be at least 9 participants that answer correctly for each test condition to be considered statistically significant ($p \leqslant 0.05$).

Table 7.7 shows the results for comparisons of the modelled concert halls for low and high ST_{early} levels and for the Bassoon and Flugelhorn samples. It can be seen that most participants could correctly distinguish between Hall 1 and Hall 3 when the ST_{early} level was set high. Furthermore, participants could distinguish hall 2 and hall 3 when ST_{early} was set low, however this only appears to be the case for the Flugelhorn and not the Bassoon. Table 7.8 shows similar results for the measured concert halls. It can be seen that Hall 5 and Hall 6 could be distinguished at high levels of ST_{early} but only for the Flugelhorn and not the Bassoon.

Bassoon			
Hall X	AB= (1,2) or (2,1)	AB= (1,3) or (3,1)	AB= (2,3) or (3,2)
High Support			
Hall 1	8	(11)	-
Hall 2	8	-	3
Hall 3	-	(10)	7
Low Support			
Hall 1	6	6	-
Hall 2	6	-	8
Hall 3	-	7	6

Flugelhorn			
Hall X	AB= (1,2) or (2,1)	AB= (1,3) or (3,1)	AB= (2,3) or (3,2)
High Support			
Hall 1	8	(9)	-
Hall 2	6	-	6
Hall 3	-	(9)	6
Low Support			
Hall 1	6	8	-
Hall 2	(9)	-	(10)
Hall 3	-	4	6

TABLE 7.7: *Number of correct responses for modelled concert halls energised with both sound sources. Circled values indicate statistically significant number of detections where* $p \leqslant 0.05$

In order to gain further insight, the participant responses were collapsed across each level of the experiment (ST_{early}, Musical sample and Modelled/Measured Halls) to evaluate the results over a higher sample of responses. Table 7.9 shows the number of correct responses for both instrument types, Bassoon and Flugelhorn and the associated p-value. It can be seen that the detection rate is very similar for both instrument types which suggests the musical sample did not significantly affect the results.

	Bassoon		
Hall X	AB= (4,5) or (5,4)	AB= (4,6) or (6,4)	AB= (5,6) or (6,5)
	High ST_{early}		
Hall 4 (LH)	6	8	-
Hall 5 (RH)	8	-	6
Hall 6 (YH)	-	8	8
	Low ST_{early}		
Hall 4 (LH)	5	3	-
Hall 5 (RH)	8	-	4
Hall 6 (YH)	-	6	7

	Flugelhorn		
Hall X	AB= (4,5) or (5,4)	AB= (4,6) or (6,4)	AB= (5,6) or (6,5)
	High ST_{early}		
Hall 4 (LH)	6	8	-
Hall 5 (RH)	(9)	-	(10)
Hall 6 (YH)	-	4	6
	Low ST_{early}		
Hall 4 (LH)	8	5	-
Hall 5 (RH)	6	-	8
Hall 6 (YH)	-	5	5

TABLE 7.8: *Number of correct responses for measured concert halls energised with both sound sources. Circled values indicate statistically significant number of detections where* $p \leqslant 0.05$

Source	N correct	N total	$p - value$
Bassoon	163	312	0.198
Flugelhorn	164	312	0.168

TABLE 7.9: *Number of correct responses for Bassoon and Flugelhorn where all other levels have been collapsed*

Table 7.10 shows the number of correct responses for both levels of ST_{early}, High and Low. It can be seen that halls rendered at high levels of ST_{early} could be detected to a statistically significant level ($p \leqslant 0.05$), however this was not the case for halls rendered at the low level of ST_{early}. This suggests that participants may have been guessing at the low level of ST_{early}.

ST_{early}	N correct	N total	$p - value$
High	178	312	0.005
Low	149	312	0.769

TABLE 7.10: *Number of correct responses for concert halls rendered at High and Low levels of* ST_{early}*, where all other levels have been collapsed*

Table 7.11 shows the number of correct responses for Modelled and Measured concert halls. It can be seen that the number of correct responses for Modelled halls is higher than for Measured

halls and shows a statistically significant detection rate ($p \leqslant 0.05$). This suggests that the differences were more apparent for Modelled concert halls than Measured concert halls.

Hall type	N correct	N total	$p - value$
Measured	157	312	0.43
Modelled	170	312	0.05

TABLE 7.11: *Number of correct responses for Modelled and Measured concert halls where all other levels have been collapsed*

7.5.10 Discussion

In summary, the results suggest that it was easier for participants to distinguish differences in the shape of the stage enclosure with higher ST_{early} levels in comparison to the low ST_{early} levels. While there were some significant number of detections for both musical samples, when the results of the whole panel were evaluated, it can be seen that the number of correct responses for each instrument were very similar, which implies the musical sample had little effect on how easy the differences were to detect. The results also show that the different shapes of stage enclosure were easier to detect when the musical phrases were auralised using computer models rather than using measured data.

The results suggest that the difference between Hall 1 and Hall 3 was relatively easy to detect at high levels of ST_{early}. From the objective analysis performed on each impulse response it can be seen that these halls differ mainly in terms of spatial parameters θ_{mean}, ϕ_{mean} and spread. This implies that the differences heard are mainly due to spatial effects.

There were fewer statistically significant results for the auralisations constructed with measured impulse responses rather than modelled impulse responses. It is feasible that these results were due to assumptions of the acoustic model (i.e. very low scattering coefficient applied to the walls or high frequency emphasis from geometrical techniques) or due to shortcomings in the recording and decoding technique applied to the measured data (i.e. bin based processing coupled with a maximum accurate frequency of the Soundfield microphone).

It was demonstrated in previous chapters that, when measured with a directional sound source, reflections appear more readily in the direction in which the sound source was facing. As the sound source was facing out into the audience in all cases, it is clear that there will be a reduction in the influence of reflections occurring from the stage enclosure. Measuring or modelling each space with an omnidirectional loudspeaker may have further highlighted the differences between impulse responses. However, it is uncertain if this is more or less representative of a realistic situation of a musician playing on stage.

When the number of correct responses was collapsed across ST_{early}, it was found that the shape of a stage enclosure was easier to hear at high levels of ST_{early} to a statistically significant level. There is a reasonable likelihood that participants guessed for halls rendered at the low level of ST_{early}.

The high setting of $ST_{early} = -8dB$ was higher than many measurements made in performance spaces which may have made the differences more apparent than in typical concert halls. The lower setting of -14dB was also comparable to the threshold of detection of early reflections for some instrument families (Gade, 1989), therefore it is unsurprising that fewer differences were detected at this level. The results must therefore be interpreted with great care.

Informal feedback from the test participants after each test provided some valuable insights. The majority of test participants found the test very challenging with many listening to each test stimulus a large number of times before making a decision. This caused many of the participants to feel moderately fatigued by the end of the test as their eagerness to get the correct answer elongated the duration of the test. Due to the subtlety of some of the comparisons, this may have contributed to some minor frustration from participants.

Participants remarked that differences between some stimuli were more obvious than others and most admitted to guessing many of the trials. Some participants remarked they had heard slight changes in timbre during the legato Flugelhorn passage with one participant remarking that they had thought the instrument had been recorded at a closer distance. A number of the participants specifically remarked that the width or size of the space had changed in a number of questions which seemed more obvious during staccato phrasing from the Bassoon. Some participants made reference to a feeling that *'something had changed'* although they were not always certain what and a few made references to aspects they referred to as *'spaciness'*. One participant remarked that they felt the reverberation time had changed slightly over the course of a number of questions. This is particularly interesting as there has been speculation that the ST_{early} (along with EDT) is linked with a musician's sense of reverberation time. Another participant remarked that some of the comparisons felt as if they had been listening to recordings made in a practice room or in a concert hall. This is also of particular interest as the higher setting for ST_{early} is more likely to be found in smaller practice rooms rather than in concert halls and so may have contributed to this effect. Another participant made reference to slight changes in note length or pitch which could indicate subtle changes were detected. Many participants remarked that they found certain parts of each phrase easier to judge differences, specifically the spaced out notes of the bassoon phrase.

The test procedure was found to be appropriate as a pilot methodology as it enabled the researcher to investigate numerous variables related to the perception of stage acoustic conditions while taking advantage of an increased available sample size. It is clear however that while the use of naive listeners allowed an increased sample size (due to easily available participants) and greater control over the stimuli, the results cannot be used to infer any aspects of a larger population as the sample is not wholly representative of the population. As the test participants were not engaged in playing their instrument as well as listening to the stimuli it could be argued that they may have an unfair advantage and may have been able to pick out certain details of the acoustic response that musicians may not be aware of. However, it could also be asserted that some of the participants may not have a similar depth of training and so may not be as sensitive to details of the acoustic response.

In summary, the results and feedback from participants did not provide sufficient evidence to support the null hypothesis that the impression of stage enclosure acoustics is independent of

the spatio-temporal distribution of early reflections. The results, in conjunction with informal feedback from the participants, imply that the spatio-temporal distribution of early reflections may be audible for musician test subjects. As discussed, the subjective responses were gathered from participants with varying abilities and experiences and so cannot be used to infer a pattern for the larger musician population. Therefore, it was determined that additional tests with experienced musicians were required in order to gain additional insight.

7.6 Summary and discussion

This chapter has described two pilot tests which aimed to help refine the main experimental work and to test various aspects of the auralisation system during its development.

Both tests were conducted as passive auralisations where participants were not required to operate a musical instrument. Instead, participants were exposed to acoustic conditions that were considered similar to those experienced by a musician, where the direct sound was emulated using a loudspeaker placed near the participant. This sound was auralised with acoustic responses that were rendered over the loudspeaker array, to emulate the acoustic response of the hall. This approach allowed participants to be recruited from the department where they were more readily available. In addition, the experiment methodology allowed the stimulus (i.e. the direct sound) to be held constant for every trial, something which is difficult to guarantee in interactive listening tests. This allowed the participants to focus entirely on the acoustic response.

The first pilot test focussed primarily on the rendering technique used to auralise both early reflections and the entire acoustic response. In addition, the experiment aimed to determine if aspects such as musical phrasing and angle of arrival of reflections produces audible differences in what the musician heard. The results indicated that an auralisation rendered using FOA and SIRR were perceived similarly in the context of stage acoustic laboratory experiments. In addition, the results suggested that the audibility of an early reflection can be affected by the type of musical phrasing. In this case, early reflections were more audible when the sound source was a clarinet playing staccato phrasing, suggesting that gaps of silence between notes provided an opportunity to detect the reflection.It was also shown that early reflections were more easily heard when arriving from certain angles, implying that the spatial distribution of early reflections could influence the musician's impression of the venue acoustics.

The second pilot test was more focussed on the main hypothesis of this research. Specifically, this test aimed to determine if participants could discern differences between different shaped concert halls that exhibit identical levels of ST_{early} in the presence of a distractor; the distractor refers to the direct sound emulated by the loudspeaker. In this experiment, participants compared sounds auralised in concert halls that had different stage geometries which caused differences in spatial or temporal distributions of early reflections. The concert halls were auralised from computer models or measured data and were presented to the participants at high and low levels of ST_{early}.

The results suggest that the shape of the stage enclosure could be detected to a statistically significant level ($p \leqslant 0.05$) when ST_{early} was set high, however this was not the case at low

levels of ST_{early}. It was also found that participant could detect differences more easily when the comparison was made between modelled concert halls. It is possible that differences between modelled concert halls were more audible to the participants due to the assumptions of how sound propagates within a geometric acoustic model. It is also possible that shortcomings in the recording and decoding technique applied to the measured data caused certain differences to be made less obvious. The Soundfield microphone, for instance, can accurately resolve the direction of sound up to approximately 5kHz (as demonstrated in Appendix D). Additionally, the time-frequency resolution of the STFT used in the parametric decoder may have been insufficient to faithfully render aspects of certain reflections.

In this experiment it was found that the musical sample did not significantly affect detection rates, which suggests that perceived changes in stage enclosure shape are not dependent on the type of musical phrase. This appears to contradict the results of the first pilot test which found the audible effect of a single reflection was more apparent when staccato phrasing was used. It is feasible that these results are a consequence of different experimental methodology or some other attribute of the musical samples used.

Overall, the results of the pilot tests have indicated that participants experiencing similar acoustic conditions to a performing musician could discern spatial or temporal variations in early reflections. These changes were experienced by some participants as a change in room size or a sense of spaciousness. It is acknowledged that, due to the experiment methodology, these results cannot be used to infer anything about the larger musician population. Therefore, it was determined that additional tests with experienced musicians were required in order to gain additional insight. These tests should allow each musician to play into the space freely and to listen back to the acoustic response therefore the auralisation system should be adapted for this purpose. The interactive auralisation system is described in the following chapter.

Chapter 8

Interactive auralisation system

Chapter 6 compared the various options for interactively auralising the stage acoustic conditions in response to the sound from a musician's instrument. In addition, a stage acoustic auralisation system was proposed which was based on Ambisonic capture of room impulse responses decoded over a loudspeaker array using a parametric decoding scheme. It was considered that the use of parametric decoding techniques could provide a perceptually enhanced auralisation in comparison to first-order Ambisonic measurements. This was deemed necessary for stage acoustic experiments where one of the key variables is the spatial distribution of early reflections. In addition, it was considered that parametric processing methods could provide some powerful soundfield manipulation techniques which could be used for future research. In Chapter 7, parametric decoding was used in two pilot tests which aimed to develop the auralisation system. The results suggested that parametric decoding was perceived similarly to Ambisonics. Given the potential advantages of parametric decoding it was decided to use this for the main interactive listening tests for musician.

This chapter will describe the final stage acoustic auralisation system in detail, including the various signal processing techniques used to render the stage acoustic impulse response. This will include an implementation of the parametric decoding techniques described in the previous chapter. In addition, it will characterise the performance of the auralisation system and laboratory via a series of measurements. This chapter will also describe the calibration procedure used to ensure that the auralisation was rendered accurately. The results of a number of objective tests will demonstrate the accuracy of the final system.

8.1 Auralisation system

There are several signal processing stages required to render an interactive auralisation for stage acoustic auralisation as shown in Figure 8.1. The sound generated by the musician is captured by a nearby microphone which is then convolved with two sets of impulse responses which contain the early reflections and late reverberation respectively. The resultant signals are summed

and rendered over an equalised loudspeaker array. The early and late convolvers consist of 16 individual convolution operations for each loudspeaker of the SoundLab.

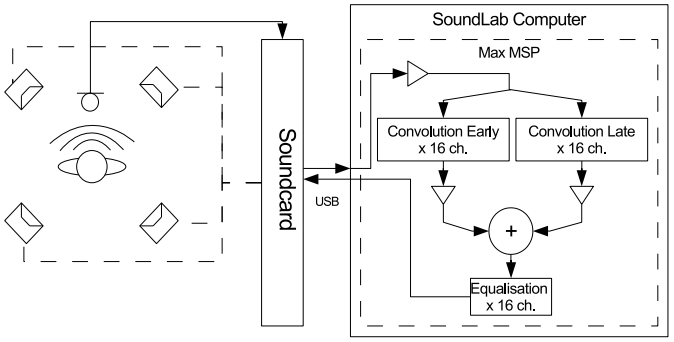


FIGURE 8.1: *System diagram of interactive auralisation system. Diagram shows only four speakers for clarity.*

It can be seen that the main parts of the system consist of the sound source capture, real-time convolution with the room impulse response and rendering over an equalised loudspeaker array. Each of these aspects will be discussed individually in the following sections.

8.1.1 Sound source capture

The position of the microphone relative to the musician is shown in Figure 8.2 where a musician can be seen seated with the microphone positioned directly in front. The picture also shows an iPad interface allowing the musician to control the auralisation and a music stand. The microphone is a Behringer ECM8000 omnidirectional measurement microphone which is positioned 30cm directly in front of the musician's instrument at a height of 1m above the floor.

As discussed in Chapter 6, it was acknowledged that the effect of unwanted feedback may have been further reduced by using a directional microphone with appropriate spectral corrections for low frequency boost (caused by the proximity effect). However it was considered that an omnidirectional microphone was more capable of faithfully capturing the direct sound from the instrument with minimal timbral artefacts.

Chapter 6 also discussed the use of an array of microphones to capture the directional radiation characteristics of the instrument or to capture a more complete timbre from the musical instrument. However, in this experiment, it was considered that this added additional complexity and processing load to the system, making it less reliable for listening tests. The use of an array of microphones could assist in enhancing interactive auralisation systems and is therefore recommended for future research.

8.1.2 Digital signal processing

The signal from the microphone is routed, via the soundcard and internal preamplifier, to Max MSP. Real-time convolution was achieved using the HISSTOOLS Max MSP external objects

FIGURE 8.2: *Photo showing the position of the microphone relative to a musician playing in the interactive auralisation system. The microphone is positioned directly in front of the musician at a radius of 30cm and a height of 1m from the floor. Also positioned around the musician are an iPad interface and a music stand.*

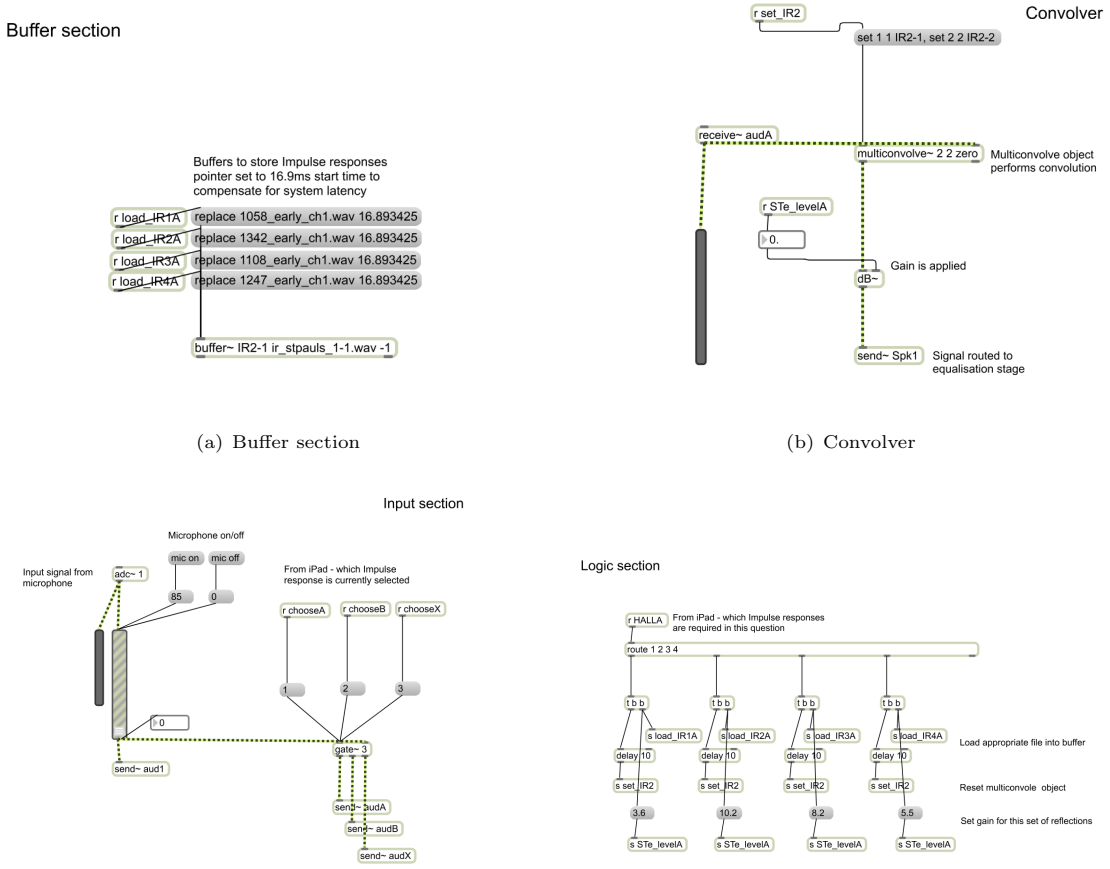
(Harker and Tremblay, 2012). The Max MSP patch loads 16 impulse responses per stage which have been decoded 'offline' by a Matlab script as described in Section 8.1.4.

The Max MSP patch was built to accommodate four sets of real-time convolvers, each with 16 channels each (one per loudspeaker) which ran in parallel. One convolver set was used for the late reverberation. The remaining three convolver sets were used to auralise early reflections. When the user selected one particular hall, the input signal would be routed to the associated convolver set. By doing so, the likelihood of audible switching artefacts was reduced. Three convolver sets were chosen so that the user could be presented with three halls per question. When the user selected the next question, the patch would retrieve the appropriate impulse responses for each convolver set and adjust the gain of each to a level based on previous calibration. A single convolver set was used for the reverberation as it was decided to keep this constant throughout the listening tests (this will be discussed further in Chapter 9).

Figure 8.3(a) shows a screen grab of each part of the Max MSP patch. For clarity, this example shows the basic processing required for a single channel of audio. There are four distinct sections of the patch. **The input section** (8.3(c)) shows the audio arriving on the object [*adc~*1] which is sent through a gain control and a gate. In the listening tests, the musician was presented with a maximum of three separate stages in each trial. They were asked to perform on each one and to answer questions based on what they had experienced.

The gate was used to route the input signal to the appropriate convolver set. **The logic section** (8.3(d)) ensures that the correct impulse responses are loaded for each convolver set and that

the gains are set so that the auralisation is rendered at the calibrated level. The logic section receives data regarding which hall should be loaded into each decoder and instructs the associated objects to load the relevant data. **The buffer section** (8.3(a)) loads individual audio files into a uniquely named buffer. The message to the buffer object includes the name of the audio file followed by a pointer time (in milliseconds) which denotes the start point of the buffer. The buffer identifier is used by the convolver (shown in the top right) to identify which impulse response the input signal should be convolved with. The input audio signal is routed to the **convolver section** (8.3(b)). Once the signal has been processed, it passes through a gain stage and is then sent through the equalisation section.



(a) Buffer section

(b) Convolver

(c) Input section

(d) Logic section

FIGURE 8.3: *Screenshot showing the four main sections of the auralisation patch, shown for a single channel only for clarity. This consists of the input section (8.3(c)) which determines which routes the input signal to specific convolvers; the logic section (8.3(d)) and buffer sections 8.3(a), which load specific IR files into a buffer; and the convolver section (8.3(b)), which convolves the input signal with the IR and routes to the equalisation section.*

Using MATLAB, SIRR-processed impulse responses were obtained for each convolver from a measured B-format impulse response. These impulse responses were then imported into the Max MSP patch. Eight, two-channel, 'multiconvolve' objects were used to perform the convolution operation for each loudspeaker. The resultant signals were processed by a loudspeaker equalisation stage prior to being rendered over the loudspeaker array.

8.1.3 Auralisation control

It was determined that the musician should have full control over the auralisation and listening tests using an iPad interface positioned nearby. The iPad was connected to the SoundLab computer using a wireless network. Communication between the iPad and the Max MSP patches were facilitated through the use of the C74 control app and associated Max MSP objects (van der veen, 2013). These objects allow two-way communication with user interface elements on the iPad allowing certain parameters to be changed on screen and for the interface to control aspects of the auralisation.

The C74 app responds directly to the Max MSP patch and allows Open Sound Control (OSC) messages to be sent to the patch. These messages were then used to trigger events such as loading certain impulse responses or recording user responses. Typically, the interface would allow musicians to select which hall they were currently playing in (i.e. Hall A or Hall B), to record their response to specific questions (i.e. Which hall did you prefer?) and move to the next question.

Figure 8.4 shows a block diagram of how the interface communicates with the Max MSP patches. In the listening tests, the individual tasks varied but utilised the same data for auralisation. Therefore, the auralisation patch is kept running throughout the entirety of the listening test whereas the test patch changes for each task. It can be seen that the test patch receives instructions from the interface and loads the appropriate auralisation data for that specific question. The test patch also records the user responses in addition to providing user feedback (i.e. changing the colour of buttons to show which hall was currently selected).

This design was considered the most reliable way of administering the listening test with minimal researcher input and reduced likelihood of switching artefacts. The individual test patches and the interfaces will be discussed in further detail in Chapter 9.

8.1.4 Parametric decoder

As discussed previously, a MATLAB script was developed in order to parametrically decode the measured impulse responses. The MATLAB script produced sixteen rendered impulse response files (one per loudspeaker). These impulse responses were then auralised using the real-time convolution system and the results rendered over the loudspeaker array. While it is possible to implement SIRR and DirAC in real-time, performing the analysis-synthesis process 'offline' ensured that the required processing demands of the auralisation system were minimised allowing more elaborate comparisons to take place in stage acoustic laboratory tests.

Analysis and synthesis of the impulse responses was performed in the time-frequency domain, using the Short Time Fourier Transform (STFT). The analysis and synthesis processes worked by transforming a short section of the impulse response to the frequency domain, applying the required processes and then synthesising that section of audio before transferring back into the time domain. The script would repeat this process for the duration of the impulse response as shown in Figure 8.5 illustrates this process.

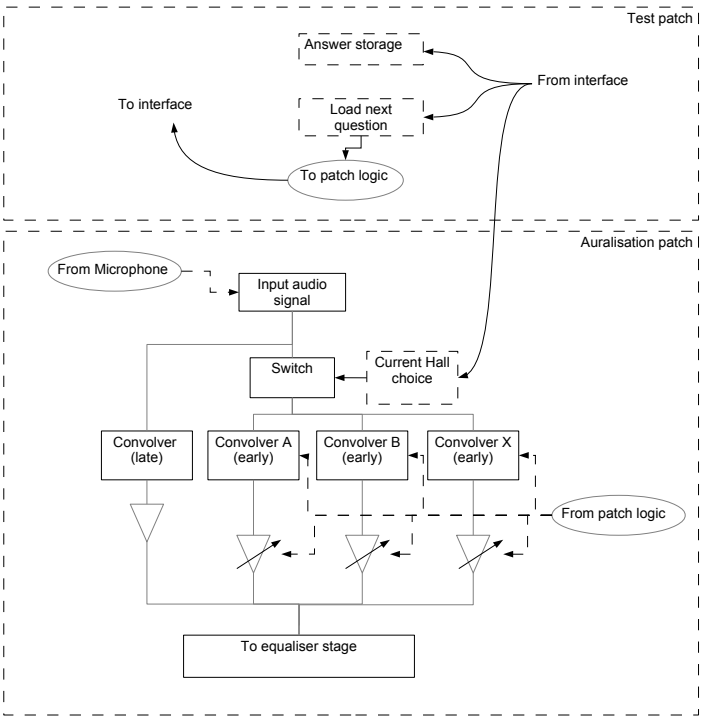


FIGURE 8.4: *Block diagram of how the iPad, test patch and auralisation patch communicated with each other*

Firstly, a region of audio is isolated and a Hanning-shaped window applied. Zero-padding was applied before and after the windowed region of audio so that the length was twice that of the selected region. A circular shift was then applied before transferring into the frequency domain. Zero-padding the signal to twice the region length, allows an FFT of twice the size to be used to calculate the frequency domain signal. While this does not add any further information to the signal, it ensures that each frequency bin is represented by two values, increasing the resolution of the frequency domain representation. This also ensures that the synthesised signals are not aliased in time. The circular shift swaps each half of the region so that the signal is mirrored around $N = 0$ and the zero-padding is moved to the centre of the region. This ensures that the analysis and synthesis are zero-phase (Smith, 2011).

Once in the frequency domain, the parametric decoding technique (SIRR) described in Section 6.5.5 was applied to synthesise the appropriate signals for each loudspeaker channel. The signal was then transferred into the time-domain using the IFFT. The signal was circular shifted and the same Hanning window was applied. The signal was then reconstructed by overlapping this region with the previous region and adding, performing the so called Overlap-Add procedure (OLA).

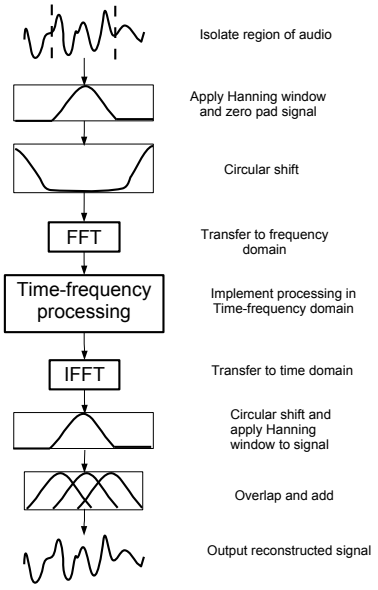


FIGURE 8.5: *Block diagram of a single cycle of the analysis and synthesis method showing how the STFT was implemented. In this diagram, the parametric decoding occurs in the section marked 'Time-frequency processing'.*

Consider an implementation of the above process where no time-frequency processing takes place, i.e. the aim is to transfer the signal into the time-frequency domain and reconstruct the original signal as closely as possible. This is possible using OLA provided that the sum of the overlapping windows is unity (Zolzer, 2011). Due to zero-padding, the resulting signal for each block is longer than the original block. This allows each block result to be overlapped and summed, reconstructing the signal correctly. Figure 8.6 shows the effect of summing overlapping Hanning windows where the overlap is 75% of the window size. It can be seen that the windows sum to produce a constant amplitude envelope of unity amplitude. The amount of overlap applied depends heavily on the type of window used. For example, when a Hanning window is used, an overlap of 50% would result in amplitude modulation of the signals.

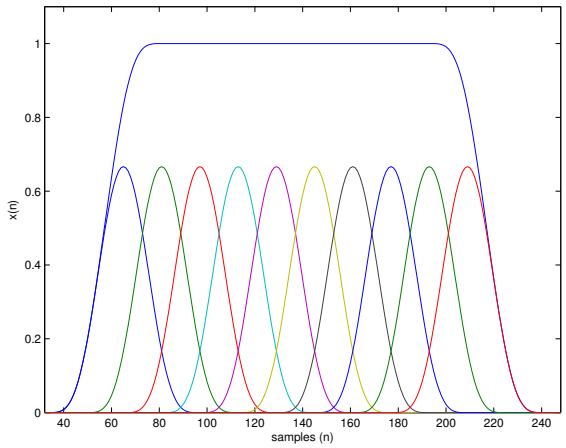


FIGURE 8.6: *Summing numerous, overlapping Hanning windows produces a constant amplitude envelope.*

It is widely understood that the STFT can operate at a high temporal resolution or a high frequency resolution (Weeks, 2007) but not both. The resolution of the STFT is dictated by the window length and the hop size. Values are often chosen to be most appropriate for a particular application. Given the transient nature of reflections in an impulse response, the STFT was implemented with a small window size (32 samples) and a low hop size (8 samples) to ensure the auralisation was rendered with a high degree of temporal resolution. It has been demonstrated previously how a multi-resolution STFT can be implemented in DirAC to provide a fine degree of control of the temporal and frequency resolution (Pihlajamaki, 2009). This has not been implemented in this study and has been left for future work.

The time-frequency processing block shown in Figure 8.5 indicates where the parametric decoding took place. For each iteration, the following operations were performed in order to analyse and synthesise the appropriate signals.

1. Compute intensity vectors for each time-frequency bin by multiplying pressure and velocity channels
2. Estimate azimuth and elevation of intensity vectors for each time-frequency bin
3. Estimate Diffuseness for each time-frequency bin
4. Use low frequency intensity vectors ($f < 5kHz$) to extrapolate the high frequency direction of arrival ($f > 5kHz$)
5. Use virtual microphone gains to decode B-format to loudspeakers
6. Apply VBAP gains according to intensity vector and weight according to diffuseness
7. Decorrelate loudspeaker signals and weight according to diffuseness
8. Apply gain correction and weighting to diffuse and non-diffuse signals

9. Summation of diffuse and non-diffuse sound

Once in the frequency domain, the B-format signals are processed by an Ambisonic decoder matrix which creates a signal for each loudspeaker based on it's physical location relative to the sweetspot and the directivity of the virtual microphone. In this case, the virtual microphone is of dipole directivity (figure-of-eight) as it has a suitably narrow frontal lobe. When the VBAP gains are applied to each loudspeaker channel, the gain of the opposing channel will be close to zero, therefore the contribution from the opposing loudspeaker will be suppressed.

Parallel to this process, the B-format signals are analysed to determine the direction of arrival and diffuseness of each frequency bin. To reduce any localisation errors introduced by the microphone, the non-diffuse sound is synthesised assuming all frequency components of a reflection have arrived from the same direction i.e. they are completely non-diffuse. High frequency components ($f > 5kHz$) are aligned by spatialising them according to the angle extrapolated from the low frequency components ($f < 5kHz$). It is assumed that the correct direction of arrival is equal to the most frequently occurring angle obtained below 5kHz. This angle is computed by calculating a circular Kernel Density Estimate (KDE) on the intensity vectors in each window and finding the angle associated with the maximum value of this estimate. This was applied to both azimuth and elevation data to determine the correct angle of arrival. This process artificially aligns the direction of arrival of all frequency bins. As the high frequency inaccuracies of the Ambisonic microphone result in an over-estimated diffuseness, especially at high frequencies, these high frequency components will be attenuated when the diffuseness crossover gain is applied. As it was not possible to compensate for the over-estimated diffuseness, the audible effect of aligning the direction of arrival vectors in this way is likely to be minimal.

Once the direction of arrival has been computed for each frequency bin and for each loudspeaker according to the direction of arrival (VBAP gain), a weighting factor was also applied which compensated for the virtual microphone directivity pattern. These weighting factors are described in further detail in Section 6.5.5.

Parallel to this, the diffuse signal for each loudspeaker was decorrelated using phase randomisation. This was implemented by replacing the imaginary parts of the signal with randomised values between 0 and 2π. The amplitude of the diffuse sound was weighted according to the directivity pattern of the virtual microphone. The non-diffuse and diffuse sounds were then weighted according to the analysed diffuseness and summed together before being transferred back to the time-domain.

An amplitude envelope was applied to the synthesised signals in order to split them into early and late parts which were auralised separately. This amplitude envelope also ensured that the first 20ms of the impulse response were silenced. The early and late parts of the synthesised impulse response for each loudspeaker were written as separate 44.1kHz, 32-bit floating point wave files. The audio files containing the early reflections were 110ms in length whereas the audio files containing the late reverberation were 2.5s in length.

8.1.5 Auralisation calibration

Each stimulus is calibrated as demonstrated in Chapter 7. This involves positioning the measurement system in the centre of the loudspeaker array and measuring the impulse response of the virtual space. Settings are then adjusted to ensure that reflections are rendered at the correct time (by correcting for system latency) and at the correct level (using ST_{early}).

The system latency was measured and corrected for using the process described in Section 7.3.1. When measuring system latency, an artificial impulse response was auralised which consisted of synthesised, high amplitude specular reflections. Comparing the time of arrival of the auralised peaks with the original impulse response allows the system latency to be computed. By truncating the impulse response by this amount, reflections will arrive at the correct time of arrival.

The latency of the system was measured with different buffer sizes to determine the contribution of the DSP processing. These results were obtained with a 2.5s long impulse response. The length of the impulse response was chosen to be representative of the length of the venue impulse responses measured previously. The results are shown in Table 8.1. These results were obtained when the auralisation system was operating at a sampling frequency of 44.1kHz.

Buffer size	Latency (samples)	Latency (ms)
1024	2538	57.6
512	1514	34.3
256	1002	22.7
128	745	16.9
64	613	13.9
32	554	12.6
16	516	11.7

TABLE 8.1: *Measured latency for different sizes of I/O buffer.* $F_s = 44.1kHz$

In this research, it is desirable for the auralisation system to be capable of rendering reflections from $20ms$ onwards. It is also required that at least three auralisations can be run in parallel so that the test participant can switch between them without any switching artefacts or loading delays. Therefore, the I/O buffer is set at the largest possible size that produces less than $20ms$ latency. From Table 8.1 it can be seen that an appropriate buffer size is 128 samples.

These measurements also give an approximate guide as to minimum allowable distance for correct auralisation of reflections (in terms of latency alone). If the lowest possible truncation size is $16.9ms$ then there must be no reflections occurring before this point for them to be rendered by the system. A delay of $16.9ms$ is equal to an acoustic propagation distance of $5.81m$. Therefore, the system is capable of auralising measurement positions that are a minimum of $2.91m$ away from the nearest reflecting surface.

8.1.6 Background noise

The background noise level of concert halls is often carefully controlled to ensure the space is as silent as possible. The background noise however is often higher in amplitude than is found in the majority of acoustic laboratories or anechoic chambers. When impulse responses are auralised, it is common to truncate the impulse response once it has fallen below the noise floor. This ensures that the auralisation system is not attempting to process an impulse response that is any longer than required for the auralisation.

As the noise floor is much higher than in the acoustic laboratory, the truncation point of the impulse response becomes audible. Additionally, the low background noise of the SoundLab may cause test participants to adjust to that background noise level which may make the resulting auralisation sound unnaturally loud. In order to circumvent this problem, a recording of the background noise in a performance space was played on a repeated loop, alongside the auralisation.

In order to reduce the likelihood of audible comb-filtering effects, caused by the musician moving around the sweetspot, the background noise was decorrelated before being rendered over the loudspeaker array. This removed any spatial aspect of the background noise. However this was not considered important in the context of this experiment.

Figure 8.7 shows the section of the Max MSP patch used to decorrelate and render the background noise to each loudspeaker. The audio signal is set to play on a loop of five minutes duration. The loop was edited such that the beginning and end of the the audio file were as similar as possible and that the beginning and end of the file occurred at zero-crossing points. These edits and the length of the loop minimised the possibility that the loop points were audible for test participants.

The decorrelator object renders the background noise loop to all 16 loudspeakers. Each speaker feed is decorrelated using a separate all-pass filter which feature a flat frequency response but have a complex phase-response. Using different settings for each all-pass filter ensured that all loudspeaker signals were suitably decorrelated. Each all-pass filter had a delay and feedback parameter. In this case, the feedback was set arbitrarily to a value of 0.3 and the delay value for each loudspeaker varied between $0ms$ and $35ms$. The level of the background noise was set to an appropriate level by measuring at the sweetspot with a SLM.

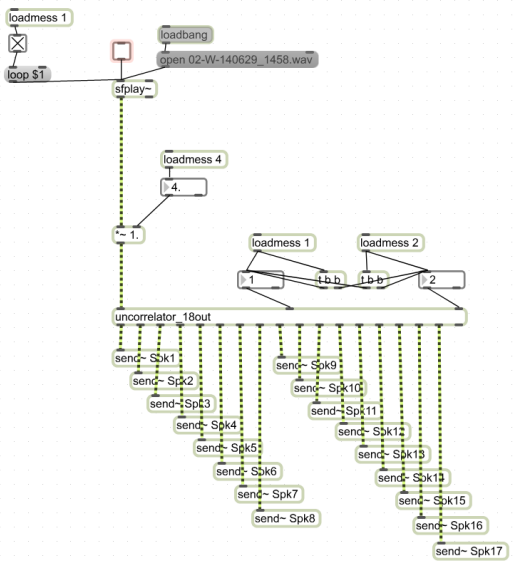


FIGURE 8.7: *Screenshot of the background noise player with decorrelation. The W-channel would play on a loop and was distributed to all 16 loudspeakers. Decorrelation was applied using all-pass filters with different delay values applied.*

Figure 8.8 shows the background noise level of the recording used in the auralisation. In addition, the plot shows the background noise level measured of the auralised background noise in the sweet spot of the SoundLab and the background noise levels of the SoundLab itself. The three measurements are shown in relation to standard Noise Rating (NR) curves (International Organisation of Standardisation, 1999). This particular recording was made at the stage centre location in the Ledger Recital Room (LRR). This recording was used as the background noise levels were typical of most recital spaces (NR20). All measurements show the L_{90} levels in octave bands where all measurements had a duration of 5 minutes. The background noise is comprised mainly of noise from the lighting system and air handling units as well as some low-level creaking from the wooden floor of the venue.

It can be seen that there is a close match between the measured and auralised background noise. At a frequency of $1kHz$ the original background noise level is $15.9dB$ whereas the auralised level is $15.0dB$. It can also be seen that these exceed the background noise level of the SoundLab, in particular at low frequencies.

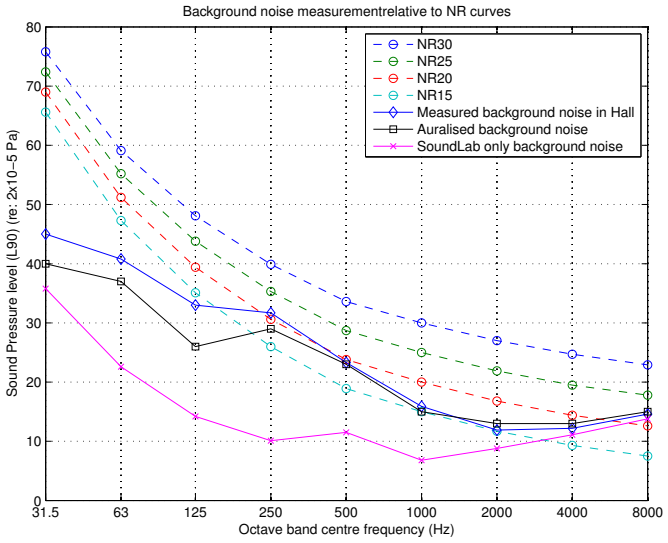


FIGURE 8.8: *Background noise levels of auralised background noise (black) in relation to the original recording (blue) and the background noise of the Soundab (magenta). Also shown are the standardised NR curves (International Organisation of Standardisation, 1999). In this case, the background noise of the auralisation does not exceed NR20.*

8.2 Accuracy of auralisation system

Before listening tests were conducted it was of interest to determine objectively how well the auralisation system was operating and to confirm that the objective parameters of the virtual soundfield correspond with the target values. Ideally, the measured impulse response of the virtual space measured at the sweetspot of the loudspeaker array would be identical to that measured in the real space. Whilst it is not expected to achieve an exact physical reconstruction of the target space with the described system, it is required that the system is capable of reproducing the space with similar acoustic parameters as measured in the real space. By aiming for equivalent parameters it is possible to objectively ensure the virtual space is a true representation of the target.

By repeating the measurement and analysis techniques described in the previous chapter for the calibrated virtual reconstruction of the space, the general accuracy of the system can be demonstrated. This section will demonstrate this with a single example virtual reconstruction of a performance space in the SoundLab.

8.2.1 Spatialisation accuracy of artificial reflections

In previous interactive auralisation systems devised for stage acoustic research, the spatial audio system used a relatively small number of loudspeakers to render the acoustic response to the musician. The consequence of this is that the spatial characteristics of these reflections may not have been representative of the target. For instance, specular reflections may have been

reproduced with a larger spatial extent than intended or may have arrived from an incorrect direction.

Therefore, it is of interest to determine how accurate the interactive auralisation system is in terms of reflection spatialisation. A test was devised where an artificial impulse response was created which featured synthesised reflections arriving at specific times and from specific angles of arrival. This impulse response was imported into the interactive auralisation system and a series of measurements were made in the sweetspot of the loudspeaker array. The objective was to determine if the auralisation system could recreate the reflections at the correct time and angle of arrival.

Three test signals were created using MATLAB which featured 16 regularly spaced impulses (every 0.15 seconds). The reflections were spatialised to regularly spaced angles of azimuth in an anticlockwise direction between 0° and 337.5° in 22.5° increments. This ensures that reflections appeared from both the direction of the loudspeakers and also in between the loudspeakers. The three test impulse responses were identical with the exception that the reflections were panned to different angles of elevation i.e. high, low and ear height. Each test impulse response was exactly 2.5 seconds long which is representative of the maximum length of a measured stage impulse response. The temporal arrangement of reflections was chosen as equal to the maximum measured T_{30} of the SoundLab, this ensured that the acoustic decay created by each reflection would not overlap with the next reflection which would otherwise add noise to the angle estimation. Figure 8.9 shows a plot of the generated signal (W-channel only).

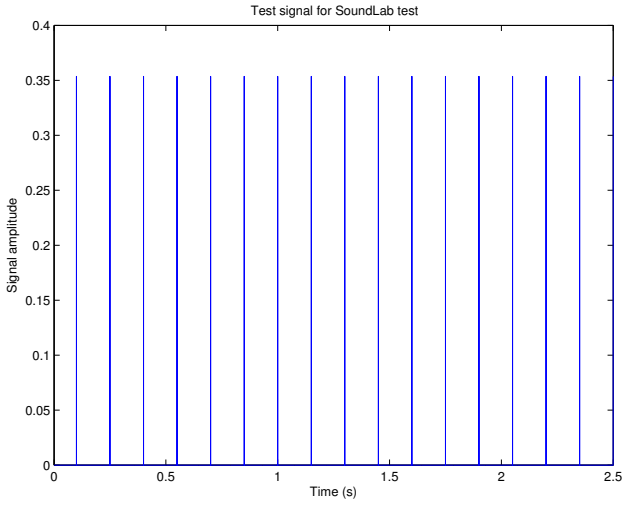


FIGURE 8.9: *Plot of the signal used as an impulse response to test the interactive auralisation system. Artificial reflections are spaced 0.15 seconds apart and are panned in* 22.5° *increments at set angles of elevation. This plot shows the W-channel only.*

The resulting B-format signal was processed as described using SIRR and imported into the interactive auralisation system. A measurement system, identical to that used in the performance space surveys, was constructed in the centre of the loudspeaker array. The aim was to measure the impulse response of the virtual space as rendered at the sweetspot of the loudspeaker array as shown in Figure 8.10. A microphone was positioned 30cm in front of the loudspeaker which

captured the sounds produced by the measurement loudspeaker. A 10-second log sine sweep was played through the measurement loudspeaker that was auralised with the test impulse response and rendered over the loudspeaker array.

FIGURE 8.10: *Measurement system positioned in the sweetspot of the loudspeaker array. A Genelec 1029A loudspeaker is mounted at a height of 0.93m (floor- top of LF driver height) and the Soundfield ST350 ambisonic microphone is mounted at a height of 1.29m above the wooden floor. The sound from the loudspeaker is captured by a Behringer ECM1000 omnidirectional measurement microphone positioned 30cm from the measurement loudspeaker. Absorption was applied to the LCD screen and a wooden floor deployed*

The resulting auralisation was recorded at the centre of the array using the Soundfield ST350 Ambisonic microphone and then analysed to determine if the direction and time of arrival was correct. Figure 8.11 shows a plot of the impulse response as measured in the sweetspot of the loudspeaker array (W-channel only). The direct sound can be seen to occur at $t = 0$s followed by each of the auralised reflections. The acoustic decay of the SoundLab can be seen to occur on each of the auralised reflections. The temporal spacing of each reflection ensures the decay of the SoundLab does not interfere with the reflection analysis.

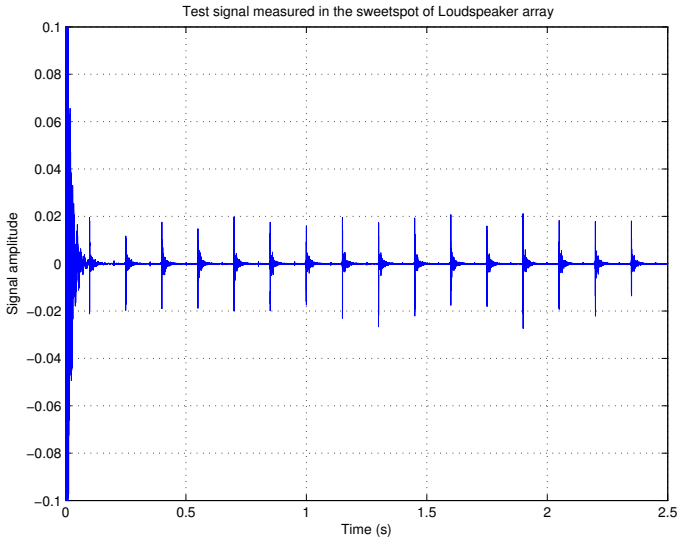


FIGURE 8.11: *Plot of one of the auralised impulse responses as measured in the sweetspot of the Loudspeaker array. Each reflection produces its own acoustic decay. The direct sound from the measurement loudspeaker can be seen to occur at* $t = 0s$.

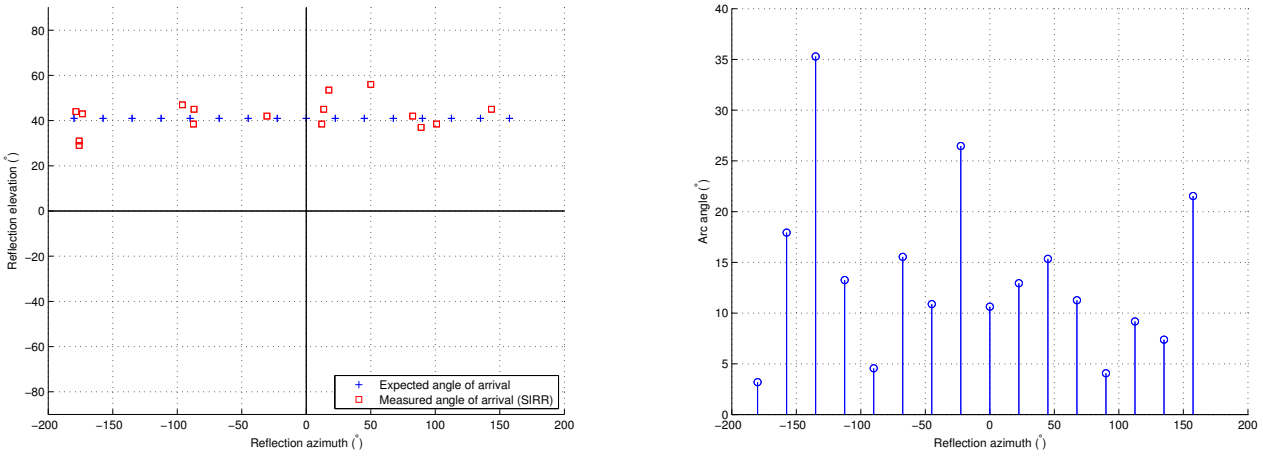
Prior to each measurement, the gain of each test impulse response was calibrated to an ST_{late} value of $-23.5dB$ to ensure that each set of reflections had the same overall amplitude. Furthermore, the latency of the system was measured and reduced as described previously in this chapter. The peak of each measured reflection was taken as the time of arrival of the reflection and compared with the expected time of arrival. In this experiment, an i/o buffer size of 32 samples was used with a sample frequency of $44.1kHz$. This produced a measured latency of 564 samples which equates to $12.8ms$. The initial pointer in each of the buffers in the auralisation system were set to begin at $t = 12.8ms$ thus reducing the measured system latency to 0 samples.

The difference between the expected and measured angle of arrival for both rendering techniques was calculated using equation (8.1) which shows the norm of the ratio of the cross product and the dot product of two vectors $\vec{a}$ and $\vec{b}$. In this case, $\vec{a}$ and $\vec{b}$ represent the direction of the expected and measured reflections. These vectors are unit length and are expressed in Cartesian form by deriving directional cosines.

$$\theta_e = \arctan\left(\left\|\frac{\vec{a}\times\vec{b}}{\vec{a}\cdot\vec{b}}\right\|\right) \tag{8.1}$$

Figure 8.12(a) shows the azimuth and elevation of the synthesised reflections as they were panned to an angle equal to the ceiling loudspeakers in the SoundLab (shown here as blue crosses). Also shown are the location of the reflections measured at the sweetspot when rendered using SIRR (shown as red squares). It can be seen that, in terms of elevation, the reflections rendered with SIRR are reasonably close to the expected elevation values. A striking feature of the rendered reflections is that they tend to cluster around azimuth angles of $0°, \pm 90°$ and $180°$ which correspond to the location of the loudspeakers at this elevation.

Figure 8.12(b) quantifies the angular error between expected and measured directions of arrival. The x-axis shows the expected angle of arrival and the y-axis displays the difference between the expected and measured angles of arrival. The angular error exhibited by the SIRR auralisation shows a lower angular error, however the error increases when the reflection arrives from between two loudspeakers. The largest error is shown to be 35.3° (when the intended angle of arrival was −135°) and the smallest error is shown to be 3.2° (when the intended angle of arrival was −180°). The average error is 13.7°($\sigma = 8.5°$)

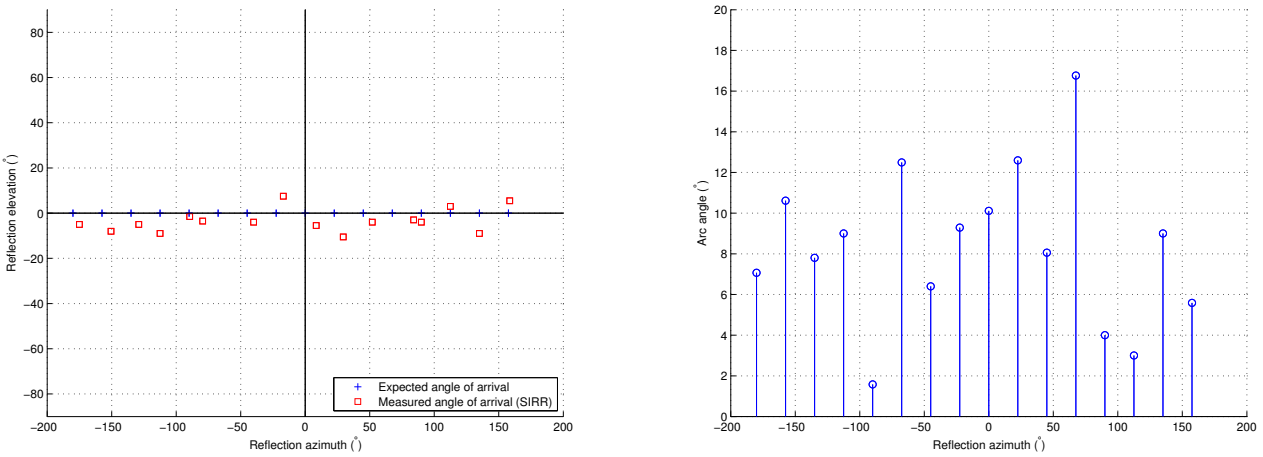


(a) Direction of arrival

(b) Angular error

FIGURE 8.12: *Measured angle of arrival and associated angular error for synthesised reflections produced in the SoundLab at a high elevation.Figure 8.12(a) shows the intended angle of arrival (blue crosses) and the measured angle of arrival (red squares). The associated angular error is shown in Figure 8.12(b). The largest error tends to occur when the reflections arrive between two loudspeakers.*

Figure 8.13 shows similar results for when the reflections are panned at the same angles of arrival but with an elevation of 0 degrees. The reflections appear to arrive from very similar azimuths and elevations to the expected angle of arrival. In some cases the reflections appear to be clustered together in azimuth, for example around ±90°. The largest error for this impulse response is shown to be 16.8° (when the intended angle of arrival was 67.5°) and the smallest error is shown to be 1.6° (when the intended angle of arrival was −90°). The average error is 8.3°($\sigma = 3.6°$)

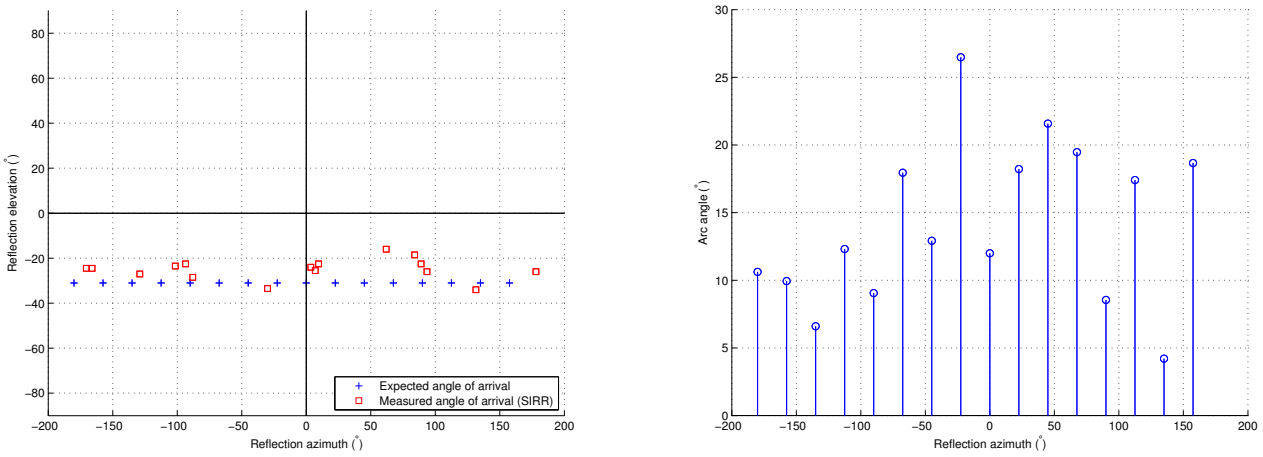


(a) Direction of arrival

(b) Angular error

FIGURE 8.13: *Measured angle of arrival and associated angular error for synthesised reflections produced in the SoundLab with no elevation.Figure 8.12(a) shows the intended angle of arrival (blue crosses) and the measured angle of arrival (red squares). The associated angular error is shown in Figure 8.12(b). The largest error tends to occur when the reflections arrive between two loudspeakers.*

Figure 8.14 shows similar results for when the reflections are panned at the same angles of arrival but with an elevation of -31 degrees. It can be seen that the angular error has increased to similar extents as when the reflections were panned to a high elevation. Similarly, the angular error is lowest when the reflections are panned towards the direction of the loudspeakers and highest when the reflections are panned in between. The measured elevation of the reflections is very similar to the expected values. The largest error is shown to be $26.5°$ (when the intended angle of arrival was $-22.5°$) and the smallest error is shown to be $4.2°$ (when the intended angle of arrival was $-135°$). The average error is $14.1°(\sigma = 6.1°)$



(a) Direction of arrival

(b) Angular error

FIGURE 8.14: *Measured angle of arrival and associated angular error for synthesised reflections produced in the SoundLab at a low elevation.Figure 8.12(a) shows the intended angle of arrival (blue crosses) and the measured angle of arrival (red squares). The associated angular error is shown in Figure 8.12(b). The largest error tends to occur when the reflections arrive between two loudspeakers.*

It is acknowledged that in this experiment the measured angle of arrival could be influenced by the correct positioning and orientation of the soundfield microphone relative to the loudspeaker array. As analogue tools were used to position the microphone relative to the loudspeaker array it may be possible to obtain improved results for the same auralisation by performing more accurate alignment of the loudspeakers and microphone. It is also possible that inaccuracies of the Ambisonic microphone contributed additional angular error. As discussed in Appendix B the Ambisonic microphone is considered to accurately localise reflections to a maximum frequency of $5kHz$. In this analysis, each reflection is localised using intensity vectors below this frequency to reduce this error.

These results have objectively characterised the spatialisation accuracy of the interactive auralisation system. It was shown that reflections were spatialised more accurately when they occupied the lateral plane as opposed to arriving with a high or low degree of elevation.

The loudspeaker density is much lower at high and low elevations and so reflections arriving from between loudspeakers, at these angles of elevation, will tend to group towards individual loudspeakers. The largest error was found to be 35.3° when the reflection was panned between loudspeakers at a high elevation. When reflections arrived in the lateral plane the largest error was found to be 16.8°.

It is also feasible that interfering reflections were introduced by the presence of the wooden floor. If these interfering reflections occurred within a short time-delay of the original impulse, then this would introduce an error into the angle of arrive estimation.

8.2.2 Accuracy of measured impulse response

A further test of the auralisation system involves auralising a stage acoustic impulse response measured in a real venue and comparing the auralisation with the measured data. The measured and auralised impulse responses can be compared using both monaural acoustic parameters, such as EDT and T_{30}, and spatial and temporal parameters such as the mean time of arrival.

In the following example, a stage acoustic impulse response was auralised using the interactive auralisation system. Measurements, similar to those in the previous section, were made in the sweetspot of the loudspeaker array. Sine sweeps of 10-second length were generated by a loudspeaker, auralised and then rendered over the loudspeaker system. The resultant impulse response was measured with an Ambisonic microphone. By analysing the impulse response measured at the sweetspot, adjustments could be made to ensure the system was rendering the impulse response accurately.

For this test, the I/O buffer was set to a size of 32 samples ($F_s = 44.1kHz$). The total system latency was measured at the sweetspot using the synthesised impulse response described in the previous section. The time of arrival of the reflections was compared with their expected time of arrival. The system latency was found to be 564 samples which equates to $12.8ms$. This latency was compensated as described earlier in this chapter.

The impulse response was measured at the stage centre position of the Younger Hall St Andrews. The details of this concert hall survey and the measurement details are included in Chapter 3. The impulse response used in this example had a source orientation of 180° relative to stage front. This impulse response was used as it was likely to contain a large number of high amplitude reflections which were easier to observe. The impulse response was processed using SIRR as described previously. In this test, the impulse response was rendered as a set of $2.5s$ long audio files rather than splitting them into early and late parts at $100ms$.

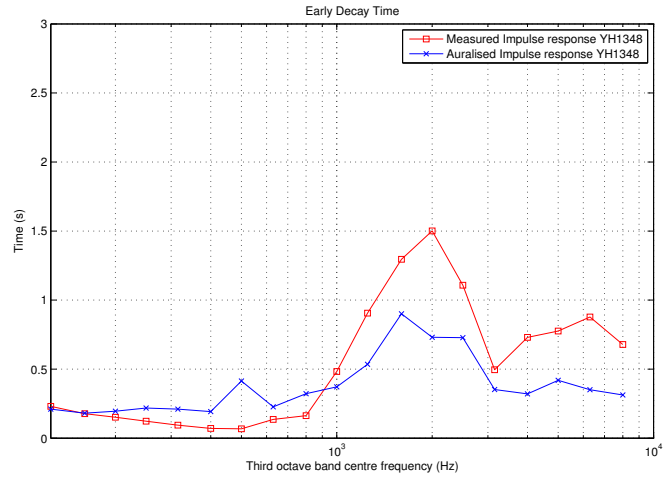
After the latency had been compensated, the gain of the impulse response was set. This was achieved by deriving ST_{late} from the measured and auralised impulse response and adjusting the gain of the system until the auralised ST_{late} matched the measured value. A single figure for ST_{late} was obtained by averaging octave band levels between $250Hz$ and $2kHz$ as described by ISO 3382:-1 (International Organisation of Standardisation, 2009). It can be seen in Table 8.2 that ST_{early} is $4.3dB$ higher than the original measurement. This is thought to be due to the contribution of early energy from the SoundLab acoustic response.

IR	$\mathrm{ST}_{early}(dB)$	$\mathrm{ST}_{late}(dB)$
virtual	-9.1	-15.9
measured	-13.4	-15.9

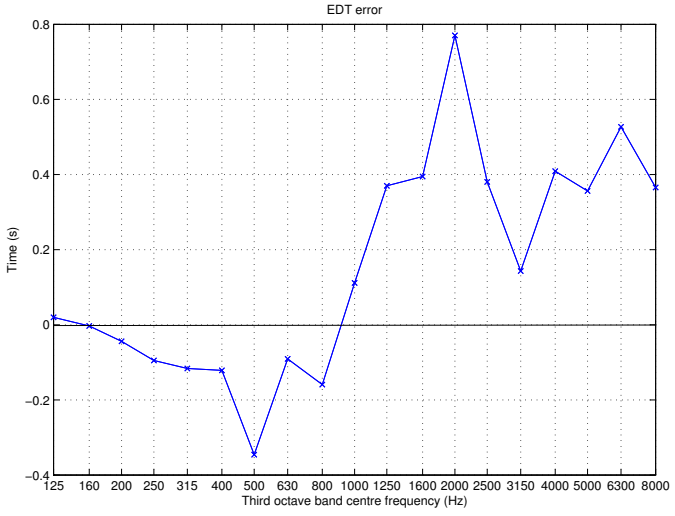
TABLE 8.2: *ST_{early} and ST_{late} for the measured impulse response and for the virtual impulse response after calibration.*

Figure 8.15(a) shows a comparison of the Early Decay time obtained from measured and auralised results. The measured EDT is shown in red while the auralised EDT is shown in blue, results are displayed in third octave bands. The difference between each set of results is shown in Figure

8.15(b). It can be seen that overall, the measured and auralised impulse responses show similar EDT characteristics in terms of frequency content both showing a peak at $2kHz$. However, the auralised EDT is much shorter than the measured EDT at $2kHz$ with the largest difference occurring in this band of $0.77s$. The average error for this auralisation is $0.15s$ ($\sigma = 0.29s$).



(a)

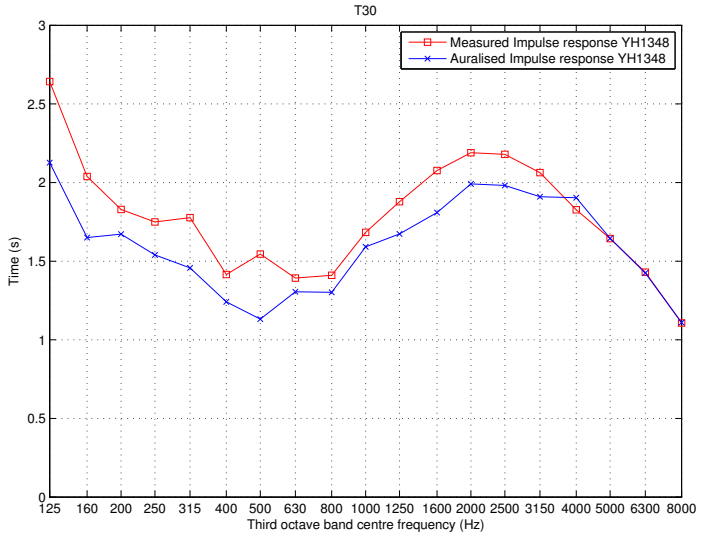


(b)

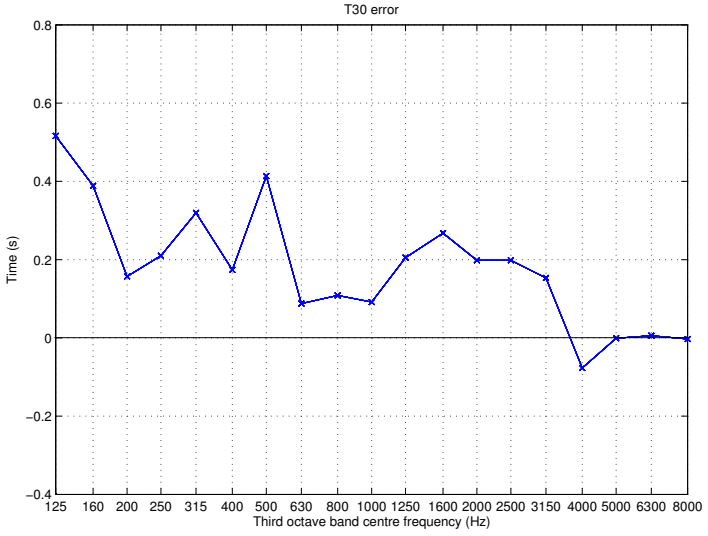
FIGURE 8.15: *Figure 8.15(a) compares measured EDT and auralised EDT for an impulse response measured in the Younger Hall St Andrews. The red trace shows the measured EDT while the blue trace shows the auralised EDT. Figure 8.15(b) shows the difference between measured and auralised results. All results are displayed in third octave bands.*

The reverberation time estimated from the T_{30} curves is shown in Figure 8.16. In Figure 8.16(a) a much closer match can be seen in terms of T_{30} in comparison with the results for EDT. As with EDT, both auralised and measured impulse responses exhibit very similar characteristics in the frequency domain, showing a peak value of 2.19s at 2kHz for the measured impulse response and 1.99s at 2kHz for the auralised impulse response. In general, the reverberation time appears slightly shorter in the auralised impulse response. Figure 8.16(b) displays the differences between

the two sets of results. On average, there is a 0.18s error between the plots ($\sigma = 0.29s$) with the largest difference of 0.52s occurring at 125Hz.



(a)



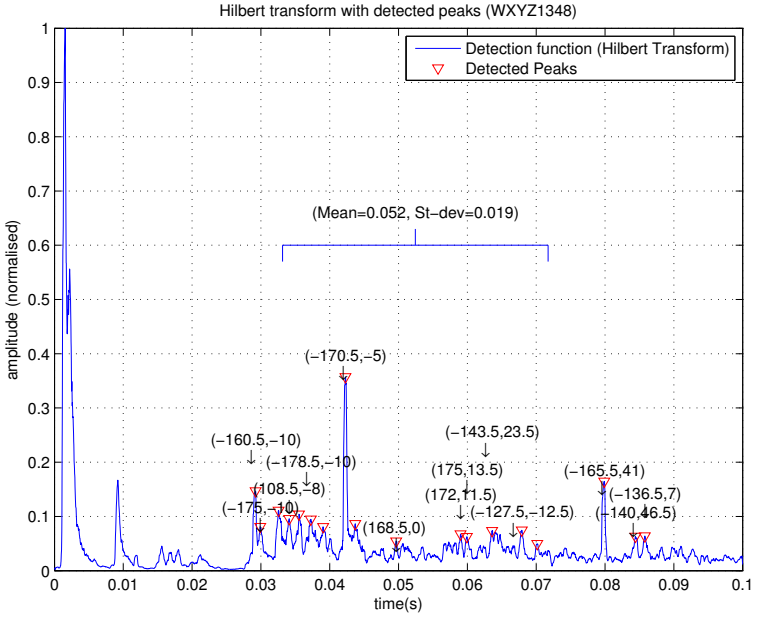
(b)

FIGURE 8.16: *Figure 8.16(a) compares measured T_{30} and auralised T_{30} for an impulse response measured in the Younger Hall St Andrews. The red trace shows the measured T_{30} while the blue trace shows the auralised T_{30}. Figure 8.16(b) shows the difference between measured and auralised results. All results are displayed in third octave bands.*

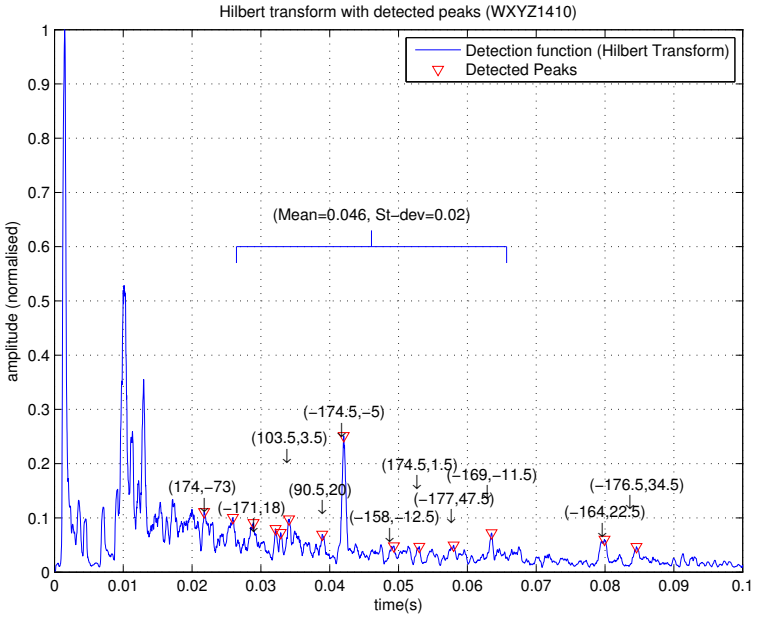
A more detailed inspection of the impulse responses can be performed using the Hilbert Transform to observe when reflections arrive. Figure 8.17 shows the Hilbert transform of the first 0.1s of the measured and auralised impulse responses. In both impulse responses, reflections are only detected after $t = 0.02s$ and are shown by the red markers. The plots show the mean and standard deviation of the time of arrival of early reflections between 20ms and 100ms. These parameters are used to compare the temporal characteristics of early reflections in this research.

The measured impulse response exhibits a $t_{mean} = 52ms$ ($t_\sigma = 19ms$) whereas the auralised impulse response shows a $t_{mean} = 46ms$ ($t_\sigma = 20ms$) indicating a higher amount of reflections arriving earlier and slightly more spread out over time. These values suggest that the auralised impulse response contains significantly more energy between $t = 0.01s$ and $t = 0.03s$ than in the measured impulse response. This additional energy is the contribution of the SoundLab acoustic response.

There are notable similarities between the two impulse responses, for example, the distinct reflection arriving at $t = 43ms$ is observable in both the measured and auralised impulse response. As a single example, this reflection arrives from very similar directions in both measured and auralised impulse responses, as can be seen from the azimuth and elevation annotated above it. For this particular reflection the angular error is as low as 4° in azimuth. Reflections occurring later in the impulse response can also be seen to have similar characteristics, i.e. between $t = 0.05s$ and $t = 0.09s$. The amplitude of these reflections appears to be lower in the auralised impulse response than in the measured impulse response.



(a) Measured impulse response



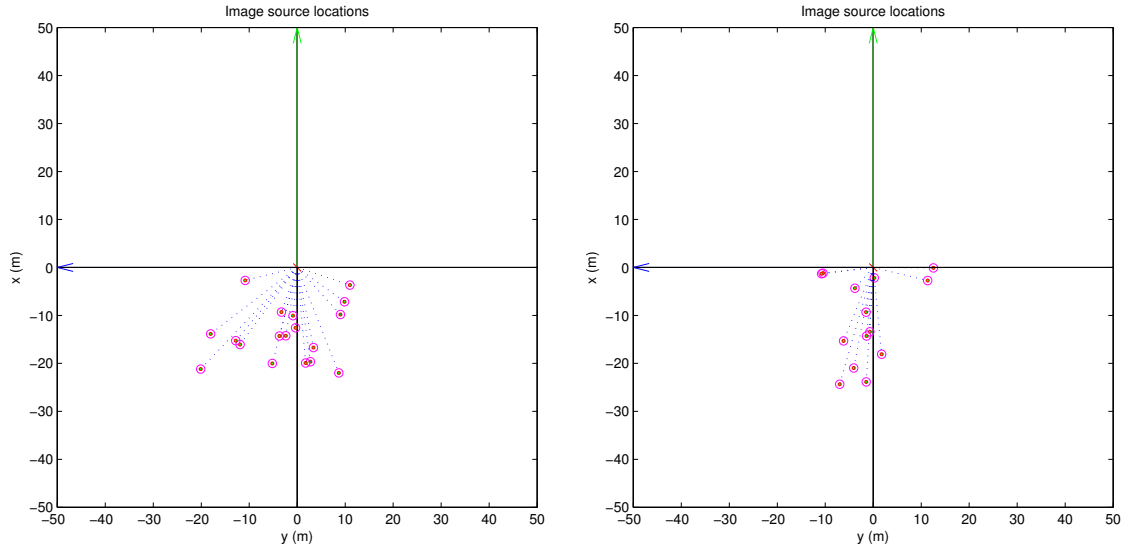
(b) Auralised impulse response

FIGURE 8.17: *Comparison of the Hilbert transform of measured and virtual impulse responses. The annotation on each peak show the azimuth and elevation of each detected reflection. In the auralised impulse response, the reflections occurring between 10ms and 25ms are caused by the acoustic response of the SoundLab.*

The spatial and temporal distribution of early reflections were compared using an image source plot (as described in Chapter 4).Figure 8.18 shows the image source plots for the virtual and measured spaces. Both plots are viewed in plan with the green arrow pointing towards the front

of the stage and the blue arrow pointing towards stage left. Each reflection is shown as an image source connected to the origin by a blue line. The image sources correspond to the detected reflections shown in Figure 8.17.

It can be seen that in both measured impulse response that the reflections appear mainly from a rearward direction as would be expected for this particular impulse response (where the source was oriented towards the rear stage wall). In the auralised impulse response, the reflections appear more clustered towards the rear whereas in the measured impulse response they are slightly more spread out. This is likely to be due to the spatialisation errors inherent in VBAP spatialisation highlighted in the previous section.



(a) Measured impulse response

(b) Auralised impulse response

FIGURE 8.18: *Comparison of the Image source locations of the measured and virtual impulse responses shown in plan. Notice that both plots show the early reflections to the rear being reproduced with reasonable accuracy.*

Table 8.3 shows the spatial parameters for both measured and auralised impulse responses. These parameters, discussed in detail in Chapter 4, show the azimuth and elevation of the mean resultant vector and the spread of the image sources where a value of 0 indicates reflections are arriving from a single direction and a value of 1 indicates reflections are spread equally around the unit sphere.

It can be seen that the values θ_{mean} and ϕ_{mean} are very similar with a difference of 4.5° in azimuth and 7.4° in elevation. The spread values are similar however the difference of 0.165 is reflective of the tighter clustering of the reflections in the auralised impulse response. These differences are likely to be caused by reflections being pulled towards the nearest loudspeaker as they are spatialised using VBAP. This effect was demonstrated in the previous section.

IR	$\theta_{mean}(^\circ)$	$\phi_{mean}(^\circ)$	Spread
Auralised	-170.3	-2.4	0.611
Measured	-174.8	5.0	0.776

TABLE 8.3: *Comparison of spatial parameters of early reflections detected between 20-100ms for measured and auralised impulse responses. θ_{mean} and ϕ_{mean} denote the direction of the mean resultant vector whereas spread indicates how spread out the image sources are (where a value of 1 indicates no spread and 0 indicates equal spread around the unit sphere)*

Overall, it can be seen from these results that the spatial and temporal characteristics of the measured impulse response are largely preserved in the virtual version rendered using SIRR and measured in the SoundLab. As the method of spatialisation is perceptually motivated it was not expected to obtain a precise physical reproduction of the impulse response however the similarities observed between the measured and virtual impulse responses provide a positive indication that the auralisation system is operating as expected and that the result will be considered plausible by a listener. In addition, the results show that the monaural characteristics of the rendered auralisation, such as T_{30} and EDT, closely match the parameters obtained from the original measurements. However, it was shown that the difference between measured and auralised EDT was higher than that measured for T_{30}.

As discussed in the previous section, the use of VBAP introduces some spatialisation errors depending on the angle of arrival of each reflection. The spatialisation errors are larger when the reflections arrive from an elevated direction due to the relatively sparse arrangement of the elevated loudspeakers in the SoundLab. It may be possible to obtain more accurate results by using a more uniformly distributed loudspeaker array. Further advances may be made by altering the panning functions used to provide improved localisation between the loudspeakers.

The main difference observed between the two impulse responses is the overall energy within the first $50ms$ where the impulse response of the virtual space includes the contribution of the SoundLab. This contribution has been controlled as far as possible and is considered to be suitable for listening tests with musician test subjects. It is clear though that further improvements could be made with the loudspeaker array positioned within anechoic conditions.

8.3 Summary and discussion

This chapter has described the interactive auralisation system used for stage acoustic experiments with musicians. The system was implemented in an acoustically treated laboratory (SoundLab) and rendered the stage acoustic response using parametric decoding techniques. The system utilised a single omnidirectional microphone to pick up the direct sound from the musician which was convolved with a multi-channel impulse response and rendered over a 16 channel loudspeaker array. The specific implementation of these stages of the auralisation were described in detail and the methods used to calibrate the loudspeaker system were demonstrated.

The auralisation was assessed in a similar manner to the analysis performed for different stage enclosures which used existing room acoustic parameters in addition to spatial and temporal parameters developed for this research. Once calibrated, the room acoustic parameters suggested that a virtual version of measured stage acoustic conditions could be rendered accurately. However, as demonstrated in Section 8.2.1, the spatial accuracy of the auralisation was found to be better in the lateral plane (where there are more loudspeakers) than when reflections are elevated with elevated reflections tending to pull towards the nearest loudspeaker. It was considered that the measured spatial distribution of early reflections was a close approximation to that measured in the target stage enclosures and that this would be considered a plausible representation for musician test subjects.

With the operation of the interactive auralisation system verified, listening tests can now be performed to determine the effect of spatial and temporal distributions of early reflections on musicians. These will be described in the following chapter.

Chapter 9

Interactive Listening test

The responses from the pilot tests (described in Chapter 8) provided an indication that test participants, exposed to similar acoustic conditions to that of a performer, could discern differences between concert halls where the spatial or temporal distribution of early reflections was varied. However, the varying background of the participants meant these responses could not be used to infer anything about the wider musician population. In addition, the tests did not attempt to determine the subjective effect of these variations. Therefore, a further set of listening tests was designed to allow musicians to play on virtual concert hall stages with varying spatial or temporal distribution of early reflections. Responses from the musicians were collected to determine if these variables were drivers of musician preference. In addition, the responses were analysed to indicate the subjective effect of these variables.

This chapter will therefore present the findings of the main listening test which was conducted with experienced musicians. The chapter will begin by reviewing the background theory and approach behind each listening test. These listening tests were designed to explore the main hypotheses of this thesis as described in Chapter 1 and so a discussion of the main results will conclude this chapter.

9.1 Background theory

By observing how the stage acoustic characteristics influence musician preference, a stage enclsoure could be designed to accentuate any favourable aspects, thus optimising the conditions for a performer. The process of evaluating subjective responses to different product formulations (Quantitative sensory evaluation) aims to measure the intensity of sensory attributes, using a group of panellists as the instrument of measurement (Le and Worch, 2015).

Quantitative sensory evaluation methods have been used previously in the field of stage acoustics. Authors such as Gade (1982), Guthrie (2014) and Brereton (2014) have introduced musicians into virtual stage acoustic environments and evaluated their preference towards each environment. By analysis of the preference responses, in relation to the physical attributes of each stage, it

was possible to determine the objective characteristics of the stage acoustic conditions that drive musician preference.

In order to conduct this type of experiment, it is useful to consider the different subjective domains in which sound is perceived. The filter model, proposed by Bech and Zacharov (2006), is a widely accepted model for how sound is perceived by humans. This is shown in Figure 9.1.

Sounds can be characterised by physical measurements, perceptual measurements and affective measurements. The filter model begins with a complex acoustic stimulus which encounters the human hearing system. This stimulus can be characterised in terms of physical acoustic parameters such as $T30$ or EDT. As described in Chapter 2, the human hearing system transforms the physical soundfield into a sensory stimulus which comprises the first 'filter'. The sensory stimulus is then interpreted by the brain creating a perceptual event where it is analysed by the individual based on a number of subjective attributes. The sound can be characterised by these attributes and compared i.e. a stimulus could have a bright or dark timbre. The formation of these attributes results in the second 'filter'. Using these attributes, a person is able to form an overall impression of the sound allowing affective judgements (preference, liking etc) to be made. For example, a sound may produce a pleasing effect due to the perceived sense of envelopment.

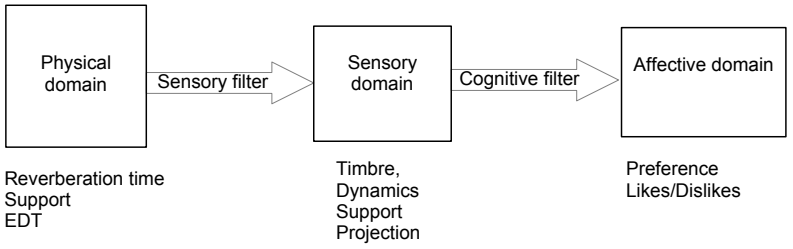


FIGURE 9.1: *The filter model of human hearing demonstrates how physical stimuli are filtered by the senses and are judged by the individual making use of perceptual attributes (such as timbre, loudness etc). Finally, the stimuli is judged in terms of preference in the so called affective domain. Reproduced from Bech and Zacharov (2006).*

In this research, it is of interest to determine the objective acoustic characteristics that drive musician preference. In addition, it is necessary to determine the subjective consequence of these acoustic conditions to assist in understanding why certain conditions are favourable.

It is possible to gain an insight into this by introducing musicians into different stage acoustic environments and recording their responses in the sensory and affective domains. The results can be used to determine which physical attributes are the dominant drivers of musician preference. In this research, the physical attributes of interest have already been identified as the spatial and temporal distribution of early reflections. Therefore, a number of tests will be conducted to determine if these variables are significant drivers of musician preference.

In the specific case of a performer's preference to stage acoustic conditions, different musicians may not be able to discern certain aspects due to the differences in masking their instrument creates. In addition, a performer may feel more satisfied playing under one set of conditions because it suits their particular instrument. Therefore, it is possible that musicians of similar

instrument families make their preference judgements based on a different set of sensory attributes than other instrument families. This is of particular interest in this research as, for example, optimising a stage environment for brass instruments may cause detrimental effects for string instruments. To observe these effects a multivariate statistical analysis technique known as *'Preference Mapping'* will be used.

9.2 Objectives

As discussed in Chapter 1, the principal aim of this research was to determine if the spatio-temporal distribution of early reflections is a driver of musician preference towards performing in a particular concert hall. If this is found to be a salient factor for performing musicians then it may be necessary to include this in future concert hall design.

Of further interest is how the distribution of early reflections influences the musician's impression of a concert hall, in terms of subjective attributes. This would highlight the subjective impact of varying the distribution of early reflections. Furthermore, by relating the musician's preference responses to these subjective attributes it would be possible to determine the salient subjective attributes that drive musician preference.

The multiple areas of interest indicate the need for an interactive listening test where musicians play on virtual stages with a range of spatial and temporal attributes, which can be characterised using the physical parameters described in Chapter 4. By asking musicians to describe each virtual hall in terms of preference and specific subjective attributes, it is possible to obtain a complete objective and subjective description of each hall. These data could then be used to gain insight into the aforementioned areas of interest.

The main listening test consisted of a number of tasks completed by each musician. Experienced musicians were recruited to play their instrument in a number of virtual acoustic environments that were auralised interactively in response to the sound of their instrument. They were asked to respond to a series of questions regarding their experience. The results of the main listening test were used to test the main hypothesis of this research as outlined in Chapter 1. Like previous research in this field (Brereton, 2014, Gade, 1982, Guthrie, 2014), Multivariate analysis techniques will be used to assist in determining the main dimensions of musician preference. Before describing the experiment, the background theory behind this technique will be reviewed.

9.2.1 Analysis of preference patterns

An individual's preference towards a particular product can be driven by numerous subjective dimensions. In this research it is of interest to reveal if the spatial and temporal distribution of early reflections is a driving factor for preference for stage acoustic conditions and furthermore is the effect consistent amongst different instrument families. In this type of experiment, a panel of consumers is presented with a number of products which contain a range of formulations of a specific physical attribute. The panel of consumers is asked to sample each product and rate on a scale how much they like each product or order the products in terms of preference. In addition, a

panel of expert assessors may be asked to rate each product in terms of sensory attributes, rating the intensity of each attribute. These data can then be evaluated simultaneously to indicate the subjective dimensions which are most likely to be drivers of user preference.

This type of analysis is generally referred to as preference mapping, a set of multivariate statistical analysis techniques originally conceived by Chang and Carroll (1968). These techniques produce visual representations of the perceptual space from which it is possible to observe the dominant drivers of user preference and the arrangements of products within that space.

There are two types of preference mapping techniques, referred to as internal or external preference mapping. Internal preference mapping (MDPREF - Multi Dimensional PREFerence scaling) considers only the hedonic responses (i.e. preference) from each participant. The term MDPREF originally referred to a specific computer program developed for this purpose but now refers more generally to this type of technique. It makes use of Principal Component Analysis (PCA) (Le and Worch, 2015) to reveal how the stimuli are arranged in a multi-dimensional perceptual space. This allows the researchers to observe how stimuli are grouped in terms of preference and also how many subjective dimensions are being used to arrive at these choices.

External Preference Mapping (PREFMAP) processes the attribute descriptors of the products and regresses the participant's preference responses into that product space. This main dataset could consist of attribute intensity data from an expert panel which would aid the identification of each subjective dimension of the perceptual map. Similarly, the relationship between the physical characteristics of the stimuli and consumer preference could be evaluated by regression of objective attribute data onto the perceptual space. This approach was utilised by Gade (1982) in stage acoustic research, who used it to determine that musicians tended to prefer halls with high energy early reflections.

The use of PCA in this analysis is particularly advantageous in situations where the salient objective and subjective dimensions are unknown. PCA re-orients the distribution of preference data in terms of variation in particular dimensions. By doing so, it is possible to reduce the fully dimensional space (N variables = N dimensions) into a lower dimensional reconstruction without a significant loss of information (Naes et al., 2010). This is particularly useful in this context as it aids in refining which dimensions are most important to the participants. While the preference judgements might be made using numerous dimensions, they are commonly reduced to or displayed in 3 or less dimensions where data may be scaled or reoriented as necessary for clarity.

Increasing the intensity of a particular attribute may produce a linear preference response where 'more is better' (Le and Worch, 2015). Often, increasing the intensity of an attribute may produce increasing preference up to a certain point (sometimes referred to as the 'bliss point') where any further increases reduce preference. This model of preference is sometimes referred to as a quadratic, 'Danzart' or 'ideal point' model (Le and Worch, 2015). Once the perceptual space has been derived, it is possible to observe how the stimuli are arranged on each subjective or objective dimension in terms of preference. By performing linear or quadratic regression, it is possible to gain additional insight into the ideal amount of a particular attribute.

9.3 Methdology

A listening test was developed, using the interactive auralisation system, where musicians were asked to play on different virtual stage environments and respond to a series of questions. Experienced musician test participants were invited to attend a single session in the SoundLab where they would complete three distinct tasks. These were:

- A discrimination test
- A preference test
- A perceptual attribute test

The same eight concert hall stimuli were used in each test which resulted in three datasets describing each hall. These included the objective attributes, obtained using the analysis techniques described in Chapter 4, the intensity of four subjective attributes and preference data. These data were processed using the preference mapping techniques described earlier in this chapter; allowing the preference for each hall to be expressed in terms of objective and subjective attributes. The discrimination test acted primarily as a post-hoc screening test but also provides an opportunity for each participant to acclimatise to the parameters of the test.

9.3.1 General arrangements

The responses recorded from each participant can be adversely affected by listener fatigue. In passive listening tests, fatigue can increase due to the length, repetitiveness and complexity of the task. In this listening test, the participants were also tasked with playing their instrument in addition to listening to the resulting stimuli. The general arrangements for the listening test were designed with these factors in mind.

Each participant was asked to attend a single, 2-hour session at the SoundLab where the three tests would be administered with 10-minute breaks in between. During the breaks, participants were allowed to leave the SoundLab if they wished. Each participant was informed that each task should take around 30-40 minutes to complete however they were free to work through each exercise at their own pace. During each test, participants were interrupted in 10-minute intervals to remind them of the time elapsed. This was to ensure the experiment did not take much longer than planned. The interruptions took place only when the participant was moving onto the next trial.

In addition to the number of trials and overall length of each test, the type of musical phrase performed by the musician could introduce bias into the responses. For example, musicians may become bored if asked to play one very simple phrase for all trials. Conversely, if the phrase was too complicated or unfamiliar they may find it more difficult to perceive changes in the acoustic conditions. Supplying a single phrase for every musician may also be problematic as the physical action of playing this phrase may be more difficult on some instruments than others. Therefore, each participant was free to choose which phrases to play in order to excite the space but it

was recommended that they keep the phrases consistent when comparing pairs of venues. Some participants opted to play very short phrases in some comparisons, sometimes single notes or scales. Each participant was asked to bring sheet music of a piece they were comfortable playing and identify five phrases of less than eight bars length to play in each hall.

The participants were asked to play, seated, in the sweetspot of the loudspeaker array and were instructed to face the microphone at all times, though they were not physically restricted in any way. Figure 9.2 shows the experimental set up during a dry run of the test. The participant was seated on a stool in the sweetspot of the SoundLab. The image also shows the location of the music stand, microphone and iPad interface in addition to the wooden floor. The auralisation system remained active for the whole duration of the experiment so that background noise and concert hall response were auralised throughout. This was to avoid the participants acclimatising to the acoustic conditions of the SoundLab.

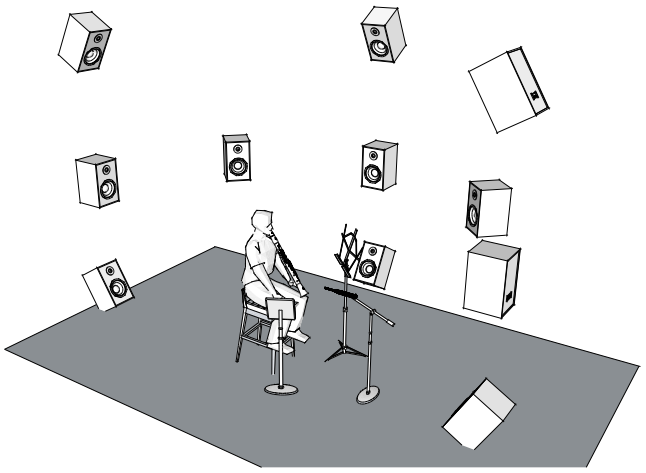


FIGURE 9.2: *Model of a musician playing their instrument in the interactive auralisation system. These show the arrangement of music stand, microphone and iPad interface. A number of loudspeakers have been removed from the model for clarity.*

The experiment was concluded by asking each participant to provide written responses to a small number of questions which aimed to gather further information about their experiences of performing in different acoustic environments and some feedback regarding the test itself. After the experiment was complete, the researcher informally discussed the details of the experiment with the participant. This discussion typically lasted approximately 15-20 minutes.

9.3.2 Participants

In this experiment, it was important that participants had sufficient experience of playing their instrument to a high standard but also performing classical/contemporary music to the public. This would ensure that they were intimately familiar with the sound of their instrument and had the capacity to use extended techniques if required. In addition, this would ensure that musicians had experience of playing in multiple venues and may already have formed opinions regarding how the acoustic conditions affect their performance.

Musicians were recruited predominantly from local Universities that had a music course or from local professional and amateur orchestras and ensembles. The experiment was also advertised on a number of online forums and social media. The requirements stated suitable candidates should have attained a minimum of Grade 8 standard (ABRSM) on their instrument (or equivalent) or have a demonstrable record of public performance. It was also required that the musicians had, to their knowledge, a good standard of hearing (although formal audiometry tests were not conducted). Suitable participants were also considered to play the following musical instruments:

- Violin, Viola, Spanish/Classical guitar
- Clarinet, Oboe, Cor Anglais, Saxophone, Flute
- Trumpet, Trombone, Cornet, Flugel Horn

These instruments were chosen as they could all be easily carried into the SoundLab, generally projected sound forward (as opposed to instruments like the French Horn) and didn't rely on floor resonances (for example low strings). Limiting the type of instruments also provided the opportunity to compare different musicians of the same instrument.

Each participant was sent an instruction sheet prior to arriving at the SoundLab which included a description of the SoundLab, broad details of the test, instructions for each experiment, travel details and a consent form. Before each task, the researcher reviewed the test instructions and demonstrated the user interface. The participant was given the opportunity to ask questions prior to each task. When they were happy to proceed, the researcher retreated behind the screen and the participant was instructed to begin the test. Each participant was rewarded with a £10 cinema voucher for their time in addition to a sealed bottle of still mineral water to ensure they were comfortable throughout the test.

A total of sixteen (16) participants took part in the listening test (10 male/6 female) aged between 19 and 48 years of age (mean = 28.5 years (σ = 10.5)). The participants had a range of experience between 9 and 40 years (mean = 22.6 years (σ = 10.1)). The instruments used in this test consisted of four trombones, four violins, three saxophones (2 Alto and 1 Tenor), two clarinets one flute and two classical guitars. All participants had a wide range of experience including both solo and orchestral performance.

9.3.3 Stimuli

The stimuli consisted of self-generated sounds from the instrument, auralised with impulse responses measured in different performance spaces as documented in Chapter 3. The acoustic response of each venue was rendered over the loudspeaker array in the SoundLab.

As this research is mainly concerned with the effect of early reflections, the late reverberation was kept constant throughout all the comparisons. As described in Chapter 7, this was achieved by splitting the impulse responses into early and late regions and auralising separately. Auralising in this way ensured that ST_{early} could be controlled more easily. The independent variables of

the test were therefore the spatial and temporal distribution of early reflections and the level of early objective support. The dependent variable in each of the three tests was considered to be either the correct response from the musician, the preference judgements of the musician or the subjective intensity of the chosen attributes.

It was critical that the impulse responses selected for this experiment featured a sufficient range of spatio-temporal distributions of early reflections in order to explore the subjective effects. In addition, it was important that the impulse responses were representative of the acoustic conditions experienced by soloist musicians. Therefore, the selection process consisted of two stages. The first stage was based on the physical characteristics of the measurement, i.e. the location and orientation of the sound source. The second stage further refined the list of impulse responses by comparing the spatial and temporal distribution of early reflections.

It was first necessary to determine an appropriate number of impulse responses so the independent variables could be explored in sufficient detail. It was considered important that the entire exercise should not take any more than 2 hours to complete and each task must consist of less than 30 trials. This was to ensure that participant fatigue was minimised. By using the same set of stimuli in each experiment, the responses from each test could be combined allowing more advanced, multivariate analysis to take place.

The selected experimental paradigms were discussed previously in this section. It was determined that an appropriate paradigm for the discrimination test was the ABX test (duo-trio in balanced reference mode). This test aimed to determine if the participant could detect differences in the spatio-temporal distribution of early reflections when ST_{early} was held constant. The test was repeated at a high and low level of ST_{early} therefore low amplitude reflections were not compared with high amplitude reflections. It was determined that a comparison of four stimuli at two levels of objective support results in 24 questions in total.

The preference test was conducted as a pair-wise comparison of all impulse responses, where the participant is asked to determine which one they prefer. The results of the test produced a score for each set of acoustic condition in rank order of preference. It was believed that a rank ordering system in this context would make the experiment much too complicated as the participant would be required to remember a large number of acoustic conditions. If the participant compared each set of acoustic conditions to every other only once, the test would consist of $\frac{N(N-1)}{2}$ trials (where N is the number of stimuli). A comparison of four stimuli at two levels of ST_{early} will result in 28 trials in total.

The attribute test presented each impulse response to the participant sequentially. The participant was asked to rate the intensity of four subjective attributes for each trial. Assuming the same impulse responses were used as in the other two experiments, the total number of trials would be eight. It was evident that eight different impulse responses was an appropriate number of stimuli for each experiment. By repeating the comparisons at two levels of ST_{early} this would allow the subjective effect spatio-temporal distribution and level of early reflections to be studied in sufficient detail. Therefore, four stage impulse responses were chosen for presentation at two levels of ST_{early}.

The impulse responses were chosen to be representative of a realistic performance scenario. When a soloist performs on stage, it is common for them to occupy a down-stage centre position as they are the main focus of the performance. A soloist will often move and gesture while playing in ways that are either necessary for the performance or in order to convey a specific emotion in their performance. While some gestures include a wide range of movement, it is unlikely the musician will, for example, turn away from the audience. The impulse responses were sourced from the measurements described in Chapter 3. Therefore, the impulse responses used in the test should face either 0° or ±45° in orientation and be sourced from a down-stage centre position.

Two of the measured venues (The Caird Hall and Stevenson Hall) featured exceptionally large stages which were considered less likely to host a soloist playing alone so they have not been included in this selection. The Recital Room measurements were made very close to a rear wall and so will be difficult to auralise so they were not considered to be suitable for use in the experiment. Measurements made in the Glasgow University Concert Hall were made in close proximity to a number of pianos and harpsichords. Furthermore, this venue was not set up in a representative performance configuration upon measurement and so these impulse responses were not considered suitable for this listening test.

The impulse responses were therefore selected from measurements made at the down-stage centre position in the Younger Hall, Ledger Recital Room, Glasgow City Halls (Grand Hall) and the Reid Hall. To further aid selection, impulse responses measured in these venues were analysed in terms of various spatial and temporal parameters to determine which were the most suitable for use in this listening test.

The impulse responses were analysed using the techniques described in Chapter 6. The extracted parameters consisted of: mean time of arrival (t_{mean}), standard deviation of time of arrival t_{σ}, azimuth (θ_{mean}) and elevation (ϕ_{mean}) of mean resultant vector and angular spread ($Spread$). These parameters were plotted against each other and assessed to determine which choice of stimuli would produce the most variance across each variable.

Figure 9.3 shows a scatter plot of the average time of arrival against the angular spread of early reflections. Each point represents a measured impulse response that was chosen as a possible candidate for the listening test. Each impulse response has been labeled according to the Hall and the angle of orientation i.e. LH315 refers to the Ledger Hall with a source orientation of 315°.

It can be seen that the majority of impulse responses form a cluster with a centre at the approximate coordinates of $(x, y) = (0.06s, 0.65)$. The objective was to select four impulse responses where two are angled to 0° and two are angled to ±45°. This would mean the auralisations were representative of the performer facing straight out into the audience or obliquely to either side of the hall. Additionally, it was desirable for the stimuli to occupy a range of values across each dimension. It can be seen that LH315, RH0, GCH45 and YH0 are the most extreme values from impulse responses measured in these halls. It can also be seen from Figure 9.3 that these stimuli produce a range of values across each dimension.

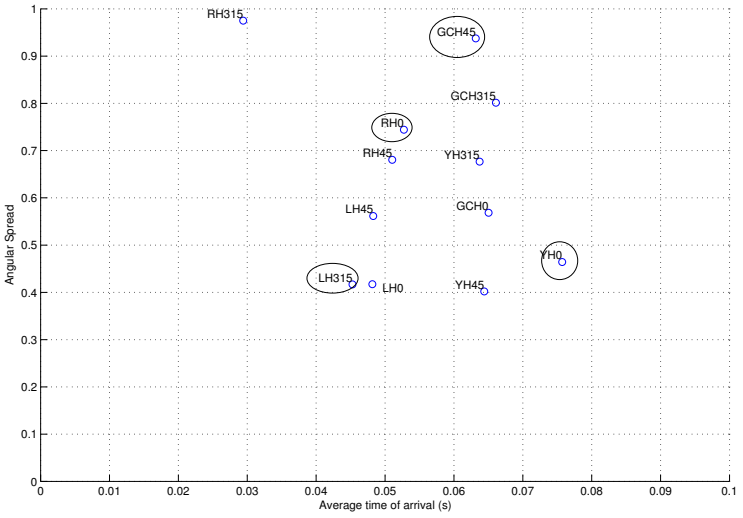


FIGURE 9.3: *Scatter plot showing the average time of arrival versus the angular spread of reflections for measured impulse responses identified as being suitable for auralisation. Encircled points identify impulse responses that were selected for the listening tests.*

The circled impulse responses in Figure 9.3 indicate which of the stimuli fit the criteria described. Table 9.1 summarises the numerical values obtained from each of the selected impulse responses.

Hall	t_{mean} (ms)	t_σ(ms)	θ_{mean} (°)	ϕ_{mean}(°)	spread
GCH 45° (GCH45)	63.2	11.3	78.92	5.05	0.937
LH 315° (LH315)	45.3	21.3	-73.32	-18.11	0.4168
RH 0° (RH0)	52.7	19.5	27.20	-3.45	0.744
YH 0° (YH0)	75.7	23.5	26.35	0.32	0.464

TABLE 9.1: *Spatial and temporal parameters for selected impulse responses.*

Figure 9.4 shows the average time of arrival plotted against the standard deviation of time of arrival for the same impulse response data. In the majority of cases, the standard deviation varies between 11ms and 28ms. The average time of arrival varies between 30ms and 76ms. The impulse responses identified previously have been circled to show where they lie on this scale. It can be seen that in terms of average time of arrival there is a range of $30.4ms$ between the lowest and highest values. The standard deviation occupies a range of $12.2ms$.

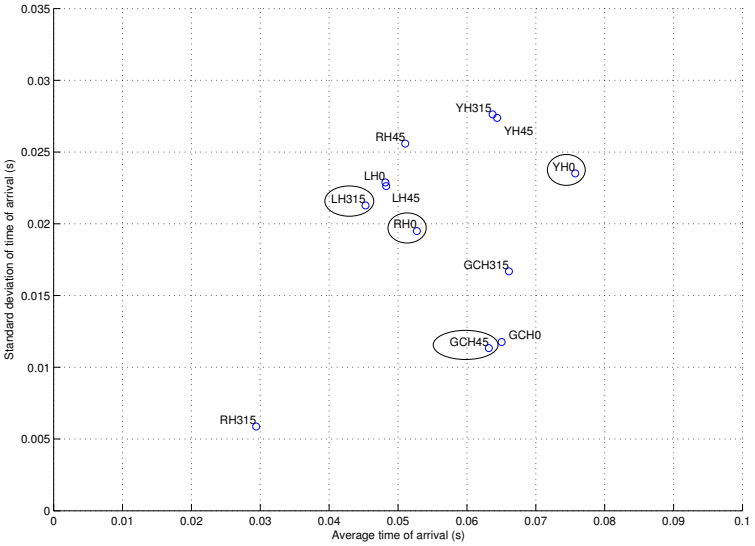


FIGURE 9.4: *Scatter plot showing the average time of arrival versus the standard deviation of time of arrival for measured impulse responses identified as being suitable for auralisation. Encircled points identify impulse responses that were selected for the listening tests.*

Figure 9.5 displays the azimuth and elevation of the mean resultant vector for each stimulus. It can be seen that the elevation values fall within $\pm 20°$. The azimuth varies more widely between approximately $\pm 100°$. Similarly, the impulse responses identified previously have been circled. It can be seen that the impulse responses do not vary a great deal in terms of elevation but occupy a wide range of values in terms of azimuth. As discussed in previous reports, the mean azimuth appears to vary with source orientation which can be seen here also. For example, measurements made with a source orientation of 315° appear on the left of the plot while measurements made at an orientation of 45° appear on the right of the plot.

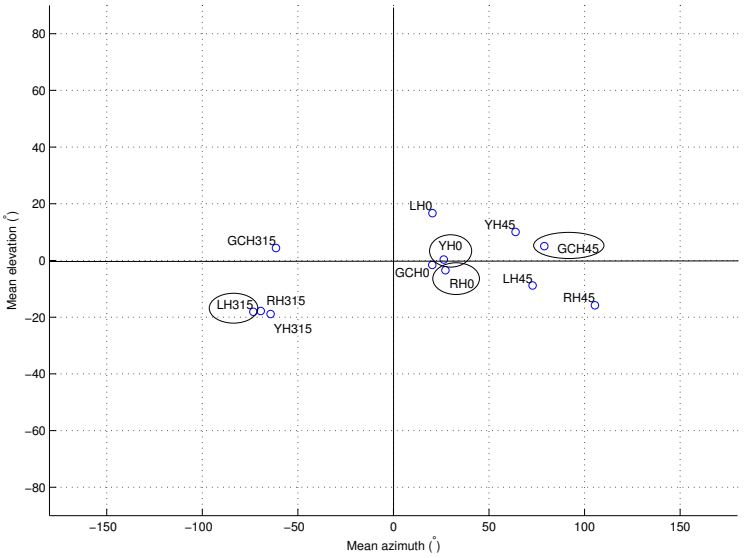
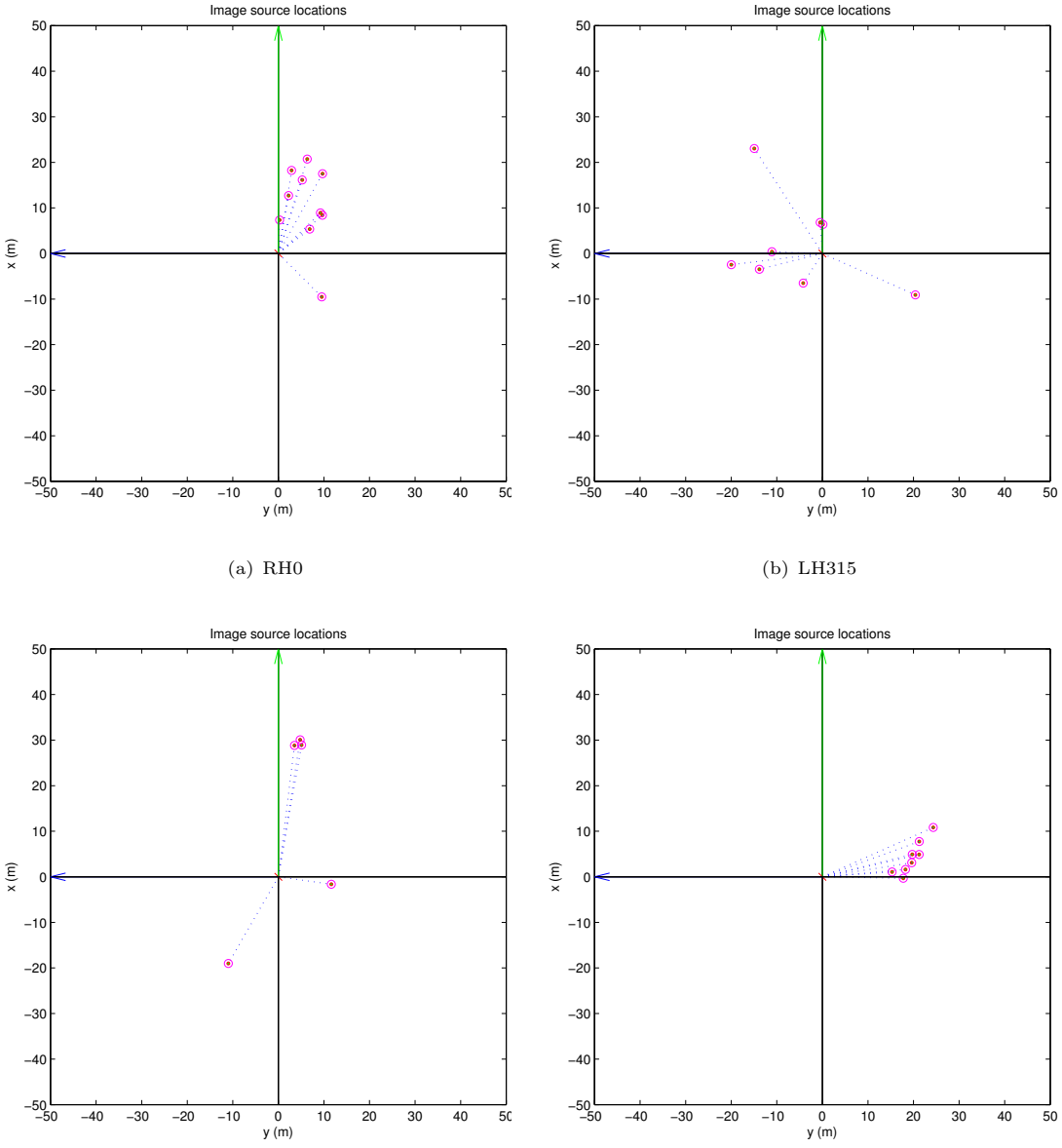


FIGURE 9.5: *Scatter plot showing the mean azimuth vs mean elevation of reflections for measured impulse responses identified as being suitable for auralisation. Encircled points identify impulse responses that were selected for the listening tests.*

Image source plots for the selected stimuli are shown in Figure 9.6 in plan view where the y-axis (shown in green) points to the stage front direction and the x-axis (shown in blue) points towards stage left. In all plots the points represent the spatio-temporal location of a reflection. Both axes are in metres where the distance from the origin represents the time of flight of the reflection.

As reflected in Figure 9.4, impulse responses LH315 and YH0 display the largest angular spread. This can be seen in Figures 9.6(a) and 9.6(d) where reflections are clustered towards a particular direction. Figures 9.6(b) and 9.6(c) however show reflections are spread out more in space.



(a) RH0

(b) LH315

(c) YH0

(d) GCH45

FIGURE 9.6: *Image source plots of the selected stimuli shown in plan. The green axis points to the stage front direction and the blue axis points towards stage left. In all plots the points represent the spatio-temporal location of a reflection.*

Hence, for this test, four impulse responses were selected that were presented at two levels of ST_{early}. These impulse responses were chosen from appropriate locations, source orientations and venues. The parameters that vary the most are t_{mean}, θ_{mean} and angular spread. The standard deviation of time of arrival (t_σ) and mean resultant elevation ϕ_{mean} vary far less. Overall, these impulse responses were considered appropriate for the main listening test.

9.3.4 Stimulus creation and calibration

The stimuli were processed offline using SIRR as described in Chapter 8. Impulse responses for each loudspeaker feed were synthesised based on an intensity analysis of the measured B-format impulse response. Impulse responses were rendered at a sampling frequency of 44.1kHz and a 32-bit floating point resolution. The analysis was computed using an STFT with a window size of 32 samples and a hop size of 8 samples. The direction of arrival estimate was performed on frequencies below 5kHz and the high frequency parts of the signal spatialised to a direction extrapolated from the lower frequencies. The synthesised impulse responses utilised VBAP to spatialise the non-diffuse reflections and phase randomisation to render diffuse parts of the impulse response. Each impulse response was silenced from $t = 0$ms to $t = 20$ms and the early and late parts of the impulse response separated using 5ms linear crossfades. The late part of the impulse response was truncated at $t = 2.5$s which was significantly longer than the impulse responses used. The separated sections of the impulse responses were used by different sets of convolvers to enable the early reflections to change independently from the late response.

The auralisation system was set to an I/O buffer size of 128 samples which was determined to be the highest possible value (to encourage system stability) if virtual reflections as early as 20ms were to be recreated.

The stimuli were calibrated in a similar manner described in Chapter 7 where a loudspeaker and Ambisonic microphone were used to excite the virtual space. Adjustments were made to ensure that system latency was reduced as far as possible and reflections were produced at the required level.

The level of ST_{early} was calibrated for each impulse response to be representative of halls that had high and low levels of ST_{early}. It was necessary for the high and low settings to be suitably far apart but occupy a range that would be found in real concert halls. In Chapter 7 it was shown that, due to the limited acoustic response of the SoundLab, there was a lower limit to ST_{early} of -15.4dB when the wooden floor and screen absorption were applied. In Chapter 5 it was shown that small halls tended to exhibit ST_{early} levels that occupied a range of approximately -16dB to -10dB when measured at the down-stage centre position. Furthermore, as discussed in Chapter 2 an informal JND for ST_{early} is generally considered to be 2dB. Therefore a low setting of -14dB and a high setting of -10dB was chosen. The late response, where $t > 100ms$ was calibrated to an ST_{late} level of -14dB which was considered an appropriate level of reverberation found in smaller recital venues.

The levels for early reflections and late reflections were calibrated separately. Table 9.2 shows the ST_{early} levels after the early part of the impulse response was calibrated. These values are averaged over octave bands between 250Hz and 2kHz. It can be seen that the highest deviance from the target level for the low setting of IR2 which produced a value of -14.09dB. The maximum allowable deviance for calibration was set as 0.1dB. The broadband ST_{early} values are also shown for comparison. It can be seen that the broadband level is higher than the average level. This is thought to be due to the contribution above 2kHz which is not included in the average level.

Hall	ST_{early} average (250Hz-2kHz) (dB)	ST_{early} Broadband (dB)
RH (High)	-9.97 ($\sigma = 1.91$)	-9.97
YH (High)	-10.05 ($\sigma = 1.97$)	-9.98
LRR (High)	-9.99 ($\sigma = 2.56$)	-9.91
GCH (High)	-10.05 ($\sigma = 2.28$)	-9.89
RH (Low)	-13.94 ($\sigma = 2.07$)	-13.35
YH (Low)	-14.09 ($\sigma = 2.45$)	-13.31
LRR (Low)	-13.99 ($\sigma = 2.61$)	-13.22
GCH (Low)	-14.04 ($\sigma = 2.14$)	-13.40

TABLE 9.2: *Measured ST_{early} levels after calibration. ST_{early} values shown are averaged over octave bands between 250Hz and 2kHz.*

The ST_{early} levels were chosen to occupy high and low values observed by measurement (as described in Chapter 5). In addition, the low ST_{early} value was limited by the acoustic contribution of the SoundLab (as described in Appendix D). Figure 9.7 shows the measured ST_{early} for the stimuli as measured in the sweetspot of the SoundLab loudspeaker array. It can be seen that at high frequencies the spread of ST_{early} is much greater when the responses are set high. For example at a frequency of 8kHz, the range of values is 5.2dB for the high setting whereas it is 1.4dB when reflections are set low.

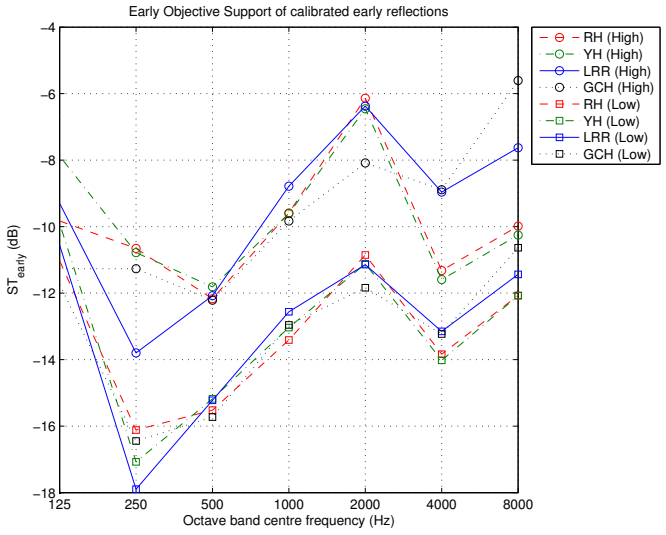


FIGURE 9.7: *This plot shows the measured ST_{early} of each hall after the early reflections have been calibrated. The calibration was achieved by adjusting the gain so that the average ST between 250Hz and 2kHz was either -10dB (when set high) or -14dB (when set low).*

9.3.5 Discrimination test

The discrimination test aimed to determine if musicians can distinguish variations in spatial or temporal distributions of early reflections. This test was also administered to provide post-test screening data as it was not possible for the musicians to attend more than one session.

This test utilised an ABX paradigm (duo-trio test in balanced reference mode (International Organisation of Standardisation, 2004)). In each trial the participant is presented with three stimuli where one is clearly marked as reference X. A test stimulus and a hidden reference are randomly referred to as either 'A' or 'B'. After listening to all stimuli the participant is asked to determine which sample, 'A' or 'B', is the same as the reference X. In this task, the participant was asked to compare different virtual concert hall stages by playing their instrument in each one. The participant used an iPad interface to select the hall they were playing and to record their answer. A screen shot of the interface is shown in Figure 9.8. It can be seen that the buttons changed colour depending on which hall was selected.

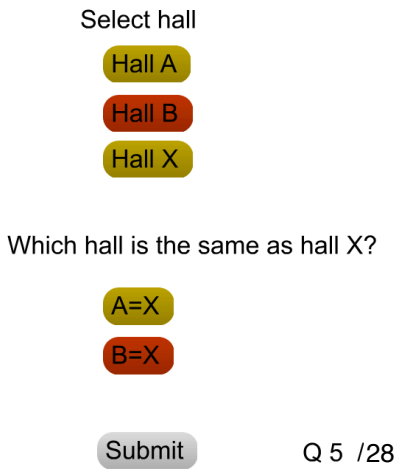


FIGURE 9.8: *Screenshot of the iPad interface used in the discrimination test. The participant uses the upper buttons to select which hall they are playing in and the lower row to record their answer.*

Four additional 'null' tests were added in this experiment which presented participants with exaggerated differences between the concert halls. These comparisons varied both the spatio-temporal distribution of early reflections and the level of ST_{early}. The responses from these tests were not included in the analysis.

The number of correct responses can be considered for each question to determine the likelihood that all participants could perceive a difference between the samples. It is also possible to compare the number of correct responses per participant against a fixed threshold to evaluate their ability to discern differences. The responses from test participants were collected as nominal data i.e. correct or incorrect. Therefore it was most appropriate to summarise the test results by the number of participants that correctly identified the test stimulus in each question. For each participant, the preference and attribute responses were considered valid if they could demonstrate that they could discern the differences between halls. Therefore, it was considered that each participant must obtain a score of 50% or higher on comparisons of halls at a high level of ST_{early}.

9.3.6 Preference test

The preference test aimed to determine if spatial or temporal distributions of early reflections are a salient perceptual dimension driving musician preference for a particular venue. The test allowed a comparison of preference patterns for a subpopulation of musicians to determine if they had similar reactions to the stimuli. Furthermore, the results were processed to determine if musician preference patterns correlated with a particular spatial/temporal parameter. Musicians compared pairs of concert halls and were asked to decide which one they preferred, resulting in a rank order of stimuli. Multivariate analysis methods were then used to determine individual preference patterns for each musician. Similar techniques will be used to determine a correlation between musician preference and objective descriptions of the stimuli. The participant used the iPad interface to select the hall they were playing and to record their answer. A screen shot of the interface is shown in Figure 9.9. When a hall or answer is selected, the button turns red.

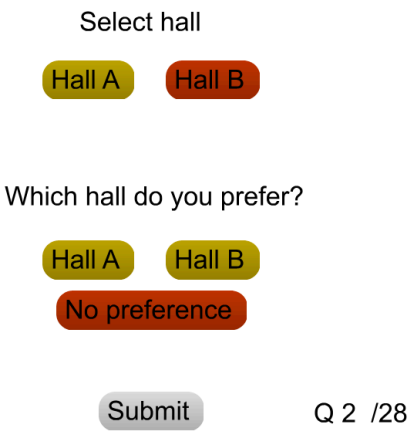


FIGURE 9.9: *Screenshot of the iPad interface used in the preference test. The participant uses the upper buttons to select which hall they are playing in and the lower row to record their answer.*

The pair-wise method of comparing each virtual stage was considered appropriate in this context due to the number of stimuli and the required effort for the participant to remember each hall. A ranking paradigm may have lengthened the test considerably and would have required a slightly more complex question.

In order to compare each hall with every other hall once, $\frac{N(N-1)}{2}$ comparisons were required, where N is the number of halls. This resulted in 28 comparisons for this task. Unlike the previous test, comparisons were presented where the early objective support varied in addition to the spatio-temporal distribution of early reflections.

9.3.7 Attribute test

The perceptual attribute test aimed to study the effect of early reflection distribution on a series of predefined attributes. Each participant was presented with concert hall stimuli, one at a time,

and asked to rate the intensity of different subjective attributes on a semantic differential scale. The scales were implemented as faders which varied as float values between values of 0 and 100, a numerical value was displayed as the fader was moved. At the beginning of each new trial, all faders were reset to a midway position. The participant used the iPad interface to control the test. A screen shot of the interface is shown in Figure 9.10.

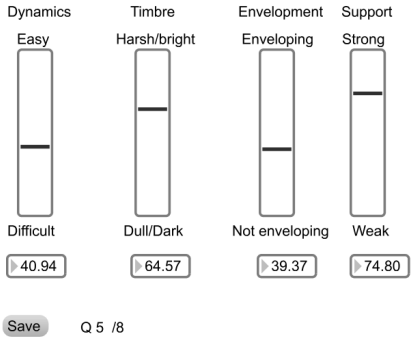


FIGURE 9.10: *Screenshot of the iPad interface used in the Attribute test. The participant uses the sliders to record the subjective intensity of each attribute.*

The participant was instructed to play as long as they liked in each hall before recording their responses. In order to minimise the test length, the participant was not allowed to go back and change their answers. In this test, there were eight trials in addition to three null questions at the beginning of the test. The null trials were identified as practice questions so that the participant could get used to using the scales prior to responding. The null trials were selected so that they occupied a range of objective attributes.

The participant was presented with textual definitions of each attribute prior to the beginning of the experiment. It was common for the participant to have studied the definitions prior to arriving at SoundLab or during the preceding 10 minute break. These subjective attributes were chosen due to their importance for soloists as described by Gade (1982) with the exception of *Envelopment* which was included to account for the spatial variation of early reflections. The definition of each subjective attribute is as follows and were inspired by previous experiments performed by Guthrie (2014):

- **Dynamics** - Ease of varying the dynamic range of your instrument (does *forte* sound loud and *piano* sound soft?)
- **Timbre** -The tonal quality of your instrument and the hall acoustics combined.
- **Envelopment** -The extent to which you feel your are enveloped by the hall response (Does the acoustic response surround you or come mainly from a single direction?)
- **Support** -Sense of how strongly your efforts feel supported by the stage (A weak sense of Support makes it feel as though you have to exert more effort for what you hear back from the hall).

The anchors used for each attribute were as follows and inspired by previous work by Gade (2013):

Difficult - **Dynamics** - Easy

Dull/Dark - **Timbre** - Harsh/bright

Not enveloping - **Envelopment** - Enveloping

Weak - **Support** - Strong

Normally, in this type of test, the attributes are assessed using a panel of expert assessors while the the preference judgements are assessed using a sample of consumers. In some cases, this is argued to be an invalid approach as there could be a disconnect with how the expert assessors and the consumers perceive each stimulus. In this context, it was considered that, due to their previous experience performing with their instrument, that the musicians represented both expert listeners and the consumer population.

9.3.8 Written responses

After the main part of the experiment, each participant was asked to provide some textual feedback regarding their experiences in the test and with stage acoustics in general. Participants were normally left alone to complete the form which took approximately 10 minutes.

These questions were selected to gain additional insight as to the type of acoustics the musicians of different instruments find favourable and how they tend to adjust their sound in acoustic environments. In addition, this gave the participants an opportunity to feedback any views that were not possible to collect during the listening test. Each participant was asked the following questions:

- *From your previous experience, which venue has the best acoustic conditions for a solo performance? Please explain why you prefer this particular venue.*
- *Do you find it necessary to adjust any aspects of your technique when performing in different venues? If so, please describe the type of adjustments you might make.*
- *In the listening test, did you notice any differences between the concert halls you played in? If so, please describe any differences you heard.*
- *In relation to stage acoustics or the experiment you have just completed, do you have any further comments you would like to add?*

9.4 Results

The results of the listening test are presented in the following sections. The results of each test will be summarised before presenting the results of a multidimensional analysis technique which makes use of multiple datasets.

The results of the listening test are presented in the following sections where each section of the test will be discussed separately. The results of the preference test contribute to two parts of the experiment where these results are considered in relation to the objective attributes of each concert hall in addition to the subjective attributes obtained in the listening test.

9.4.1 Discrimination test

The results of the discrimination test provide will help determine if varying particular aspects of the impulse response are audible to a performing musician. In addition, the results can serve as a post-hoc screening test, removing participants from further tests who could not discern differences between halls when rendered at a high level of ST_{early}

Table 9.3 shows the number of correct answers for each comparison when the halls are rendered at a high level of ST_{early}. For example, when the reference X is RH(High) and is compared with itself and GCH(High), there are 9 correct detections.

Circled values indicate the number of detections is significant to $p < 0.05$. By using the binomial distribution as discussed in Chapter 8, the number of correct detections for a sample size of 16 was calculated to be 11 or greater to be considered statistically significant. Significant detection rates were found when GCH(High) was compared with YH(High) and also when YH(High) was compared with LRR(High) and GCH(High).

Test Hall	Hall X			
	RH(High)	YH(High)	LRR(High)	GCH(High)
RH(High)	-	10	7	7
YH(High)	10	-	8	(12)
LRR(High)	6	(13)	-	8
GCH(High)	9	(14)	8	-
$mean = 9.8(\sigma = 2.53)$				

TABLE 9.3: *Number of correct answers for each hall rendered at a high setting of ST_{early}. Circled values indicate significant detections ($p < 0.05$)*

Similarly, Table 9.4 shows the number of correct detections when halls are rendered at a low setting of ST_{early}. It can be seen that at lower levels of ST_{early} there are no statistically significant detections. However, the results do not show a simple reduction in detection rates among all comparisons. For example, a comparison of LRR (Low) and RH (Low) show more detections when ST_{early} was set low than when ST_{early} was set high.

Test Hall LRR(Low)	GCH(Low)	Hall X	RH(Low)	YH(Low)
RH(Low)	-	6	10	7
YH(Low)	9	-	8	6
LRR(Low)	8	5	-	7
GCH(Low)	10	7	10	-
$mean = 7.8(\sigma = 1.71)$				

TABLE 9.4: *Number of correct answers for each hall rendered at a low setting of ST_{early}. Circled values indicate significant detections ($p < 0.05$)*

These results also show the mean result for both settings and it can be seen that a high level of ST_{early} results in a slightly higher mean detection rate ($mean = 9.8$) than when the responses are rendered at a low level of ST_{early} ($mean = 7.8$). This suggests that the effect of varying the distribution of early reflections is more prominent at high levels of ST_{early}. To test this further, the results were collapsed over all halls at each level of ST_{early} and tested to determine if the different levels of ST_{early} result in significant detection rates. The results, shown in Table 9.5, indicate that halls rendered at either level of ST_{early} could not be detected to a statistically significant level ($p < 0.05$).

ST_{early}	N correct	N total	$p - value$
High	112	384	1.0
Low	98	384	1.0

TABLE 9.5: *Number of correct responses for Bassoon and Flugelhorn where all other levels have been collapsed*

The post-hoc screening criterion was based on the comparisons where ST_{early} was set high as these comparisons should be easier for participants to detect. If participants could not detect differences at a high level of ST_{early} it is unlikely that they will be able to do so at low levels. Therefore, participants with less than 50% correct (less than 6) at high levels of ST_{early} will be excluded from further analysis. Table 9.6 shows the number of correct answers for each participant for high and low settings of ST_{early}. It can be seen that at the high setting all participants except from P1 and P8 attained a score of 50% or higher. Participant P8 was the only one however to attain a higher number of correct responses for the low setting. From these results it can be seen that participants P1 and P8 could not reliably detect differences between the stimuli and so will be excluded from the remainder of the analysis.

Participants	High ST_{early}	Low ST_{early}
(P1)	(4)	(4)
P2	8	9
P3	6	5
P4	9	4
P5	8	8
P6	8	8
P7	7	5
(P8)	(2)	(7)
P9	7	11
P10	8	6
P11	6	2
P12	11	7
P13	7	7
P14	7	5
P15	6	7
P16	8	10

TABLE 9.6: *Number of correct responses for each participant when halls were set to a low and high ST_{early} level. For musicians to be included in the results they must attain 50% (i.e. 6 or more) or higher when ST_{early} is set high. It can be seen that participants P1 and P8 score below this threshold and so will not be included in the remaining analysis*

It can be seen that the highest score was obtained by P12 who played Classical Guitar. Due to the way in which this instrument is played, it was possible for this participant to play a sustained chord and then use their free hand to switch between halls while the chord sounded. In addition, this participant was able to move their head relative to the instrument which may have aided in determining the difference between halls. It should be noted that P14 also played Classical Guitar, however this participant did not change the concert hall during a sustained note.

In summary, this test allowed the participants to be screened to ensure that the remainder of the analysis is applied only to participants who could hear differences between reflection distributions. The results also indicate that participants could perceive differences between certain concert halls that have the same level of ST_{early} but different distributions of early reflections. When results were considered for all concert halls together at the two levels of ST_{early}, no significant detection rates were found which implies that, even at high levels of ST_{early} differences between concert halls can be difficult to detect.

It is of interest to determine if the different concert halls presented caused a change in musician preference, especially in relation to their objective and subjective attributes. The following analysis will therefore focus on the participant's preference towards each concert hall.

9.4.2 Preference test

The preference test aims to determine how differences in ST_{early} and reflection distribution affect musician preference. From a series of paired comparisons, it is possible to determine a preference score and rank order of each hall for each participant. This was achieved using an adaption of a method demonstrated by Gade (1982).

For each musician, the preference responses are entered into a matrix $\boldsymbol{X_{i,j}}$ where columns and rows both show each concert hall. An example of this is shown in Table 9.7, where a preferred hall receives a score of 1 and if neither hall is preferred both halls receive a score of 0.5. In this example, Hall 5 was preferred over Hall 4 ($\boldsymbol{X_{4,5}} = 1$). The columns of the matrix are summed and normalised by the number of stimuli to provide the preference score for each hall for each participant. The responses in each column were summed to give s_j which describes the number of times soundfield j has been preferred to other soundfields subtracted by the number of times it has been rejected. These values are normalised by $M - 1$ where M is the number of halls to give the preference score, a_j which varies between 0 and 1.

Halls	1	2	3	4	5	6	7	8
1	0	0	0.5	0.5	1	1	0.5	1
2	1	0	0	0	1	0.5	0	0.5
3	0.5	1	0	0.5	1	0	0.5	0
4	0.5	1	0.5	0	1	1	1	0.5
5	0	0	0	0	0	0.5	1	0
6	0	0.5	1	0	0.5	0	1	1
7	0.5	1	0.5	0	0	0	0	1
8	0	0.5	1	0.5	1	0	0	0
s_j	2.5	4	3.5	1.5	5.5	3	4	4
a_j	0.36	0.57	0.5	0.21	0.79	0.43	0.57	0.57

TABLE 9.7: *Example of how pair-wise preference comparisons can be converted into a rank score using a score matrix. Scores are entered into the matrix indicating the participants preference in each comparison. The columns are summed and normalised to obtain a value between* 0 *and* 1 *for each hall.*

By using Thurstone's Law of Comparative Judgment (Thurstone, 1994), it is possible to determine if one hall was preferred significantly more than others. This method is based on the notion that the proportion of times a stimulus is judged greater than another is determined by the degree to which sensation A and sensation B differ. It further assumes that repeated presentation of a stimulus will result in a range of responses that follow a normal distribution (discriminal process). Therefore, the perceived difference between two stimuli can be evaluated by the distance between the means of each stimulus response distribution. This model requires specific knowledge of the correlation and standard deviation which cannot be measured directly. A number of simplifications have been developed which make certain assumptions about the shape of each distribution. Case V is the most often used which assumes that each stimulus produces an equal standard deviation and is not correlated with the other stimulus in each pair. This allows the distance between each discriminal process to be calculated using (Ramamurti, 2014):

$$\mu_B - \mu_A = z_{BA}\sqrt{2} \tag{9.1}$$

where μ_A and μ_B are estimates of the scale position of stimulus A and B, and z_{BA} is the z-score computed from the proportion of times stimulus B is chosen over A. This method requires a summation of all the preference matrices for each participant and dividing by the number of

participants. This allows each stimulus to be positioned on the continuum by converting the observed proportions of preference to z-scores. The average z-score is calculated for each stimuli, which indicates how consistently each stimulus has been rated against all others. These are then adjusted to ensure all scores are positive.

Figure 9.11, summarises the preference responses from the participants for each virtual concert hall as mean and standard deviation. It can be seen that the mean responses for each hall are close to a value of 0.5 and generally show a wide spread of responses. This implies that there was a high degree of disagreement between participants regarding which hall was preferred. Despite this, the results show that GCH(L) was the most preferred hall and YH(H) was the least preferred.

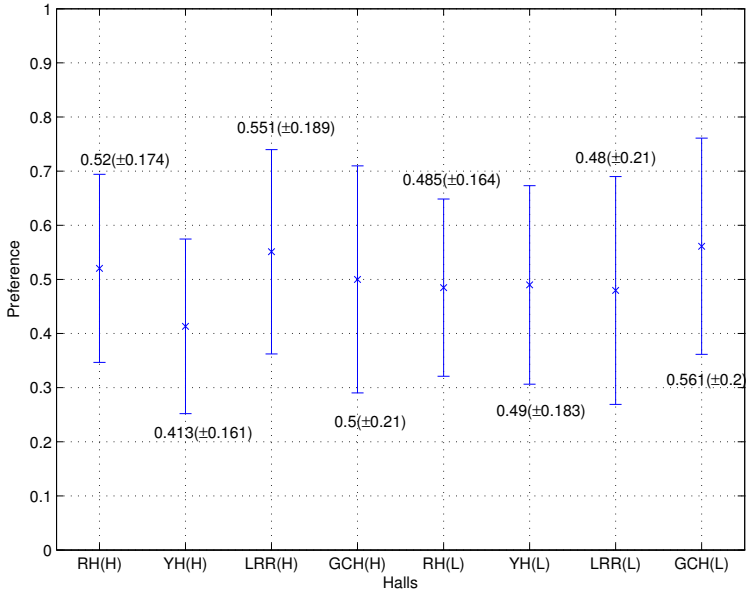


FIGURE 9.11: *Summary of preference responses for panel for each virtual concert hall shown as mean and standard deviation.*

Figure 9.12 shows the results for z-score of preference of the whole panel and represents how consistently each hall has been preferred over each of the others. As discussed, the results are adjusted according to the lowest z-score so that all values are positive and the lowest value equals zero. As demonstrated by Wankling et al. (2012), the z-score can be used to determine if there is a significant difference in preference between concert halls. A z-score of less than 1.96 will result in a p-value of greater than 0.05 whereas a z-score of greater than 1.96 indicates a p-value of less than 0.05. In Figure 9.12, it can be seen that there are no values greater than 1.96 which indicates that this panel of musicians did not prefer one hall significantly more than the others.

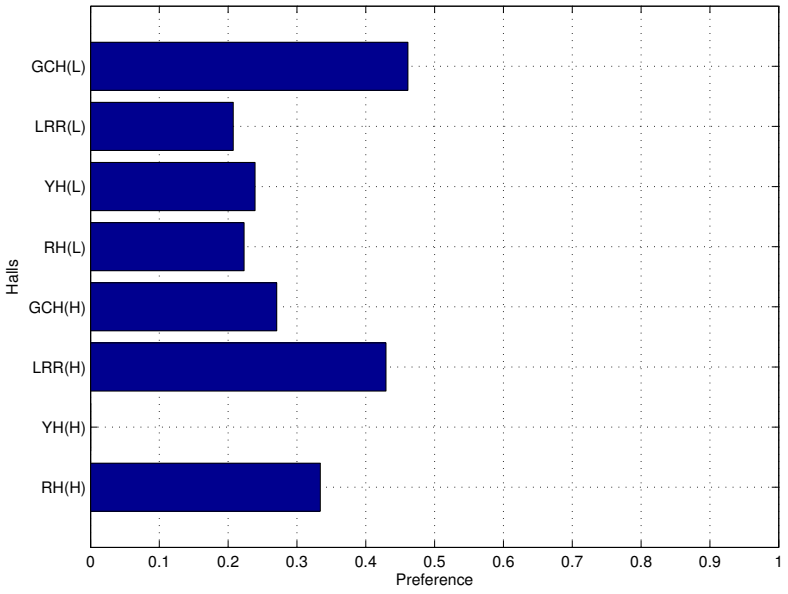


FIGURE 9.12: *Summary of preference responses to each stimuli as computed using Thurstone's Law of Comparative Judgement (Case V). Scores have been adjusted such that the lowest score is* 0 *and all others are positive.*

To determine if the spatio-temporal characteristics or ST_{early} had a significant effect on musician preference, a two-way ANOVA was conducted where the factors were the Hall (four levels, each with different spatiotemporal characteristics) and ST_{early} (two level, high and low settings).

Prior to performing the ANOVA analysis, the preference responses for each hall were checked for normality and homogeneity of variance using Bartlett's Test. The results of Bartlett's test produced a value of (1.904) with an associated p-value of (0.965). This tests the null hypothesis that the data (for each hall) comes from a normal distribution with equal variance. The high p-value supports this null hypothesis. This suggests that the data is appropriate for analysis using ANOVA.

Table 9.8 shows the results of a two-way ANOVA where the independent variables are Hall (Columns) and ST_{early} (rows). The dependent variable is musician preference. It can be seen from the $Probs > F$ column that there is no statistically significant effect of either ST_{early} or Hall type on the preference judgements of the panel. Therefore, the null hypothesis that musician preference is affected by ST_{early} or the spatio-temporal distribution of reflections cannot be rejected. This reflects the responses summarised previously in Section 9.4.2 where the mean and standard deviation of preference responses was seen to be very similar for all stimuli.

Source	SS	df	MS	F	$Probs > F$	η_p^2	$Power(1-\beta)$
Halls	0.099	3	0.033	0.94	0.425	0.025	0.062
ST_{early}	0.002	1	0.002	0.05	0.829	0.0004	0.050
Halls * ST_{early}	0.110	3	0.037	1.05	0.375	0.028	0.064
Error	3.647	104	0.035				
Total	3.857	111					

TABLE 9.8: *Results of two-way ANOVA where Columns are each hall and rows are settings of ST_{early}. SS denotes the Sum of Squares, df denotes degrees of freedom, MS denotes Mean Squares (SS/df), F is the F statistic, $Prob > F$ is the p-value associated with a hypothesis test that the row, column or interaction effects are all the same. η_p^2 is a measure of the proportion of variance for each effect. $Power(1-\beta)$ is an estimate of the test power.*

A post-hoc power analysis was conducted with the program G*Power (Faul et al., 2007) with the results displayed in Table 9.8. The effect size, $\eta_p^2 = \frac{SS_{effect}}{SS_{effect}+SS_{error}}$ was calculated for each effect and interaction which allowed an estimation of test power. It was found that the largest power for the chosen significance level ($\alpha = 0.05$) was $Power(1-\beta) = 0.064$ which suggests the non-significant results are due to insufficient test power, rather than through a measured non-effect. This may have been primarily due to the modest sample size of the panel or the musicians using different criteria for deciding preference for a particular hall.

To examine the preference responses of this panel further, external preference mapping was used to project each concert hall into a product space based on their objective or subjective attributes. In a product space, a close proximity between halls suggests those halls were preferred similarly by the panel. By regressing in the objective and subjective attributes as vectors, it is possible to identify the main differences between each stimuli. Then, by regressing each participant's responses into this space as vectors (where a participants preference towards a hall can be projected onto each vector), it is possible to determine the dimensions with which preference judgements were being made.

9.4.3 Objective parameters

An external preference map was produced by performing a standardised PCA on a matrix, $\boldsymbol{X_{i,j}}$, containing the objective attributes (i - columns) for each hall (j -rows). Each participant's preference for each hall was regressed onto this product space as a supplementary quantitative variable. This preference map produces a visualisation of participant preference in relation to the objective parameters which will help to evaluate the likelihood that preference was related to a particular objective parameter. The processing was performed using the *FactoMineR* package (Le et al., 2008) in *R* (R Core Team, 2013).

Table 9.9 shows the eigenvalue and percentage variance accounted for by each dimension. Two higher dimensions were found to account for negligible variance. It can be seen that nearly 98% of the variance is accounted for by the first three dimensions, in addition to the eigenvalue dropping below 1.00 at Dimension 4. This suggests that the product space can be described by three dimensions without any significant loss of information.

Dimension	Eigenvalue	(%) Variance	(%) Cumulative variance
1	3.51	58.5	58.5
2	1.36	22.7	81.17
3	1.00	16.7	97.83
4	0.13	2.16	100.00

TABLE 9.9: *Summary of variance for each dimension of the PCA. Table shows the eigenvalue, % variance of each dimension and the cumulative variance.*

In order to identify which objective attributes might be driving the participant's preference, the results of the PCA are shown in Figure 9.13 which show the location of each hall in terms of each objective parameter. It is possible to identify these dimensions by evaluating the position of each hall in relation to the objective parameters (highlighted in Table 9.1). By regressing the participant's preference vectors onto the product space,

On Dimension 1, RH and YH are very similar, yet there is a large difference between LRR and GCH. This appears to reflect the values of θ_{mean} where halls were chosen that had reflections arriving mainly from stage left or stage right or directly ahead. On Dimension 2, GCH and LRR appear to be almost identical with RH located close to these halls. In contrast YH is very different, occupying a position on the positive side of Dimension 2. This appears to reflect how the halls differ in terms of t_{mean} where YH has a much higher value than the others. Figures 9.13(a) and 9.13(b) clearly show halls are separated by ST_{early} along Dimension 3.

Figures 9.13(c) and 9.13(d) show that Dimension 1 has the highest correlation with θ_{mean} ($Corr(Dim.1, t_{mean} = 0.98, p = 0.00001)$, Dimension 2 has the highest correlation with t_{mean} ($Corr(Dim.2, t_{mean}) = 0.82, p = 0.012$) and Dimension 3 has the highest correlation with ST_{early} ($Corr(Dim.3, ST_{early}) = 1, p = 0$. This suggests that Dimension 1 is θ_{mean}, Dimension 2 is t_{mean} and Dimension 3 is ST_{early}

In reference to Dimension 1, it can be seen that approximately half of the participant preference vectors point in the positive direction and half point in the negative direction. It can be seen that in either direction there is a mix of different types of instrument (i.e. strings, brass etc). This suggests that some of the panel preferred reflections arriving from a particular direction and that this was not driven by musical instrument family. The panel also appeared to be divided regarding their preference towards high or low ST_{early} (Dimension 3) and similarly, this did not appear to be related to the musical instrument family.

On Dimension 2, it can be seen that 9 of the 14 participants have preference vectors that are oriented in the negative direction of Dimension 2. This suggests that these participants preferred reflections to arrive earlier rather than later. Overall, it can be seen that the majority of preference vectors for the panel are oriented away from YH and towards RH, GCH and LRR. *Vln4* and *Sax1* are the exceptions where they appeared to show a stronger preference for YH(L).

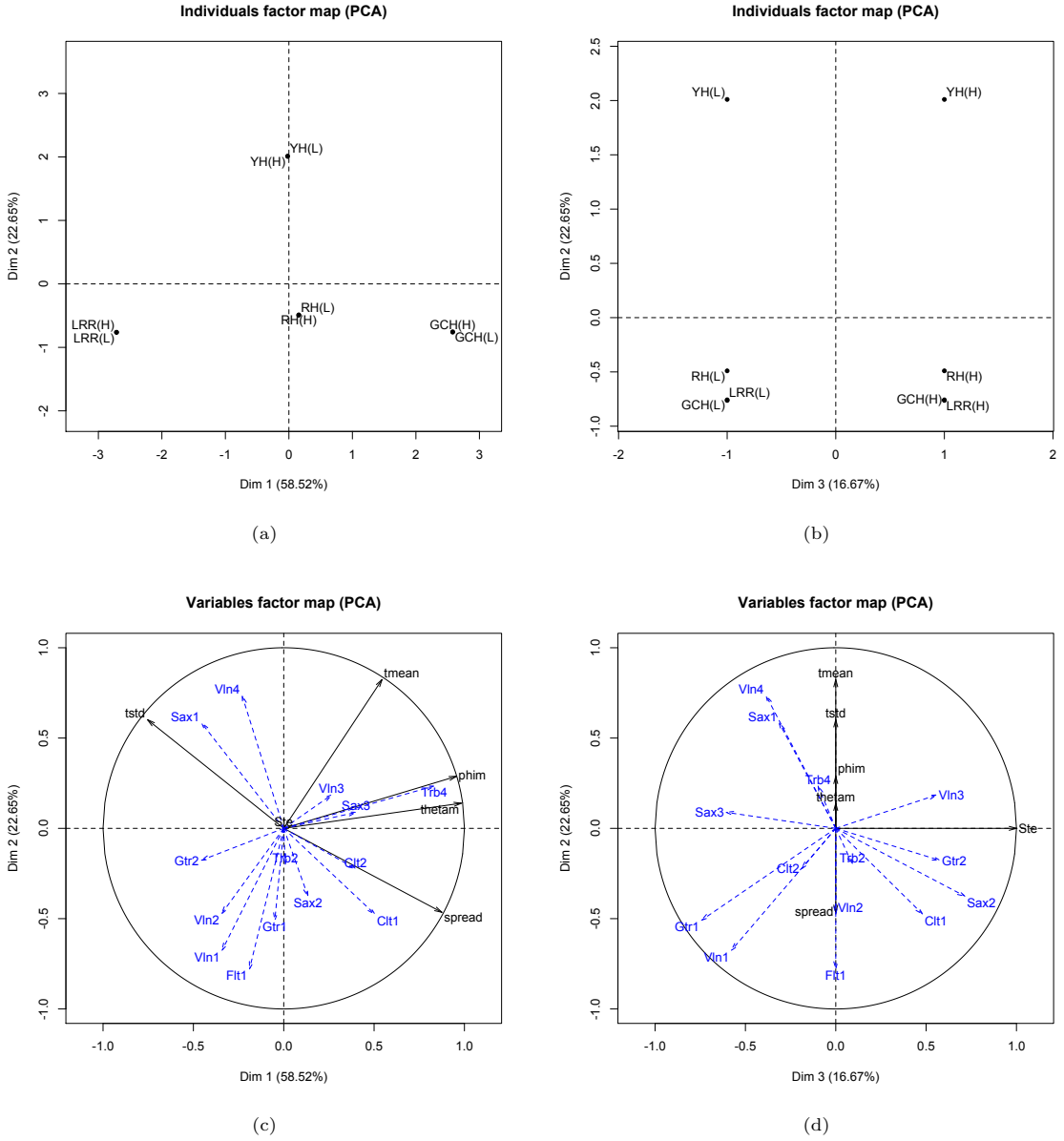


FIGURE 9.13: *PCA of objective parameters with regressed preference data for each participant. 9.13(a) and 9.13(b) show the location of each hall in the product space and 9.13(c) and 9.13(d) show the preference of each participant regressed into the product space. Each participant is shown as a dashed vector whereas the parameters are shown as solid vectors.*

By plotting the preference scores in relation each of these parameters it is possible to determine how these objective acoustic parameters influenced the preference of the participants. Figure 9.14 shows the mean preference scores plotted against θ_{mean} in addition a quadratic fit has been applied. The plot shows that lower preference scores occur when the majority of reflections arrive from stage front and that that higher preference scores appear to occur when reflections arrive from lateral directions. The quadratic regression results in a coefficient of determination, $R^2 = 0.299$, which suggests a quadratic model does not fully describe participant preference in response to θ_{mean}.

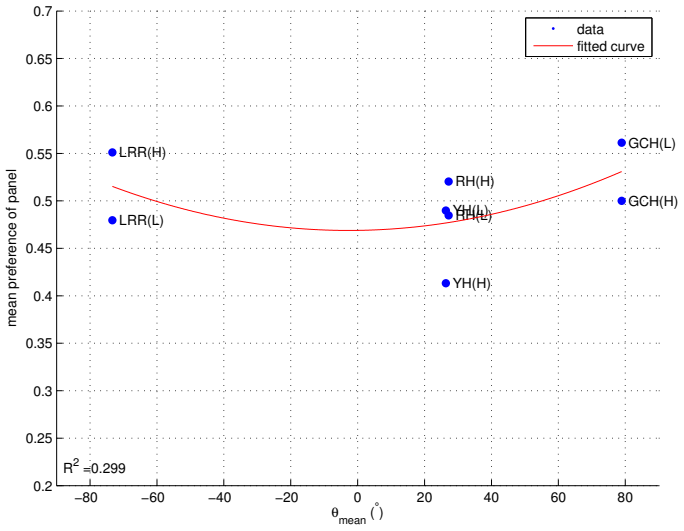


FIGURE 9.14: *Plots showing mean musician preference for each concert hall plotted against the objective parameter θ_{mean}. A quadratic fit has been applied where the coefficient of determination, R^2, is equal to 0.299.*

Figure 9.15 shows the mean preference scores plotted against t_{mean}. This parameter was found to correlate well with Dimension 2 of the PCA. It can be seen that the highest preference score tends to correspond with halls that have a lower value of t_{mean}. This implies that the panel preferred conditions where reflections arrived earlier in time. This plot is also overlaid with a quadratic fit which can be seen to decrease after $t = 60ms$. This highlights that reflections arriving close to 60ms are most preferred but reflections arriving later than this are less favoured. Likewise, the coefficient of determination, ($R^2 = 0.366$), indicates a quadratic model is a poor descriptor of the results. This may be in part due to the differences in preference for low and high ST_{early} but also due to the relatively low value of preference for RH in comparison with GCH and LRR.

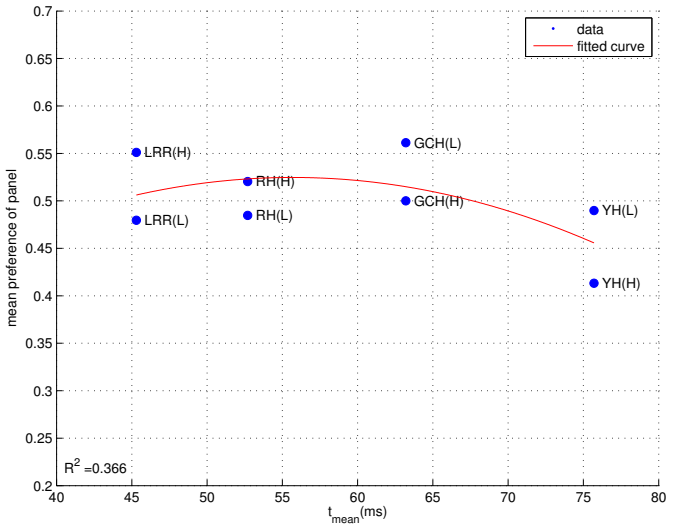


FIGURE 9.15: *Plots showing mean musician preference for each concert hall plotted against the objective parameter t_{mean}. A quadratic fit has been applied where the coefficient of determination, R^2, is equal to 0.366.*

Figure 9.16 shows the mean preference values for each concert hall plotted against the objective parameter, ST_{early}. It can be seen that the panel preference is very similar for both settings of ST_{early}. It can be seen that three of the halls presented at a low ST_{early} (RH(L), LRR(L) and GCH(L)) have almost identical preference values and GCH(L) is preferred most. Halls presented at a high level of ST_{early} appear to produce a higher spread in preference. The variation in preference between halls rendered at the same level of ST_{early} imply that other properties of the soundfield are influencing musician preference.

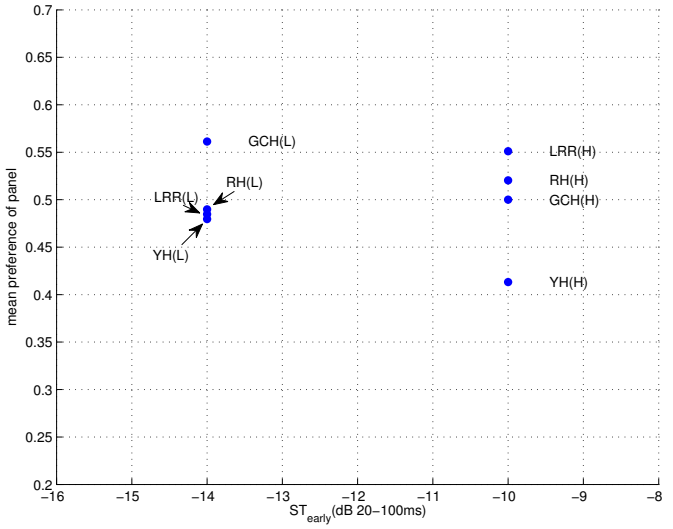
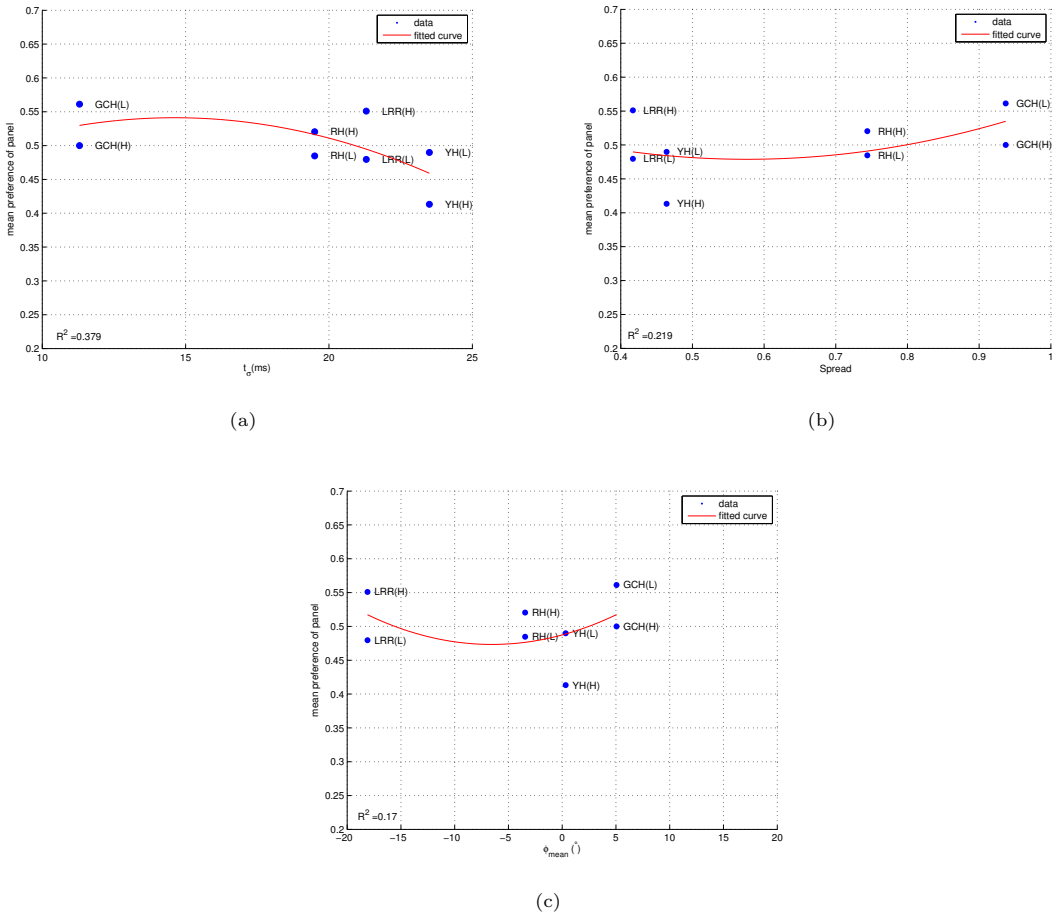


FIGURE 9.16: *Plot showing mean musician preference for each concert hall plotted against ST_{early}.*

For completeness, the remaining objective attributes are plotted against mean preference and are shown in Figure 9.17. In Figure 9.17(a) it can be seen that there is a general decrease in preference as reflections become more spread out in time. Figure 9.17(b) shows a slight increase in preference towards halls where reflections arrive from a single dominant direction (as indicated by a high value of *spread*) rather than from many directions. It can also be seen that there is a trend for an increase in preference towards halls where reflections are elevated above or below the musician as shown in Figure 9.17(c)



(a) (b)

(c)

FIGURE 9.17: *Preference data plotted against t_σ, Spread, and ϕ_{mean}. Each plot is overlaid with a quadratic fit and the coefficient of determination, R^2 shown.*

In summary, applying PCA to the objective attributes of each hall and mapping on each musician's preference response helps to reveal the objective attributes of the concert halls that resulted in the preference responses of the panel. It can be seen that the panel produced a range of preference responses in relation to each subjective attribute and this segmentation was consistent over all musical instrument families. However, it was found that there was a slight increase in agreement when preference responses were observed in relation to t_{mean}. This could account for the panel's general preference towards GCH(L) and away from YH(H).

9.4.4 Sensory Attribute test

A similar analysis can be performed to determine how the halls varied in terms of subjective attributes and how the panel's preference responses related to these attributes. The following will first describe how the panel rated each concert hall in terms of each subjective attribute. External preference mapping will then be applied to examine the responses of each panel member to determine if there was any agreement between them.

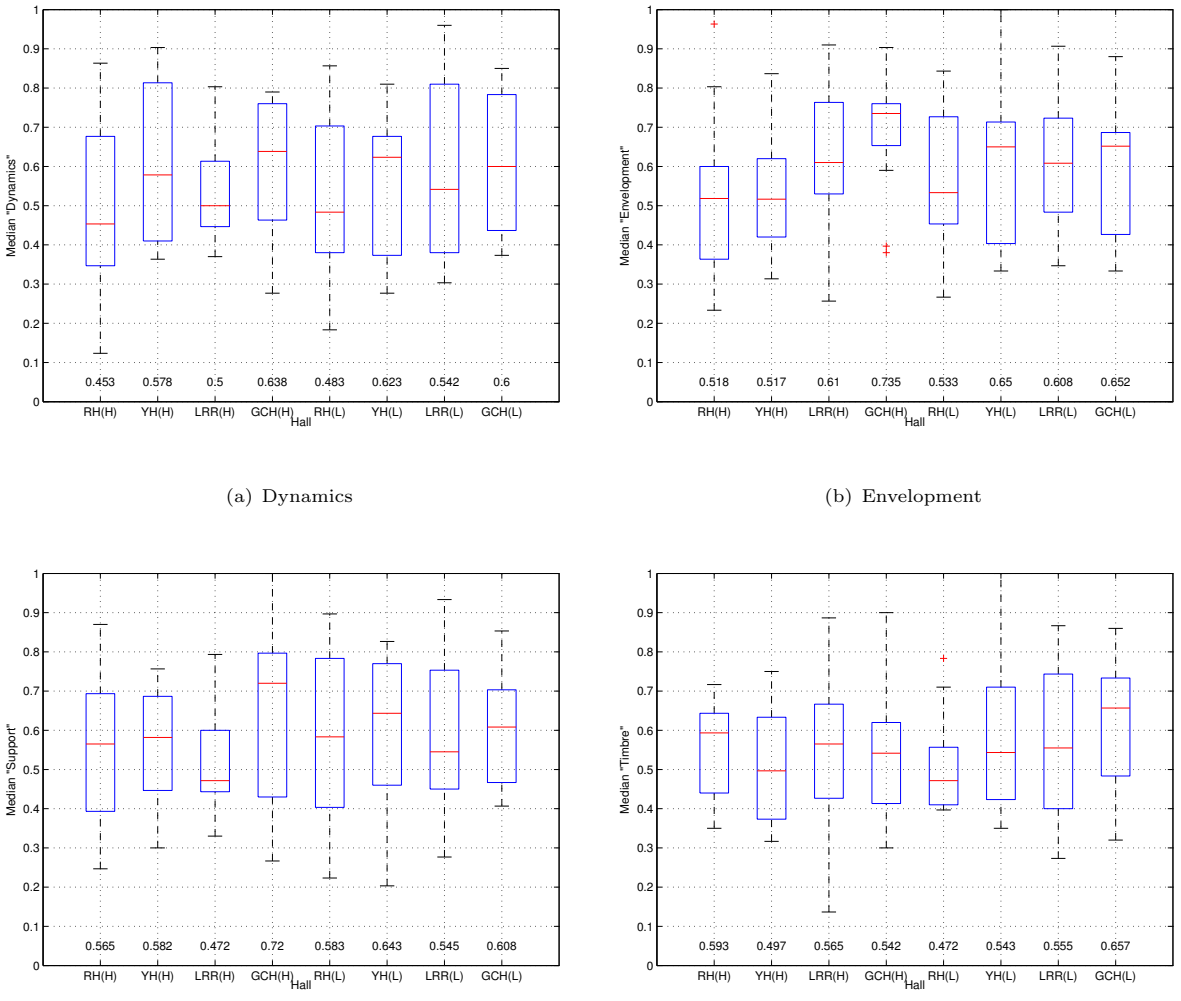
Figure 9.18 shows the panel results for each subjective attribute, summarised as the median (red line) and 25th/75th percentiles (bottom and top of each box). Outliers are shown as red crosses and the median value shown at the bottom of each box plot.

Figure 9.18(a) shows the responses for the subjective attribute *Dynamics*. For this attribute, the anchor points are defined as 0- *Difficult* and 1- *Easy* which refer to how easy or difficult the participants felt it was to vary the dynamic range of their instrument. It can be seen that *YH(L)* and *GCH(H)* were perceived as being the easiest to control the dynamic range, whereas *RH(H)* and *RH(L)* were perceived as being the most difficult.

In Figure 9.18(b) the results are summarised for the subjective attribute, *Envelopment*, where 0- *Not enveloping* and 1 - *Enveloping*. The results suggest that *GCH(H)* was perceived as being the most enveloping and *RH(H)* and *YH(H)* were perceived as being the least enveloping. This is of particular interest as the objective analysis of *GCH(H)* showed that the early reflections appeared to arrive mainly from a lateral direction.

The responses for *Support* are shown in Figure 9.18(c) where 0 - *Weak* and 1 - *Strong* which refers to the extent to which the musician feels supported by the space. It can be seen that GCH(H) was felt by most participants to provide the strongest support. LRR(H) was perceived to provide the least support. It is of interest that the perceived level of support was found to be very similar for halls presented at high and low ST_{early}, suggesting that a change of 4dB is still very subtle. Both the highest and lowest perceived support were found in halls with a high level of ST_{early} which suggests that perceived support is affected by other properties of the soundfield in addition to the overall energy of early reflections. The results are similar to those shown for *Dynamics*. This suggests that these two attributes were perceived to be very similar by the panel.

Figure 9.18(d) summarises the musicians' responses for *Timbre* where 0- *Dull/Dark* and 1- *Harsh/Bright*. It can be seen that GCH(L) appears to produce the harshest/brightest timbre whereas RH(L) produces the Dullest/Darkest timbre. It should be noted that the median response for all halls lies close to 0.5 which suggests that halls were perceived by most participants as being fairly balanced in timbre or slightly bright sounding.



(a) Dynamics

(b) Envelopment

(c) Support

(d) Timbre

FIGURE 9.18: *Summary of musician responses to attribute intensity. Results are summarised as the median (red line) and 25th/75th percentiles (bottom and top of each box). Outliers are shown as red crosses. Figures 9.18(a), 9.18(b), 9.18(c) and 9.18(d) show results for Dynamics, Envelopment, Support and Timbre respectively.*

It can be seen that the spread of responses for each hall is relatively high which implies each participant perceived the hall in different ways or that each scale was being utilised in a different way. As the participants did not have a reference hall for this test, it is possible that some musicians were drawing on comparisons with previous experience or comparing each hall with the previous one they had just played in. Many of the participants commented that this test was somewhat easier than the discrimination or preference tests as they were being instructed what to listen for.

As with the preference responses, it is also of interest to determine if the independent variables have a statistically significant effect on each of the subjective attributes. Therefore, a two-way ANOVA was performed to determine if either the Hall or ST_{early} had a significant effect on panel's judgement of each subjective parameter. Prior to this analysis, the responses for

each subjective attribute were checked to ensure that responses from each hall were normally distributed and had equal variance. This was achieved using Bartlett's test. The results for each parameter are shown below in Table 9.10. The high p-values suggest that the panel's response for each subjective attribute are normally distributed and have equal variance. This indicates that ANOVA can be used to analyse the results.

Attribute	Bartlett's statistic	df	p-value
Dynamics	3.99	7	0.78
Timbre	6.16	7	0.52
Envelopment	1.86	7	0.97
Support	5.15	7	0.64

TABLE 9.10: *Results of Bartlett's test for normality and homogeneity of variance. p-values greater than 0.05 indicate that the results for each hall are normal and are of equal variance. This supports the use of ANOVA.*

The results of the two-way ANOVA for each subjective attribute are presented in Table 9.11. It can be seen from each set of results that there is no statistically significant link between ST_{early} or Hall on any of the subjective attributes presented. Of these results, the lowest p-value was recorded for the subjective attribute *Envelopment* when the variable is *Hall*. This could indicate that the differences in reflection distribution were producing a perceivable change in Envelopment, however the test was not sensitive enough to confirm this. This is confirmed by the results of a post-hoc power analysis conducted with the program G*Power (Faul et al., 2007) with the results displayed in Table 9.11. It can be seen that the test power is less than 0.07 for all the tests.

Source	SS	df	MS	F	$Probs > F$	η_p^2	$Power(1-\beta)$
Dynamics							
Halls	0.126	3	0.042	1.22	0.307	0.032	0.06
ST_{early}	0.001	1	0.001	0.02	0.894	0.0002	0.05
Halls * ST_{early}	0.039	3	0.013	0.38	0.771	0.01	0.055
Error	3.58	104	0.034				
Total	3.74	111					
Timbre							
Halls	0.051	3	0.017	0.62	0.604	0.017	0.06
ST_{early}	0.0225	1	0.022	0.82	0.368	0.007	0.057
Halls * ST_{early}	0.055	3	0.018	0.67	0.571	0.018	0.06
Error	2.86	104	0.027				
Total	2.99	111					
Envelopment							
Halls	0.156	3	0.0522	1.72	0.167	0.04	0.07
ST_{early}	0.0002	1	0.002	0.01	0.938	0.0005	0.05
Halls * ST_{early}	0.139	3	0.046	1.53	0.211	0.039	0.07
Error	3.15	104	0.03				
Total	3.44	111					
Support							
Halls	0.079	3	0.026	0.75	0.523	0.02	0.06
ST_{early}	0.006	1	0.006	0.17	0.680	0.002	0.05
Halls * ST_{early}	0.053	3	0.018	0.5	0.683	0.014	0.06
Error	3.64	104	0.035				
Total	3.78	111					

TABLE 9.11: *Results of two-way ANOVA where Columns are each hall and rows are settings of ST_{early}. The results are shown for all subjective attributes. SS denotes the Sum of Squares, df denotes degrees of freedom, MS denotes Mean Squares (SS/df), F is the F statistic, $Prob > F$ is the p-value associated with a hypothesis test that the row, column or interaction effects are all the same. η_p^2 is a measure of the proportion of variance for each effect. $Power(1-\beta)$ is an estimate of the test power.*

As with the objective parameters, it is of interest to determine if the panel's preference judgements can be accounted for by any particular subjective dimension and to examine how well the participant's agreed with each other. Therefore, the subjective attributes for each hall were analysed using a standardised PCA and the preference responses regressed on to the product space. As before, the processing was performed using the *FactoMineR* package (Le et al., 2008) in *R* (R Core Team, 2013). The subjective attributes for each hall were arranged as a matrix where columns denoted each subjective attribute and rows denoted each concert hall.

Table 9.12 shows the eigenvalue and percentage variance accounted for by each dimension. It can be seen that nearly 92% of the variance is accounted for by the first three dimensions, in addition to the eigenvalue dropping below 1.00 at Dimension 2. This suggests that the product space can be described by three dimensions without any significant loss of information.

Dimension	Eigenvalue	(%) Variance	(%) Cumulative variance
1	2.09	52.5	52.5
2	0.86	21.5	74.0
3	0.72	17.9	91.95
4	0.32	8.04	100.00

TABLE 9.12: *Summary of variance for each dimension of the PCA. Table shows the eigenvalue, % variance of each dimension and the cumulative variance.*

Table 9.13 shows how well each subjective attribute is correlated with each dimension. Dimension 1 appears to be correlated with two sensory attributes which suggests that halls that were perceived to have higher envelopment were also perceived to have higher perceived support. Dimension 2 appears to be well correlated with perceived Timbre such that halls on the positive axis of Dimension 2 were perceived to have a brighter Timbre. Dimension 3 is correlated with Dynamics and so halls on the positive axis of Dimension 3 were perceived to be easier to control in terms of Dynamics.

Parameter	Correlation	p-value
Dimension 1		
Support	0.839	0.009
Envelopment	0.826	0.012
Dimension 2		
Timbre	0.842	0.009
Dimension 3		
Dynamics	0.696	0.05

TABLE 9.13: *Correlation of each subjective attribute with each dimension. The p-values indicate the significance level that the correlation value is different from 0. Results are shown for p-values that are less than or equal to 0.05.*

Figure 9.19(a) and 9.19(b) show the product space in terms of the subjective attributes. It can be seen that LRR(H) and RH(H) have been judged to be similar in terms of the sensory attributes, owing to their proximity in all three dimensions. Similarly, LRR (L) and YH(L) have been judged to be similar, especially on Dimension 2. GCH (L) appears to have been perceived quite differently from the others owing to its distance from the other halls.

Figure 9.19(c) and 9.19(d) show the same product space with the individual preference vectors regressed on (dashed vectors), in addition to the sensory attributes shown as solid vectors. It can be seen that the preference vectors are oriented widely across the product space, which suggests that participants were basing their preference responses on different combinations of subjective attributes.

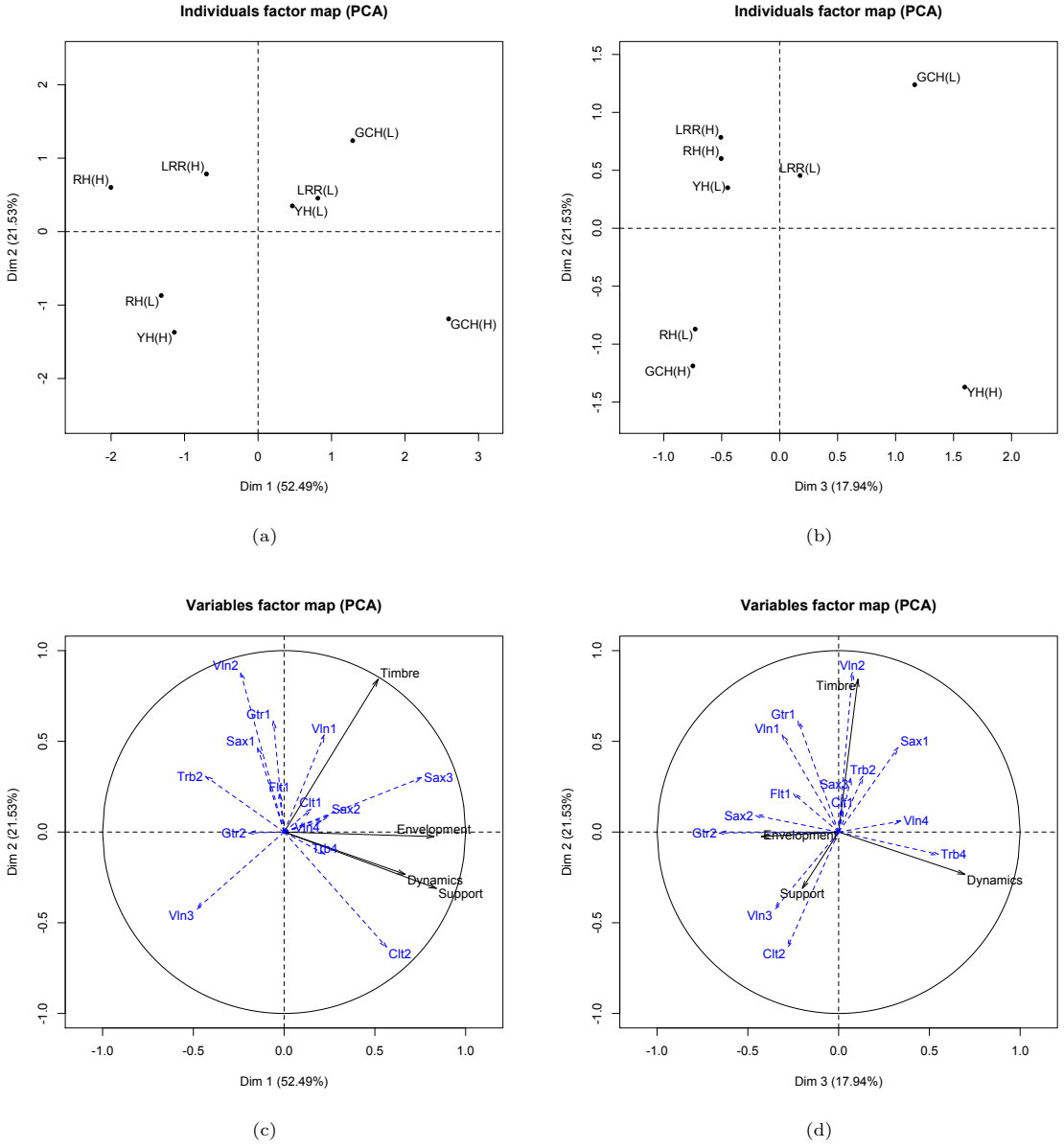


FIGURE 9.19: *PCA of preference data with mapped subjective attribute shown as blue dashed vectors. The preference pattern for each individual is shown as a solid vector.*

In Figure 9.20, the hall that is perceived to have the strongest support (GCH(L)) does not coincide with the most preferred hall, nor does the least preferred hall (YH(H)) coincide with the lowest perceived support. It is also of interest that the hall with the lowest perceived support (LRR(H)) was set at a high setting of ST_{early}. These results imply that the perception of support is influenced by other factors in addition to the overall energy of early reflections.

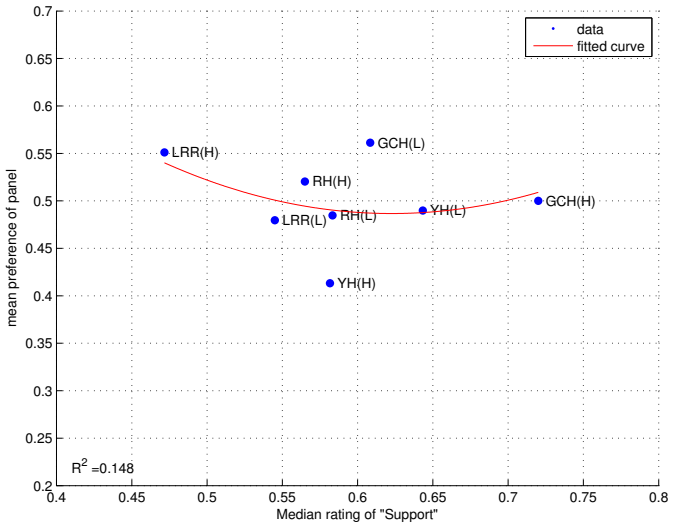


FIGURE 9.20: *Mean preference plotted against median "Support". A quadratic fit is applied and is shown as a red trace. The coefficient of determination, R^2 is 0.148.*

Similarly, Envelopment is plotted against preference in Figure 9.21. It can be seen that the most preferred hall (GCH(L)) has one of the highest ratings for envelopment. The hall perceived as most enveloping coincides with the hall that has the most spatially clustered arrangement of early reflections (GCH(H)). These results suggest that the panel considered halls perceived as more enveloping to be slightly more favourable.

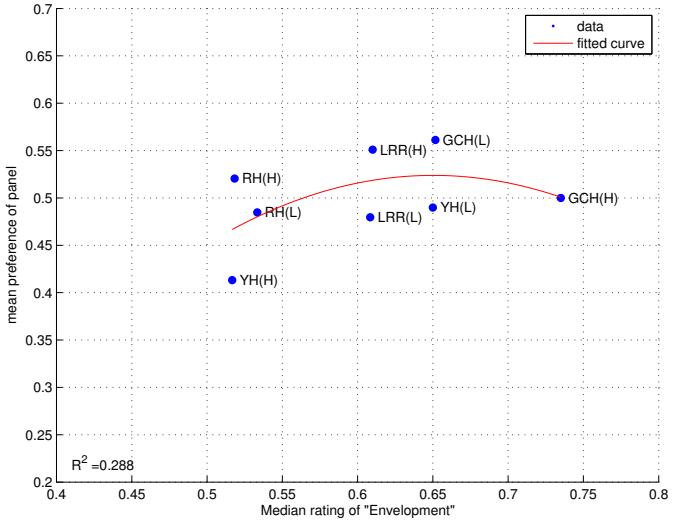


FIGURE 9.21: *Mean preference plotted against median "Envelopment". A quadratic fit is applied and is shown as a red trace. The coefficient of determination, R^2 is 0.288.*

Timbre is plotted against preference in Figure 9.22. It can be seen that the most preferred hall (GCH(L)) appears to produce a Timbre that is perceived as relatively Harsh/Bright and the least preferred hall (YH(H)) was perceived to have a more balanced timbre between Dark/Dull

and Harsh/Bright. It can be seen that the relationship between preference and Timbre is nearly linear. While the median preference value does not vary significantly, this implies that Timbre is a relatively important subjective attribute of the acoustic conditions.

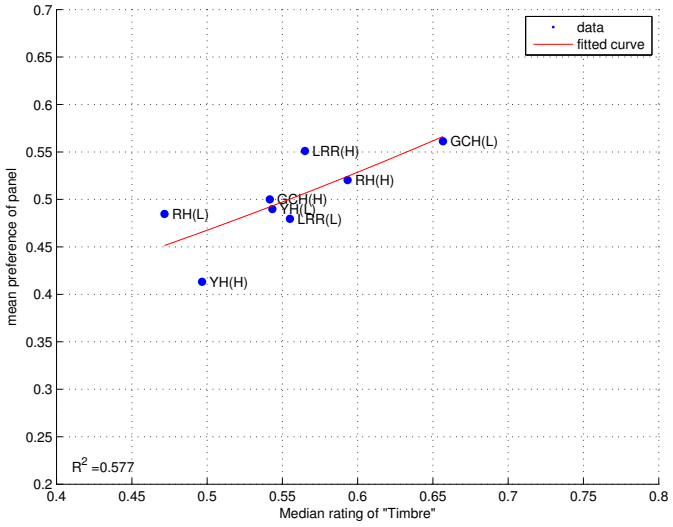


FIGURE 9.22: *Mean preference plotted against median "Timbre". A quadratic fit is applied and is shown as a red trace. The coefficient of determination, R^2 is 0.577.*

Figure 9.23 shows how preference varies with the attribute *Dynamics*. It can be seen that the halls that are perceived as the easiest and hardest to control (GCH(H) and RH(H)) have very similar median preference values. This suggests that the panel's preference pattern was not driven strongly by ease of control of dynamic range.

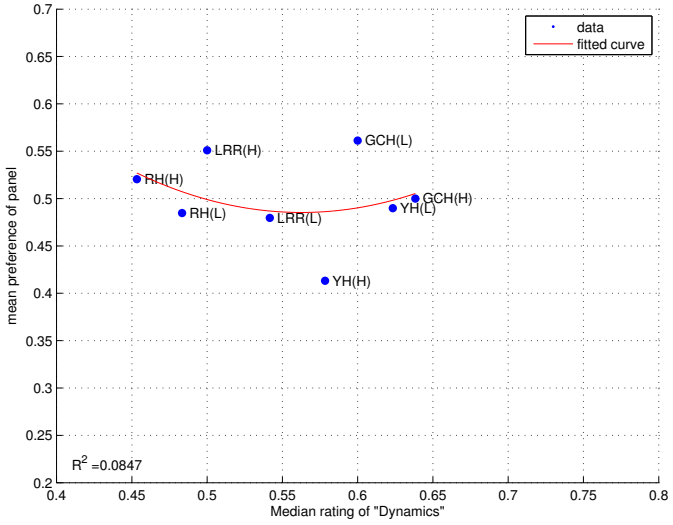


FIGURE 9.23: *Mean preference plotted against median "Dynamics". A quadratic fit is applied and is shown as a red trace. The coefficient of determination, R^2 is 0.0847.*

In summary, the results show that the halls presented to the panel were perceived to be very similar in terms of each subjective attribute and that there was considerable disagreement between panelists regarding the intensity of these attributes. However, it was found that there was slightly more agreement in the subjective attribute, *Timbre*, where more participants tended to prefer concert halls perceived as being brighter/harsher in Timbre.

It is possible that the subjective attribute responses were influenced by participants using the scale in different ways. In addition, participants may have been basing their responses in comparison to either previous experience or other questions in the test. It is also possible that the definitions of each subjective attribute were not clear enough, or the anchors too simplified to make clear judgements. Future experiments of this type should prioritise defining the subjective attributes used. This could be achieved using Individual Vocabulary Profiling (IVP) as demonstrated by Lokki et al. (2012).

9.4.5 Written responses

After the listening test was completed, each participant was invited to record written responses to four additional questions which aimed to gain further insight into their preferred stage acoustic in addition to thoughts regarding the test itself. The responses for each question are summarised and discussed as follows:

Do you find it necessary to adjust any aspects of your technique when performing in different venues? If so, please describe the type of adjustments you might make.

This question referred to the previous experience of each participant when playing in different venues. A number of participants made reference to how loud our soft they play in different sized venues. Clt1 referred to this as *"gauging the intimacy of a space - how soft can I play and still be heard"*. Many participants also responded that they varied their dynamics depending on the size of each space. For example, Flt1 reported that *"Softer phrases may need to be louder in a bigger venue and quieter in a smaller venue"*

Many participants responded specifically regarding changes to their articulation made in very dry or very reverberant spaces. Trb2 noted that they: *"must make changes to length of notes, start and end of notes and speed of changing notes. In dry rooms I think about how much 'articulation noise' might be coming from my tongue."*. Similarly, Clt1 noted that: *"short notes need to be shorter in wet spaces and dry spaces require the opposite"*. Sax2 also referred to adjusting the vibrato: *I need to control it more [in larger venues] if I don't, the sound can start to "merge" and it can be quite harsh on the ears"*. Gtr1 also responded that they used: " *More vibrato if room is less responsive*". Vln2 referred to adjusting *" Projection (exaggerate articulation), such as getting clearer 'consonants'; or adjust the balance between the high and low registers."*. Vln4 also reported that *"More resonant acoustics can call for slower tempi"*.

All trombonists made favourable references towards big or reverberant spaces being optimal for brass instruments. Trb1 elaborated on this by saying: *"in an extremely resonant hall you*

have to articulate harder to get more clarity and in a dry acoustic you need to work harder to support a phrase...The resonance makes playing long phrases easier". This suggests that for trombonists, the physical effort of sustaining a phrase is of high priority. Vln4 similarly reported that *"Different acoustics require more effort for a consistent sustained note"*.

A number of participants referred to very specific technical adjustments they make in different halls. For example, Gtr1 referred to "*Altering angle of picking hand to increase/decrease amount of nail included in finger strike*". Clt1 and Sax3 mentioned that they select a higher strength of reed in larger halls and Trb4 made changes to their "*air column*". Another Sax3 reported that they "*chose different mouthpieces based on the size of the room...sacrificing some tonal quality for projection.*"

In summary, it is evident that musicians are very aware of the stage acoustic conditions they inhabit and each have individual strategies with how they play in different spaces. These adjustments are wide ranging and include aspects such as how loud they are playing, the length and separation of notes and dynamic range. In addition, musicians may make extensive technical adjustments to their instrument depending on the space they are playing in. It was found that most of these changes appear to result from the perceived size or reverberance of the space with musician often referring to extremes as playing outside or in a practice room or recording studio in contrast to playing in a cathedral or large venue. Many participants reported how the acoustics can affect how much effort is required to play in a particular hall. This implies that poor stage acoustics can tire a musician during a performance which may also affect how they play.

It is clear from the comments that the stage acoustic conditions can dramatically change how a musician approaches a performance, highlighting the importance of good quality conditions for the musician.

In the listening test, did you notice any differences between the concert halls you played in? If so, please describe any differences you heard

---

This question gave participants the opportunity to report on differences they perceived in each hall and allowed them to speculate on the variables that were changing.

The majority of participants reported that the tests appeared to decrease in difficulty where the discrimination test was the hardest. Most participants found the difference between some halls to be much more subtle than others. This may have been due to variables being hidden from the participants as many participants reported that they felt the attribute test was the easiest of all the tasks as they were asked to listen for particular attributes. Most of the participants remarked that they were sometimes unsure whether the differences they heard between halls was due to the hall, or due to variations in their performance of a particular phrase.

Most participants reported hearing a change in reverberation time for example Sax1 referred to the *"length of reverb"* or (Vln1) *"response time"* varying in the halls they played. This is particularly interesting given that the reverberant sound did not change throughout the experiment.

Clt1 and Trb2 reported that they perceived differences in timbre between some of the halls while others were considered to be *"more responsive"* dynamics wise. Trb1 reported that some halls were *"boomy"* while others were more *"muffled"*. Vln4 reported that the: *"Brightness/mellowness of the sound carried. Length/shape of the resonance after a sound"*. These responses could be interpreted as references to timbre and dynamics. Vln2 reported that differences were perceived *"Most prominently in reverberation & Support. The differences became more clear in the last section of the test when I started listening to specific categories"*.

Trb3 reported that they could discern *"The amount and spread of decay. The quality of tone, the amount of dynamic contrast I could achieved"*. This implies that this participant could hear many aspects of the acoustic conditions varying between concert halls.

It is evident that participants could hear differences between halls but that these differences were more evident towards the end of the test when they were instructed to listen for specific variables. Most participants interpreted the differences between each hall as changes in reverberation which is of particular interest as the reverberant decay was constant throughout all comparisons. It is feasible that the variations in early reflection influence the musician's perception of reverberation in a similar manner to EDT.

From your previous experience, which venue has the best acoustic conditions for a solo performance? Please explain why you prefer this particular venue.

This question gave an opportunity for the participants to give examples of places they enjoyed playing in to determine if there were any similarities between their responses. Participants were invited to refer to specific venues or types of venues.

All of the trombone players said that they preferred to play in churches or cathedrals as it required far less effort to play during the performance. Many participants cited playing outside or in studio recording sessions as being their most uncomfortable places to play. Many participants also mentioned that the acoustics of practice spaces were often very unpleasant as it allowed them to hear every single detail of their sound which had the effect of making them lose confidence. Nevertheless all of the participants that commented on this understood the importance of developing their performance in environments such as these for the same reasons.

A number of participants also recognised that playing in very reverberant spaces was very challenging and that a space which is too reverberant was as problematic as one that is too dry. For example, *Flt1* reported that they liked performing in *"Renfield St Stephen's Church - It's big for the sound to travel but not too big for the sound to disappear"*. Similarly, *Gtr1* noted that they liked to plain in *"St Mary's Cathedral, Edinburgh - Great reverb that doesn't intrude too much"*. Also *Trb3* reported that they enjoyed playing in *"New England Conservatory's Jordan Hall - Live but has a quick decay. Provides a full sound for brass instruments. Isn't so live it compromises clarity"*.

It would appear from these comments that many of the musicians in this sample considered reverberation to be a very important factor when performing. It appears that a longer reverberation time can assist their playing with regards to effort. However, a very long reverberation time can compromise their perception of clarity or feel intrusive.

In relation to stage acoustics or the experiment you have just completed, do you have any further comments you would like to add?

This question allowed the participants to provide any feedback related to the test or research. Their responses provided some useful insight into how future stage acoustic experiments should be conducted.

Some participants responded regarding the length and difficulty of the test for example Trb1 responded: ''...although found the test quite long stamina wise as was feeling quite tired towards the end". Vln2 responded: *"Increasingly interesting tests. I found the first part quite challenging and required quite a lot of attention. Would maybe have liked it a bit shorter"*. From this it is clear that, even with breaks, a 2 hour long test is quite challenging for the participants. This is thought to be due in part to the difficulty in the first test where participants may have been expecting larger changes and were required to listen very carefully to variables they may not have been familiar with.

Vln3 responded specifically to the Timbre parameter in test 3, *"I wasn't sure if the given opposition (harsh/bright - dull/dark) was always appropriate. There were halls where 'rich - simple' might have been added"*. This highlights the definition of each attribute wasn't always clear for different participants and that further work is required to more clearly define these attributes in a language that is consistent with musicians. In previous questions, many participants referred to a hall's *'resonance'* when they were referring to *'reverberance'*. In doing this, it is possible more accurate results could be obtained from listening tests.

Vln4 referred to the visual feedback received on stage: *"I wonder if the visual information you have from a real concert hall, as opposed to a virtual environment, has more of an effect on how I perceive the acoustic..."*. This follows as a laboratory context is very much artificial and lacks certain aspects of the performing environment. In addition to the visual impact of the hall, it is possible that the temperature or glare from stage lights can affect how they feel during a performance. Future stage acoustic laboratory experiments should consider carefully which aspects are included.

It is evident that further development is required for this type of experiment which should focus on making the virtual environment more realistic and defining subjective attributes more accurately for each participant to ensure they are presented with a scale that relates to their experience. It is also clear that single session tests are very tiring for musicians especially if the differences are very subtle. Multiple sessions would allow more opportunity for training participants and reduce their fatigue throughout each experiment.

9.5 Summary and discussion

This chapter has described the main listening test of this research which aimed to determine if the spatial or temporal distribution of early reflections was a significant driver of preference and the resulting effect on perceived subjective attributes for a given hall. Experienced musicians were invited to perform their instrument in different auralised concert halls and asked to respond to a series of questions. Each hall was based upon measured data and varied only in terms of the early reflections. The independent variables of the test were the spatial and temporal distribution of early reflections and ST_{early}. Sixteen participants took part in the experiment and two participants were screened as they could not consistently determine differences between halls at a high level of ST_{early}. It is acknowledged that this is a low sample size but is relatively close to similar research, for example, by Guthrie (2014). The majority of participants found the experiment very difficult however some reported that comparisons between some halls were much easier to discern than others. The difficulty of the test may have been due to the variables being concealed from the participants as many were unsure what to listen for until the attribute test.

The discrimination test found that the difference between particular halls at high levels of ST_{early} could be detected to a statistically significant level however the same comparisons made at low levels of ST_{early} could not. However, when the responses were assessed only between the levels of ST_{early}, it was found that the panel could not consistently detect differences between halls at either level. This suggests that the differences between particular halls was difficult to detect, even when rendered at a high level of ST_{early}. However, this is not reflected by the written responses from the participants who appeared to be able to highlight and describe specific differences between the halls. It is possible that by the end of the test participants had accumulated enough experience to determine what aspects of the hall were changing. Additionally, it is possible that the subjective attribute definitions given towards the end of the test aided the participants in hearing how each hall was changing, rather than at the beginning of the listening test when no indication given to each participant.

The results of the preference test suggest that each hall was perceived in a highly personalised manner or the differences between each hall were so subtle that the differences did not result in variations in preference. This was evident due to the large variation in individual preference responses for each hall. The preference results showed that YH(H) was the least preferred hall whereas GCH(L) was most preferred by the panel, but not significantly more than any other halls presented. This was confirmed by Thurstone's Law of Comparative Judgement which found that no single hall was preferred significantly more than the others by the panel. In addition, a two-way ANOVA could not detect any significant effects of the Hall or ST_{early} on musician preference which highlights that the difference in preference was very small. As with the discrimination test, the written responses demonstrate that participants could perceive specific differences between concert halls which suggests that the wide variation in preference responses was related to differing interpretations or requirements rather than uncertainty.

The attribute test also implied that the participants did not agree on how the acoustic properties of each virtual hall influenced specific subjective parameters, due to the high variation in each

attribute and for each hall. It is also the case that participants may have used the scale in different ways, suggesting that future stage acoustic experiments should be carried out over a number of sessions allowing each participant to be trained. A two-way ANOVA did not reveal any significant effect of the Hall or ST_{early} on any of the subjective attributes.

The objective parameters and subjective attribute data were analysed in tandem with the preference data to determine if particular objective parameters were influencing musician preference towards each hall and if preference was associated with an increase in a perceived subjective aspect of the hall response. It was found that the preference responses from this panel varied significantly, suggesting that each participant was using different combinations of objective and subjective attributes to determine their preference towards each concert hall. There also did not appear to be any consistent judgements across musical instrument families. Despite this, it was found that there was a slight majority of participants who preferred halls with lower values of t_{mean} which may account for the slight preference towards GCH(L) and away from YH(H). In addition, it was found that a slight majority of participants tended to prefer halls that were perceived to have a brighter/harsher Timbre.

It was found that the most preferred hall, GCH(L) featured reflections that were rendered at a low level of ST_{early} and had reflections that arrived from a single lateral direction, away from the stage front direction. It was also found that this hall had reflections which were clustered together in time, as opposed to being spread out. It appears that, for this panel, preference was driven slightly more by the temporal parameter t_{mean}. Halls with different t_{mean} values are likely to have differing Early Decay Times (EDT). This agrees with the conclusions of Guthrie (2014) who found that preference responses of soloists were influenced by EDT. However, the results of this experiment found musicians prefer a shorter t_{mean} (and hence a shorter EDT) whereas Guthrie's results suggest the opposite.

It is important to acknowledge that Thurstone's Law of Comparative Judgement (Case V) found that no one hall was preferred more than the others to a statistically significant level. Therefore, these outcomes are reported mainly to indicate a possible area for future research rather than to indicate a firm causal relationship.

Written feedback from the participants suggested that participants could hear differences between each hall as they were found to be able to describe specific changes that they had perceived. This contrasts with the results obtained in the discrimination test which suggest that detection rates were below statistically significant levels. It was found that a number of participants perceived the differences between halls as variations in reverberation time, which is particularly interesting as the reverberant decay was fixed in this experiment. Therefore, it is possible that variations in early reflection distribution could influence how reverberant a hall is perceived to be.

In summary, the results do not provide sufficient evidence that the spatial or temporal distribution of early reflections are a major driver of musician preference towards a particular hall. Nor do they show a consistent preference towards halls that were perceived to have certain subjective characteristics. The results of the test did, however, provide some initial indication that the temporal distribution of early reflections may have a stronger impact on preference and how musicians perceive the on-stage acoustic response. Halls with earlier arriving reflections appear

to be slightly more preferred over those where reflections are spread out over time. It is likely that the small panel size contributed to the low power of this experiment, further research would benefit from concentrating on fewer acoustic parameters and aim to recruit a larger number of participants to increase the confidence in these conclusions. On the basis of these results, it is suggested that future research should concentrate on the temporal distribution of early reflections to determine how this affects musician preference towards an on-stage acoustic response. In addition, based on written feedback, it is considered highly beneficial to more clearly define the salient subjective attributes used in judging concert halls from a musician's perspective. This will aid further research in determining the subjective impact of different acoustic environments.

Chapter 10

Conclusions

The initial motivation for this research was to determine how the on-stage acoustic conditions affect a musician's impression of the venue. The stage acoustic conditions are highly relevant to a musician and can significantly influence their approach to a performance, for example, influencing how they articulate phrases or how loud they play. Ultimately, what the musician hears from the venue will affect what the audience hear and so it is critical that a musician performs in conditions that are assistive and not hindering.

Relatively little is understood about what aspects of the stage acoustic conditions are of importance to musicians or favourable ranges for existing acoustic parameters. By isolating these attributes, it will be possible to design future concert halls to provide favourable acoustic conditions for the performer, assisting them to give a high quality performance.

As discussed in Chapter 2, It was found that current stage acoustic parameters (such as ST_{early}) were independent of the spatial or temporal distribution of early reflections. It was proposed that these factors could influence a musician's impression of a performance space and thus influence how they perform a piece of music. This is important as, ultimately, this could affect how the audience hear and enjoy the performance.

However, the presence of the musical instrument and the stressors of performance can produce significant masking effects and so it is not certain if the spatial or temporal distribution of early reflections are audible to the performer. If these variables are shown to produce a significant effect on the musician's impression of the space then it was proposed that additional acoustic parameters may be required to assess stage acoustic conditions.

The main hypothesis that guided this thesis is stated as follows:

In the context of a performing soloist, the preferred acoustic conditions on stage are strongly dependent on the spatio-temporal distribution of early reflections in addition to their overall level relative to the direct sound.

This hypothesis was tested using a series of objective and subjective tests. Measurements were made on the stages of eight venues to determine how the spatial or temporal distribution of early

reflections varied in response to various physical variables. In addition, an interactive listening test was conducted which involved musician test participants performing in different acoustic environments and responding to a series of questions about their experience. The listening tests were conducted using an interactive auralisation system which allowed the participants to perform in and compare different virtual stage acoustic environments under controlled conditions.

The main findings of this research are summarised in the following sections.

10.1 Measurement of stage acoustic conditions

In Chapter 2, it was discussed how the spatial and temporal distribution of early reflections may be affected by performer-related factors (such as instrument directivity/orientation and position on stage) or hall-related factors (such as the geometry of the stage enclosure).

It was demonstrated that the current method of capturing stage acoustic impulse responses did not allow the spatial or temporal distribution of early reflections to be observed. Moreover, it was discussed how the acoustic response, as measured using an omnidirectional sound source, may not be representative of what is experienced by the musician.

Therefore, a measurement procedure was designed (in Chapter 3) so that the spatial and temporal characteristics of the acoustic response could be captured and analysed in detail. In addition, the measurement system allowed the captured acoustic responses to be used for auralisation. This measurement procedure was used to survey the stages of eight different performance spaces.

The measurement system comprised an Ambisonic microphone positioned directly over a directional loudspeaker. The loudspeaker was rotated in 45° increments after every measurement, resulting in eight measurements per stage location. The Ambisonic microphone encoded the spatial and temporal characteristics of the stage acoustic conditions within each measurement, allowing the spatial distribution to be analysed and the spatial characteristics to be included in resulting auralisations. This approach also allowed established acoustic parameters, such as T30 and ST_{early}, to be analysed from the same measurements. The use of a directional sound source, orientated in numerous directions, allowed the effect of source directivity/orientation on the acoustic conditions to be studied and also included in resulting auralisations. This approach was found to be advantageous when analysing the spatial characteristics of impulse responses as it reduced the likelihood of temporally simultaneous reflections producing localisation errors as shown previously by Tervo (2012).

In general, it is considered that omnidirectional sound sources are poor representations of musical sound sources and objective parameters and auralisations could benefit from including the directional characteristics of musical instruments. The measurement technique used in this research is aligned with other recent studies in Auditorium Acoustics which utilise directional loudspeakers or complex arrays of loudspeakers to more closely emulate the directivity pattern of musical instruments (Tervo et al., 2013b).

10.2 Analysis of acoustic conditions

In order to characterise the spatial and temporal characteristics of the stage acoustic response, a number of recent analysis techniques were reviewed in Chapter 4. It was found that an image source plot of the early reflections could express both spatial and temporal data and provide an intuitive visual display of the reflections. An analysis technique was developed which isolated each reflection in time using a peak-picking algorithm (which determined their time-of-flight) and used parametric analysis techniques to determine their angle of arrival.

A set of parameters were created which characterised how the reflections were distributed in time and space. These parameters were used, in addition to existing stage acoustic parameters, to observe how the acoustic condition on stage varied in response to performer-related and hall-related variables.

The analysis, as described in Chapter 5, found that the acoustic surveys represented a very wide range of acoustic conditions. For example larger halls, such as the Caird Hall, produced ST_{early} levels as low as -20dB whereas smaller halls such as the Recital Room produced levels as high as -10dB.

In assessing hall-related variables, it was found that ST_{early} appeared to decrease linearly when the surface area of the stage was increased. It was also found that stages of a larger physical volume, produced later arriving reflections that were less spread out in time than smaller stages. It was also found that stages of a lower physical volume resulted in early reflections that were more spread out spatially, whereas larger stages tended to produce reflections clustered towards a single direction.

It was found that ST_{early} varies significantly with source orientation on stage. In general, ST_{early} increases when the sound source is facing the nearest reflecting surface. In contrast, ST_{late} varies much less. These findings were also found when the measurement position was varied, i.e. ST_{early} varied much more than ST_{late}. In this case, it was found that the increase in level was observed mainly at high frequencies where the sound source was more directional. It was also found that the spatial distribution of early reflections varied significantly with source orientation. In general, reflections tend to appear mainly from the direction the sound source is pointing in. This is regardless of the measurement position. The temporal distribution of early reflections varies much less however.

In summary, these measurements suggested that the spatial and temporal distribution of early reflections, in addition to existing stage acoustic parameters, are influenced by both the hall and the performer. The extent of these variations suggests that these aspects may have audible consequences for the performer.

10.3 Auralisation of stage acoustic conditions

In Chapter 6 it was discussed how an interactive auralisation of stage acoustic conditions could be used as a way of introducing musicians into a controlled environment to allow them to compare and contrast different halls. Current spatial audio techniques were reviewed in relation to previous stage acoustic auralisations to determine the most appropriate method for this research. It was considered that parametric decoding techniques offered a promising solution that allowed first-order Ambisonic impulse responses to be presented to a musician at an increased perceived spatial resolution. These techniques could render the soundfield in such a way that a musician moving around the sweetspot would not perceive any phasing/timbral artefacts that are commonly heard in Ambisonic systems.

In Chapter 7, It was discussed how different aspects of the auralisation system could contribute to a perceived lack of realism of the auralisation. It was found that these compounding aspects resulted in a perceived *'PA effect'* where the musician felt as if they were performing through a PA system in a space rather than naturally. It was found that the microphone technique used to capture the direct sound from the musician was a significant contributor to this effect. It was discussed how the number, layout and type of microphone could introduce different artefacts onto the rendered acoustic response.

Chapter 8 described the interactive auralisation system constructed in the Arup-DDS SoundLab and the procedure used to create and calibrate each auralisation. A series of objective measurements were used to determine how accurately the auralisation system could render the stage acoustic response.

It was found that the auralisation technique could more accurately spatialise reflections on the lateral plane than when elevated as there were fewer loudspeakers in that direction. It was also demonstrated how the non-anechoic conditions of the SoundLab affected the accuracy of the auralisation. The system was considered to render the stage acoustic conditions with enough accuracy so that it could be used to test musicians in relation to the main research hypothesis.

10.4 Perception of acoustic conditions

Chapter 7 reported two pilot tests which were conducted during the development of the auralisation system. These tests aimed to examine the operation of the system and to gain some initial support for the main hypothesis of this research.

These pilot tests introduced non-musician test participants into similar acoustic conditions experienced by a performer. This was achieved by replacing the musical instrument with a loudspeaker through which anechoically recorded musical phrases were rendered. These phrases were also auralised over a 3D loudspeaker system, emulating the direct sound and acoustic response respectively.

The first test focussed mainly on the spatial audio technique used to auralise early reflections and stage acoustic impulse responses. Additionally, the experiment aimed to determine if musical

phrasing and angle of arrival of reflections produced audible differences in what the participants heard. The results suggested that participants perceived auralisations as being very similar when auralised using first-order Ambisonics when compared to SIRR. Due to the way it renders the impulse response, SIRR does not introduce phasing artefacts if the listener moves with respect to the sweetspot. SIRR can be used to perform complex transformations on impulse responses which may assist in future stage acoustic research. In addition, SIRR can render first-order Ambisonic impulse responses at a higher perceived spatial resolution. Therefore, it was considered that SIRR was an appropriate spatial audio technique to render stage acoustic auralisations. The results also suggested that the audibility of a single early reflections could be affected by the type of musical phrasing (staccato or legato) and the angle of arrival of the reflection.

The second pilot test aimed to determine if participants could discern differences between different shaped concert halls that exhibited identical levels of ST_{early}. The different shaped stage enclosures would produce different spatial and temporal distributions in early reflections. In this experiment, half of the halls were auralised based on measured impulse responses whereas the other half were based on acoustically modelled impulse responses. It was found that participants could discern differences between acoustic responses where the spatio-temporal distribution of reflections varied but ST_{early} remained constant. It was found that the number of positive detections was affected by the level of ST_{early} and that differences were easier to discern when modelled impulse responses were used.

Chapter 8 described the interactive auralisation system developed as a result of these pilot tests. The system was adapted to auralise the direct sound of the musician's instrument. In addition, the parametric decoding technique was adapted to reduce the impact of localisation errors caused by the Ambisonic microphone. These localisation errors are reported in more detail in Appendix B.

Chapter 9 described an interactive listening test where experienced musicians were invited to play their instrument in different acoustic environments and respond to a series of questions. The spatial and temporal distributions of early reflections were varied in addition to the level of ST_{early} while the reverberant tail was kept constant throughout the experiment.

The results of the test suggested that participants could discern variations between acoustic responses where the distribution of early reflections was varied. It was found that the preference patterns for individual participants varied greatly, suggesting that they used different aspects of the impulse response to judge their preference. The results did not show any preference trends for participants of the same instrument suggesting that the preference towards a particular acoustic response is highly personal. It was also found that, on average, participants did not tend to prefer either the high or low level of ST_{early} which suggests that a 4dB increase in ST_{early} is very subtle.

Overall, the most preferred hall featured early reflections that were rendered at a low level of ST_{early} and had reflections that arrived from a single lateral direction, away from the stage front direction. It was also found that this hall had reflections which were clustered together in time, as opposed to being spread out. In terms of subjective attributes, participants perceived this hall to have the brightest/harshest timbre out of all tested. Crucially however, it was found

that this slight preference towards one particular concert hall was insufficient to be considered significantly different from the others to a statistically significant level. Additionally, a two-way ANOVA did not provide sufficient evidence to prove that either the spatio-temporal distribution of early reflections or ST_{early} had a significant impact on musician preference. Similarly, there was insufficient evidence to prove that these factors had a significant impact on any of the subjective attributes explored in this experiment.

Temporal parameters were found to result in the most agreement amongst the panel in terms of overall preference with participants preferring reflections to arrive earlier in time. In addition, written feedback from the participants found that many perceived the difference between concert halls as a change in reverberation time, despite the reverberant decay being constant throughout. This implies that the distribution of reflections could influence the reverberation time as perceived by the musician. This may be due to the temporal distribution of reflections affecting the EDT. This fits with previous research by Guthrie (2014) which found that EDT was a significant driver of preference for soloists. This could indicate a possible area for future study.

In summary, this research did not find sufficient evidence to support the alternative hypothesis that the spatio-temporal distribution of early reflections was a significant driver of musician preference.

The results of the interactive listening test showed that an increase in ST_{early} did not produce a corresponding increase in preference. It is possible that a 4dB variation is very subtle, however, it is equally possible that other subjective factors relating to the early reflections also influence musician preference. This supports the premise that additional acoustic parameters are required to assess and design stage enclosures. The subjective tests produced some initial indications that the distribution of early reflections affects how reverberant musicians perceive the venue. Therefore, future tests should attempt to confirm this effect and determine if it is a salient aspect of stage acoustic conditions for musicians.

It is likely that the test was not sufficiently sensitive to obtain evidence that the spatio-temporal distribution of early reflections was a driver of musician preference. Further testing should re-evaluate the testing methods and gather responses from a larger sample of musicians. Future tests should also aim to more accurately define appropriate subjective attributes for musicians. This will aid in obtaining more reliable responses from musician test participants but will also assist in further understanding what factors influence musician preference to on-stage concert hall acoustics.

10.5 Novel contributions

This research offers a number of novel contributions to the field of stage acoustic research:

- This research implemented a parametric analysis techniques in order to determine the spatial and temporal distribution of early reflections from a measured impulse response. This builds on existing approaches proposed by authors such as Guthrie (2014) and McCarthy

et al. (2008) which adapt existing stage acoustic parameters to include spatial information. A parametric analysis allows a detailed decomposition of the spatio-temporal distribution of early reflections which can then be assessed using statistical parameters (such as those presented in this thesis) or visualised as a 3D point cloud. This analysis method could be further improved by including the amplitude of each reflection in the visualisation or by exploring additional statistical processing to describe the distribution of reflections.

- A detailed analysis of the stage acoustic conditions found on eight concert halls was presented which described how hall-related and musician-related variables could influence the acoustic response. The results demonstrated that it was possible for two halls to feature the same level of ST_{early} but have different spatial or temporal distributions of early reflections which highlights a potential weakness for ST_{early} as a design parameter. It was also shown that stage geometry and the musician's position on stage could affect the distribution of early reflections, aspects which could potentially be refined to help create more successful concert hall stages.

- The parametric decoding methods used in the interactive auralisation system were found to be highly effective at auralising the stage acoustic conditions for musicians. This was especially successful as it reduced the likelihood of the musician experiencing timbral artefacts as they move around the sweetspot. In addition, this technique allows a perceptually enhanced rendering of measured acoustic responses made with a first-order Ambisonic microphone. This is important as first-order Ambisonic microphones are more widely available than HOA microphones (as shown by Guthrie (2014)) and therefore allows a more convincing auralisation of an existing space to be produced. Parametric decoding methods are also compatible with soundfields measured using HOA, allowing further enhancements to a more detailed capture of the soundfield. This method represents a significant step forward in terms of the techniques used in stage acoustic laboratory tests.

- The interactive listening test could not produce sufficient evidence to prove that the spatial or temporal distribution of early reflections is a strong driver of musician preference. However, the analysis did find some initial evidence to suggest that the temporal distribution of early reflections may have been influencing the preference responses of the participants. In addition, it found that the different levels of ST_{early} did not significantly influence musician preference, this will aid in further understanding the subjective impact of this parameter. This is a significant finding, which supports conclusions by Guthrie (2014) and Jeon et al. (2014) that could direct a more focussed research project and potentially find additional acoustic parameters to assist in designing concert hall stages.

10.6 Future work

This thesis has discussed many different facets of stage acoustic research including the measurement, analysis and auralisation of stage acoustic responses. In addition, this research utilised sensory testing to gain further insight into the main drivers of preference towards particular acoustic conditions. It was found that each one of these areas would benefit from future research which are highlighted in this section.

Research directions

This research did not find sufficient evidence to show that the spatial or temporal distribution of early reflections was a significant driver of musician preference. It is acknowledged that in this research, the relatively small sample size and limited population of musicians may have affected this. Therefore, future research should work collaboratively with a conservatoire or orchestra.

Some responses from musicians provided an initial suggestion that the temporal distribution of reflections could influence musician preference. This fits with conclusions from Guthrie (2014). Therefore, future research could focus specifically on this aspect of the acoustic response to determine the effect on musician preference. This research focussed on the temporal and spatial distribution of early reflections. However it did not assess the effect of temporal diffuseness on these reflections and the associated subjective effect. It was found previously by Bermond and Davies (2001) that the temporal diffuseness of early reflections could be discerned by musicians. Through the use of interactive auralisation systems in tandem with the synthesis of diffuse reflections (as demonstrated by Robinson et al. (2013a)) it may be possible to explore this effect in more detail.

It is also the case that this research focussed on a single context of a musician performing solo on a stage. This situation does not often occur in real life and so future research should investigate the experience of many different performance scenarios. This could include orchestras or small ensembles or could involve some other genres of music as it is likely that musicians of different genres will have different requirements for their performance conditions.

So far, stage acoustic research has focussed on isolating the best acoustic conditions for a musician to perform in. However, it is considered equally useful to determine the acoustic conditions that are considered to be poor. This will help determine what acoustic conditions are best avoided.

In this research, the early reflections were defined as those that arrived between 20ms and 100ms as is the case with other stage acoustic parameters. However, in some situations it is possible that reflections arriving before or after these times may influence the musician's impression of the hall. Therefore, a future stage acoustics study should attempt to determine if these time values are appropriate for use. For example, is 100ms representative of the perceived mixing time for musicians?

As described in Chapter 2, the JND for ST_{early} is only weakly defined as being approximately 2dB (Hak et al., 2012). Future research could more formally define the JND for ST_{early} for different instruments. This will help understand the implications of increasing or decreasing ST_{early}.

Sensory testing methods

As discussed in Chapter 9, some of the participants commented on the appropriateness of the subjective attributes used in the listening test. For example, once participant remarked that they felt *'timbre'* was much more complex than the unidimensional scale used in the test; offering terms such as *'richness'* to describe how they thought timbre was varying in the test. It is possible that the attribute definitions used in stage acoustic research are not yet fully refined for subjective testing. By using an approach called *Individual Vocabulary Profiling* (IVP), as proposed by Lokki et al. (2012), it may be possible to decompose these basic attributes further

to provide more detailed insight into how musicians experience the acoustic conditions on stage. This may also aid in communication between musicians and acousticians which, in the long term, will help improve the acoustic conditions in venues.

While MDPREF and PREFMAP methods have been shown to be successful in this field, more recently developed sensory testing methods may be more appropriate, such as Partial-Least Square Regression (PLS-R).

The testing procedure was considered appropriate for obtaining appropriate responses from musicians in a short space of time while attempting to minimise fatigue. It was found that most participants found the test quite difficult and so future testing should consider inviting musicians to attend multiple sessions to allow for participant training. It is recommended that these tests focus on one or two acoustic parameters such that a more detailed analysis can be produced for the same number of trials.

Stage acoustic measurements

In this research it was hoped that the approach to measuring stage acoustic impulse responses would allow arbitrary source directivity patterns to be synthesised. It was discussed how this could potentially improve the plausibility of the auralisation and the relevance of stage acoustic measurements. When attempting to do this, it was found that the arrangement of the loudspeaker and Ambisonic microphone caused the gain of the W-channel to vary relative to the X, Y and Z channels. This had the effect of distorting the spatial encoding of reflections and so was not applied in the listening tests or objective analyses. Given the potential advantages, it is proposed that this technique be investigated further. A promising technique, developed by Pollow et al. (2013),uses an array of transducers to generate sound as spherical harmonics from which it is possible to synthesise various directivity patterns.

The use of a first-order Ambisonic microphone was found to be very successful in localising the direction of arrival of early reflections. This type of microphone is readily available and so is straightforward to implement in future studies. However, in this thesis, it was demonstrated that the Ambisonic microphone could provide accurate localisation of early reflections only below a frequency of approximately 5kHz. It is possible that the use of HOA microphones or intensity probes could encode the spatial distribution of reflections more accurately.

Stage acoustic analysis

The use of the image source plot to visualise the spatial and temporal distribution of early reflections proved very useful in this research. In addition, the parameters developed to describe this distribution appeared to neatly describe the distribution of reflections. As image source models and plots are used in concert hall development this method should be highly intuitive to concert hall designers. Further improvements would be to weight the image sources according to their relative amplitude so that the distribution in amplitude can also be observed.

Future stage acoustic impulse responses may be able to decompose an impulse response into a series of image sources with a high degree of accuracy. However, being able to quantify the distribution of reflections will be a key component in both research and consultancy. Therefore, any future analysis should consider carefully how reflections are quantified. It may be possible

to use techniques from the field of computer vision to describe the distribution of a point cloud i.e. centroid or moment etc.

Interactive auralisation

Many participants commented that the interactive auralisation system could be used in musician practice, where musicians could refine their performance technique in virtual versions of concert halls prior to performance. There is also scope to study other aspects of musician performance such as the management of stage anxiety as demonstrated by Williamon et al. (2014). For use in this way, further work may be required to generalise the system to work with all types of instrument. As described by Williamon, a simulation could include other aspects of the performance environment such as bright lights, audience members etc to add to the simulation's plausibility. The interactive auralisation system could also be adapted to record a musician's performance and auralise it from different perspectives in the concert hall. This could prove an effective teaching tool to demonstrate to a musician how certain articulation is perceived from the audience.

The use of parametric decoding techniques in this research produced encouraging results in terms of auralisation accuracy. A clear advantage of this technique is that it reduces the likelihood of rendering artefacts that may be produced by a musician moving around the sweetspot. Parametric decoding techniques also allow complex transformations of the rendered soundfield which could prove effective in future stage acoustic research projects. This could be achieved by adjusting the parametric data prior to re-synthesis of the soundfield. Potentially, this could be used to adjust the frequency content of specific reflections or alter their temporal or spatial distribution. The interested reader is encouraged to review work by Politis et al. (2012) and Kallinger et al. (2009) for details regarding this technique. Similarly, other spatial audio techniques such as WFS might allow small ensembles to play together in virtual acoustic environments due to the absence of a sweetspot. Developments in the stage acoustic measurement technique will also provide improvements in the auralisation system as both should be considered simultaneously.

As discussed in Chapter 7, the microphone technique used to capture the direct sound from a musician's instrument can have a significant impact on the plausibility of the auralisation. Future work in this area could investigate the use of a surrounding microphone array to capture the direct sound which may allow the musician to turn and gesture in different ways. It could also investigate the use of different microphone or signal processing techniques which could reduce the likelihood of unwanted feedback through the system.

10.7 Concluding remarks

This research aimed to determine if the spatio-temporal distribution of early reflections was a dominant driver of musician preference towards stage acoustic conditions. A novel stage acoustic measurement technique was developed and deployed in eight local concert halls. The captured data allowed the spatio-temporal distribution of reflections to be observed in relation to hall-related and performer-related variables. Analysis of the measurements found that the temporal

and spatial distribution of early reflections varies in response to performer-related variables and hall-related variables.

In order to determine the subjective impact of these variations, an interactive listening test was developed where experienced musicians were asked to perform in virtual versions of concert halls. Their responses to a series of questions allowed the subjective effect of varying these aspects to be explored. The results of this research suggest that the spatio-temporal distribution of early reflections does produce a perceivable subjective effect but also that this effect is very subtle and is dependent on the musician's personal taste.

This research highlighted that the perception of venue acoustics from a musician's perspective is highly personalised, even amongst musicians of the same instruments. It is also evident that the definitions of salient subjective attributes are not clear for musicians and may be more complex. Therefore, future stage acoustic research should focus more on defining a common language between acousticians and musicians. This will aid in finding properties of the soundfield that are favourable for all musicians. The approach to measurement and auralisation of stage acoustic conditions, presented in this thesis, could easily be used for a larger scale project such as this.

Appendix A

Loudspeaker directivity measurement

The loudspeaker used in the venue measurements for this research was a Genelec 1029A loudspeaker. Prior to conducting the venue surveys, it was of interest to determine the directional radiation characteristics of this loudspeaker. By measuring the directivity pattern it was possible to synthesise arbitrary source radiation patterns from the measurements made in the venue surveys. While this was not completed for this research, it is useful to characterise the directivity of this loudspeaker. The following is a brief description of how the loudspeaker was measured.

The loudspeaker was positioned in front of an omnidirectional measurement microphone in the SoundLab. The loudspeaker was rotated in regular increments and an impulse response was measured at each angle of azimuth. The resulting impulse responses were truncated in time to ensure the influence of the SoundLab acoustic response was minimised. The impulse responses were then analysed in third octave bands to observe how each frequency band varies with source azimuth.

As shown in Figure A.1, the loudspeaker was mounted on a tripod so that the height from the floor to the top of the low frequency driver was 155cm. A guide was fixed to the top of the loudspeaker to ensure it was aimed at the correct angle relative to the microphone. A Behringer ECM8000 omnidirectional measurement microphone was mounted on a microphone stand at a height of 155cm. The distance between the two transducers was 1m. The closest reflecting surface to either transducer was the ceiling which was 115cm away. All reflecting surfaces in the SoundLab were treated with acoustic absorption to minimise the effect of reflected sound on the measurement. The loudspeaker was rotated by 10° (clockwise) after each measurement, resulting in a total of 36 measurements.

FIGURE A.1: Image showing the relative positions of the loudspeaker and microphone. The radius between the transducers was 1m. The other loudspeakers shown in the image were not used in the experiment. The loudspeaker was rotated in 10° increments after each measurement.

This loudspeaker was connected via XLR leads to the line output of a Behringer ADA8000 A/DAC which was connected via ADAT to an M-Audio Profire Lightbridge soundcard. The soundcard was connected to a Macbook via Firewire cables. The rotary gain control on the loudspeaker was set to a 12 o' clock position and the signal level controlled digitally using Reaper. The output level was set to -16dBFS.

Impulse responses were obtained by playing a 10-second, logarithmically swept sinusoidal signal (0 - 22050Hz) through the loudspeaker and measuring with the microphone. This signal was generated using MATLAB (Mathworks, 2013, Wells, 2012). This code generates a sine sweep with an 100ms amplitude ramp at the start and the end to avoid any unwanted transient signals. Furthermore, it generates an inverse sweep by time-reversing and applying a -6dB per octave envelope. The sine sweep was played back and measured simultaneously in the Reaper DAW system (Reaper, 2013) at a sampling frequency of 44100Hz and a 32-bit floating-point bit depth.

The initial analysis was achieved by truncating the measured impulse response to 470 samples (approx. 10.6ms) to reduce the impact of the SoundLab acoustic decay. Each impulse response was then time aligned (where necessary) and the propagation time removed (approx. 240 samples were removed) and transferred into the frequency domain using a Fast Fourier Transform (FFT) of length 1024 samples. Figure A.2 shows the directional characteristics of the loudspeaker at different angles of azimuth. The magnitude is shown in dBFS (relative to Full Scale) with the colour bar range shown adjacent.

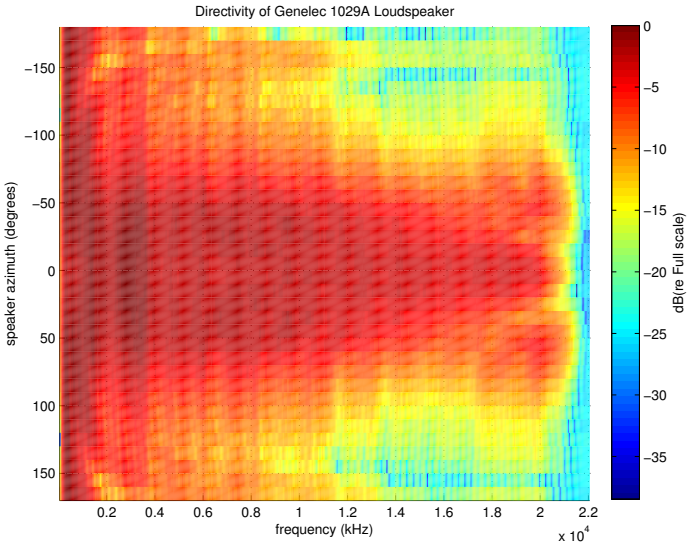


FIGURE A.2: Directivity characteristics of the Genelec 1029A loudspeaker measured in the medial plane. Positive angles represent angles in the clockwise direction. As expected the loudspeaker contains more high frequency content at angles close to 0° (on-axis)

It can be seen from the plot above that at low frequencies the energy from the loudspeaker does not vary much with speaker orientation. Between 2kHz and 10kHz, the energy is concentrated between $\pm 50^\circ$. As the frequency increases, the physical width of this band reduces to approximately $\pm 10^\circ$.

Appendix B

Directional accuracy of ST350 Ambisonic microphone

Throughout this research, a Soundfield ST350 Ambisonic microphone was used to capture spatial room impulse responses. Reflections in the impulse response were detected using a peak finding algorithm and localised spatially by estimating the active intensity of the reflection from which a direction of arrival can be obtained.

It has been observed that, when viewed in the frequency domain, intensity vectors often do not point in the same direction for a single reflection. A contributing factor to this is the Ambisonic microphone itself. It has been reported by Vilkamo (2008) and Protheroe and Guillemin (2013) that an angular error is introduced by the microphone which tends to increase in frequency. This error is caused by the capsule spacing of the Soundfield microphone.

This has two important implications for this research. Firstly, the intensity vectors are used to analyse the direction of arrival of early reflections which will cause the objective analysis to be flawed. Secondly, the intensity vectors are used to resynthesise early reflections for auralisation of stage acoustic responses. If there exists an angular error at higher frequencies, the virtual acoustic response may not be rendered correctly. This will impact the results of listening tests with musician test subjects.

The aim of this experiment was to determine a cut-off frequency below which the directional analysis is considered reasonably accurate. The cut-off frequency obtained in this experiment was used to improve the accuracy of the spatial analysis and auralisation of room impulse responses.

The Soundfield ST350 microphone was positioned at the sweetspot of the Arup-DDS SoundLab at a height of 1.30m above the floor. This position was approximately 2.2m away from any lateral reflecting surfaces and approximately 1.4m away from the ceiling. The microphone was aligned with the sweetspot using a plumb-bob and visual aids so that the microphone was positioned correctly and angled towards the front-middle speaker of the array.

An auralisation system was set up so that high amplitude impulse responses would emanate from each loudspeaker after a predetermined delay. This would emulate a reflection occurring

in the direction of the loudspeaker. A swept sine wave technique was then used to measure the auralised impulse response at the centre of the loudspeaker array. The impulse response was measured with the Ambisonic microphone.

The impulse response was then analysed to determine the time and direction of arrival of the artificial reflections. The intensity vectors used to determine the direction of arrival were compared in the frequency domain with the expected direction of arrival in order to determine the angular error.

The measurement set up is shown in Figure B.1 below. The soundfield microphone is positioned directly above a Genelec 1029A loudspeaker. The soundfield microphone is also positioned in the centre of the loudspeaker array. The loudspeaker array is linked to a computer system capable of rendering reflections using real-time convolution. The sound from the loudspeaker was picked up by the M-Audio Luna microphone and convolved with an impulse response containing 'artificial' reflections (described in more detail below). The auralised reflections would then be measured by the Soundfield microphone. A swept sine signal was used to measure the auralised space and its impulse response was extracted from the recording made by the soundfield microphone.

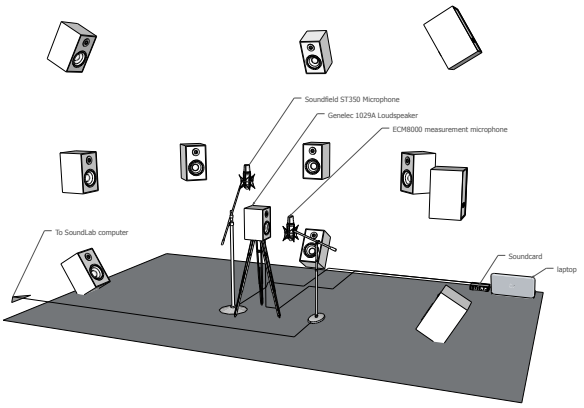


FIGURE B.1: *Diagram showing the apparatus used in this test. For clarity, some of the loudspeakers in the array have been removed.*

The apparatus used in this experiment was set up to test numerous variables of the real-time auralisation system including system latency and the acoustic response of the SoundLab. It is acknowledged that the test set up could be improved by performing the test in anechoic conditions and by using a single sound source with the Soundfield microphone rotating in fixed angular increments.

The artificial reflections were synthesised over the loudspeaker array using a real-time convolution system developed as part of this research. The auralisation system was implemented in Max MSP and utilised Alex Harker's real-time convolution objects. A single convolution engine was set up for each loudspeaker, therefore the Max MSP patch consisted of sixteen convolution objects in total. The convolution objects were fed from a single input channel and output to each individual loudspeaker.

A B-format impulse response was created where artificial reflections were spaced in 0.15 second intervals beginning at $t = 0.15s$. Each reflection was panned to a different angle beginning at 0° and ending at 337.5° panning in 22.5° angular increments in an anticlockwise direction. These angles were chosen so that half of the artificial reflections would appear directly from loudspeakers in the array and half would appear directly in-between the loudspeakers. The time delay between each reflection was chosen as it is approximately equal to the reverberation time of the SoundLab. This means the analysis of any given artificial reflection would not be include reflections from the SoundLab itself.

This impulse response was processed using SIRR to obtain a 16-channel impulse response where the artificial reflections were panned using amplitude panning techniques. This 16-channel impulse response was split into 16 mono channels and inserted into the auralisation system. The impulse response utilised in this experiment panned all artificial reflections to an elevation of 0°. For the purposes of this experiment, artificial reflections panned between the loudspeakers were not included in the analysis, only those that were panned directly to each loudspeaker directions.

The impulse response measurements were made by exciting the space with a 10-second logarithmically swept sine wave which was captured by a microphone near the loudspeaker and input to the auralisation system which consequently played over the loudspeaker array where the artificial room response was measured using the Soundfield Microphone. The measured impulse response was extracted using inverse filtering as demonstrated by Farina (2000).

The measured impulse responses were analysed with an adapted SIRR script which detected the presence of early reflections. At time delays equal to the detected reflections, the SIRR script would obtain the direction of arrival of the reflections by estimating the intensity vector in the time-frequency domain.

The arc angle between the measured and predicted direction of arrival was assessed using equation (B.1) where θ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ which in this case represent the intended and measured intensity vectors for each artificial reflection (Bourne, 2014).

$$\cos(\theta) = \frac{\boldsymbol{ab}}{|\boldsymbol{a}||\boldsymbol{b}|} \tag{B.1}$$

This comparison was performed for each frequency bin of the measured reflections, producing a frequency dependent error for all eight reflections included in the analysis. The results were then averaged to obtain a mean spatial error and also assessed to obtain the maximum error at all frequencies.

Figure B.2 shows the raw results of the experiment. While this plot is not very easy to read, it clearly shows how the directional accuracy of the Soundfield microphone reduces with frequency.

The plot shows the analysed azimuth of the artificial reflections estimated at different frequencies. The different traces, show the results for artificial reflections arriving at different angles. The legend shows the intended angle of arrival. A visual inspection reveals that the angle of arrival is reasonably accurate below 5kHz. Above this point the estimation is erratic. As an example,

the reflection intended to arrive at 45° (shown in red) displays results very close to this angle below 5kHz. The data hint shown at 1.89kHz shows an angle of arrival of 38.19°.

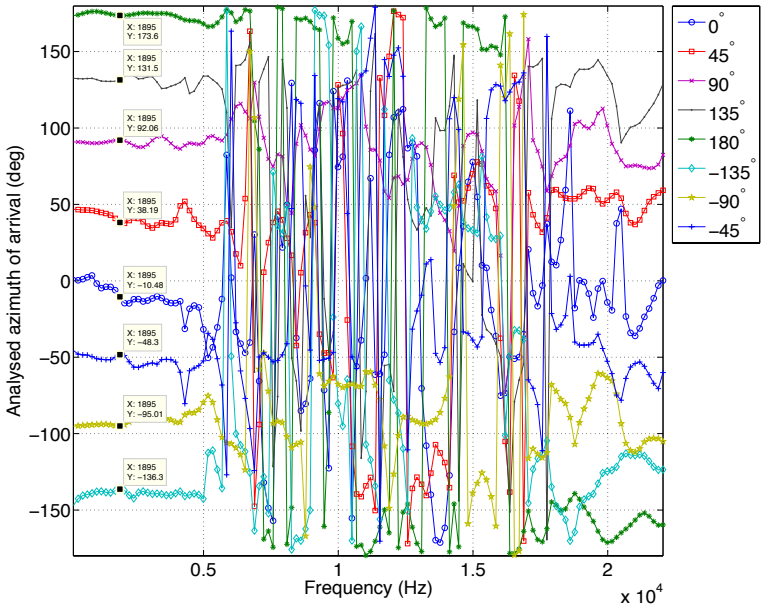


FIGURE B.2: *Analysed angle of azimuth of artificial reflections created by the SoundLab at intended angles of arrival shown in the legend. These raw results demonstrate that the correct angle of arrival was obtained only below 5kHz*

In order to demonstrate this more clearly, the angular error was derived for each artificial reflection at all frequencies. The angular error includes both azimuth and elevation data and represented the angle between the intended reflection vector and the measured reflection vector.

The results shown in Figure B.3 shows the mean (shown in blue) and maximum angular error (shown in red dashed) measured from artificial reflections occurring from loudspeakers surrounding the Soundfield microphone. It can be seen that in general, the angular error increases with frequency until approximately 17kHz below which the error appears to reduce.

It can be seen that the mean angular error is below 20° below a cut off frequency of approximately 5kHz. Below this frequency the mean and maximum angular error can be seen to be very similar whereas above 6kHz the maximum angular error is considerably higher than the mean value.

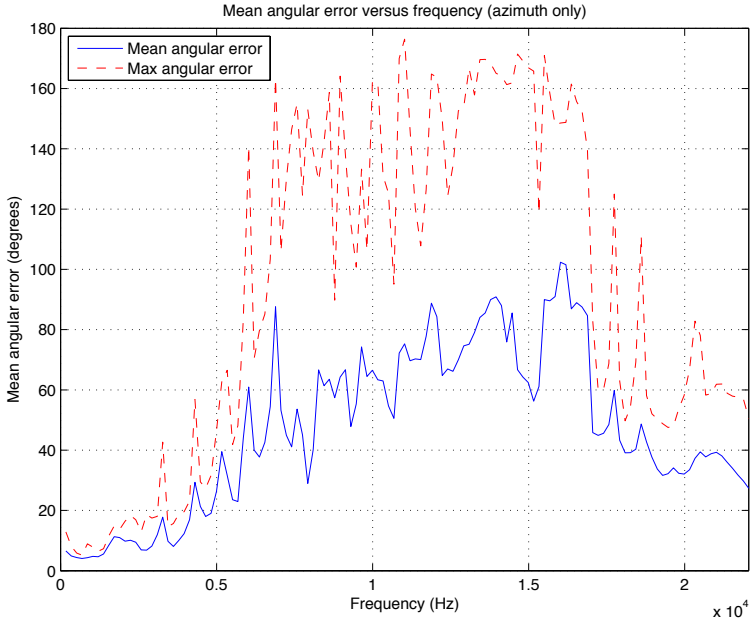


FIGURE B.3: *Plot showing the mean (blue) and Maximum angular error versus frequency. The data includes 8 artificial reflections from loudspeaker occurring at 45° intervals around the lateral plane.*

The reduction in error observed above 17kHz is not likely to be an indication that localisation has improved, rather the relative placement of the loudspeaker and microphone may have produced an attenuation at high frequencies (due to the directivity of the tweeter not aligning with the microphone).

The maximum error observed is approaching 180° at above 10kHz which is considerably high, representing a full reversal of the sound localisation. This magnitude is similar to that measured by Vilkamo (2008).

The experimental apparatus and test procedure was set up to test a number of different aspects of the soundfield reproduction system and while great care was taken to ensure the Soundfield microphone was positioned in the centre of the loudspeaker array it is likely that the angular location of the loudspeakers relative to microphone deviated slightly from the expected angles, introducing an error into the results.

It is acknowledged that this test was performed with reflections panned to the lateral plane. A more comprehensive set of results may have been obtained by repeating the measurements at different angles of elevation.

The Soundfield microphone was positioned at a height of 1.30m which corresponds to the height of the loudspeakers measured up to the top of the low frequency driver. This would result in the tweeters of the loudspeaker array being slightly higher than the microphone possibly introducing further high frequency errors.

The SoundLab (while heavily acoustically treated) is not an anechoic space and so it is possible that the room acoustic response introduced further error into the angular estimation. The

directional analysis was performed on the regions of the acoustic response that corresponded with the time of arrival of the artificial reflections. Furthermore, the artificial reflections were timed so that the SoundLab acoustic response had decayed sufficiently before the next reflection began. These factors minimised the error as far as possible however it is acknowledged that more reliable results could be produced in an anechoic chamber using precision tools to align the Soundfield microphone to the loudspeakers.

The estimated cut off frequency is similar to those recommended by other authors and therefore these results are considered to be a reasonable estimate of the angular error produced by the Soundfield microphone for the purposes of this research.

From the results discussed previously, it can be seen that the measured angle of arrival begins to increase beyond a frequency of approximately 3.5kHz and exceeds an error of 20° degrees beyond a frequency of 5kHz. This agrees with similar tests performed by Protheroe and Guillemin (2013) and Vilkamo (2008). Currently, there is no data regarding the spatial acuity of musicians to early reflections and so the cut off frequency chosen is arbitrary. In this research, a cut off frequency of 5kHz is proposed as a reasonable level of accuracy for both image source analysis and re-synthesis of impulse responses for auralisation.

The intensity analysis technique has been used previously to obtain image source plots showing the spatio-temporal distribution of early reflections. The angle of arrival is normally obtained by finding the maximum value of a histogram of the analysed angle of arrival. Analysing the direction of arrival using intensity vectors below a cut off frequency of 5kHz should reduce the noise in the histogram allowing the direction of arrival to be obtained with higher accuracy. Initial experiments in this area have shown the elevation of early reflections is obtained more accurately with a slight improvement in azimuth estimation.

The results of this experiment have shown that the angular estimation can not be relied upon at frequencies above 5kHz. Therefore, early reflections are likely being spatialised incorrectly at high frequencies. As demonstrated by Vilkamo (2008) the error observed at high frequencies will also result in SIRR overestimating the diffuseness of the soundfield at any given time. In the synthesis phase of SIRR, the diffuseness rating acts as a cross-fade between a fully diffuse soundfield (rendered with decorrelated speaker feeds) and a non-diffuse soundfield (rendered with VBAP). Therefore, at high frequencies, the non-diffuse components are likely being rendered at a lower amplitude than expected with the diffuse soundfield dominating.

In summary, a B-format soundfield recorded with a Soundfield ST350 microphone that is processed by SIRR will produce reflections where the high frequencies are likely to be incorrectly spatialised and of lower amplitude than in the recorded scenario. This problem can be overcome by using a different 'front-end' microphone system which is more accurate at higher frequencies. Often SIRR (or similar) reproduction techniques use an intensity probe with spaced omni directional transducers to record room impulse responses. While the microphone spacing can affect the accuracy at different frequencies, this approach has been demonstrated to be successful.

This study aimed to determine how accurately artificial reflections could be localised in space using a Soundfield ST350 Ambisonic microphone. The results showed that the angle of arrival was determined to a reasonable level of accuracy (below approximately 20°) below a frequency

of 5kHz. These results agree with previous studies in this area. A cut off frequency of 5kHz is proposed for both analysis and re-synthesis of early reflections. This study has suggested that the high frequency parts of early reflections (when synthesised using SIRR) should be spatialised in a direction extrapolated from the data below 5kHz, using the Mean resultant vector.

Appendix C

SoundLab calibration

To accurately spatialise the direction of a sound source over a loudspeaker array, it must be confirmed that the contribution of sound pressure level from each loudspeaker is equal at the sweetspot. It is also critical that the loudspeakers do not introduce any spectral colouration as a consequence of their frequency response. Furthermore, it is necessary that there is no significant delay between signals emanating from each loudspeaker. Failing to address these factors can introduce localisation errors and timbral colouration in the resulting auralisation. Therefore it is necessary to equalise the contribution of each loudspeaker prior to rendering any auralisations.

The SoundLab features a 16-channel 3D loudspeaker array comprising of three rings of Yamaha MSP5A active loudspeakers arranged where there are eight loudspeakers on the lateral plane and four speakers in the top and bottom rings. The loudspeakers are connected (via balanced underfloor cable routing) to either a SoniCore A16 ultra D-A Converter or the analogue outputs of a Motu 896 USB/Firewire Soundcard. This soundcard is connected via USB 3.0 to a Mac Pro running OS X Maverick.

The layout of the loudspeaker array is shown in Figure C.1 and described in Table C.1 where the angular location of each loudspeaker, relative to the sweetspot, is given. It can be seen that the ceiling loudspeakers have a higher angle of elevation than the floor loudspeakers, resulting in a non-spherical loudspeaker layout. This is primarily due to the height of the middle ring of loudspeakers where the height was limited by the loudspeaker stands available. This results in the height of the sweetspot being closer to the floor than the ceiling.

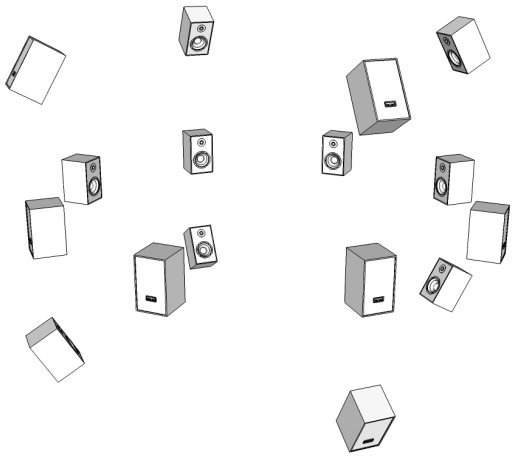


FIGURE C.1: *Layout of loudspeakers in the SoundLab*

Description	θ	ϕ	$r(m)$
Upper - Rear	181.2	41.5	1.62
Upper - Left	90.2	41.5	1.64
Upper - Front	0.6	41.2	1.65
Upper - Right	-91.1	41.6	1.64
Middle - Rear Right	-136.9	0	1.65
Middle - Rear	180.9	0	1.66
Middle - Rear Left	135.3	0	1.66
Middle - Left	92.0	0	1.65
Middle - Front Left	46.4	0	1.65
Middle - Front	1.9	0	1.66
Middle - Front Right	-44.4	0	1.66
Middle - Right	-90.6	0	1.65
Lower - Rear	180.0	-31.2	1.68
Lower - Left	91.7	-31.2	1.68
Lower - Front	3.7	-31.7	1.67
Lower - Right	-92.2	-31.4	1.68

TABLE C.1: *Location of loudspeakers shown in polar coordinates.*

In Table C.1 it can be noted that be seen that the distance to each loudspeaker is not equal causing inter-channel delay. In order to prevent this affecting the spatialisation accuracy, a delay line was introduced into each output channel of the Max patch. The delay was applied to all but the furthest loudspeaker channels such that the 'virtual' distance to each loudspeaker was the same, reducing the likelihood of inter channel delay. The delay was calculated using the radius r and multiplying by the speed of sound.

The level contribution of each loudspeaker was equalised by measuring the level of a pink noise signal played separately through each loudspeaker with a Sound Level Meter. The microphone from a B & K 2260 SLM was mounted on a microphone stand in the sweetspot of the loudspeaker array and connected to the meter via an extension cable. The SLM was set to record L_{eq} in

third octave bands with a measurement duration of 30 seconds. Prior to measurements, the SLM was calibrated, as is common practice, using the accompanying calibrator which fits over the microphone capsule and emits a $1kHz$ tone at $L_p = 94.0dB$.

A bank of parametric filters were implemented in Max for each loudspeaker channel between the output of the auralisation patch and the loudspeaker output. The filters were implemented using the *SPAT.eq* externals (IRCAM, No date). Pink noise was played through each output channel of the auralisation patch and the filters were adjusted so that the spectrum measured from each loudspeaker was visually as flat as possible above 100Hz. To avoid excessive phase rotation, only minor adjustments (i.e. no greater than $\pm 6dB$ gains) were made.

After the spectrum of each loudspeaker had been adjusted, the gain of each channel was adjusted so that each loudspeaker produces a broadband L_{eq} of 75dB at the sweetspot. Figure C.2 shows the average level (and standard deviation) measured in each third octave band for all 16 loudspeakers after the equalisation filters have been applied. Above 100Hz, the standard deviation is no larger than 2.7dB. Above a frequency of 400Hz the standard deviation does not exceed 1dB.

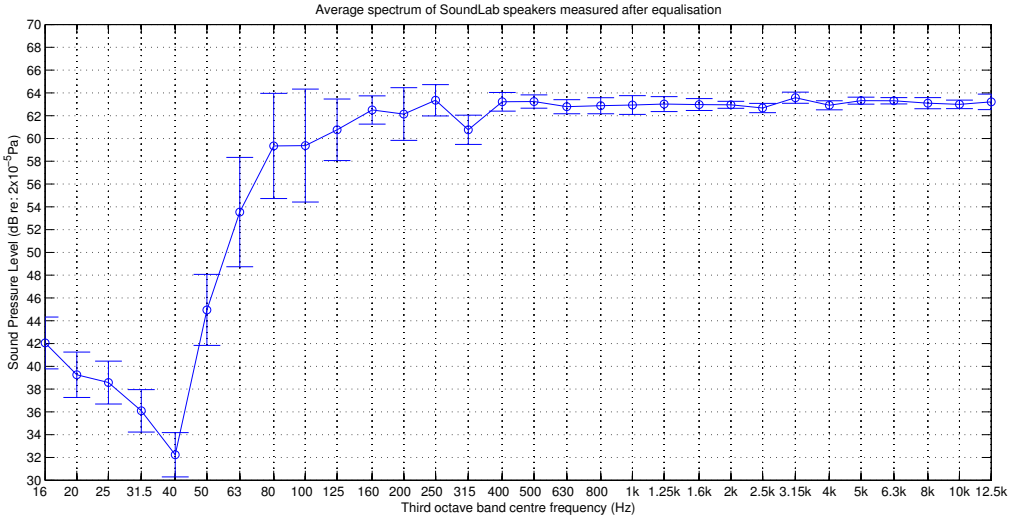


FIGURE C.2: *Mean and standard deviation of frequency response of all 16 loudspeakers after equalisation filters applied and gain adjusted so that each speaker produces* $L_{Aeq} = 75dB$.

When used for other auralisation projects, the SoundLab also makes use of a Tannoy T12 subwoofer to extend the low frequency content of the auralisation. The subwoofer was not included in the auralisation work for this thesis but, for completeness, is shown in these results.

The subwoofer is positioned at a radius of $1.7m$ from the sweetspot at an angle of 115° clockwise from the front of the loudspeaker array. Pink noise is played through the subwoofer and the cut off frequency of the internal low pass filter adjusted so there is minimal contribution above $100Hz$. The level of the subwoofer is adjusted by playing pink noise through it in addition to uncorrelated pink noise through the equalised satellite loudspeakers. The level is adjusted so that visually, the low frequency response measured at the sweetspot is visually flat to a frequency that is as low as possible. Figure C.3 shows the sound pressure level measured at the sweet spot

for the subwoofer only. It can be seen that the frequency response peaks at 31.5Hz to a level of $71.5dB$ (L_{eq}) and drops off rapidly above $100Hz$.

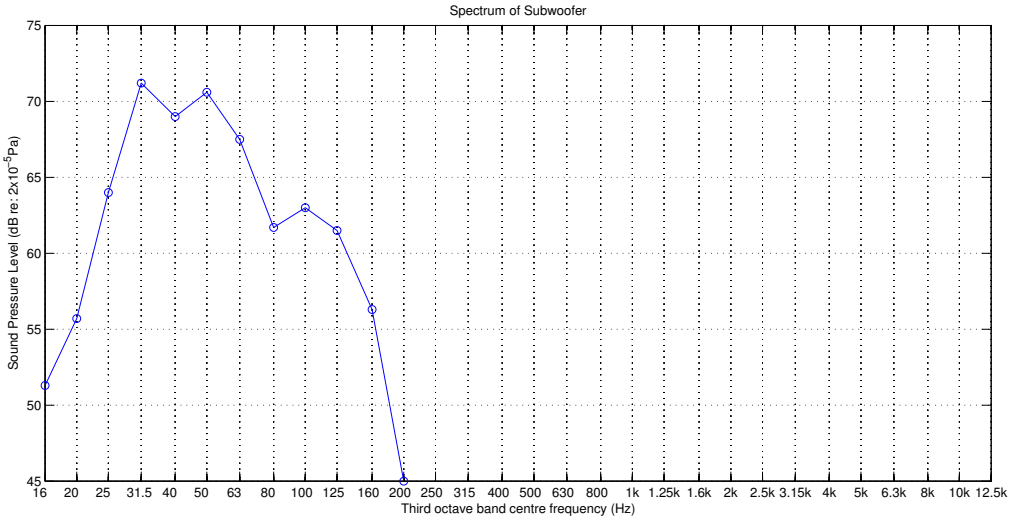


FIGURE C.3: *Sound pressure level of pink noise rendered through the subwoofer only. Measurement was made at the sweetspot of the loudspeaker array.*

A final gain adjustment was made so that when correlated pink noise was played through all loudspeaker channels (including the subwoofer), a broadband L_{Aeq} of 85dB was measured at the sweetspot. Uncorrelated pink noise was played through each loudspeaker channel and the spectrum measured at the sweetspot. Figure C.4 shows the spectrum of this measurement in third octave bands. A broadband L_{Aeq} of $82.7dB$ was measured. It can be seen that the low frequency response measured at the sweetspot, compared with Figure C.2, has been extended down to a frequency of $31.5Hz$.

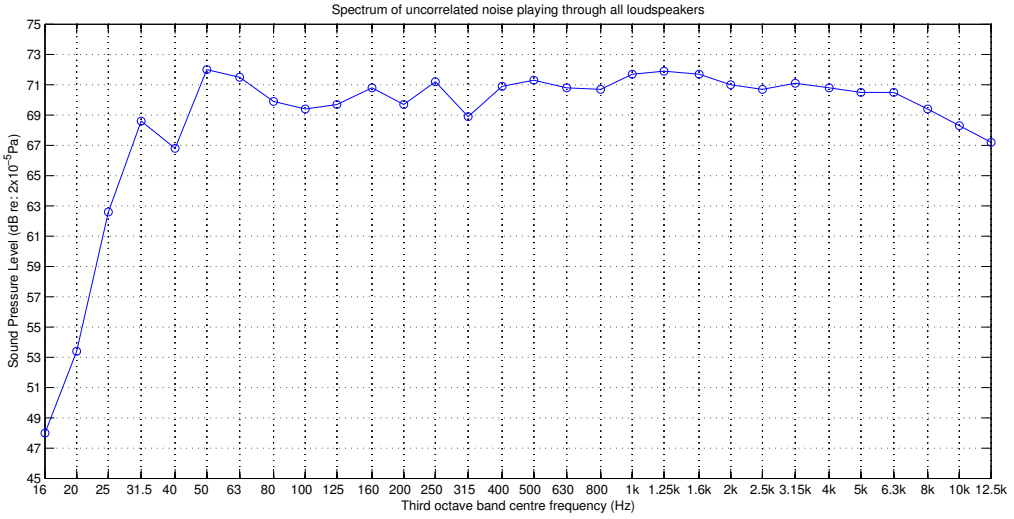


FIGURE C.4: *Spectrum measured at the sweetspot when uncorrelated pink noise was played through all loudspeakers and subwoofer simultaneously after calibration.*

Above a frequency of $100Hz$ (as the subwoofer is not used in this research) it can be seen that uncorrelated noise playing through all satellite speakers produces a flat frequency response within a tolerance of $\pm 3dB$. Furthermore, all loudspeakers are calibrated to the same broadband level. This will ensure that the loudspeaker array does not influence the spatialisation of sound sources and that the spectral characteristics of the auralisation are maintained.

Appendix D

SoundLab acoustic conditions

Previous stage acoustic laboratory experiments (such as those by Gade (1989) and Ueno and Tachibana (2003)) featured a spatial audio system housed inside an anechoic chamber. This allowed the virtual acoustic response to be heard by the test subject with no additional colouration from the room. However, in many circumstances, it is difficult to obtain access to such facilities and so auralisation systems are often housed in acoustically controlled, but non-anechoic listening rooms (such as those used by Brereton et al. (2012a) and Guthrie (2014)). The consequence of this is that the laboratory contributes to the acoustic conditions experienced by the musician and also to measurements made of virtual spaces. By ensuring that the length of the laboratory impulse response is significantly shorter than the virtual reproduction, this contribution can be minimised.

By performing measurements similar to those described in the concert hall surveys, the acoustic conditions of the SoundLab can be characterised prior to auralisation to further understand its influence. Measurements were made in the sweetspot of the loudspeaker array with an equipment set up similar to that used for the performance space surveys. Swept sine waves were played through the loudspeaker and measured by the microphone. The signal from the microphone is processed using the interactive auralisation system and played back over the loudspeaker array.

The SoundLab is an acoustically controlled space with dimensions 4m (l) x 6m (w) x 2.5m (h). The SoundLab is a box-in-box structure where the internal structure is isolated from the rest of the building by constructing it on a spring floated concrete slab. Furthermore, the majority of noisy audio equipment is kept in an adjacent control room. Therefore, the SoundLab has a very low background noise level. The SoundLab is frequently used as a commercial auralisation system and therefore features screens in front of, and behind the sweetspot. The front screen is an LCD television positioned behind angled glazing whereas the rear screen is made of plastic and mounted within a wooden frame to facilitate the use of rear-projected stereographic projections.

The floor reflection is typically removed from the measured room impulse response and is reproduced naturally by having the musician stand on a hard floor. In order to be more representative of a concert hall stage, The SoundLab floor is covered with tongue and groove wooden floorboards at the centre of the loudspeaker array. The wooden area is 1.85m by 1.8m ($3.33m^2$).

The SoundLab has an LCD screen built into the wall directly in front of the sweetspot. There is an angled glass cover in front. The distance from the sweetspot to the LCD screen is $2.05m$. When sound is generated at the sweetspot, a reflection arrives from this direction shortly after the direct sound. Porous acoustic treatment is used to cover the glass cover when experiments are taking place to reduce the effect of this reflection. The panels are held loosely in place over the surface of the screen with tape. The SoundLab also has a projector screen directly behind the loudspeaker array. The plastic panel is normally covered with heavy absorbing curtain. The distance to the projector screen is $2.25m$

Figure D.1 show the SoundLab in two of the measured configurations where Figure D.1(a) shows the SoundLab with carpeted floor with porous absorption applied to the LCD screen whereas D.1(b) shows the SoundLab floor covered in a layer of tongue and groove slats and the screen absorption removed.

(a) Without floor - With Screen absorption

(b) With floor - Without Screen absorption

FIGURE D.1: *Photograph of SoundLab with wooden floor and screen absorption*

The acoustic characteristics of the SoundLab in each configuration were measured in a similar fashion to the performance space surveys as described in Chapter 4. This method involves an Ambisonic microphone positioned at the sweet spot of the loudspeaker array directly above a measurement loudspeaker. Impulse responses were measured using the swept sine wave technique with the source rotated in 45° increments after each measurement. Measurements were made with and without a wooden floor present and also with and without the TV screen treatment applied.

Figure D.2 shows the measured EDT in the SoundLab expressed in octave bands. The EDT is shown as the mean value obtained from all eight measurements in each case (where the sound source has been rotated by 45° after each measurement). The standard deviation is also shown. It is clear that EDT is reduced when the additional room treatment is applied. Furthermore, the

standard deviation is reduced when the treatment is applied which implies that EDT has become more uniform with varying source direction. This is to be expected as the TV absorption in the direction of the TV screen has been increased towards that of the side walls of the SoundLab. The largest reduction in mean EDT is 13ms at a frequency of 1kHz. An increase of 5ms at 2kHz is observed when the treatment is applied.

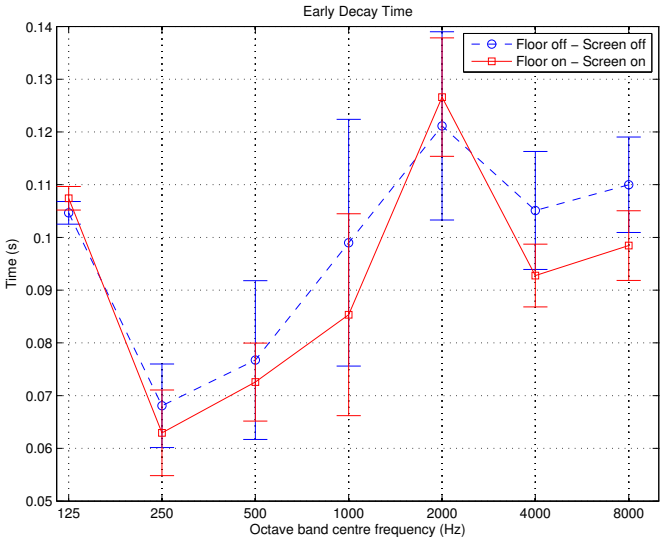


FIGURE D.2: *Octave band EDT measured at the sweetspot of the SoundLab with various configurations.*

Figure D.3 shows a similar set of results showing how the reverberation time (T_{30}) varies with and without additional room treatment. It can be seen that T_{30} reduces when the additional room treatment is applied, mainly below a frequency of 8kHz. The largest change of 17ms occurs at a frequency of 500Hz. Like EDT, it can be seen that the standard deviation of T_{30} is reduced at all frequencies when the additional room treatment is applied, implying that T_{30} has become more uniform with source direction.

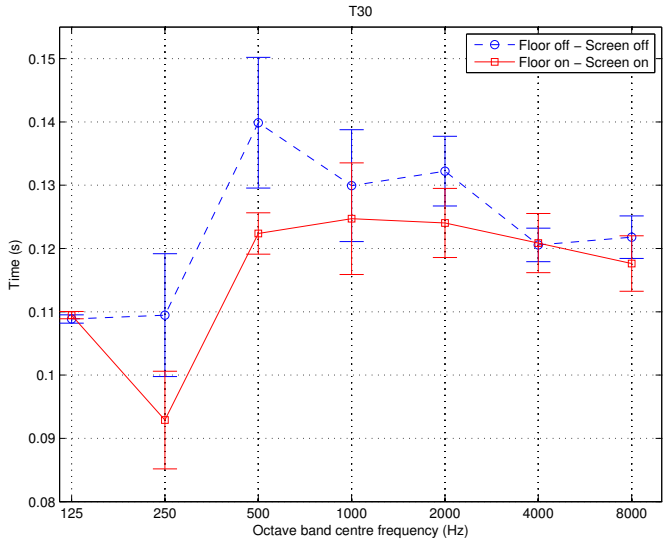


FIGURE D.3: *Octave band T_{30} measured at the sweetspot of the SoundLab with various configurations.*

Figure D.4 shows the measured values of ST_{early} for different source orientations in the SoundLab, with (red) and without (blue) the additional room treatment applied. ST_{early} is shown as the mean value obtained in octave bands between 250Hz and 2kHz. It can be seen that in both cases, the highest values of ST_{early} occurs when the source is facing away from the LCD screen (source angle of 180°). It can also be seen that higher values occur when the additional room treatment is not in place. When the source is facing 0° orientation ST_{early} reduces by 0.86dB when the room treatment is applied to an ST_{early} of -14.36dB.

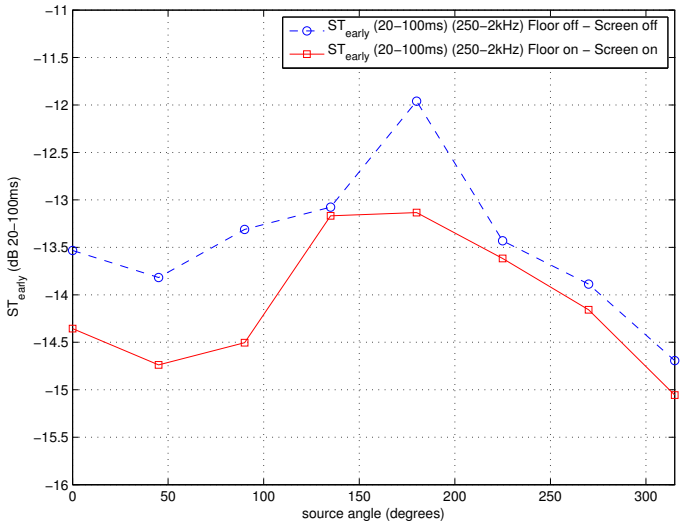


FIGURE D.4: *ST_{early} measured at the sweetspot of the SoundLab with (red) and without (blue) additional room treatment in place. This plot shows the mean value of ST_{early} obtained in octave bands between 250Hz and 2kHz.*

A similar comparison is shown in Figure D.5 for ST_{late} showing the variation with source angle, with (red) and without (blue) the additional room treatment. It can be seen that when the room treatment is applied, ST_{late} varies comparatively little and is attenuated by more than 3dB. When the source is facing 0° orientation ST_{late} is reduced by 4.19dB when the room treatment is applied to an ST_{late} of -52.84dB.

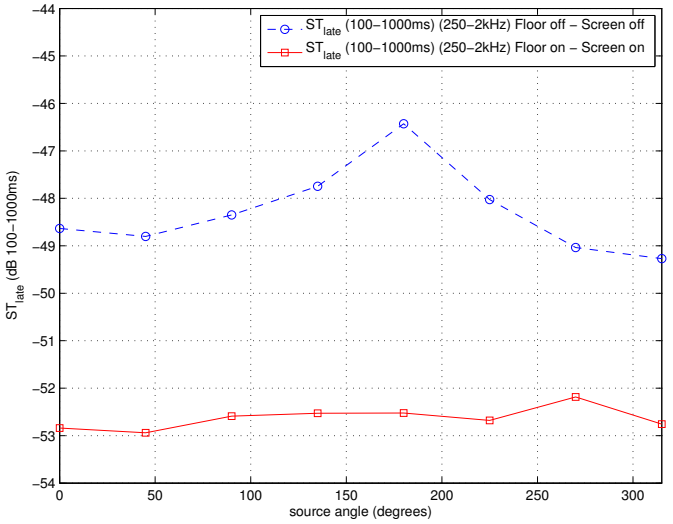


FIGURE D.5: *ST_{late} measured at the sweetspot of the SoundLab with and without absorption over the LCD screen and a wooden floor in place. This plot shows the mean value of ST_{late} obtained in octave bands between 250Hz and 2kHz.*

The effect of the different configurations can be seen by observing the early part of the impulse response measured in the SoundLab. Figure D.6 compares the impulse responses (where the loudspeaker is facing forward) when the floor and screen treatment are applied and when they are taken away. It can be seen that the amplitude of the reflection from the LCD screen is significantly reduced when porous absorption is applied to the LCD screen.

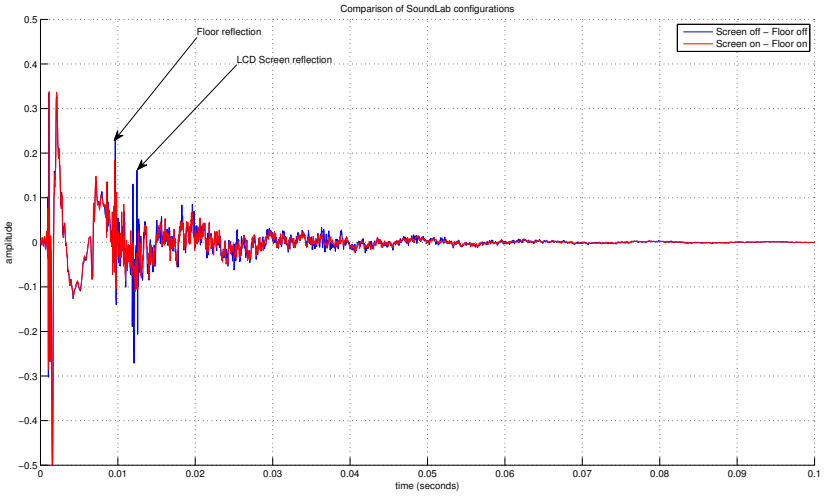


FIGURE D.6: *Comparison of impulse response measured in SoundLab under different configurations. The blue trace shows the acoustic response of the SoundLab with no treatment and the loudspeaker oriented forward, towards the LCD screen. The red trace shows an identical measurement with a wooden floor and porous absorption applied*

While anechoic conditions would be preferred, the acoustic response of the SoundLab is considered to be reduced as far as possible so that it is suitable for auralisation of most virtual stage acoustic environments.

These measurements also provide a basic indication of the the minimum values of acoustic parameters that can be auralised in the SoundLab. For example, it is not recommended to attempt to auralise early reflections with an ST_{early} of less than -14dB. From measurements obtained in performance space surveys it was found that ST_{early} values in small recital spaces were often above a value of -14dB. Much larger stages featured ST_{early} values of less than this. Therefore, the auralisation system will not be able to reliably emulate the acoustics of large stages.

It was also of interest to determine the background noise of the SoundLab. Ideally, the background noise of the SoundLab should not exceed the background noise observed in the concert hall measurements. A SLM was used to measure the background noise spectrum using the L_{90} metric which shows the level of noise exceeded over 90% of the measurement time. Measurements of 5 minutes duration were made at the sweetspot. Figure D.7 compares the background noise measured at the sweet spot of the SoundLab when all equipment was switched on, versus all equipment switched off (not including the lights). The plot shows the L_{90} in octave bands as measured over the 5 minute period. It can be seen that in both plots the background noise only exceeds NR20 at a frequency of 8kHz. Below a frequency of 2kHz, it can be seen that the background noise does not exceed NR10. This is significantly lower than the background noise measured in most concert halls.

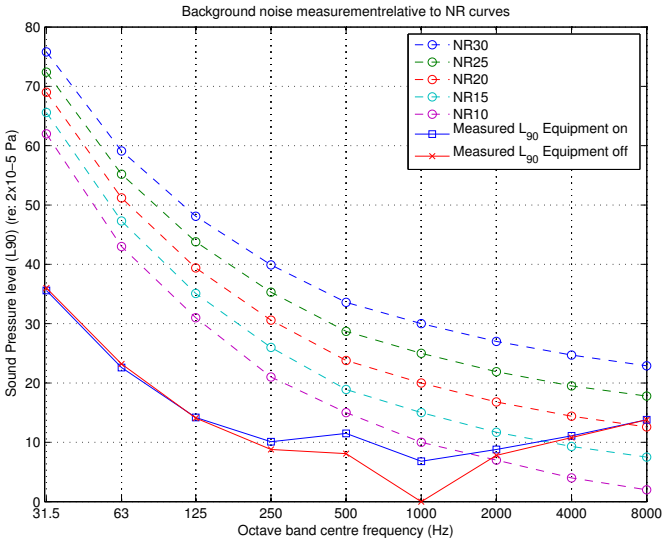


FIGURE D.7: *Background noise spectrum relative to NR curves (International Organisation of Standardisation, 1999). The blue trace shows the background noise with all equipment switched on. Whilst the red trace shows the background noise measured with all equipment (with the exception of lights) switched off.*

The lack of background noise is desirable so that it does not influence the musician's impression of the auralisation. However, a low background noise can make artefacts in the auralisation appear audible. In order to avoid this, background noise measured in the venue is added back into the auralisation to ensure it is more representative of the target space.

Appendix E

Instructions to participants

Test Protocol

Thank you very much for agreeing to be part of this study. Your time is very much appreciated.

The activity you are about to take part in is an interactive listening test where you will be asked to play short musical phrases in different concert hall stages recreated using an array of loudspeakers. In each trial, you will be asked to answer questions based on what you have heard. The test has been designed to learn more about how musicians perceive the acoustic conditions they experience on stage during a performance. This document will describe how the test is conducted and what you are being asked to do. Please read through carefully prior to the test and indicate to the researcher if you have any queries on arrival.

The test is composed of three different exercises, which are expected to take approximately 30 minutes each. The exercises are separated by 10-minute breaks. You may proceed with each exercise at your own pace, however, it is expected that the test will take approximately 2 hours to complete.

The table below shows an approximate timetable for the test.

Time from arrival	**Activity**
0-10 minutes	Arrival and warm up
10-40 minutes	First test
40-50 minutes	*Break 1*
50 – 80 minutes	Second test
80-90 minutes	*Break 2*
90-120 minutes	Third test

Please ensure that you select some musical phrases to play throughout the experiment. Feel free to bring along sheet music, a music stand will be provided.

You are free to choose the musical phrases you play in each virtual hall, however it is recommended to keep them short (no more than 8 bars) and to ensure you are comfortable performing them. As you will be asked to play in a number of virtual halls, it is recommended that you identify around five different phrases for variety.

Checklist

- ✓ Please ensure you bring your instrument with you
- ✓ Please complete the attached consent form
- ✓ Please bring 5 short musical phrases

The SoundLab

The listening test will take place in the SoundLab at the Digital Design Studio, Pacific Quay. The SoundLab is an acoustically treated laboratory and is acoustically isolated from the building and therefore has a very low background noise.

Figure 1 Image of musician playing in virtual hall

Inside the SoundLab you will see an array of loudspeakers on stands, on the ceiling and on the floor arranged in a sphere all pointing to the centre of the room. One of the loudspeakers is directly in front of the door, please be careful when moving past it or any cables or equipment you see in the room.

You will be asked to sit on a stool in the centre of the loudspeaker array with a microphone positioned in front of you (see Figure 1). You will also see a music stand and a touch screen interface nearby. You will use the interface to control each test and record your responses. The microphone picks up the sound from your instrument in order to emulate the sound of you playing in a hall. The sounds you make are not being recorded.

Once you arrive at the SoundLab you will be asked to warm up your instrument for a short period before beginning the test. Please leave any coats or bags outside the loudspeaker array and please switch off your mobile phone.

Rules of the test

In order to ensure the reliability of the responses and for your safety, please observe the following rules when you are in the SoundLab.

- Please avoid touching or moving any audio equipment in the room.
- During the test, please face the front at all times and keep movements (relative to the microphone in front of you) to a minimum.
- The only persons in the room will be yourself and the researcher who will be controlling the test. The only record of the test will be your responses, which are stored as text files, and your completed consent form.
- During each exercise please only address the researcher in the event you have a problem or if you wish to stop the test.
- In case of any emergency, the test will cease and the researcher will guide you out of the SoundLab.
- Please ensure you complete and sign the consent form provided.

Listening tests

The following text describes each task including how to operate the interface. Please read the descriptions carefully prior to each test and feel free to ask the researcher for more detail as required.

Task 1
In this exercise there are 28 questions in total.

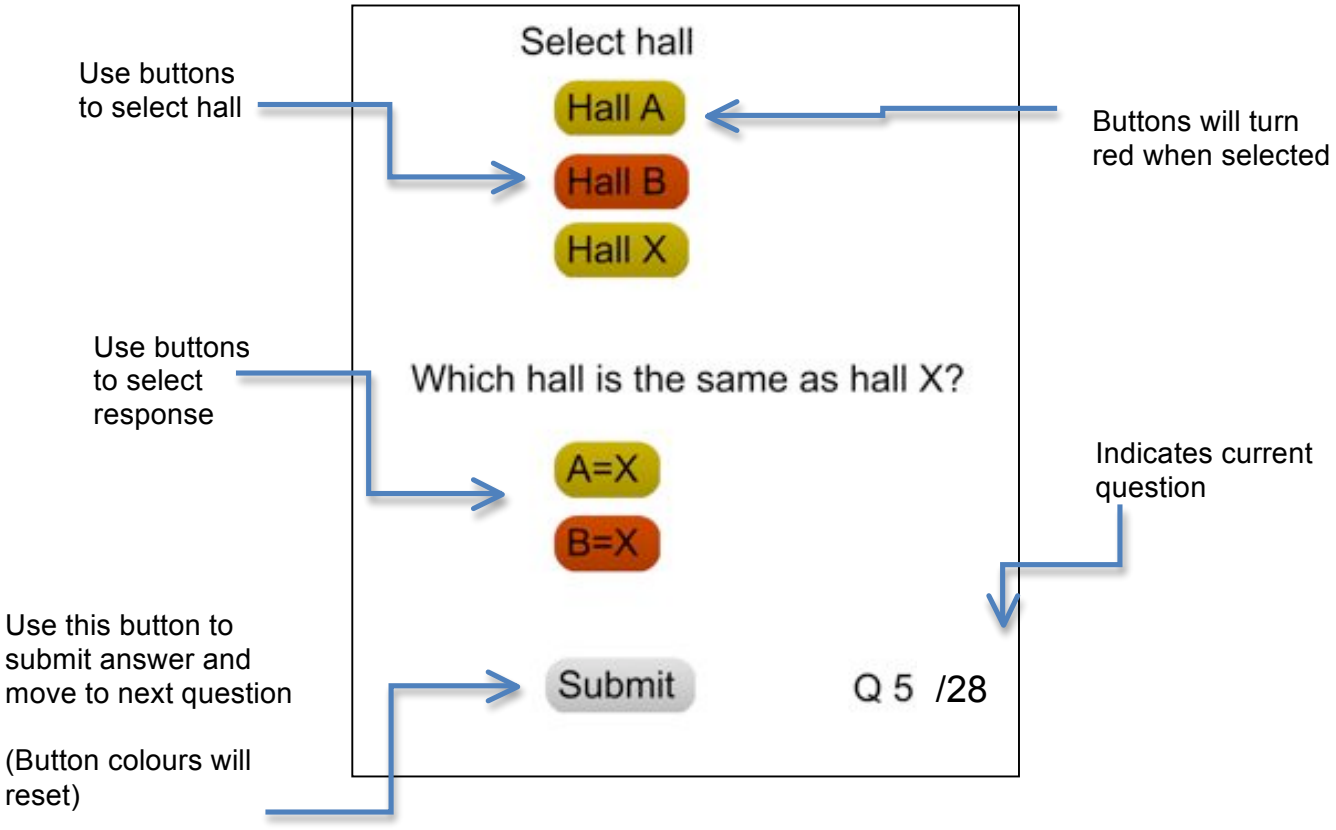
For each question, you will be presented with a choice of three virtual stages to play on. These are identified as A, B or X. You can choose which stage you are playing on by pressing the related button. When you select a concert hall, the related button will change colour to red.

After you have played in all three stages, please indicate which hall you thought sounded the same as hall X:

- *Press 'A=X' if you thought A sounded the same as X*
- *Press 'B=X' if you thought B sounded the same as X*

Once you have made your choice press the button marked 'Submit' to submit your answer and move on to the next question. When the answer is submitted, the interface will reset and you may proceed with the next question. The current hall will always revert back to Hall X after your response has been submitted.

Please note that once an answer has been recorded you will not be permitted to change it. You are allowed to play in each hall as many times as you wish before making your decision. Your progress in the test is indicated by a counter located in the bottom right hand corner of the interface.



Task 2
In this exercise there are 28 questions in total.

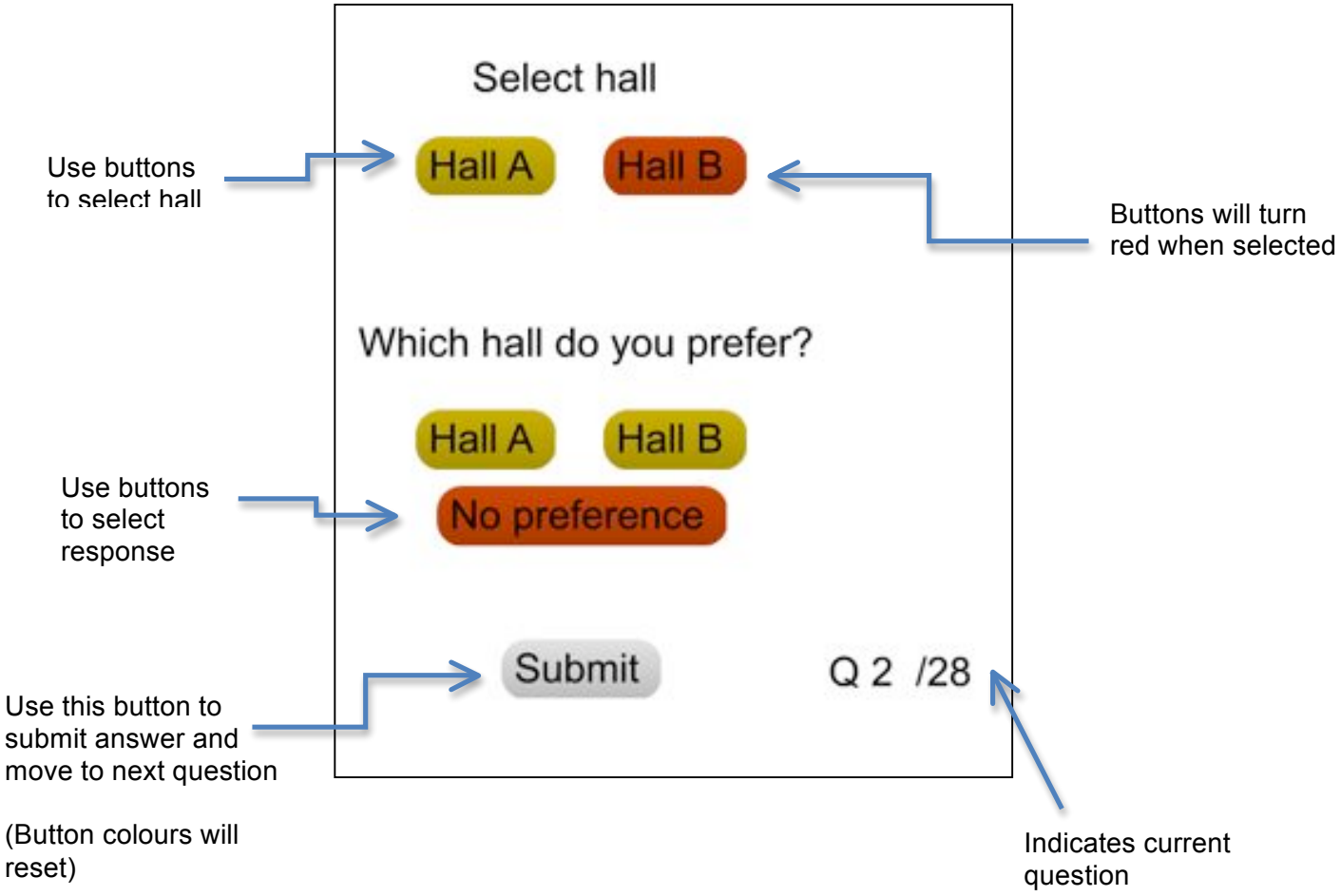
In this test you will be presented with pairs of virtual stages to play on. These are identified as Hall A and Hall B. As before, you can choose which hall you are playing in by pressing the related button. The buttons will change colour to red to indicate which hall you are currently playing in. Please ensure you select a hall to play in when you begin each question.

Once you have played in each hall, please indicate which one you most preferred playing in.

- *Press 'Hall A' if you preferred playing in Hall A*
- *Press 'Hall B' if you preferred playing in Hall B*
- *Press 'No preference' if you did not prefer one hall over the other.*

Once you have made your choice press the button marked 'Submit' to submit your answer and move on to the next question.

Please note that once an answer has been recorded you will not be permitted to change it. You are allowed to play in each hall as many times as you wish before making your decision. Your progress in the test is indicated by a counter located in the bottom right hand corner of the interface.



Test 3

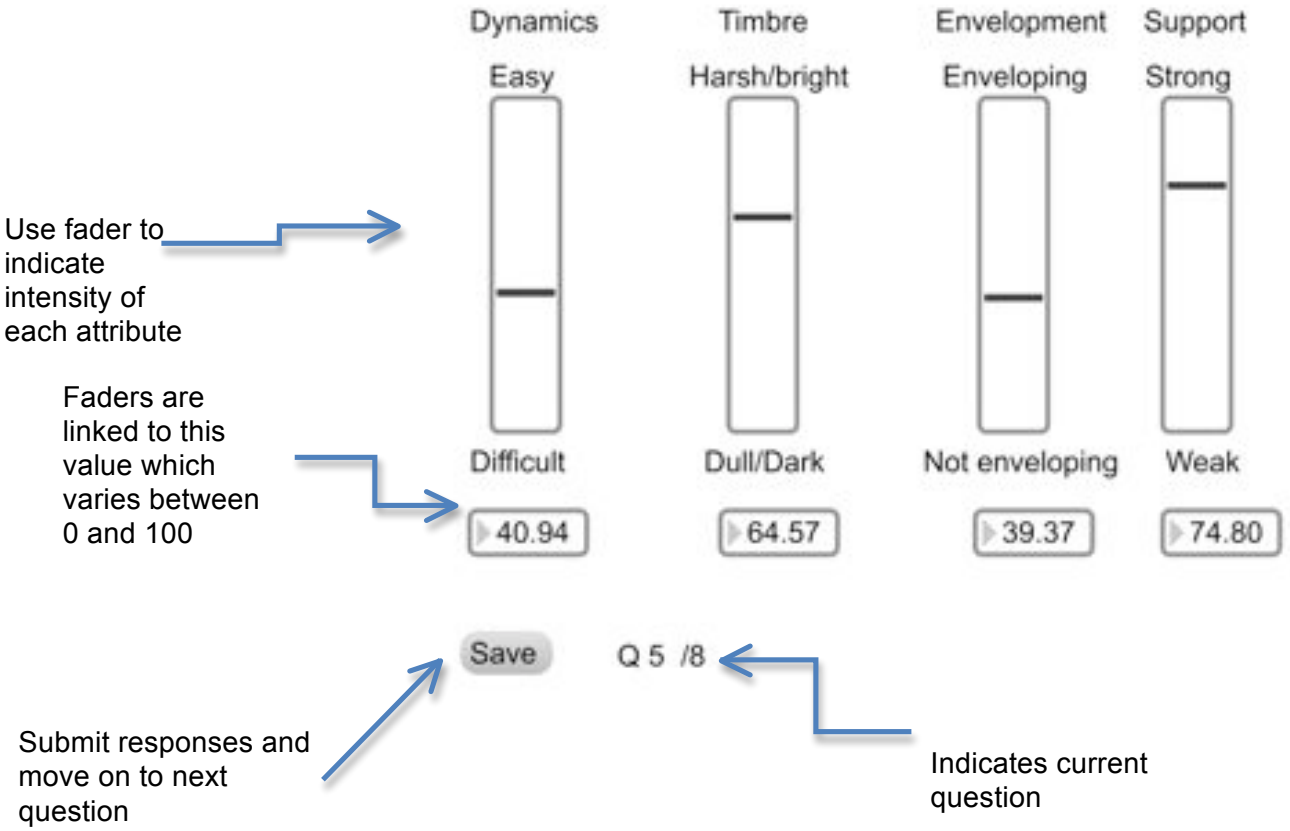
In this test you will be presented with a series of virtual stages to play on, one after another. After playing in the hall, you will be asked to rate the intensity of various attributes described below. You will record your responses for each virtual hall by adjusting a slider for each attribute. There are eight virtual halls in total, indicated by the counter at the bottom left of the display.

Each slider is linked to a display, which provides a numerical value between 0 and 100.

Once you have played in the hall, please use the sliders to indicate the intensity of each attribute. Once you are satisfied with your responses, press 'Save' to record your answer and move on to the next hall. Please ensure that you have recorded a response for each slider.

You are allowed to play in each hall as many times as you wish before making your decision, however you will not be permitted to go back and change your answers once they are submitted.

There are three practice questions at the beginning of this task which will not be recorded.



The attributes are defined below. Please read these definitions carefully prior to beginning this task.

Dynamics
Ease of varying the dynamic range of your instrument (does forte sound loud and piano sound soft?).

Timbre
The tonal quality of your instrument and the hall acoustics combined.

Envelopment
The extent to which you feel your are enveloped by the hall response (Does the acoustic response surround you or come mainly from a single direction?)

Support
Sense of how strongly your efforts feel supported by the stage (A weak sense of Support makes it feel as though you have to exert more effort for what you hear back from the hall).

Participant number:

Project title: The effect of spatio-temporal distributions of early reflections for performers

Name, position and contact address of Researcher:
Iain Laird, PhD Candidate, Digital Design Studio, Glasgow School of Art, The Hub, Pacific Quay, Pacific Drive, G511EA I.Laird1@student.gsa.ac.uk

Age:

Please detail below if you have any known hearing difficulties(e.g. Tinnitus, hearing loss etc) :

Please identify the musical instrument you will be playing today and how many years experience you have with the instrument:

Please indicate approximately how many years of public performing experience you have and the broad nature of those performances (e.g. 5 years symphonic orchestra, 5 years solo/accompanied etc.):

Consent form - Researcher Copy

Participant number:

Before you decide to take part in this research project it is important for you to understand why the research is being done and what will be involved. Please take time to read the attached information sheet carefully. Ask if anything is unclear or if you would like more information.

This research is partially funded by Arup Acoustics and the researcher holds a part-time position within the company. The responses obtained from this listening test will be used only as part of academic research at the Glasgow School of Art and will only published anonymously as part of academic submissions (i.e. PhD Thesis, Journal submissions etc)

Please Initial Box

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason. ☐
3. I agree to take part in the above study. ☐
4. I agree to the use and presentation of my anonymous results in academic publications. ☐
5. I have received my gift voucher as incentive for taking part in the test. ☐

Name of Participant (block capitals)	Date	Signature
Name of Researcher (block capitals)	Date	Signature

Consent form - Participant Copy

Participant number:

Before you decide to take part in this research project it is important for you to understand why the research is being done and what will be involved. Please take time to read the attached information sheet carefully. Ask if anything is unclear or if you would like more information.

This research is partially funded by Arup Acoustics and the researcher holds a part-time position within the company. The responses obtained from this listening test will be used only as part of academic research at the Glasgow School of Art and will only published anonymously as part of academic submissions (i.e. PhD Thesis, Journal submissions etc)

Please Initial Box

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason. ☐
3. I agree to take part in the above study. ☐
4. I agree to the use and presentation of my anonymous results in academic publications. ☐
5. I have received my gift voucher as incentive for taking part in the test. ☐

Name of Participant (block capitals)	Date	Signature
Name of Researcher (block capitals)	Date	Signature

Additional questions

From your previous experience, which venue has the best acoustic conditions for a solo performance? Please explain why you prefer this particular venue.

Do you find it necessary to adjust any aspects of your technique when performing in different venues? If so, please describe the type of adjustments you might make.

In the listening test, did you notice any differences between the concert halls you played in? If so, please describe any differences you heard.

In relation to stage acoustics or the experiment you have just completed, do you have any further comments you would like to add?

Bibliography

J. S. Abel and P. Huang. A simple, robust measured of reverberation echo density. San Francisco, USA, October 2006. Proceedings of the 121st Audio Engineering Society Convention.

J. S. Abel, N. J. Bryan, P. Huang, M. A. Kolar, and B. V. Pentcheva. Estimating Room Impulse Responses from Recorded Balloon Pops. San Francisco, USA, November 2010. Proceedings of the 129th Audio Engineering Society Convention.

J. Ahonen. *Microphone front-ends for spatial sound analysis and synthesis with Directional Audio Coding*. PhD thesis, School of Electrical Engineering, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland, March 2013.

J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustic Society of America*, 66(4), 1979.

Y. Ando. *Architectural Acoustics: Blending sound sources, soundfields, and listeners*. Modern Acoustics and Signal Processing, Springer-Verlag, 1st edition, 1998.

Arup. Glasgow city halls project sheet. `http://www.arup.com/Projects/Glasgow_City_Halls.aspx`, September 2013. URL `http://www.arup.com/Projects/Glasgow_City_Halls.aspx`. Accessed 23/09/2013.

L. Aspock, S. Pelzer, F. Wefers, and M. Vorlander. A real-time auralization plugin for architectural design and education. Berlin, Germany, April 2014. Proceedings of the EAA Joint Symposium on Auralization and Ambisonics.

M. Barron. *Auditorium acoustics and architectural design*. Spon Press, 2nd edition, 2009.

A. Bassuet. New acoustical parameters and visualization techniques to analyze the spatial distribution of sound in music spaces. Melbourne, Australia, August 2010. Proceedings of the International Symposium on Room Acoustics (ISRA).

E. Battenberg and R. Avizienis. Implementing real-tme partitioned convolution algorithms on conventional operating systems. Paris, France, September 2011. Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11).

S. Bech and N. Zacharov. *Perceptual audio evaluation - Theory, method and application*. John Wiley and sons, 1st edition, 2006.

M. Benjamin, Eric, R. Lee, and J. Heller, Aaron. Localization in horizontal-only ambisonic systems. San Francisco, USA, October 2006. Proceedings of the 121st Audio Engineering Society Convention.

L. Beranek. *Concert halls and opera houses, music, acoustics and architecture*. Springer, 2nd edition, 2004.

S. Berge and N. Barrett. A new method for B-format to binaural transcoding. Tokyo, Japan, October 2010a. Proceedings of the 40th international Audio Engineering Society Conference.

S. Berge and N. Barrett. High angular resolution planewave expansion. Paris, France, May 2010b. Proceedings of the 2nd international symposium on Ambisonics and Spherical Acoustics.

R. Bermond and W. Davies. Influence of diffuse reflections on the playing of musicians. Rome, September 2001. Proceedings of the 17th International Congress on Acoustics, 4D.10.01.

S. Bertet, J. Daniel, E. Parizet, L. Gros, and O. Warusfel. Investigation of the perceived spatial resolution of higher order ambisonics soundfields: A subjective evaluation involving virtual and real 3d microphones. Saariselka, Finland, March 2007. Proceedings of the 30th Audio Engineering Society Conference.

D. A. Bies and C. H. Hansen. *Engineering noise control: Theory and practice*. Spon Press, 3rd edition, 2003.

J. Blauert. *Spatial hearing - The psychophysics of human sound localization*. 2nd edition, 1996.

J. Blauert and B. Jonas. Acoustic communication: The precedence effect. *Forum Acusticum*, 2005.

E. Bourdillat. Auralization of sound fields in auditoria using wave field synthesis. Master's thesis, Laboratory of Acoustical Imaging and Sound Control, Department of Applied Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, July 2001.

M. Bourne. 7. Vectors in 3-D space. `http://www.intmath.com/vectors/7-vectors-in-3d-space.php`, November 2014. URL `http://www.intmath.com/vectors/7-vectors-in-3d-space.php`. Accessed 10/02/2015.

J. Bradley. Review of objective room acoustics measures and future needs. Melbourne, Australia, August 2010. Proceedings of the International Symposium on Room Acoustics (ISRA).

J. Brereton, D. Murphy, and D. Howard. A loudspeaker-based room acoustics simulation for real-time musical performance. Proceedings of the 25th UK Audio Engineering Society conference in association with the 4th International Symposium on Ambisonics and Spherical Acoustics, 2012a.

J. S. Brereton. *Singing in Space(s): Singing performance in real and virtual acoustic environments - Singers' evaluation, performance analysis and listeners' perception*. PhD thesis, Department of Electronics, University of York, York, UK, August 2014.

J. S. Brereton, D. T. Murphy, and D. M. Howard. The virtual singing studio: A loudspeaker-based room acoustics simulation for real-time musical performance. Odense, Denmark, June 2012b. Proceedings of the Joint Baltic-Nordic Acoustics Meeting (BNAM).

Brüel & Kjær. 4128C - Head and Torso Simulator (HATS). http://www.bksv.com/Products/transducers/ear-simulators/head-and-torso/hats-type-4128c?tab=descriptions#D7A0F12401CB41B9920601C54CB9874A, September No date. URL http://www.bksv.com/Products/transducers/ear-simulators/head-and-torso/hats-type-4128c?tab=descriptions#D7A0F12401CB41B9920601C54CB9874A. Accessed 23/01/2015.

D. S. Brungart, B. D. Simpson, M. A. Ericson, and S. K. R. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustic Society of America*, 110 (5) Pt. 1, 2001.

D. Cabrera, D. Lee, R. Collins, B. Hartmann, W. L. Martens, and H. Sato. Characterising the variation in oral-binaural room impulse responses for horizontal rotations of a head and torso simulator. Proceedings of the International Symposium on Room Acoustics, Melbourne, Australia, August 2010.

D. Cabrera, L. Miranda, C. Jin, and N. Epain. Spatial analysis of acoustic support on auditorium stages: modelling and measurement using high-order transducers. Fremantle, Australia, November 2012. Proceedings of Acoustics 2012.

D. Cabrera, M. Yadav, L. Miranda, R. Collins, and W. L. Martens. The sound of one's own voice in auditoria and other rooms. Toronto, Canada, 2013. Proceedings of the International Symposium on Room Acoustics.

Caird Hall. Caird Hall technical specifications. http://www.cairdhall.co.uk/caird/, No date. URL http://www.cairdhall.co.uk/caird/. Accessed 15/10/2014.

J. Chang and D. J. Carroll. How to use mdpref, a computer program for multidimensional analysis of preference data. Unpublished report, Bell Telephone Laboratories., 1968.

W. Chiang, S.-t. Chen, and C.-t. Huang. Subjective assessment of stage acoustics for solo and chamber music performances. *Acta Acustica united with Acustica*, 89(5):848–856, 2003.

A. Clifford and J. D. Reiss. Proximity effect detection for directional microphones. New York, USA, 2011. Proceedings of the 131st Audio Engineering Society Convention.

D. Courville. Spherical harmonics. http://web.uniovi.es/qcg/harmonics/harmonics.html, September 2007. URL http://web.uniovi.es/qcg/harmonics/harmonics.html. Accessed 30/01/2015.

T. Cox and P. D'Antonio. *Acoustic absorbers and diffusers: Theory, design and application.* Spon Press, 1st edition, 2004.

Cycling74. MaxMSP. https://cycling74.com/, September 2013. URL https://cycling74.com/. Accessed 08/11/2011.

B.-I. Dalenback. CATT Acoustics Modelling Software. http://www.catt.se/, No date. URL http://www.catt.se/. Accessed 20/11/2010.

J. J. Dammerud. *Stage acoustics for symphony orchestras in concert halls.* PhD thesis, University of Bath, Bath, UK, 2009.

J. J. Dammerud. Stage acoustics for symphony orchestras - just black magic? June 2011.

J. J. Dammerud. Details on stage and ensemble geometry related to masking effects. June 2012.

J. J. Dammerud and M. Baron. Early subjective and objective studies of concert hall stage conditions for orchestra performance. In *19th International Congress on Acoustics*, September 2007. URL `http://opus.bath.ac.uk/532/`.

J. Daniel, R. Nicol, and S. Moreau. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. Amsterdam, The Netherlands, March 2003. Proceedings of the 114th Audio Engineering Society Convention.

G. Defrance and J.-D. Polack. Measuring the mixing time in auditoria. *Journal of the Acoustical Society of America*, 123(5), June 2008.

G. Defrance, L. Daudet, and J.-D. Polack. Detecting arrivals within room impulse responses using matching pursuit. Espoo, Finland, September 2008. Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08).

Edinburgh Guide. Reid Concert Hall description. `http://www.edinburghguide.com/venue/reidconcerthall`, September 2014. URL `http://www.edinburghguide.com/venue/reidconcerthall`. Accessed 13/10/2014.

R. Essert. Progress in concert hall design. *EBU Technical Review*, 1997.

A. F. Everest and K. C. Pohlmann. *The Master Handbook of Acoustics*. McGraw-Hill, 3rd edition, 2009.

A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. Paris, France, 2000. Proceedings of the 108th Audio Engineering Society Conference.

A. Farina and R. Ayalon. Recording concert hall acoustics for posterity. Proceedings of the 24th AES International Conference on Multichannel audio, June 2003.

A. Farina and L. Tronchin. Measurement and reproduction of spatial sound characteristics of auditoria. *Acoustical Science and Technology*, 26(2):193–199, 2005.

F. Faul, E. Erdfelder, A. Lang, and A. Buchner. G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences, 2007.

N. H. Fletcher and T. D. Rossing. *The physics of musical instruments*. Springer, 2nd edition, 1998.

From MathWorld–A Wolfram Web Resource. Moore-penrose matrix inverse. `http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html`, September 2012. URL `http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html`. Accessed 23/08/2014.

A. C. Gade. Subjective room acoustic experiments with musicians. ISSN 0105-3027 32, Technical University of Denmark, 1982.

A. C. Gade. Investigations of musicians' room acoustic conditions in concert halls. part1: methods and laboratory experiments. *Acustica*, 69:193–203, 1989.

A. C. Gade. Acoustics for symphony orchestras; status after three decades of experimental research. Proceedings of the International Symposium on Room Acoustics, August 2010.

A. C. Gade. Subjective and objective measures of relevance for the description of acoustics conditions on orchestra stages. Toronto, Canada, June 2013. Proceedings of the International Symposium on Room Acoustics.

M. Gerzon. General metatheory of auditory localisation. Vienna, March 1992. Proceedings of the 92nd Audio Engineering Society Convention.

M. A. Gerzon. Periphony: With-height sound reproduction. Munich, Germany, 1972. Proceedings of the 2nd Audio Engineering Society Convention.

Gerzonic. DecoPro Ambisonic Decoder. `http://www.gerzonic.net/`, No date. URL `http://www.gerzonic.net/`. Accessed 23/01/2015.

M. Giovannini and A. Arianna. The acoustical characterization of orchestra platforms and uncertainty estimation of the results. *Applied Acoustics*, 71:889–901, 2010.

B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

Glasgow City of Music. Recital Room technical specifications. `http://www.glasgowcityofmusic.com/find_a/129_city_halls_recital_rooms`, No date. URL `http://www.glasgowcityofmusic.com/find_a/129_city_halls_recital_rooms`. Accessed 24/08/2014.

Glasgow Life. Glasgow City Halls. `http://www.glasgowconcerthalls.com/city-halls`, September 2013. URL `http://www.glasgowconcerthalls.com/city-halls`. Accessed 27/08/2010.

Glasgow University. Glasgow University Concert Hall Description. `%http://www.gla.ac.uk/schools/cca/facilities/musicfacilities/`, September 2014. URL `http://www.gla.ac.uk/schools/cca/facilities/musicfacilities/`. Accessed 16/11/2014.

G.R.A.S. Sound Vibration. G.R.A.S. 50VI-1 Vector intensity probe. `http://www.gras.dk/products/special-microphone/intensity-probes/50vi-1.html`, 2013. URL `http://www.gras.dk/products/special-microphone/intensity-probes/50vi-1.html`. Accessed 13/09/2014.

D. Griesinger. Further investigation into the loudness of running reverberation. *Proceedings of the Institute of Acoustics*, 17:35, 1995.

A. Guthrie. *Stage acoustics for musicians: A multidimensional approach using 3D Ambisonic technology*. PhD thesis, Rensselaer Polytechnic Institute, Troy, New York, April 2014.

A. Guthrie, S. Clapp, and N. Xiang. Using ambisonics for stage acoustics research. Toronto, Canada, 2013. Proceedings of the International Symposium on Room Acoustics.

C. C. J. M. Hak, R. Wenmaekers, J. P. M. Hak, and A. P. D. van Luxemburg. The source directivity of a dodecahedron sound source determined by stepwise rotation. pages 1875–1879, Aalborg, Denmark, June 2011. Proceedings of Fourm Acusticum.

C. C. J. M. Hak, R. Wenmaekers, and L. C. J. Luxemburg. Measuring room impulse responses: Impact of the decay range on derived room acoustic parameters. *Acta Acustica united with Acustica*, 98(2012):907–915, 2012.

T. Halmrast. Sound coloration from (very) early reflections. *Journal of the Acoustical Society of America*, 109(5):2303, 2001.

T. Halmrast. Coloration due to reflections. further investigations. Proceedings of the international congress on acoustics, September 2007.

T. Halmrast. When source is also receiver. Toronto, Canada, June 2013. Proceedings of the International Symposium on Room Acoustics.

A. Harker and P. A. Tremblay. The HISSTools Impulse Response Toolbox: Convolution for the Masses. Ljubljana, Slovenia, September 2012. Proceedings of the International Computer Music Conference.

J. Heller, Aaron, M. Benjamin, Eric, and R. Lee. Design of ambisonic decoders for irregular arrays of loudspeakers by non-linear optimization. San Francisco, USA, 2010. Proceedings of the 129th Audio Engineering Society Convention.

J. Heller, Aaron, M. Benjamin, Eric, and R. Lee. A toolkit for the design of ambisonic decoders. CCRMA, Standford, USA, April 2012. Proceedings of the Linux audio conference.

F. Hollerweger. An introduction to Higher Order Ambisonic. `flo.mur.at/writings/HOA-intro.pdf`, October 2008. URL `http://flo.mur.at/writings/HOA-intro.pdf`. Accessed 13/09/2014.

M. Holter, T. Corbach, and U. Zolzer. Impulse response measurement techniques and their applicability in the real world. Como, Italy, September 2009. Proceedings of the 12th international conference on Digital Audio Effects (DAFx-09).

D. M. Howard and J. Angus. *Acoustics and Psychoacoustics*. Focal Press, 2nd edition, 2001.

S. J. Hyung, H. K. Yong, and Y. J. Jin. Effects of stage volume and absorption on acoustics of concert halls. Sydney, Australia, August 2010. Proceedings of the 20th international congress on acoustics (ICA).

International Organisation of Standardisation. BS EN 61672-1:2013: Electroacoustics. Sound Level Meters - Specifications, 2013 .

International Organisation of Standardisation. BS EN ISO 8233:1999 - Sound insulation and noise reduction for buildings. Code of practice, 1999.

International Organisation of Standardisation. BS ISO 10399:2004 Sensory analysis - Methodology - Duo-trio test, 2004.

International Organisation of Standardisation. BS EN ISO 3382-1:2009 Acoustics- Measurement of room acoustic parameters, 2009.

International Telecommunications Union (ITU). *ITU-R BS.775-3 Multichannel stereophonic sound system with and without accompanying picture*. International Telecommunications Union (ITU), Geneva, 2012.

IRCAM. Spatialisateur product home page. `http://forumnet.ircam.fr/product/spat/`, No date. URL `http://forumnet.ircam.fr/product/spat/`. Accessed 05/05/2012.

J. Y. Jeon and M. Barron. Evaluation of stage acoustics in seoul arts center concert hall by measuring stage support. *The Journal of the Acoustical Society of America*, 117(1):232–239, 2005.

J. Y. Jeon, L. Hansol, and K. Y. Sub. Preferred acoustical conditions for musicians on stage in solo duet and ensemble performances. Krakow, Poland, September 2014. Proceedings of Fourm Acusticum.

E. Kahle. Room acoustical quality of concert halls: perceptual factors and acoustic criteria - return from experience. Toronto, Canada, June 2013. International Symposium on Room Acoustics (ISRA).

Z. S. Kalkandjiev and S. Weinzierl. Room acoustics viewed from the stage: Solo performers' adjustments to the acoustical environment. Toronto, Canada, June 2013. International Symposium on Room Acoustics (ISRA).

M. Kallinger, G. Del Galdo, F. Kuech, D. Mahne, and R. Schultz-Amling. Spatial filtering using directional audio coding parameters. Taipei, Taiwan, April 2009. Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing.

K. Kato, K. Ueno, and K. Kawai. Musicians' adjustment of performance to room acoustics part 2: Acoustical analysis of performed sound signals. Madrid, September 2007. 19th International Congress on Acoustics.

K. Kato, K. Ueno, and K. Kawai. Musicians' adjustment of performance to room acoustics part 3: Understanding the variations in musical expressions. Paris, July 2008. Euronoise 2008.

G. Kearney. *Auditory Scene Synthesis using Virtual Acoustic Recording and Reproduction*. PhD thesis, Trinity College, University of Dublin, Dublin, October 2009.

L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders. *Fundamentals of Acoustics*. Wiley, 3rd edition, 1982.

C. Knufinke. SIR2 Audio Tools - Reverb. `http://www.siraudiotools.com/sir2.php`, 2010. URL `http://www.siraudiotools.com/sir2.php`. Accessed 05/03/2010.

D. Ko, W. Woszczyk, J. Hong, and S. Levine. Augmented stage support in ensemble performance using virtual acoustics technology. Montreal, QC, Canada, 2013. Proceedings of Meetings on Acoustics (ICA 2013).

M. Kunkemoller, P. Dietrich, and M. Pollow. Synthesis of room impulse responses for variable source characteristics. *Acta polytechnica*, 51, No. 5, 2011.

H. Kuttruff. *Room Acoustics*. Applied Science Publishers, London, 2nd edition, 1979.

H. Kuttruff. *Acoustics:An introduction*. Number ISBN 0-203097089-6. Taylor & Francis, 1st edition, 2007.

I. Laird, D. Murphy, P. Chapman, and S. Jouan. Development of a virtual performance studio with application of virtual acoustic recording methods. In *Audio Engineering Society Convention 130*. Audio Engineering Society, 2011.

M.-V. Laitinen. *Techniques for versatile spatial-audio reproduction in time-frequency domain*. PhD thesis, School of Electrical Engineering, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland, February 2014.

M. Lautenbach and M. Vercammen. Stage Acoustics, ISO 3382 and beyond. Toronto, Canada, June 2013. Proceedings of the International Symposium on Room Acoustics (ISRA).

S. Le and T. Worch. *Analysing sensory data in R*. Number ISBN-13:978-1-4665-6572-2. CRC Press, Taylor Francis Group, 1st edition, 2015.

S. Le, J. Josse, and F. Husson. FactoMineR: An R Package for Multivariate Analysis, 2008.

P. Lennox and T. Myatt. Perceptual cartoonification in multi-spatial sound systems. Budapest, Hungary, June 2011. Proceedings of the 17th International Conference on Auditory Display (ICAD-2011).

A. Lindau, L. Kosanke, and S. Weinzierl. Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. London, UK, May 2010. Proceedings of the 128th Audio Engineering Society Convention.

A. Lindau, L. Kosanke, and S. Weinzierl. Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses. *Journal of the Audio Engineering Society*, 60. No 11, 2012.

O. Linfors, J. Ahonen, and H. Moller. Strength measurements with in-situ reference. Toronto, Canada, June 2013. Proceedings of the International Symposium on Room Acoustics.

T. Lokki and J. Patynen. Applying anechoic recordings in auralisation. page 5. Proceedings of the EAA Symposium on Auralization, June 2009.

T. Lokki, J. Patynen, and V. Pulkki. Recording of anechoic symphony music. page 6. Euronoise, Paris 2008, July 2008.

T. Lokki, J. Patynen, S. Tervo, S. Siltanen, and L. Savioja. Engaging concert hall acoustics is made up of temporal envelope preserving reflections. *Journal of the Acoustic Society of America, JASA Express Letters*, 129 (6), 2011.

T. Lokki, J. Patynen, A. Kuusinen, and S. Tervo. Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *Journal of the Acoustic Society of America*, 132 (5):3148–3161, 2012.

D. Malham. Space in music - music in space. Abstracted passage from MPhil Thesis, University of York, York, UK, December 2003.

B. Masiero and M. Pollow. A review of the compressive sampling framework in the lights of spherical harmonics: applications to distributed spherical arrays. Ambisonics Symposium and spherical acoustics, May 2010.

Mathworks. MATLAB. `http://www.mathworks.co.uk/products/matlab/`, September 2013. URL `http://www.mathworks.co.uk/products/matlab/`. Accessed 01/02/2011.

R. McCarthy, D. Cabrera, and J. Basset. Directional assessment of acoustic stage support in a drama theatre. Geelong, Australia, 2008. Proceedings of Acoustics 2008.

D. Menzies. Parametric representation of complex sources in reflective environments. Proceedings of the 128th Audio Engineering Society Convention, May 2010.

D. Menzies and M. Al-Akaidi. Ambisonic synthesis of complex sources. *Journal of the Audio Engineering Society*, 55(10), October 2007.

J. Merimaa and V. Pulkki. Spatial impulse response rendering. Naples, Italy, October 2004. Proceedings of the 7th International Conference on Digital Audio Effects (DAFx'04).

J. Meyer. *Acoustics and the performance of music*. Springer, 5th edition, 2009.

MH Acoustics. Eigenmike microphone. `http://http://www.mhacoustics.com/products`, November 2014. URL `http://www.http://www.mhacoustics.com/products`. Accessed 10/02/2015.

L. Miranda Jofre, D. Cabrera, M. Yadav, A. Sygulska, and W. Martens. Evaluation of stage acoustics preference for a singer using oral-binuaral room impulse responses. Montreal, Canada, June 2013. Proc. of Meetings on Acoustics, International congress on acoustics.

A. R. Moller. *Hearing: Anatomy, Physiology and Disorders of the auditory system*. Elseviert, 2nd edition, 2006.

B. C. J. Moore. *An introduction to the psychology of hearing*. Academic Press, 4th edition, 1997.

Morset Sound Development. WINMLS. `http://www.winmls.com/`, September 2012. URL `http://www.winmls.com/`. Accessed 20/10/2012.

C. Nachbar, G. Nistelberger, and Z. Franz. Listening to the direct sound of musical instruments in freely adjustable surround directions. page 3. Proceedings of the 2nd International Symposium on Ambisonics Symposium and spherical acoustics, May 2010.

C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi. AMBIX - A suggested Ambisonics format. Lexington, Kentucky, USA, June 2011. Proceedings of the Ambisonics Symposium.

T. Naes, P. B. Brockhoff, and O. Tomic. *Statistics for sensory and consumer science*. Number ISBN: 978-0-470-51821-2. John Wiley and sons, 1st edition, 2010.

F. Ortolani. Introduction to Ambisonics: A tutorial for beginners in 3D audio - Rev. 2014. Tutorial document, Ironbridge Electronics, 2014.

F. Otondo and J. H. Rindel. The influence of the directivity of musical instruments in a room. *Acta Acustica*, 90:1178–1184, 2004.

F. Otondo, J. H. Rindel, R. Causse, N. Misdariis, and P. dela Cuadra. Directivity of musical instruments in a real performance situation. pages 312–318, Mexico, City, Mexico, 2002. Proceedings of the International Symposium on Musical Acoustics.

J. Patynen. Virtual acoustics in practice rooms. Master's thesis, Department of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland, October 2007.

J. Patynen and T. Lokki. Evaluation of concert hall auralization with virtual symphony orchestra. Melbourne, Australia, August 2010. Proceedings of the International Symposium on Room Acoustics (ISRA).

J. Patynen, S. Tervo, and T. Lokki. Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses. *Journal of the Acoustic Society of America*, 133 (2), 2013.

T. Pihlajamaki. Multi-resolution short-time fourier transform implementation of directional audio coding. Master's thesis, Helsinki University of Technology, Faculty of Electronics, Communications and Automation, Department of Signal Processing and Acoustics, Espoo, Helsinki, August 2009.

A. Politis, T. Pihlajamaki, and V. Pulkki. Parametric spatial audio effects. York, UK, September 2012. Proc. of the 15th International Conference on Digital Audio Effects (DAFx-12).

M. Pollow and M. B. Gottfried Behler. Measuring directivities of natural sound sources with a spherical microphone array. page 6. Ambisonics Symposium, Graz, June 2009.

M. Pollow, M. Klein, and M. Vorlander. Including directivity patterns in room acoustical measurements. volume 19, Montreal, Canada, 2013. International congress on acoustics.

D. Protheroe and B. Guillemin. 3D impulse response measurments of spaces using an inexpensive microphone array. Toronto, Canada, June 2013. Proceedings of the International Symposium on Room Acoustics (ISRA).

V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6), June 1997.

V. Pulkki. Applications of directional audio coding in audio. Madrid, Spain, September 2007. Proceedings of the 19th International Congress on Acoustics.

V. Pulkki, M.-V. Laitinen, J. Vilkamo, J. Ahonen, T. Lokki, and T. Pihlajamaki. Directional audio coding - perception-based reproduction of spatial sound. Zao, Miyagi, Japan, November 2009. Proceedings of the Internation workshop on the principles and applications of spatial hearing.

R Core Team. R: A Language and Environment for Statistical Computing, 2013. URL `http://www.R-project.org/`.

P. V. Ramamurti. *An introduction to psychological measurements*. Number ISBN-978-81-203-4881-3. Asoke K. Ghosh, PHI Learning Private Ltd, Delhi, 1st edition edition, 2014.

Reaper. REAPER Digital Audio Workstation. http://www.reaper.fm, September 2013. URL www.reaper.fm. Accessed 18/05/2011.

J. H. Rindel and F. Otondo. The interaction between room and musical instruments studied by multi-channel auralisation. Budapest, 2005. Forum Acusticum.

P. Robinson, J. Patynen, and T. Lokki. The effect of diffuse reflections on spatial discrimination in a simulated conert hall. *Journal of the Acoustic Society of America*, 133(5), 2013a.

P. Robinson, A. Walther, C. Faller, and B. Jonas. Echo thresholds for reflections from acoustically diffusive architectural surfaces. *Journal of the Acoustic Society of America*, 134 (4), 2013b.

T. D. Rossing, editor. *Springer Handbook of Acoustics*. Springer, 1st edition, 2007.

Royal Conservatoire of Scotland. Ledger Recital Room technical specifications. http://www.rcs.ac.uk/about_us/campusandfacilities/renfrewstreet/ledgerroom/, No datea. URL http://www.rcs.ac.uk/about_us/campusandfacilities/renfrewstreet/ledgerroom/. Accessed 07/10/2014.

Royal Conservatoire of Scotland. Stevenson Hall technical specifications. http://www.rcs.ac.uk/about_us/campusandfacilities/renfrewstreet/stevenson/, No dateb. URL http://www.rcs.ac.uk/about_us/campusandfacilities/renfrewstreet/stevenson/. Accessed 15/08/2014.

RPG. RPG VAMPS system. http://www.rpgeurope.com/products/product/vamps.html. URL http://www.rpgeurope.com/products/product/vamps.html. Accessed 21/10/2014.

F. Rumsey. *Spatial Audio*. Number ISBN 0 240 51623 0. Focal Press, 1st edition, 2001.

F. Rumsey and T. McCormick. *Sound and Recording: An introduction*. Number ISBN 0 240 51680 X. Focal Press, 4th edition, 2003.

M. Skavelik. Diffusivity of performance spaces - its significance to perceived sound quality from directive source. page 21. Joint Baltic-Nordic Acoustics Meeting, Sweden, November 2006.

B. J. Smith, R. J. Peters, and S. Owen. *Acoustics and Noise Control*. Number ISBN-10: 0-582-08804-6. Pearson Education Limited, 2nd edition edition, 1996.

J. O. Smith. *Spectral Audio Signal Processing*. http://ccrma.stanford.edu/~jos/sasp/, 2011. online book, 2011 edition.

J. O. Smith. *Introduction to Digital Filters with Audio Applications*. http://-ccrma.stanford.edu/~jos/filters/, accessed (17/11/14). online book.

Sound on Sound magazine. You are surrounded: Surround sound explained - part 3. http://www.soundonsound.com/sos/Oct01/articles/surroundsound3.asp, 2001. URL http://www.soundonsound.com/sos/Oct01/articles/surroundsound3.asp. Accessed 30/01/2015.

Studiocare. Soundfield microphone. http://www.studiocare.com/soundfield-mark-iv-microphone-system-used.html, No date. URL http://www.studiocare.com/soundfield-mark-iv-microphone-system-used.html. Accessed 02/12/2014.

S. Tervo. *Localization and tracing of early acoustic reflections in enclosures*. PhD thesis, School of Science, Department of Media Technology, Aalto University, January 2012.

S. Tervo, T. Korhonen, and T. Lokki. Estimation of reflections from impulse responses. Melbourne, Australia, August 2010. Proc. of the International Symposium on Room Acoustics.

S. Tervo, J. Patynen, and A. Kuusinen. Spatial decomposition method for room impulse responses. *Journal of the Audio Engineering Society*, 61, No. 1/2, 2013a.

S. Tervo, J. Patynen, and T. Lokki. Spatio-temporal energy measurements in renowned concert halls with a loudspeaker orchestra. *Proceedings of Meetings on Acoustics (ICA 2013)*, 19, 2013b.

The University of Edinburgh. Reid Hall homepage. `http://www.eca.ed.ac.uk/eca-home/resources/exhibition-and-performace-spaces/reid-concert-hall`, September 2014. URL `http://www.eca.ed.ac.uk/eca-home/resources/exhibition-and-performace-spaces/reid-concert-hall`. Accessed 13/10/2014.

G. Thiele. *On the localisation in the superimposed soundfield*. PhD thesis, Technische Universitat Berlin, April 1980.

L. Thurstone. A law of comparative judgement. *Psychological Review*, 101, No. 2:266–270, 1994.

H. Tuominen, J. Ramo, and V. Valimaki. Acoustic retroreflectors for music performance monitoring. Stockholm, Sweden, 2013. Proc. of the Sound and Music Computing Conference (SMC).

K. Ueno and H. Tachibana. Experimental study on the evaluation of stage acoustics by musicians using a 6- channel sound simulation system. Acoust. Sci. Tech. 24, 3 (2003), 2003.

K. Ueno and H. Tachibana. Cognitive modeling of musician's perception in concert halls. *Acoust. Sci Tech*, 26, 2, 2005.

K. Ueno and H. Tachibana. A consideration on acoustic properties on concert-hall stages. Melbourne, Australia, August 2010. Proc. of the International Symposium on Room Acoustics (ISRA).

K. Ueno, T. Kanamori, and H. Tachibana. Experimental study on stage acoustics for ensemble performance in chamber music. Acoust. Sci. Tech. 26, 4 (2005), 2005.

K. Ueno, K. Kato, and K. Lawai. Musicians' adjustment of performance to room acoustics part1: Experimental performance and interview in simulated soundfield. International confress on acoustics, Madrid, September 2007.

Unversity of St Andrews. Younger Hall technical specifications. `https://www.st-andrews.ac.uk/music/perform/facilities/youngerhall/`, No date. URL `https://www.st-andrews.ac.uk/music/perform/facilities/youngerhall/`. Accessed 07/09/2014.

J. Usher. An improved method to determine the onset timings of reflections in an acoustic impulse response. *Journal of the Acoustic Society of America, JASA Express Letters*, 127(4), 2010.

E. van den Braak and L. C. J. van Luxemburg. New (stage) parameter for conductor's acoustics? *Journal of the Acoustic Society of America*, 123(5):3198, 2008.

L. van der veen. C74 MaxMSP controller. `http://www.nr37.nl/?c=software&i=c74`, September 2013. URL `http://www.nr37.nl/?c=software&i=c74`. Accessed 25/09/2013.

V. Vetrivel. Tables of the binomial cumulative distribution. `http://mat.iitm.ac.in/home/vetri/public_html/statistics/binomial.pdf`, No date. URL `http://mat.iitm.ac.in/home/vetri/public_html/statistics/binomial.pdf(IndianInstituteofTechnology, Madras)`. Accessed 20/11/2014.

M. Vikko and R. Tomi. Designing maximum length sequence signal for frequency response measurement of switched mode converters. Nordic Workshop on Power and Industrial Electronics, June 2008.

J. Vilkamo. Spatial sound reproduction with frequency band processing of b-format audio signals. Master's thesis, Helsinki University of Technology, Faculty of Electronics, Communications and Automation, Department of Signal Processing and Acoustics, Espoo, Helsinki, May 2008.

M. Vorlander. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer-Verlag, RWTH Aachen, 1st edition, 2008.

H. Wallach, N. Edwin, and M. Rosenzweig. The precedence effect in sound localization. *The American Journal of Psychology*, 62 No. 3:315–336, 1949.

M. M. Wanderley. *Non-obvious Performer Gestures in Instrumental Music*. Springer-Verlag, 1999.

M. Wankling, B. Fazenda, and W. Davies. The assessment of low-frequency room acoustic parameters using descriptive analysis, 2012.

B. Watson and D. F. Clark. Real-time ambisonic simulation of auditorium acoustics within a digital audio workstation environment. Paris, France, May 2010. Proceedings of the 2nd international symposium on Ambisonics and Spherical Acoustics.

J. Waxman. Violin acoustic radiation synthesis: A source model for direct sound enhancement in musical acoustic environments. *jur.rochester.edu*, 4, issue 1, Fall 2005 2005.

M. Weeks. *Digital Signal Processing using MATLAB and wavelets*. Infinity Science Press LLC, 1st edition, 2007.

J. Wells. Logarithmic sweep generator. `http://www.jezwells.org/Matlab/logSweepGen.m`, June 2012. URL `http://www.jezwells.org/Matlab/logSweepGen.m`. Accessed 02/05/2010.

R. Wenmaekers and C. C. J. M. Hak. Early and late support measured over various distances: the covered versus open part of the orchestra pit. Toronto, Canada, June 2013. Proceedings of the International Symposium on Room Acoustics (ISRA).

R. Wenmaekers, C. C. J. M. Hak, and L. C. J. Luxemburg. On measurements of stage acoustic parameters: Time interval limits and various source-receiver distances. *Acta Acustica united with Acustica*, 98:776–789, 2012.

B. Wiggins. *An investigation into the real-time manipulation and control of three-dimensional soundfields*. PhD thesis, University of Derby, Derby, 2004.

Wikipedia. Caird Hall wikipedia article. `http://en.wikipedia.org/wiki/Caird_Hall`, January 2014. URL `http://en.wikipedia.org/wiki/Caird_Hall`. Accessed 15/10/2014.

A. Williamon, L. Aufegger, and H. Eiholzer. Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance. *Frontiers in Psychology*, 5(25), February 2014.

E. G. Williams. *Fourier acoustics: Sound radiation and nearfield acoustical holography*. Academic Press, 1st edition, 1999.

W. Woszczyk. Low-latency virtual acoustics for live music performance and recording. Vienna, Austria, July 2006. Proceedings of the 13th International Congress on Sound and Vibration (ICSV2013).

W. Woszczyk and W. L. Martens. Evaluation of virtual acoustic stage support for musical performance. Paris, France, June 2008. Proceedings of Acoustics '08.

W. Woszczyk, D. Ko, and B. Leonard. Virtual acoustics at the service of music performance and recording. *PAN Archives of Acoustics*, 37 No.1 (2012):109–113, 2012.

M. Yadav, D. Cabrera, and W. L. Martens. Auditory room size perceived from a room acoustic simulation with autophonic stimuli. *Acoustics Australia*, 39(3), 2011.

M. Yadav, D. Cabrera, D. Lee, L. Miranda Jofre, and W. L. Martens. The regulation of voice levels in various room acoustic conditions. Victor Harbor, Australia, November 2013a. Proceedings of Acoustics 2013.

M. Yadav, D. Cabrera, L. Miranda Jofre, W. L. Martens, D. Lee, and R. Collins. Investigating auditory room size perception with autophonic stimuli. New Paltz, New York, October 2013b. Proceedings of the 135th Audio Engineering Society Convention.

N. Yoder. Peakfinder. `http://uk.mathworks.com/matlabcentral/fileexchange/25500-peakfinder`, August 2014. URL `http://uk.mathworks.com/matlabcentral/fileexchange/25500-peakfinder`. Accessed 02/11/2014.

U. Zolzer, editor. *DAFX- Digital Audio Effects*. John Wiley and sons, 2nd edition, 2011.

F. Zotter. *Analysis and synthesis of sound-radiation with spherical arrays*. Phd thesis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Austria, Austria, September 2010.

F. Zotter, H. Pomberger, and M. Noisternig. Ambisonic decoding with and without mode-matching:a case study using the hemisphere. Paris, France, May 2010. Proceedings of the 2nd international symposium on Ambisonics and Spherical Acoustics.