Towards the Perceptual Optimisation of Virtual Room Acoustics

Dale Johnson

Applied Psychoacoustics Lab

School of Computing and Engineering

University of Huddersfield

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

September 2018

Copyright statement

Abstract

In virtual reality, it is important that the user feels immersed, and that both the visual and listening experiences are pleasant and plausible. Whilst it is now possible to accurately model room acoustics using available scene geometry in real time, the perceptual attributes may not always be optimal. Previous research has examined high level control methods over attributes, yet have only been applied to algorithmic reverberators and not geometric types, which can model the acoustics of a virtual scene more accurately. The present thesis investigates methods of perceptual control over apparent source width and tonal colouration in virtual room acoustics, and is an important step towards and intelligent optimisation method for dynamically improving the listening experience.

A review of the psychoacoustic mechanisms of spatial impression and tonal colouration was performed. Consideration was given to the effects early of reflections on these two attributes so that they can be exploited. Existing artificial reverb methods, mainly algorithmic, wave-based and geometric types, were reviewed. It was found that a geometric type was the most suitable, and so a virtual acoustics program that gave access to each reflection and their meta-data was developed. The program would allow for perceptual control methods to exploit the reflection meta-data.

Experiments were performed to find novel, directional regions to sort and group reflections by how they contribute to an attribute. The first was a region of in the horizontal plane, where any reflection arriving within it will produce maximum perceived apparent source width (ASW). Another discovered two regions of and unacceptable colouration in front of and behind the listener. Any reflection arriving within these will produce unacceptable colouration. Level adjustment of reflections within either region should manipulate the corresponding attributes, forming the basis of the control methods.

An investigation was performed where the methods were applied to binaural room impulse responses generated by the custom program in two different virtual rooms at three source-receiver distances. An elicitation test was performed to find out what perceptual differences the control methods caused using speech, guitar and orchestral sources. It was found that the largest differences were in ASW, loudness, distance and phasiness. Further investigation into the effectiveness of the control methods found that level adjustment of lateral reflections was fairly effective for controlling the degree of ASW without affecting tonal colouration. They also found that level adjustment of front-back reflections can affect ASW, yet had little effect on colouration. The final experiment compared both methods, and also investigated their effect on source loudness and distance. Again it was found that level adjustment in both regions had a significant effect on ASW yet little effect on phasiness. It was also found that they significantly affected loudness and distance. Analysis found that the changes in ASW may be linked to changes in loudness and distance.

Acknowledgements

First and foremost, I give special thanks to my supervisor Hyunkook Lee for all of his excellent levels of encouragement, support and guidance throughout my studies.

I would like give thanks to all of the past and present members of the Applied Psychoacoustics Laboratory: Rory Wallis, Chris Gribben, Tom Wilson, Tom Robotham, Oli Larkin, Jade Clarke, Maksims Mironovs, Connor Millns, Bogdan Bacila and Nikita Goddard for their various levels of support, participation and guidance; to Maddie, who gave me that final push to make it to the end of my study, and also for supplying ice-cream; to all of staff and students at the University of Huddersfield's Music Technology department and Bluerooms Studios, in particular Mark Wendl and Ben Evans for their assistance and technical support. This work would not have been possible without those who participated in my listening tests.

I thank all of the past and present staff members at Caffé Grande Abaco, formerly Java Bar Espresso; the members of the University of Huddersfield Archery Club for all of their social support; and the members of the VCV Rack community.

And finally, but by no means least, I would like to thank my amazing family and loving parents, Barbara and Michael, who have given me incredible amounts of encouragement and continuing support during my studies; and to my friends for all of their social support and encouragement.

"...when you gaze long into an abyss, the abyss also gazes into you"

Friedrich Nietzsche

Contents

List of Figures

List of Tables

List of Equations

Glossary

absorption

The process of absorbing a portion of the energy of a wave as it reflects from a surface, determined by an absorption coefficient.

AIFF

An uncompressed type of digital audio file with a fixed sample rate, developed by Apple Inc.

algorithm

Describes the steps of a procedure required to accomplish a particular task.

algorithmic reverberator

A reverberator that utilises an algorithm of delays and filters in order to simulate the reverberation of a space.

allpass filter

Similar to the comb filter, except the signal path is configured so that the filter has a flat frequency response, allowing all frequencies to pass through.

ambisonic

A type of audio format that encodes audio signals into multiple directions. An ambisonic microphone, therefore, acheives this using a tightly packed cluster of microphones.

anechoic

Without echo. An anechoic chamber for example is a room without echoes.

ASW

Apparent Source Width. A sub-paradigm of spatial impression that describes the perceived width of an auditory source.

$\mathbf{ASW_{max}}$

A region on the horizontal plane where, in the presence of front arriving direct sound, an early reflection arriving anywhere within the region will produce the maximum degree of perceived ASW.

beam tracing

A geometric method that traces sound beams from surface to surface to pre-calculate possible paths sound could take from source to receiver to a given order..

bilinear interpolation

A form of linear interpolation where the same process is applied in two dimensions between four points.

BRIR

Binaural Room Impulse Response. An impulse response of a space measured using a dummy head or rendered using a HRTF database, and designed to represent how a listener would perceive the actual space in real life.

comb filter

A type of filter that mixes a delayed signal with the original signal. The name describes the characteristic 'hair comb' shape of the filter's frequency response.

convolution

A mathematical process that applies the characteristics of a signal or impulse response onto another signal. It is commonly used in music production for applying realistic reverb to a track, as well as in spatialisation of an audio signal to simulate spatial audio over headphones.

diffraction

Occurs when the a wave passes around the edge of a surface or an obstacle such as a corner, distorting the path and causing it *bend* around the edge.

diffusion

The process of randomly scattering the sound wave in different directions by a certain degree upon reflection.

DSB

Degree of Source Broadening. A combination of the IACC and sound strength measures used to predict the perceived width of an auditory source.

DWM

Digital Waveguide Mesh. A wave based method of modelling the propagation of sound, taking into account the wave effects such as interference and diffraction.

FDN

Feedback Delay Network. A type of algorithmic reverberator that models the diffuse reverberant tail. It achieves this by feeding the output of several parallel delay units into each other via a feedback matrix.

FFT

Fast Fourier Transform. An optimised version of the Discrete Fourier Transform (DFT) which converts a signal from the time domain and into its frequency domain resprenstation, allowing for analysis of the signal's frequency response, and also an efficient method of convolution.

FIR

Finite Impulse Response. A type of digital filter that has an impulse response with a fixed length.

G$_E$

Sound Strength. Measures the ratio of sound energy radiated from a source in a reverberant space relative to the energy radiated from the same source in a free field, or anechoic space, measured at 10m.

HRTF

Head Related Transfer Function. A function that describes how sound arrives at a listener's ears from a particular direction in terms of frequency and phase response.

IACC

Interaural Cross-correlation Coeffcient. The value and associated point of maximum correlation, or similarity, between two ear signals.

IACF

Interaural Cross-correlation Function. Describes the correlation, or similarity, between two signals at different offsets from each other.

IIR

Infinite Impulse Response. A type of digital filter that has an impulse response with an infinite length.

ILD

Interaural Level Difference. The level difference between the two ear signals.

IR

Impulse Response. A signal that describes how a system such as a filter or a reverberant space reacts when excited by an extremely short impulse of energy.

ISM

Image Source Method. A geometric method of calculating the exact path sound will traverse from source to listener, taking into account each possible set of reflections the path will take.

ITD

Interaural Time Difference. The time difference of when sound arrives at each ear.

$\mathbf{L}_f$

Lateral Fraction. Predicts ASW by calculating the ratio of early lateral reflection energy to total early reflection energy.

LEV

Listener Envelopment. A sub-paradigm of spatial impression that describes the feeling being enveloped by the sound. It is generally considered an environment related concept, dependent on late reverberance.

linear interpolation

The process of linearly cross-fading or *blending* between two values or points.

localisability

The ease of being able to localise and determine the location of an auditory source.

meta-data

Information that describes the characteristics of an object or entity. A reflection, for example, is characterised by it's direction of arrival, energy and delay time.

octave

A musical interval between two frequencies spaced by a ratio of 2:1.

phasiness

Describes a type of fluctuating, metallic, comb-filter like quality in the perceived sound.

ray tracing

A geometric method that calculates all of the possible paths sound will traverse around a space by emitting a lage number of rays in all directions from around the source, with a chance that they may hit a receiver.

reverberation

A term used to characterise how the sound propagates around a space, reflecting off objects and surfaces whilst decaying over time. Is generally shortened to reverb.

RIR

Room Impulse Response. An impulse response of a room, describing how and when all of the reflections arrive and decay over time. When an accurate RIR is convolved with an audio signal, the resulting reverb closely resembles the real reverb of the space.

RIV

Raw Impulse Vector. A raw reprensentation of an impulse response that contains each captured sound ray along with meta-data about each ray. This meta-data includes the direction of arrival, distance travelled, reflection order, octave-band energy and reflection history.

sample rate

The rate that audio data is digitally captured or reproduced, usually at a frequency that is at least twice that of the average human hearing frequency range.

SI

Spatial Impression. Describes how a listener will perceive the size of an acoustic space.

specular

Describes a perfect reflection who's angle of incidence or arrival is equal to its angle of reflection. It is the opposite of a diffuse reflection.

tonal colouration

A change in the timbre of the sound due to interference between the orignal sound and reflections.

VR

Virtual Reality. Simulating reality using a computer.

WAV

An uncompressed type of digital audio file with a fixed sample rate, developed by Microsoft and IBM.

Chapter 1

Introduction

Virtual reality (VR) is currently enjoying a renaissance thanks to advances in computer hardware, allowing for immersive and entertaining experiences to be accessible to everyone. For an experience to be immersive, the simulation must convince the user that what they are perceiving is real and plausible, and causes them to forget that what they are experiencing is just a simulation. However, it should also be enjoyable, otherwise the user may become distracted by some negative quality that could diminish their experience. Putting this into the context of virtual concert hall acoustics, the listener should feel immersed in the experience and feel as though they are in a real concert hall watching and listening to an actual orchestra.

1.1 Modelling room acoustics

Consider the simulation of concert hall acoustics. In VR, to give a convincing listening experience of a concert hall, the general geometry of the space is often used to model its reverberation, either by using a simplistic approach where only the general size and volume is considered (Jot, 1997), or with a more sophisticated, geometrical method (Schröder, 2011) where the shape and materials of each surface are taken into account, modelling each path that sound takes to reach the listener.

However, what must be considered in the field of virtual concert hall acoustics is the degree of accuracy required for realistic rendering, and how this may impact the overall listening experience. To put it concisely:

A plausible experience does not necessarily need to be an accurate one.

Whilst this statement may initially suggest that perhaps using scene geometry is unnecessary, it does not rule out accurate modelling completely. Since current computer hardware now has the performance to render acoustics with greater accuracy in real-time, it is natural that the modelling algorithms should take into account the geometry of the room to help improve the plausibility, and even achieving a substantial degree of accuracy, such has been demonstrated in recent literature (Wendt, van de Par, & Ewert, 2014). For concert hall listening, the literature discussed in Chapter 2 shows that a preferable experience is one that exhibits a wide Apparent Source Width (ASW), a high degree of listener envelopment (LEV), and a low degree of unpleasant tonal artefacts and colouration. Thus, the initial question here is:

Should a virtual concert hall experience be based on an accurate model of a space with no regard given to the quality of the listening experience?

Focusing on the first question, if the virtual acoustics algorithm models what would be considered a good quality concert hall, assuming that the algorithm is considered to be highly accurate, then the virtual version of the same hall should have the same level of quality. However, as current literature notes, this quality can vary when the listener sits in different positions around the room (Beranek, 1992), potentially introducing negative qualities. This implies that accurate modelling is a "double-edged sword", in that not only are positive aspects modelled, the negative characteristics are too. Such characteristics can include distracting tonal colouration, poor localisability and inadequate clarity of the audio source. It is established in the existing literature that these effects are caused by the pattern, level, frequency components and the direction of arrival of reflections. For virtual acoustics, when models are being rendered using existing rendering methods, this reflection information is readily available, yet as discussed in the next section it is often inaccessible due to such programs being either closed-source or purposefully designed to give only certain information.

1.2 Perceptually optimised reverb

Recall that in Section 1.1 it was stated that accurate simulation methods model both the positive and negative effects. However, it was also stated that a plausible experience does not have to be accurate. This gives rise to another question:

Can the quality of the listening experience in VR be perceptually enhanced whilst still convincing the user that they are in the virtual concert hall?

This question speculates that it may be possible to compromise accuracy for the sake of improving the listening experience. In the case of a virtual concert hall, this may entail suppression or manipulation of reflections that contribute to negative aspects. It must be stressed at this point that it is possible to have both an accurate and plausible experience, and that accuracy and plausibility are independent of each other. What is being discussed here is a trade-off of between accuracy and listening experience quality whilst still retaining an acceptable degree of plausibility.

When applying this idea to a room impulse response (RIR) measured in a real room, it can be very difficult to remove or suppress any negative effects. There have been reverberators in the past that were developed with perceptual control, such as the designs proposed by Jot (1997) and later Carpentier, Noisternig, and Warufsel (2014) which provide high level control over attributes such as 'warmth', 'brilliance' and 'envelopment'. However, this design was not developed with the application for VR in mind, and has not yet been applied to geometrical virtual acoustics.

Since the reflection information is readily available in geometrical, virtual acoustics methods, it may be possible to perceptually control and optimise the acoustics of a VR experience. This could be achieved by applying the methods proposed by Carpentier et al. (2014) to a geometrical method, and adapt them to take advantage of the available meta-data so that paradigms such as 'Spatial Impression', and its sub-paradigms Apparent Source Width (ASW) and Listener Envelopment (LEV), can be intelligently optimised for the best possible listening experience.

1.2.1 Capturing reflection meta-data

Traditionally, for general measurements of room reverberation and psychoacoustic measurement, the acoustics of a real space are captured as a RIR, which represents how the space will react when excited by a sound. The space can be excited by using a variety of methods such as a starter pistol or a sine-sweep (Farina, 2000). The RIR is then usually stored in a digital audio format such as a WAV or AIFF file. Whilst this form is useful for measurement, analysis and for use in music production for adding reverb via a process known as convolution, it is in fact fairly limited. It is analogous to a photograph, where sound being detected by a microphone is much like light entering a camera lens. A photograph depicts a scene at a fixed position and moment in time, limited by the camera and lens that captured it. Similarly, an impulse response represents the reverberation at a fixed position in a room, limited by the microphone that captured it and the system that stores it, thus losing information that could otherwise be useful for perceptual processing, such as the optimisation of the perceived room acoustics.

In a physical scenario, capturing and storing information about each individual sound ray can be a difficult and unwieldy task to achieve, often requiring a complex array of multiple microphones, or a special type of microphone known as an ambisonic microphone which captures sound from many directions with a certain order of resolution. The higher the order, the higher the resolution, although a high order ambisonic microphone, such as the Eigenmike[1], is often very expensive. Physical methods to capture more information from an impulse response have been proposed, for example the 'Spatial Decompostion Method' (SDM) developed by Tervo, Pätynen, Kuusinen, and Lokki (2013), or the 'Reverberant Spatial Audio Object' (RSAO) framework by Coleman et al. (2017). These two methods capture an impulse response from many directions and employ estimation methods in an attempt to extract individual reflections, and derive their direction of arrival. However, due to the methods being fairly new at the time of writing, they are still under development and require much further testing to determine their performance in various spaces and positions. Furthermore, the main limitation of these methods which can hamper their application to VR is that, like a traditional RIR, they are measured at a *fixed*

[1] Eigenmike is a registered trademark of 'mh Acoustics LLC'

position. In order to allow for all six degrees of freedom of movement (6DOF), which includes rotation and the ability to move in all three directions around a scene, many measurements must be taken throughout all the explorable areas of the entire space. Such a task in a real space is practically unwieldy and perhaps even daunting to those who are unwilling to undertake it.

In a virtual scenario, such as in an acoustics modelling program, it would easier and more manageable to capture and store rays along with their associated meta-data. Plus, it can easily overcome the fixed position limitation of the physical scenario where batch, simulated measurements in many locations in the virtual scene can be made. Unfortunately, currently available virtual acoustics programs such as EASE, CATT and ODEON, export an RIR in the form of a digital audio file, much like a traditional RIR measurement. However, the aforementioned examples are designed for industrial applications where accurate predictions must be made, requiring long simulation times, and therefore are not ideal for real-time VR. As mentioned in the previous section, there are methods that do allow for real-time rendering, for example Funkhouser et al. (1998) and Wendt et al. (2014), although they still have the same limitation as their industrial counterparts in that they do not readily provide open access to meta-data, or do not even appear to take full advantage of it. Ideally, a computer program should allow greater access and control over each captured sound ray so that the meta-data can be used to improve the listening experience.

1.3 Research questions

From the given background information, the following primary objectives of this study are: to develop a virtual acoustics program that can provide access to reflection meta-data, and to develop methods that utilise this meta-data to perceptually control and optimise the acoustics simulated by the program. These goals are to work towards a method of enhancing the virtual listening experience whilst retaining plausibility. From this, the following research questions are proposed:

1. What reflection meta-data relates to the perception of ASW and tonal colouration?

2. Can the meta-data be used to sort and group reflections by how much they will influence each attribute?

3. What perceptually motivated methods should be used to manipulate these grouped reflections in order to control each attribute?

4. How effective are the control methods at manipulating these attributes?

5. Is it possible to control them independently?

Whilst it has been established that perceptual control methods for artificial reverb already exist, to reiterate, they are only concerned with the temporal aspects of the reflections and do not appear to utilise reflection meta-data, nor do they apply the control methods to algorithms that take into account geometry of a virtual scene. To help answer the above questions, a reverberator should meet the following requirements:

1. Give access to every possible sound path and reflection.

2. Model decay due to surface and air absorption in octave bands using existing measured coefficients, such as those listed by Vorländer (2007).

3. Store meta-data about each captured reflection including the direction of arrival.

4. Export the captured reflections and their meta-data in a raw format.

1.4 Thesis structure

The thesis describes the development of a custom virtual acoustics program, and the development of methods that utilise the reflections and their meta-data to perceptually control ASW and tonal colouration. The structure is as follows:

- Chapter 2 reviews the psychoacoustic and concert hall acoustics literature in order to understand how Spatial Impression and Tonal Colouration are perceived.

- Chapter 3 discusses the various methods of modelling room acoustics, ranging from the algorithmic types through to geometric methods that consider the geometry of a room model in detail. The purpose of the discussion was to understand the implementation of each method, and to determine how easily reflection meta-data could be extracted from each method.

- Chapter 4 introduces the custom virtual acoustics program developed for the purposes of this study. The algorithm caters to all the established requirements, namely giving access to each reflection and its meta-data, as well as exporting this raw information for use in other programs such as MATLAB. Since the program was developed from scratch with much research performed into how to implement geometrical modelling methods, it was felt necessary to benchmark the custom algorithm against an established program, which in this case was ODEON as this software is a widely used in industry applications.

- Chapter 5 investigates the existence of an angular region where manipulation of the reflection level would have the greatest influence on the perceived degree of ASW. This is based upon interpretation of the research conducted by Barron and Marshall (1981). The first part of the experiment describes a psychometric test that was performed to locate the edges of a region of maximum ASW, or ASW_{max} , where any reflection arriving within this region will produce the maximum degree of perceived ASW. As this was performed over headphones, it was followed by a verification test using loudspeakers.

- Chapter 6 investigates the relationship between reflection angle and the acceptability of tonal colouration. The objective was to find a region where an early reflection arriving in this region within 50ms of the direct sound will produce unacceptable colouration. The first experiment was an informal elicitation test that determined the characteristics of colouration at different reflection delay times and directions. The second performed a psychometric test which was designed to find the reflection angles where the colouration is most acceptable and least audible. The final part was to determine how much the level reduction was needed to be applied to reflections arriving within the regions of unacceptable colouration.

- Chapter 7 introduces two perceptual control methods that use the regions obtained in the previous two chapters. The level of reflections arriving in those regions is manipulated to enhance or suppress their associated attribute. In the case of the ASW_{max} region, a change in level results in a change in perceived

ASW. The same concept applies to the region of unacceptable colouration. Since the colouration in the previous chapter was investigated using only a single reflection, it was necessary to perform a formal elicitation test to determine the differences in colouration between BRIRs with and without ±6 dB level adjustment. Grading tests were then performed to evaluate the effectiveness of each control method. Subjects were tasked with grading the perceived ASW and 'phasiness' of stimuli with varying amounts of level adjustment applied.

- Chapter 8 compares the perceptual control methods against each other using the same grading test yet with an alternate arrangement of the stimuli to allow for a direct comparison. It also investigates the effects of the control methods on perceived loudness and distance, and then discusses how the perception of ASW, loudness and distance may be linked.

- Chapter 9 concludes and summarises the work conducted in the thesis, discusses the limitations of the study and proposes further work to overcome them.

1.4.1 Novel contributions

- A custom virtual acoustics program that provides meta-data for all the reflections captured during rendering[2].

- Discovery of a horizontal region between approximately 38.9°and 134.1°, where ASW caused by a single reflection in the presence of a direct sound is perceived to be at maximum.

[2] The source code for this program is available from https://github.com/ValleyAudio/homr

- Detailed analysis of the potential causes for this saturation found that when a reflection arrives within the $\mathrm{ASW}_{\mathrm{max}}$ region, there is a plateau in IACC_{E3} and the standard deviation of the fluctuations in inter-aural time difference in the 1 kHz octave band.

- An investigation that directly observes the relationship between reflection angle and the audibility and acceptability of colouration. The investigation found that arrival time of the reflection affected the nature of the colouration.

- Discovery of two horizontal regions where a single, early reflection in the presence of a direct sound independently causes audible and unacceptable tonal colouration. The arrival time of the reflection had little effect on the location of each region's boundaries.

- Spectral analysis of the stimuli suggests that as the reflection angle exits the unacceptable region, the severity of comb filtering in one of the ear signals may be dropping below a potential threshold point around the edges of the region.

- Evidence for an apparent relationship between perceived ASW, loudness and source distance was found, questioning the validity of the current prediction methods for ASW.

1.5 Publications

Johnson, D., Lee, H. (2016a). Investigation into the perceptual effects of image source method order. In *Audio engineering society convention 140.*

Johnson, D., Lee, H. (2016b). Taking advantage of geometric acoustics modeling using metadata. In *Interactive audio systems symposium 2016.*

Johnson, D., Lee, H. (2017). Just noticeable difference in apparent source width depending on the direction of a single reflection. In *Audio engineering society convention 142.*

Johnson, D., Lee, H. (2018). Perceptually optimised virtual acoustics. In *Proceedings of the 4th workshop on intelligent music production.*

Chapter 2

A review of psychoacoustics and concert hall acoustics

This chapter will review the existing literature that relates to psychoacoustics and concert hall acoustics, and investigate what groups or types of reflections contribute to current ′Spatial Impression′ paradigm.

In this study, it is important to understand what effects both early and late reflections have on the perceived spatial impression, focusing mainly on the effect of reflection direction on its sub-attributes, Apparent Source Width (ASW) and Listener Envelopment (LEV). The knowledge gained from this review will contribute to the development of perceptual control methods. The algorithm will take advantage of the underlying mechanisms of spatial impression in order to control ASW and LEV.

2.1 The perception of spatial impression

Spatial Impression (SI) can be defined as:

"...a characteristic spatial spreading of the auditory events..."

(Blauert & Lindemann, 1986)

"...an acoustical sensation of space in a wide sense..."

(Furuya, Fujimoto, Takeshima, & Nakamura, 1995)

"...the spatial extent of the sound image."

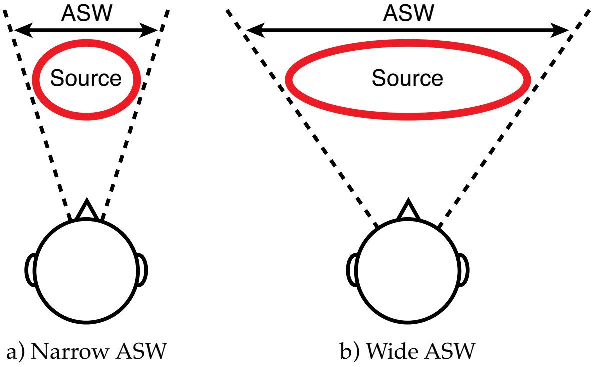
(Morimoto, Iida, & Sakagami, 2001)

From these definitions, SI can be interpreted as a psychoacoustic paradigm that deals with the perception of size and dimensions of both auditory sources and environment around the listener. Much work has been performed to understand the mechanics of SI, to measure and predict it, and how to use it to rate the quality of a concert hall. It is been established in recent literature that both temporal and directional distribution of reflections can affect the perception of SI. It was initially understood in early literature to be a singular paradigm, for example, the aforementioned definition provided by Blauert and Lindemann (1986) treats SI as being influenced solely by early reflections. However, later literature suggests that SI can be described by the two sub-paradigms: 'Apparent Source Width' (ASW) (Hidaka, Beranek, & Okano, 1995) and 'Listener Envelopment' (LEV) (Beranek, 1996). ASW, as the name suggests, defines how wide the auditory source will appear to sound to the listener, and is dependent on the characteristics of early lateral reflections (Hidaka et al., 1995). LEV on the other hand describes how enveloping the sound field will feel to a listener, and is dependent on late lateral reflections and reverberance (Bradley & Soulodre, 1995b).

This section will focus on ASW and LEV independently, and review the fundamental causes for both sub-paradigms, focusing mainly on the effects of reflection, time and level on the perception of SI. It is important to understand the relationship between the physical attributes of the reflections and SI such that they can be further studied and *'reverse engineered'* so to speak, such that a perceptual control method for SI in virtual concert halls can be developed.

2.1.1 Apparent Source Width

The term Apparent Source Width (ASW) can be described as "the apparent auditory width of the sound field created by a performing entity as perceived by a listener..." (Hidaka et al., 1995). Thus, ASW is used to describe how wide and auditory will appear to feel to a listener, see Figure 2.1. It is widely accepted that ASW is primarily the result of reflections arriving within 80ms after the direct sound (Barron & Marshall, 1981).



a) Narrow ASW

b) Wide ASW

Fig. 2.1: Visual representation of ASW, after Wallis (2017).

2.1.1.1 In relation to lateral fraction

Barron and Marshall (1981) studied the effect of delay time, direction and level of a pair of lateral reflections on the perceived spatial impression. Through a series of experiments, they developed an objective measure of ASW known as 'Lateral Fraction', or L_f. All the experiments took place in an anechoic chamber, and the reflections and reverberant field were mimicked using fixed speakers and a reverberation plate.

In the first experiment, they investigated the effect of delay time. Subjects compared the ASW between two sound fields containing a direct sound and two reflections arriving at ±40° to the listener. In one field, the level of the reflections was fixed, whilst in the other the level could be varied. The delay time of the fixed test field was varied between 5 to 80ms, whilst in the variable field it was fixed to 40ms. Subjects were tasked with adjusting the level of the variable field such that it produced ASW equivalent to that of the test field. It was found that delay time had no significant effect on degree of ASW, and so it was concluded that ASW is independent of delay time.

Barron and Marshall (1981) next experimented with the effect of reflection direction on ASW. Similar to their first experiment, subjects compared the ASW produced by a test sound field containing a pair of reflections with a variable azimuth angle $\pm\alpha^\circ$ and a fixed level of -10 dB, and a comparison sound field with reflections fixed at ±90°. Between each trial, the azimuth angle α of the reflections in the fixed test field was varied. Again, subjects were asked to adjust the level of the comparison sound field such that the degree of perceived ASW in this field was equivalent to that of the test sound field. Barron and Marshall (1981) found that in the presence of a direct sound arriving at 0°, the degree of spatial impression was found to be dependent on a reflection's angle of arrival $\alpha°$, such that as the reflection angle approached ±90° it produced maximum ASW, see Figure 2.3. In another, similar experiment that varied the angle of elevation $\beta°$, it was also found that ASW decreases as the elevation of a 90° lateral reflection increases. Finally through analysis of results produced by Reichardt and Schmidt (1967) regarding the noticeable differences in ASW from variations in ratio between lateral level to direct sound level, Barron and Marshall (1981) were able to derive a relationship between the degree of ASW and the ratio of lateral to total reflection energy, shown in Figure 2.2. From these findings, they proposed the objective measure lateral fraction (L_f), which is given by:

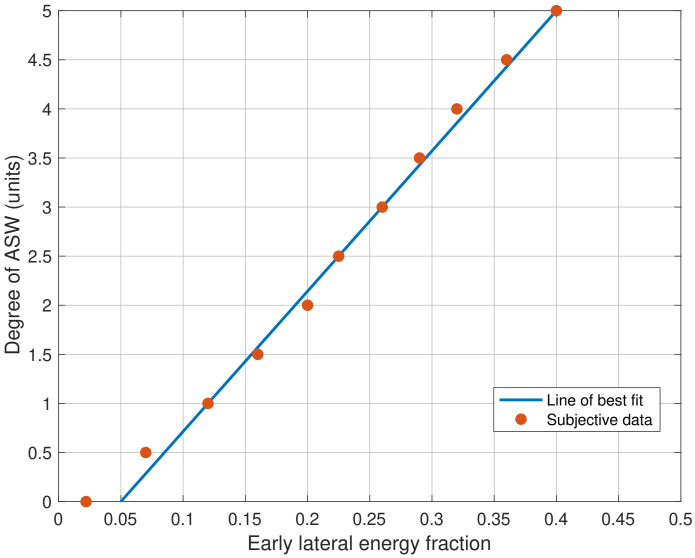


Fig. 2.2: Relationship between degree of perceived ASW and early lateral energy fraction L_f. Solid line fitted to the subjective data points, after Barron and Marshall (1981).

$$L_f = \frac{\sum_{t=5ms}^{80ms} r \cos\phi}{\sum_{t=0ms}^{80ms} r} \tag{2.1}$$

where r is the sound intensity and ϕ is the azimuth angle of the reflection. Whilst this finding initially appears valid for a single or pair of reflections, further experiments

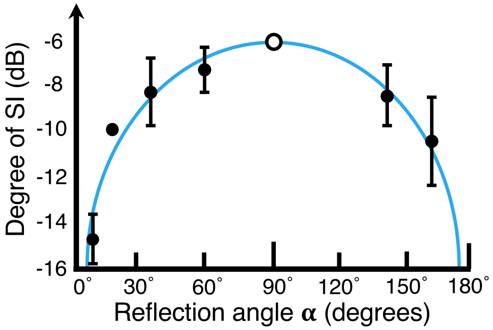


Fig. 2.3: Mean and 95% confidence intervals of subjective spatial impression versus reflection angle, adapted from Barron and Marshall (1981).

performed by Barron and Marshall (1981) found L_f to still be valid for sound fields with multiple reflections. In the context of sound fields where multiple early reflections are present, BS EN ISO 3382-1 (2009) states that L_f can be measured using

$$L_f = \frac{\int_{5ms}^{80ms} p_L^2(t)dt}{\int_{0ms}^{80ms} p^2(t)dt} \tag{2.2}$$

where p_L is the impulse response measured with a figure-of-eight pattern microphone with the null point facing towards the source, and p is an impulse response measured at the same point with an omni-directional pattern microphone. The significance of the findings of Barron and Marshall (1981) is that it shows that ASW has a dependency on the direction of arrival and level of early reflections, such that a more lateral angular distribution of reflections leads to a heightened sense of ASW. For perceptual control and optimisation, these findings will be taken into consideration for the development of the control methods as they will need to take into account direction and level.

2.1.1.2 In relation to the IACC

Whilst L_f is an accepted measure for ASW, it is also widely established that the ASW can be measured using the Inter-aural Cross-correlation Coefficient (IACC) (International Standards Organisation, 2009). By definition, the IACC is the maximum of the inter-aural cross-correlation function (IACF) between two ear signals, and it is inversely correlated with ASW. The IACF describes the similarity between the two signals, and is given by

$$IACF(\tau) = \frac{\int_{t_1}^{t_2} p_L(t) p_R(t+\tau) dt}{\sqrt{\int_{t_1}^{t_2} p_L^2(t) dt \int_{t_1}^{t_2} p_R^2(t) dt}} \tag{2.3}$$

where p_L and p_R are the left and right ear signals of a binaural room impulse response (BRIR), t_1 and t_2 are the start and end points of the measurement window, and τ is the particular offset of the right ear signal at which the IACF value was measured. The values for the IACF range between -1 and 1 where 1 indicates 100% similarity and 0 equates to 0% similarity. A value of -1 also indicates that there is 100% similarity yet one of the signals is phase inverted. The typical range of values for τ is $\pm$1ms (Hidaka et al., 1995). The IACC is then given by

$$IACC = |IACF|_{max} \tag{2.4}$$

IACC uses the same range as the IACF, and is inversely proportional to the degree of ASW. Thus the ASW can be given as $[1 - IACC]$.

Keet (1968) investigated the effects of early reflections on spatial impression and the relationship between the IACC and ASW. Keet (1968) made stereo recordings of a pre-recorded orchestral excerpt being played from a single speaker on a stage in three different concert halls. The recordings were taken at four different positions in each hall. They were then played back to ten observers using a stereophonic speaker setup in an anechoic chamber. The observers were asked to indicate the perceived ASW using an interval scale. The level of each recording was randomly varied so that the effects of listening level on ASW could also be observed. Keet (1968) found a linear relationship between ASW and playback level. To analyse the effect of coherence between the signals and ASW, Keet (1968) measured RIRs in the same recording positions and analysed the IACC of the 50ms of each RIR. When comparing the IACC with the subjective ratings of ASW, it was found that a linear relationship between $[1 - IACC]$ and ASW exists.

Hidaka et al. (1995) develops this property further as a method of measuring the acoustical quality of concert halls. However, rather than using the IACC as a broadband measure, using the research performed by Okano et al. (1994) that evaluated the relationship between ASW and IACC over several octave-bands to create ’equal ASW’ contours (see Figure 2.4), Hidaka et al. (1995) considered the 500 Hz, 1 and 2 kHz octave bands in which the IACC was most effective for evaluating ASW and assessing acoustical quality. Thus, they proposed that ASW be measured by the average IACC of early reflections measured at 500 Hz, 1 and 2 kHz, otherwise known as the IACC_{E3}. What was not explored by Hidaka et al. (1995) is the fundamental mechanisms of IACC, such as its relation to direction of arrival and delay time, which are crucial to the development of the perceptual control algorithm. The aforementioned fundamental attributes were later studied by Okano, Beranek, and Hidaka (1998). As observed by Barron and Marshall (1981), ASW changes with the direction of arrival ϕ, however, only significantly for stimuli below 500 Hz. It was then found that the measured ASW is equal at $\phi = 60°$ and $90°$. Okano et al. (1998) also found that in octave bands above 500 Hz, contrary to what is observed with L_f, the perceived ASW increases as the number of early lateral reflections

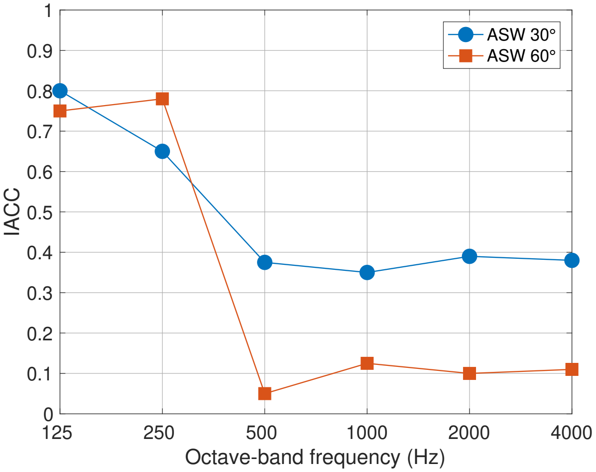


Fig. 2.4: Equal ASW contours per octave-band frequency, after Okano et al. (1994).

increases, which correlates to what is measured by $[1 - IACC_{E3}]$. Overall, Okano et al. (1998) found L_f to be less reliable at correctly predicting the perceived ASW than $[1 - IACC_{E3}]$. However, what is common between their findings and those of Barron and Marshall (1981) is that the degree of perceived ASW is dependent on the direction of arrival of the early reflections, and that the level of the lateral reflections appear to have the most influence on width.

Whilst the IACC is a widely accepted measure for ASW, it is not without its limitations. So far, the discussed literature uses the IACC for analysis of measured BRIRs alone. By simulating a rectangular hall with virtual acoustics program, de Vries, Hulsebos, and Baan (2001) tested the reliability of the traditional objective measures by measuring RIRs in 149 locations of the room spaced apart by 0.05m. Across the width of the hall, they found large fluctuations in IACC and L_f without a corresponding change in perceived ASW, deeming the predictors as "useless" (de Vries et al., 2001).

Mason, Brookes, and Rumsey (2004) further note that the method of analysing the IACC using solely the measured BRIR alone may not be the most accurate method for predicting subjective effects. This is because a measured BRIR exhibits very different properties compared to the sounds produced during a typical concert hall performance. The main difference is that the short duration of an impulse allows little interaction between the direct sound and subsequent reflections, whilst a musical signal allows for interactions that will affect the objective analysis over time. Thus, Mason et al. (2004) propose that the time-varying fluctuations in the IACC of continuous tonal signals, be it binaural recordings or anechoic sources convolved with existing BRIRs, should be analysed[1]. What this implies is that ASW is a source dependent measure, and that simply measuring the IACC from a BRIR alone does not give an adequate prediction of how ASW, and thus by extension SI, may be perceived in realistic scenario.

2.1.1.3 Effects of fluctuations in interaural time and level differences

Recent research has shown that fluctuations in inter-aural differences, namely time and level differences or ITD and ILD, have been attributed to the perception of ASW. Since the measurement of these fluctuations must be performed on continuous audio sources, they have been found to be ideal for predicting spatial qualities (Mason et al., 2004). Briefly, the ITD describes the difference in time for when a sound wave arrives between the two ears. The ILD describes the level difference between the two ear signals. It has been established that these two measures are responsible for the ability for humans to determine the location of an auditory source (Blauert, 1997).

[1] Mason et al. (2004) applied the principle of objectively analysing continuous signals to develop a measurement model.

Grantham and Wightman (1978) investigated the detectability of a time-varying, or fluctuating, ITD. These time differences were modulated with a sinusoid at different rates to determine the threshold at which subjects could detect the movement of a stimulus against a non-moving stimulus. They found that the binaural system of a human exhibits a ’lowpass’ or ’lagging’ characteristic, such that the perception of movement decreases as the rate of ITD fluctuation increases above 2-5 Hz. However, in the presence of a stationary source, movement is detected at up to 20 Hz. They further found that as the rate increases up to 20 Hz, the magnitude of this modulation must also increase in order for the movement to be reliably detected, see Figure 2.5.

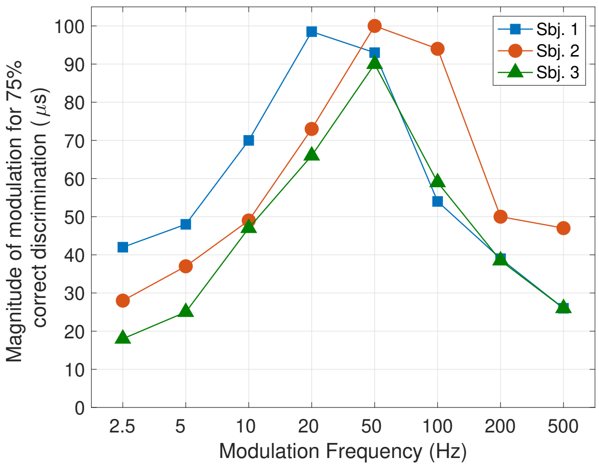


Fig. 2.5: Subjective data from experiment 1, where the peak ITD (μs) is the threshold required to reliably detect a moving stimulus, plotted against modulation frequency (Hz), after Grantham and Wightman (1978).

Blauert and Lindemann (1986) suggested that fluctuations in ITD and ILD both contribute to the perception of ASW, noting that *’lateralization’* of auditory events are caused by inter-aural time differences, and speculated that a sense of spaciousness is created if these differences vary over time. By taking running measurements of the time-varying ITD in a binaural signal that is band limited between 100 Hz to 3 kHz, Blauert and Lindemann (1986) calculated the standard deviation of the fluctuating differences. They found that when compared to their subjective scores, the standard deviation of the fluctuations showed a positive correlation with perceived spaciousness.

Mason and Rumsey (2001) reviewed how the fluctuations can manifest themselves in an acoustic environment, and performed elicitation tests to determine the subjective effects of the fluctuations. They proposed a measurement technique known as the inter-aural cross-correlation fluctuation function (IACCFF). This measure calculates the IACC of a series of measurements of a binaural signal that is filtered into several frequency bands between 50 to 2500 Hz over time. These measurements are then used to determine the changes in the ITD of each band. The resulting fluctuation signal is then further filtered by a 10 to 125 Hz bandpass filter. A Fast Fourier Transform (FFT) is applied to the filtered signal to determine the magnitude of the fluctuating signal. The IACCFF is then obtained by averaging the magnitude of the fluctuating signals for each filter band. This proposed measurement technique was applied to simple, artificial sound fields in which an audio source was placed 15m away from a virtual listener, and an artificial reflection by a virtual wall 5m to right of the listener, see Figure 2.6. To vary the angle and delay time of the reflection, and to simulate different reflection patterns, the wall position was moved away from the listener in 5m steps. The resulting sound field would allow for a fluctuating ITD over time in complex signals, for example the interaction between three sinusoids at 480, 500 and 520 Hz, shown in Figure 2.7. Mason and Rumsey (2001) progressed onto a simulated, reverberant room, in which anechoic musical sources were convolved with the BRIR of the virtual space. Through analysis of the

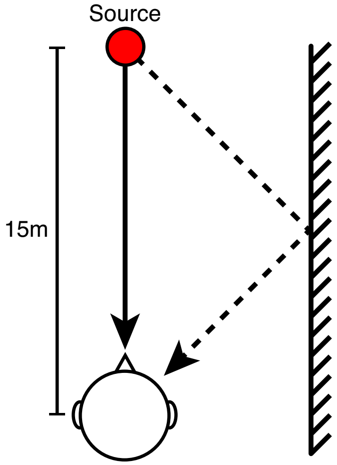


Fig. 2.6: Visualisation of the arrangement of the virtual setup with a source placed 15m in front of the listener, and wall positions 5m to the right, after Mason and Rumsey (2001).

IACCFF, they verified that the interaction between the direct sound and the resulting reflections creates time-varying fluctuations in the ITD, and that the nature of the fluctuations are significantly affected by the reflection pattern and the characteristics of the source.

The subjective experiment discussed by Mason and Rumsey (2001), investigated the subjective effects of time varying fluctuations in ITD using both verbal and non-verbal elicitation techniques. In subjective experiments performed over headphones, sine tones were delivered to each ear and phase modulated by 180° from each other over different frequencies to induce varying rates and magnitudes of phase modulation. Subjects were instructed to draw on a digital graphics tablet the perceived spatial attributes of the stimulus. It was found that: at 5 Hz, the subjects could perceive a moving image; at 10 Hz the movement can still be perceived but becomes less clear and harder to track; at 100 Hz, the movement disappears completely. What was common to all cases was that subjects could perceive both a static scene component and a moving scene component. It was further found that at a low magnitude

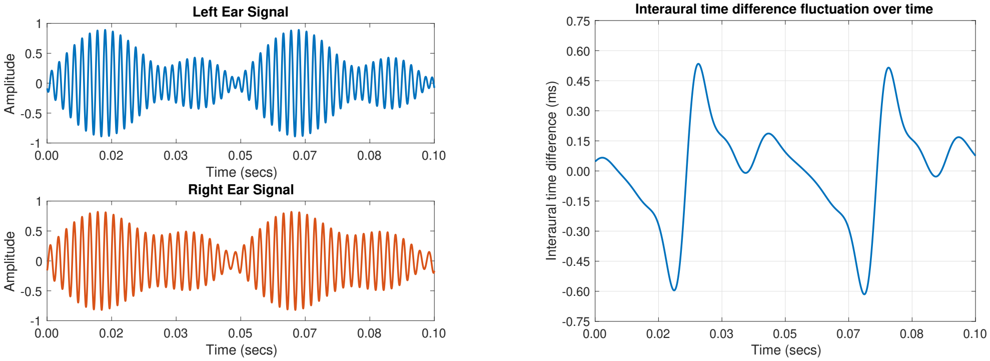


Fig. 2.7: Inter-aural time difference fluctuations over time, recreated using Max 7 and MATLAB, after Mason and Rumsey (2001).

of phase modulation, there was no perceived movement regardless of frequency. With an increasing modulation depth, however, increasing levels of movement or widening began to occur.

Mason and Rumsey (2001) summarised that the main perceptual effects of the increasing magnitude of fluctuations in ITD was an increase in width or range in movement of the auditory image. From this, it is understood that the characteristics of the sound source influence the behaviour of the fluctuations, which would affect the magnitude of the perceived width, or ASW.

2.1.1.4 Relationship with the early sound strength parameter

There has recently been research into the relationship of the early sound strength parameter, G_E, and perceived ASW. G_E is defined by ISO 3382-1 (2009) as:

$$G_E = 10 \cdot \log_{10} \frac{\int_{0ms}^{80ms} p^2(t)dt}{\int_{0ms}^{80ms} p_{10}^2(t)dt} \tag{2.5}$$

where p is an impulse response measured in at a given distance from a source, whilst p_{10} is a free field measurement of the same source from 10m away. Both measurements are taken using an omni-directional microphone and assume the source also be omni-directional. This parameter can be thought of as being equivalent to the loudness of the sound source (Beranek, 2011).

Beranek (2011) found that ASW may not only be dependent on BQI, which is equivalent to [1-IACC$_{E3}$]. It was noted that a previous study performed by Morimoto and Iida (1995) that ASW is influenced by the G_E, and presented a method for calculating ASW using the parameter. However, because the BQI is also a good predictor for ASW, Beranek (2011) proposed that the BQI and G_E parameters should be combined to form the Degree of Source Broadening (DSB) measure:

$$DSB = 31 \cdot BQI + \frac{5 \cdot G_E}{3} \tag{2.6}$$

Lee (2013) later performed a study into the relationship between both ASW and LEV, and the source-listener distance. Binaural and ambisonic RIRs of the St. Paul's concert hall space at the University of Huddersfield were measured at three distances between 3-12m. A listening test was performed on headphones, where two anechoic recordings were convolved with the BRIRs to create the stimuli. Subjects were tasked with grading the relative ASW or LEV of each stimulus between 0 and 100. It was found that between the three distances, there was a significant difference in ASW, and that there was linear reduction in ASW as the distance doubled. However, when compared to the classical measures [1-IACC$_{E3}$] and L_F, it was found that the predicted ASW did not agree with subjective results. In fact, it was found that there was a significant correlation between perceived ASW and G_E. Lee (2013) hypothesised that the energy parameters are more effective at measuring ASW, suggesting that they are directly linked to a "perception of distant-dependent" ASW. This is similar to the different types of perceived width suggested by Rumsey (2002), such as environmental or individual source / ensemble width.

2.1.2 Listener Envelopment

Listener Envelopment (LEV) has been described as the "subjective impression by a listener that (s)he is enveloped by the sound field" (Hidaka et al., 1995); "the sense of feeling surrounded by the sound" (Bradley & Soulodre, 1995a); and "the degree of fullness of sound images around the listener" (Morimoto et al., 2001). It was initially understood by Morimoto and Pösselt (1989) that spaciousness was equally affected by both reverberance and lateral reflection energy. However, more recent literature generally agrees that spaciousness is made up at least two paradigms, and that the reverberant energy is the primary contributor to the sense of envelopment.

2.1.2.1 Effects of lateral reflections

The study performed by Barron and Marshall (1981) was primarily concerned with the effects of a single pair of reflections arriving up to 80ms after the direct sound on SI. However, Bradley and Soulodre (1995a) found that late, reverberant energy arriving after 80ms at increasingly lateral directions also contributed to SI, creating a similar spacious sensation that they described as 'listener envelopment', or LEV. Bradley and Soulodre (1995a) tested the perception of LEV on up to ten subjects over five experiments using reflections with various combinations of level, direction and delay times. By implementing a similar methodology to Barron and Marshall (1981), in an anechoic chamber pairs of speakers were placed at fixed positions to mimic early reflections and a reverberant sound field, where the reverberance was created by a digital audio effects unit.

The first two experiments observed the listener's ability to perceive changes in early lateral energy and ASW at differing levels of late energy. It was found that at high levels of late energy, it was harder for subjects to distinguish differences in ASW. This suggests that there is a possible perceptual threshold of early lateral reflections in the presence of late energy. In the third experiment, the delay time and level of the early reflections was held constant whilst the late energy time and level was varied. Using 70ms long noise bursts, it was found that a large degree of LEV is perceived when high amounts of late energy are present.

The fourth experiment, like the previous, kept the early reflection time and level constant whilst the level and reverb decay time (RT) of the late reverberant energy was varied. Again, results shows that the perception of LEV increased with large values of late energy and RT. Finally, in the last experiment the location of the early reflections were kept constant whilst the angular distribution of the late energy was varied. It was found that the perception of LEV also increased as the angular distribution of the late energy increased to $\pm 90^\circ$. This strongly implies that LEV is influenced by the nature of late lateral reflections. Since the late arriving energy is not fused to the with the direct sound as with ASW, it can lead to more 'spatially distributed' effects that can give a sense of LEV. From the observed relationship with angular distribution, in a follow-up study Bradley and Soulodre (1995b) proposed that LEV was to be measured using the late lateral fraction, or LF_{80}^{∞}, given by:

$$LF_{80}^{\infty} = \frac{\int_{80ms}^{\infty} p^2(t)\cos^2(\alpha)dt}{\int_{80ms}^{\infty} p_{10}^2(t)dt} \tag{2.7}$$

where p is the measured RIR, p_{10} is the measurement of the source at 10m in a free field or anechoic environment, and α is the angle of incidence. The 80ms start point was chosen to complement the early lateral fraction measure Eq. 2.1, which ends at 80ms (Barron & Marshall, 1981).

2.1.2.2 The effects of reflections from non-lateral directions

Whilst it is generally accepted that late lateral reflections contribute to the perception of LEV, there is literature that suggests that non-lateral reflections, such as those arriving from behind and above the listener, can also have a substantial contribution.

Morimoto and Iida (1998) investigated the effectiveness of measuring LEV using the ratio of energy in front of and behind the listener, or the front-back (F/B) ratio. In their experiment, they arranged six loudspeakers around the listener on the horizontal plane and simulated a sound field using artificial early and late reflections distributed amongst the speakers. The F/B ratio of this sound field was adjusted by controlling the level of the frontal and rear speakers. They found that the reflections that arrive from behind the listener, which naturally produces a low F/B ratio, produce a great sense of envelopment. They do note, however, it is not easy to produce strong late reflections from behind the listener in a concert hall in order to create the perception of envelopment.

In a later study, Bradley, Reich, and Norcross (2000) investigated the combined effects of early and late lateral reflections on ASW and LEV. Six experiments were performed on subjects using simulated sound fields, each with varied early and late components. Each experiment was a pairwise comparison where subjects rated the difference in ASW or LEV between the two sound fields on a five-point scale. The sound fields were simulated using an 8 channel speaker setup where the speakers were arranged in a circle on the horizontal plane around the listener. The direct sound arrived from the speaker directly in front of the listener, and the early and late reflections were distributed amongst the other speakers. The reflections themselves were simulated using a digital reverberation unit, being fed by a digital equalizer that controlled the spectral components of the early and late reflections. Impulse responses of each sound field were measured such that objective parameters could be calculated.

Bradley et al. (2000) found that in the absence of late lateral reflections, LEV can be caused by late reflections arriving from non-lateral directions. However, they still found that lateral arriving late reflections still had the greatest influence over LEV, and so are the most important for "creating a strong sense of LEV" (Bradley et al., 2000). Using a similar methodology to Bradley and Soulodre (1995a), Furuya, Fujimoto, Ji, and Higa (2001) also investigated the effects non-lateral reflections on LEV. Like the earlier literature, they found that LEV is definitely influenced greatly by late arriving lateral reflections, however, they also found that the influence of late reflections arriving from behind and overhead also strongly correlated with subjective LEV scores. This then suggests that rear and overhead reflections are as important as lateral reflections for creating a great sense of LEV. Furuya et al. (2001) assert that the degree of LEV should not be measured using the level of late lateral reflections only, and that level of these should not be *exaggerated* as it could lead to a certain *unnaturalness* to the spatial impression. They suggest that LEV should be calculated using measurements taken from other directions, and that a "well balanced distribution" of late reflections from various directions should be considered in concert hall design.

Wakuda, Furuya, Fujimoto, Isogai, and Anai (2003) further investigated the effects of different reflection directions and distributions of late reverberant energy on LEV. The motive of their research was to "clarify the degrees of contribution of directional later energy to listener envelopment" (Wakuda et al., 2003). They hypothesised that the late energy arriving from directions other than lateral contributed to the perception of LEV. By varying the level of reverberant energy from either in front, to the side, behind or above the listener, subjects were asked to judge the difference in LEV between pairs of stimuli. Wakuda et al. (2003) found that late arriving energy from overhead and behind the listener did indeed have a significant contribution to LEV.

2.1.2.3 Confusion over the classification of envelopment

LEV is established as being a sensation that is caused by late arriving reflections or reverberant energy. There is literature, however, that uses the term *envelopment* in a potentially confusing manner such that it is used to describe different perceptual effects. For example, Furuya et al. (1995) investigated the effects of 'upside' early reflections on what they labelled as *auditory envelopment*, where upside reflections are those that arrive above or elevated to the listener. *Auditory envelopment* was described as 'feeling inside the music', as opposed to 'looking at it through a window'. Three experiments were performed to test the effects of a single reflection, then groups of late reflections on the *auditory size of sound image* and *envelopment*. In all experiments, sound fields were simulated by distributing artificial reflections amongst 17 speakers arranged around the subject. Furuya et al. (1995) found that the degree of perceived envelopment increases when the amount of upside energy of reflections arriving within 200ms increases. Whilst it was later found that non-lateral reflections can create and influence the sensation of envelopment (Wakuda et al., 2003), the suggestion of Furuya et al. (1995) that early reflections within the 200ms window contribute to envelopment does not agree with the widely established concept of LEV.

Morimoto and Iida (1998) also noted a significant interaction between the F/B ratio of early reflections and LEV, such that as the F/B ratio of early reflections is increased it reduces the perception of LEV. They conclude that early reflections also contribute to the LEV. However, most literature accepts that LEV is caused by late arriving energy and that early reflections are associated with ASW. It is unclear in the methodology of Morimoto and Iida (1998) as to how LEV was described or defined to subjects, thus there maybe be some potential confusion amongst listeners over the type of envelopment that was being perceived. Furthermore, the sound fields simulated by Morimoto and Iida (1998) do not contain a reverberant tail, and only discrete late reflections with a maximum delay time of 104ms were used. This is contrary to the other literature such as Bradley and Soulodre (1995a) who did make use of a reverberant tail.

It is generally accepted in more recent literature (e.g. George et al. (2010)) that early reflections tend to not be the cause of LEV, although their perceptual effects could be mistaken for another enveloping sensation, such as source envelopment Rumsey (2002), yet be described similarly to LEV. Rumsey (2002) noted that the term envelopment is quite broad and could be confused with said source envelopment. They noted that subjects can have the tendency to define it as being "surrounded by a number of dry sources". Rumsey (2002) theorised that a single sound source could become so diffused and wide that it envelops the listener. Whilst this is possible if a listener was sat in middle of an orchestra, it is not normally the case when the listener is positioned in the audience seating area, and so Rumsey (2002) hypothesised that this type of envelopment may be related to the definition given by Morimoto et al. (2001). Therefore, the broad term envelopment was split into three types: *individual* source, *ensemble source* and *environmental* envelopment, with the latter being similar to the established LEV measure. Thus, it is widely accepted that LEV is defined as an environmental effect that creates a sensation of being enveloped by the sound, and is found to be caused by late arriving reverberant energy.

2.2 The perception of tonal colouration

For the purposes of this study, and in order to control the perception of tonal colouration, this section will investigate the audibility and preference over the types of colouration. It will also explore the influence that reflection direction and level have over its perception.

In the context of room acoustics, tonal colouration has been defined in a variety of ways, such as: "[an] audible distortion, which alters the (natural) color of the sound." (Salomons, 1995); "changes in Timbre/'Klangfarbe'" (Halmrast, 2001); "... a characteristic change of timbre" (Kuttruff, 2016, p. 203). It can be understood as being the result of mixing delayed copies of a signal with itself, and is exactly what occurs in a reverberant space where reflections are perceived as copies or echoes of the original. These signals interact and interfere with the original direct sound and each other, changing the frequency response of the sound and producing an audible colouration. The most fundamental type of colouration is comb-filtering, the name of which is due to the shape of the frequency response looking similar to that of a hair comb, see Figure 2.8. This fundamental effect occurs when two signals interact with each other, and can manifest itself as an audible 'repetition pitch'. The interaction can be demonstrated using two sine waves at the same frequency. When the waves are summed, an interference will occur, as shown in Figure 2.9. If the two waves are exactly in phase, the summation will result in constructive

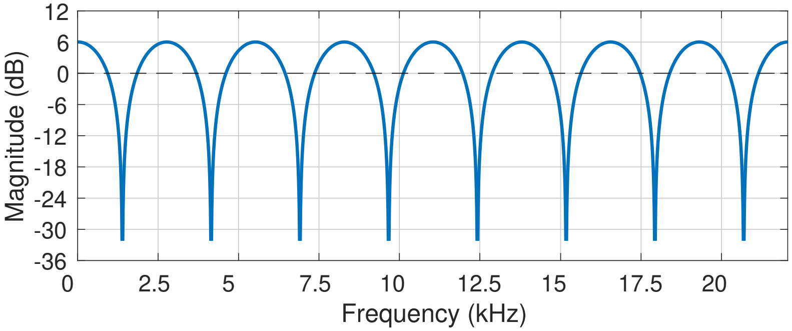


Fig. 2.8: Typical ’comb-filter’ frequency response when a delayed copy of a signal interacts with the original. The distance, or bandwidth, between the *’teeth’* is equal to $1/\Delta t$, where t is the delay time of the reflection.

interference known as *superposition*. When the phase of one sine wave shifts, which is what occurs when it is delayed by an arbitrary amount of time, summation will result in destructive interference known as *cancellation*. If the waves are out of phase by one half-wave length, or where one is the complete inverse of the other, they cancel out entirely. Halmrast (2001) demonstrates that the comb-filter effect occurs when a single reflection interacts with the direct sound. Whilst this example discusses the effect with only a single reflection, in reality it can occur when any number of reflections meet and interfere at a given point in a room. The example also demonstrates the fundamental principle of how the frequency response of a

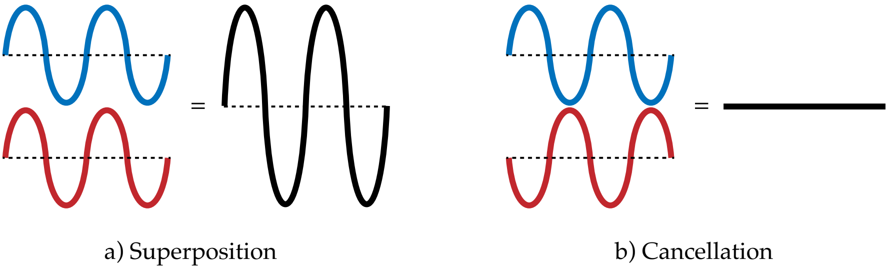


Fig. 2.9: Demonstration of the interference between two waves that leads to comb-filtering. **a)** Shows superposition when waves constructively interfere whilst **b)** shows cancellation when they destructively interfere.

signal changes when mixed with copies of itself, much like in an enclosed space where the direct sound is interacting with thousands of reflections, affecting the frequency response and 'colouring' the sound (Bech, 1996). The characteristics of the frequency response depend upon the position of either the sound source or the listener within the space. The response is also affected by numerous other factors such as the dimensions, geometry and surface material.

2.2.1 Audibility of colouration

Bech (1995) investigated the effects of several individual reflections on the timbre, or colouration, of reproduced sound in rooms, measuring both the threshold of detection and just noticeable differences of the colouration. The motive for the study was to determine which reflections individually contribute to changes in timbre, and what difference in level is required to produce a noticeable change in timbre. Bech (1995) performed an experiment where six speakers were placed at different positions in a controlled yet echoic room in order to simulate reflections of a reverberant space. The reflection paths, level and direction of arrival were calculated using a virtual acoustics program so that the effects of each artificial reflection could be classified by their attributes. Using both a speech signal and a one second long pink noise burst as the stimuli in a listening test, subjects were tasked with identifying a change in timbre when level of each reflection was reduced.

It was generally found that as the reflection level was gradually reduced, several changes could be perceived. First the image shift disappeared, followed by a change in loudness, and finally the colouration between the two sound fields. It was also found that the threshold of detection for a difference in colouration depends on the level of the reverberant field, such that when removed, the threshold will decrease by 2–5 dB. Finally, the study found that floor, ceiling, and wall reflections each have an individual influence on the changes in timbre. This suggests that audibility of colouration is expectedly dependent on reflection level, as well as the level of the reverberant field. It also suggests that the audibility threshold has a dependency on the direction of arrival of the reflection, although this will be discussed in the next sub-section.

Brunner, Maempel, and Weinzierl (2007) later investigated the relationship between reflection delay time and level and the audibility of comb-filter colouration. They conducted an experiment in which the comb-filter effect was simulated digitally using different delay times ranging from 0.1 to 15ms mixed with an unaltered direct sound signal. A psychometric 'staircase' test was performed where the subject was asked to identify the stimulus they thought was being filtered until they reached a point where they could no longer perceived the filtering effect[2]. On each trial the reflection level was changed initially in 0.5 dB steps, then reduced to 0.25 dB after the first reversal in the step direction. Several tests were performed using either a piano, speech or snare drum sample. Brunner et al. (2007) found that the average level difference that listeners were able to detect a difference in timbre was 18 dB, and the sensitivity to the difference grows with increasing delay time. They also found that listeners are sensitive to the changes in timbre of noisy signals.

2.2.2 Effect of reflection direction

One area of timbral colouration in regard to concert halls acoustics that has not been explored deeply is the effects of the reflection direction on the perception of

[2] Psychometric test methods are further discussed in Section 2.3 of Chapter 5

tonal colouration. Barron (1971) noted that the colouration effects due to reflections become less noticeable as the reflections and direct sound become more laterally separated. This suggests that depending on the level difference between the direct sound and the reflection, the audibility of the colouration may decrease as the reflection arrives at an increasingly lateral direction to the listener.

Seki and Ito (2003) investigated the effect the direction of arrival of both the reflection and direct sound on the perception of colouration. They hypothesised that the change in the colouration was the result of the directional dependency of the spectrum of the HRTF. In an anechoic chamber, four loudspeakers were placed around the listener in a circle on the horizontal plane at a distance of 1.8m, see Figure 2.10. A one second long pink noise burst was used as the source signal. The signal was split into two copies where one was delayed by 2ms and attenuated to simulate a reflection. The reflection level ranges from -22.5 to 0 dB in 1.5 dB steps. To simulate different directions that the two signals can arrive from, the direct sound and reflection signals were assigned to a pair of loudspeakers, resulting in seven combinations: FB, FR, BF, BR, RF, RB and RL. The first letter in each combination indicates which

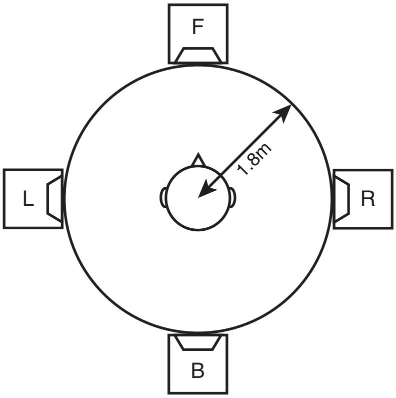


Fig. 2.10: Seki and Ito experimental setup, after Seki and Ito (2003).

speaker is assigned the direct sound, whilst the second indicates which is assigned to the reflection. For each trial, the subject was played first the reference stimulus containing only the direct sound, followed by comparison stimulus containing both the direct sound and the reflection at a random level. The subject was asked to listen to the reference followed by the comparison and say if they could hear any difference in timbre.

The results produced by Seki and Ito (2003) from the experiment showed that the subjects were more likely to perceive colouration as the reflection level increased, which is understandable as the audibility of the colouration increases with reflection level. Furthermore, the 50% threshold of the perception of the colouration was roughly -10 dB for all pairs, thus the 50% threshold does not depend on direction. However, Seki and Ito (2003) observed that when both the direct sound and reflection arrive lateral to the listener, or the RL pair, the peak likelihood at a reflection level of 0 dB is lower, suggesting that listeners are less likely to hear a difference in timbre in this scenario. They hypothesised that this is due to the differences in diffraction aspects of the ear pinna depending on the reflection direction. By analysing the HRTFs at the different direct sound and reflection direction combinations, and computing the difference spectrum between the reference and comparison stimuli to observe the degree of comb-filtering, Seki and Ito (2003) were able to numerically predict the likelihood of the perception of colouration. They found from their numerical model that the likelihood is related to the degree of comb-filtering, which agreed with the subjective results. Plus, as observed with the RL pair, they found that the predicted likelihood was lower at smaller level differences than in other pairs.

It must be noted that the experiment performed by Seki and Ito (2003) was limited to: using a single, 2ms reflection which of course limits the kind of frequency response of the colouration; and the use of a pink noise burst which is more likely to cause a perceived difference in colouration as opposed to using a speech or musical source. They also did not directly observe the change in comb-filtering as a function of reflection, and only observed differences between 90° or 180° spaced pairs of loudspeakers.

2.2.3 On the subjective preference of colouration

It has been established that the nature and characteristics of the colouration depend on several variables, mainly delay time and level. Since these can vary greatly at different positions in a concert hall, the type of colouration will vary too. Therefore, there may be some subjective preference over the types of colouration that may be encountered.

Ando (1977) tested the effects of the delay time and direction of arrival of a single reflection on the subjective preference of resulting sound field using two different orchestral motifs. Whilst it must be noted this study did not directly focus on the preference of colouration, the discussion in the study alludes to a possible subjective preference related to colouration. Ando (1977) observed that the most preferred delay times ranged between 32 to 128 ms, depending on the source material. It was also observed that the angle of arrival of the single echo influenced the preference of the perceived sound field, such that as the angle increased from 0° to 60° there was a sharp increase in preference. In comparison, the IACC expectedly drops yet 'levels off' at around 60°. Based on the subjective scores, Ando (1977) centred the range of preferred echo directions around 55°. To reiterate, although this study is not observing the preference in terms of colouration directly, it is worth mentioning because the Ando (1977) considered that the preference of sound field may be based upon the perceived colouration caused by the nature of a reflection.

In a later study performed by Halmrast (2000), it was theorised that the nature of the colouration may be the reason as to why the acoustical qualities of certain concert halls are perceived as unpleasant, even if other objective measurements may predict the quality of the acoustics of the halls to be good. However, the study also hypothesised that perhaps not all colouration is perceived as bad, and thus investigated the perception of colouration in order to deduce what is considered *good* or *bad*. This was done by measuring RIRs in different concert halls, taking measurements at many points around the orchestra platform and audience seating area, and analysing the frequency response of the early reflections using a sequence of short analysis windows. Halmrast (2000) found that comb filter effects occur at different time regions after the initial direct sound, and that the frequency response was dependent on the geometry and position of nearby surfaces.

It was concluded that certain types of colouration enhance the timbre of orchestral bass instruments, such as the double bass and timpani, by the means of constructive interference enhancing the low frequencies. This would of course then be considered an example of *good* colouration. On the other hand, discrete early reflections arriving within 5–20 ms after the direct sound can create a sense of *box-klangfarbe*, which translates to "as if the orchestra [was] placed in a small box." (Halmrast, 2000), which is considered *negative*. This can be avoided by creating more reflections through the use of either diffusers or suspended reflectors, as well as building the geometry of the space to spread the distribution, such that their inter-delay times do not fall within the *box-klangfarbe* region. To summarise, these findings define that *good* colouration is regarded as a type that enhances the perceived sound, whilst *bad* colouration is that which distracts the listener and makes the auditory source sound unpleasant.

Robotham (2016) investigated the effect of a single, vertical reflection on the relationship between subjective preference and spatial and timbral attributes. Two experiments were performed to first elicit the timbral and spatial effects of the vertical, ceiling reflection, and to them determine if the effects have either a positive or negative effect on the listening experience. To mimic a ceiling reflection, a sound field was simulated using two loudspeakers with one placed directly in front of the listener, and the other placed at half-way between the listener and the front loudspeaker, and elevated above the listening position whilst being pointed towards the listener as to simulate a ceiling reflection. The signal from the vertical loudspeaker was delayed by 1.63ms and attenuated by 4.1 dB relative to the direct sound loudspeaker to replicate the same delay time as the real ceiling reflection. Two stimuli were used in the experiments, where the reference stimulus was the direct sound signal only, and the comparison stimulus was the direct sound combined with the simulated reflection. Six auditory source types were used throughout the experiments, including continuous and transient musical signals, and a speech signal.

The first experiment was performed in a semi-anechoic chamber in order to exaggerate the vertical reflection. In the first section, subjects were tasked with comparing the two stimuli, and then grade the level of preference between the two. They then graded the perceived timbral and spatial differences between the stimuli. From this first experiment it was concluded that there was no correlation between the perceived timbral or spatial differences and subjective preference, and that there was no significant difference in spatial or timbral differences, or subjective preference, when the vertical reflection was present.

The second experiment was similar to the first experiment in terms of the stimuli used and methodology, although this time it was performed in the ITU-R.1116 (2015) compliant room at the University of Huddersfield as to investigate listener preference in a more realistic listening environment. In this experiment, subjects were tasked with grading the preference between the same reference and comparison stimuli, as well as verbally describe the reason behind their decision. From this second experiment, it was found that for 42% of the time, the timbral differences between stimuli were the basis for a subject's preference.

In conclusion, the study found that there was a high variance in subjective preference over the playback of audio with the inclusion of the vertical reflection, and that the addition of the vertical reflection resulted in both positive and negative changes in colouration. Robotham (2016) found that subjects based their preference on timbral rather than spatial attributes, such that when expressing a preference of a particular sound field, the most frequently elicited attributes were timbral. Furthermore, it was found that the changes in timbral sensations, whether perceived as positive or negative, correspond with the individual subjects increase or decrease in preference for the sound field containing the vertical reflection. What this suggests is that, overall, not all colouration is actually perceived as being *'unacceptable'*, and that in certain circumstances such as arriving from a particular direction, at a certain delay time, or depending on the source material, there is a subjective preference of the colouration, be it positive or negative.

2.3 Discussion and summary

Spatial Impression (SI) is widely accepted as being the combination of at least two sub-paradigms: Apparent Source Width (ASW) and Listener Envelopment (LEV). ASW describes the perceived width of an auditory source and is dependent on early reflections. The reviewed literature establishes that the direction of arrival and level of the early reflections are the predominant factors in the perception of ASW, such that the more lateral the reflections are to the listener, the higher the ASW. However, the correlation between the left and right ear signals also plays a major role, where the less correlated the signals are, the greater the width. From this, ASW can be predicted using the Lateral Fraction (L_f) which measures the ratio of lateral reflection energy to total energy, or the Inter-aural Cross-correlation Coefficient (IACC) of early reflections. Both are measured up to 80ms after the direct sound.

LEV, on the other hand, describes the sensation of being enveloped by sound and is generally regarded as an environmental effect. The level of late arriving, lateral reverberant energy is accepted as being the greatest contributor to LEV, thus it can be predicted using the Late Lateral Fraction (LF_{80}^{∞}), which is measured from 80ms and onwards. More recent literature suggests that LEV is further enhanced by non-lateral reverberant energy, for example from behind or above the listener, which is likely to cause the listener feel more enveloped by the sound.

Section 2.2 discussed the effects of early reflections on timbre and tonal colouration. In reverberant spaces there will naturally exist a form of tonal colouration. It was found that a single reflection interacts with the direct sound and creates timbral changes or colouration which can manifest itself as a number of subjective attributes, most notably as *comb filtering*. The nature of the colouration is linked to the delay time and level of the reflections. Furthermore, by assuming that the source position is fixed, the reflections delay time and level are thus dependent on the listener's position. There is literature that suggests that the direction of arrival can influence of the audibility and preference of colouration, although it is unclear what the exact effect that reflection direction has, other than it is possible that the influence is related to directional dependency of the head-related transfer-function (HRTF).

This colouration can also be perceived as either a *positive* effect that can support and enhance the sound, or as a *negative* effect that can distract the listener and create an unpleasant listening experience. At short delay times between 5–20ms, discrete reflections may create a *boxiness* effect, which can be perceived as negative colouration. This can be overcome by not allowing reflections to arrive too close together in time, or by spreading the reflections by physically diffusing them. The severity of the colouration is dependent on the reflection level, and this is best observed with the comb-filter effect where the depth of the 'teeth' increases with reflection level. It was in fact found that there is an audibility threshold for comb-filter colouration, such that the effect is still audible with a level difference as high as 20 dB between the reflection and the direct sound. When keeping in mind the relationship between reflection level and audibility of colouration, it is possible that there may be a threshold in which whilst the colouration is still audible, yet it is in fact acceptable. Linking this to the possibility of reflection direction affecting the audibility of colouration, it is further possible that the colouration may be perceived as being acceptable when the reflection arrives more lateral than frontal to the listener.

Chapter 3

Simulating room acoustics

In music production, video games, and in general the creation of spatial audio experiences, it is common to use artificial reverberation in order to create a sense of space. The primary objective is to make the listener think that the experience is believable, plausible, and immersive. Today, it is desirable for artificial reverberation to be efficient and perceptually plausible, such that it can be coupled with virtual reality and video games. Also, in these situations it is important for the reverb to sound pleasant and for it to positively contribute to the experience, heightening the listener's feeling of immersion.

Most methods, if not all, concentrate on accurately modelling the behaviour of sound. However, achieving high levels of accuracy in real-time simulations has been difficult to accomplish, and is only beginning to be achieved on consumer hardware now available at the time of writing. Still, trade-offs must be made to achieve reasonable efficiency whilst maintaining a good level of plausibility. This chapter looks at several methods that are commonly used to simulate reverberation, discussing their development history, implementation and limitations. This chapter also provides background information for the subsequent chapter that will introduce the custom artificial reverberator developed for this project.

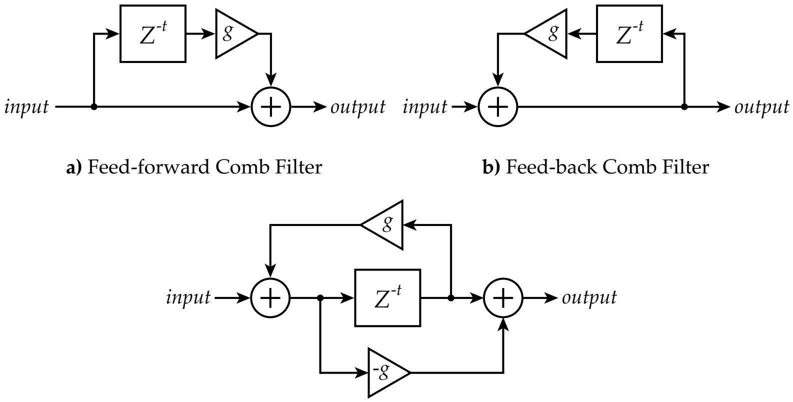
3.1 Algorithmic reverberators

An algorithmic reverberator uses a network of digital delays and filters to simulate reflections and echoes exhibited in the reverberant space. Over the course of fifty years, there has been constant research and development of various algorithms that are designed to either simulate a particular type of room such as a hall, or to create pleasing sense of space and dimension that "brings life" to an otherwise "lifeless" and dry recording (Välimäki, Parker, Savioja, Smith, & Abel, 2012). What is common to all types of algorithmic reverberators is their use of digital delays and filters, yet what is interesting about this particular class of reverberator is how the arrangement of delays and filters and their chosen parameters can result in many types of perceived reverberation.

3.1.1 Comb and all-pass filtering

To increase the echo density of a signal, the most common method is to make the use of a digital filter to make copies of the original signal (Schroeder & Logan, 1960), see Figure 3.1. The three common types are feed-forward, feed-back comb filters, and the all-pass filter. The feed-forward comb filter, characterised by its distinct 'hair comb' frequency response, feeds the original input signal ahead of the digital delay unit, Z. When the delayed signal emerges from Z, it interferes with the original signal, much like what happens in reality when a reflection interacts with the direct sound. The severity of the interference is controlled by the gain factor g. Because the output does not feed-back into the filter, its impulse response (IR) will be finite, thus making this type of comb filter a Finite Impulse Response (FIR) filter.

A feed-back comb filter meanwhile lets the original signal pass through, however, the difference is that now the delayed copy signal is fed back to the input by a fraction controlled by g. If a short impulse was processed by a feed-back comb-filter, provided that enough signal was fed back into it, the resulting impulse response (IR) would be heard repeating at a constant time interval with its amplitude decaying over time. The time interval between impulses is determined by the delay time t, whilst the rate of decay is determined by both the decay time and the gain factor g. With either a long delay time or g value close to 1.0, the signal would take longer to decay. Because the signal is fed back into the delay and is multiplied by the value of g, the IR will always decay if g is less than 1.0. When g is equal to 1.0, the IR will never decay, whilst above 1.0 the amplitude will continue to grow, or 'explode'. Because of this behaviour, and the fact the IR will decay indefinitely, the feed-back comb filter is otherwise known as an Infinite Impulse Response (IIR) filter.



a) Feed-forward Comb Filter

b) Feed-back Comb Filter

c) All-pass Filter

Fig. 3.1: The types of filters commonly used in an algorithmic reverberators. Z denotes a delay unit with a delay time given by t, after Schroeder (1961).

Schroeder and Logan (1960) noted that the comb filter imparts a metallic, timbre to the signal. As an alternative, they proposed the 'all-pass' filter, whose name is derived from its flat frequency response. This filter both feeds the delayed signal back into itself, as well as the original signal ahead of the delay. Both comb and all-pass filters form the basis of algorithmic reverberators.

3.1.2 Artificial room reverberation

Schroeder and Logan (1960) theorised that the reverb of room can be artificially simulated by using a network of comb-filters and all-pass filters. They assert that for the resulting reverb to sound natural and plausible, the network must be capable increasing of the echo density of the signal to at least 1000 echoes per second. Thus, they proposed an algorithm that used a bank of parallel, feed-back comb filters, that sum into two, cascaded all-pass filters. A flow chart of the algorithm is shown in Figure 3.2. Casual testing of this algorithm[1] found that it is generally suitable to smooth, non-transient sounds such as bowed instruments and soft vocals, yet it is completely unsuited to percussive and transient sounds, revealing an unpleasant

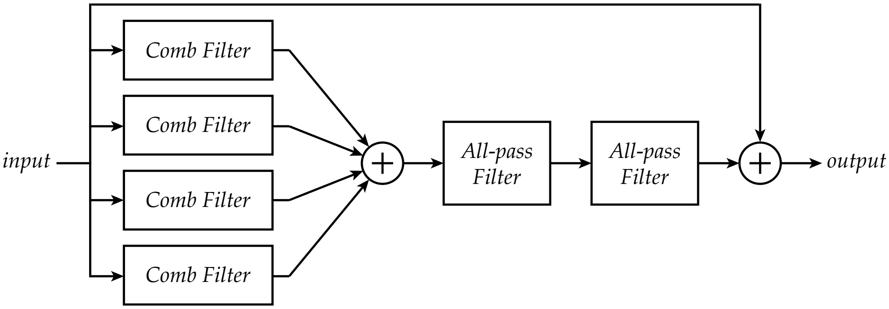


Fig. 3.2: Flow chart of the Schroeder reverb algorithm, after Schroeder (1961).

[1] All algorithms in Section 3.1 were replicated and auditioned using patches created in Cycling 74's Max 7 software, along with anechoic sound sources.

metallic, resonant quality. The algorithm is also unable to simulate air or surface material absorption, and exhibits a ’puffing’, diffused quality due to the lack of early reflection modelling. However, early reflections were later added to this algorithm by Schroeder (1970) using a multi-tap, FIR delay.

Moorer (1979) adapted the Schroeder and Logan (1960) design to simulate air and surface material absorption by adding low-pass filters into the feedback path of the comb filters. Moorer also expanded the early reflection modelling capability by calculating the delay time of each tap of the FIR using a method known as the Image Source Method (ISM), proposed by Allen and Berkley (1979). The ISM is later discussed in further detail in Section 3.2.2.

Gardner (1992) altered the Moorer (1979) design, proposing three ’diffuse’ reverberators all based around a general all-pass filter only algorithm, with each algorithm optimised for a particular room size. This design incorporates a nested all-pass filter, which is an all-pass filter whose delay core is another all-pass filter. This nested filter was designed to further increase the echo density. Gardner theorised that a single diffuse reverberator would not be able to simply span a wide range of reverberation times by linear scaling of delay time or gain, e.g. making a large room sound like a small room by decreasing the gain coefficients and reducing the delay times, resulting in poor quality and unoptimised reverberation.

Jot and Chaigne (1991) introduced an algorithm based on the Schroeder and Logan (1960) design known as a Feedback Delay Network (FDN) reverberator. Like Moorer (1979) before, they addressed many of the original’s shortcomings. Jot and Chaigne (1991) noted that both the comb-filter and all-pass filter designs proposed by Schroeder and Logan (1960) suffered from unnatural tonal colouration. However,

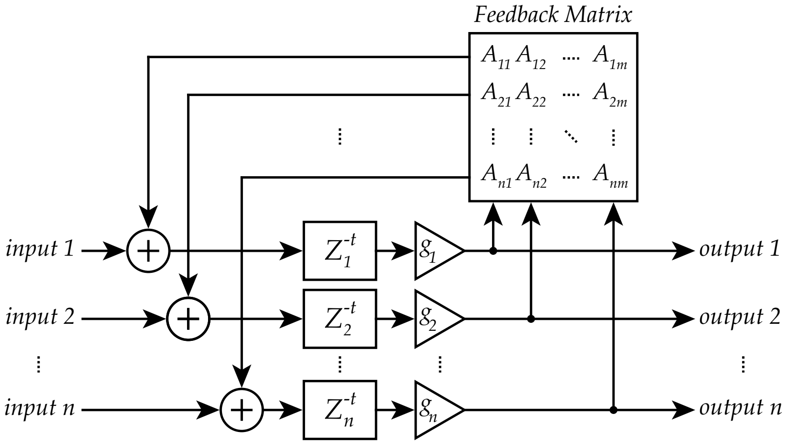


Fig. 3.3: Flow diagram of a Feedback Delay Network (FDN) reverberator, after Jot and Chaigne (1991).

they did find that the frequency response of parallel comb filter section was closer to that of a real room. They then proceeded to generalise the parallel comb-filter section into a multiple feedback delay network, where the output of each delay unit is fed back into each other via a feedback matrix, see Figure 3.3.

The choice of values in feedback matrix A is a highly debated topic for FDN reverb as they influence both its frequency response and efficiency (Jot, 1997), although the general rule is that they should not allow the reverberator to 'explode'. This gives rise to an interesting quality of the FDN reverb where, disregarding the gain coefficients g_n, with correctly chosen matrix values the reverb can be made into a lossless system. Therefore natural, frequency dependent absorption can be simulated using spectral filters in place of the gain coefficients. Finally, this reverberator alone is only able to simulate reverberation created by late reflections, and thus would also require a multi-tap FIR filter stage to model early reflections.

3.1.3 Limitations

Algorithmic reverberators are designed to model the characteristics of a particular type of room, such as a concert hall, a large room, or even an abstract model that gives a general impression of reverb for a musical or artistic purpose. The discussed methods proposed by Moorer (1979), Gardner (1992) and Jot and Chaigne (1991) can model attributes such as the decay and diffusion of the reverb tail, and can do so with great plausibility. However, algorithmic reverberators are, in essence, a type of complex digital filter made up of several smaller filters that give the impression of reverb, yet do not mimic an exact impulse response of a real space. Thus, they are not suitable for analysis of exact room models which require much more accurate modelling techniques such as geometric methods, which will be discussed later in Section 3.2. Needless to say, for real time applications such as virtual reality (VR) that require low CPU usage, algorithmic reverberators have been successfully coupled with geometric methods under the assumption that only the early reflections need to be accurately modelled, and that the late reverberant tail only needs to be modelled generically in order to be plausible (Wendt et al., 2014). Again, these hybrid approaches will be discussed later in Section 3.4.

However, this study calls for a reverberator that is able to provide access to each individual reflection, including metadata such as the direction of arrival and octave band energy, which are properties that algorithmic reverberators are currently unable to provide.

3.2 Geometric

Geometric reverberators approximate the paths of sound waves using rays under the assumption that sound behaves in the same way as light. There are several approaches to this, the main methods being Ray Tracing, the Image Source Method, and Beam Tracing,

3.2.1 Ray tracing

Compared to the other geometric reverberation methods that will be discussed later, ray tracing is the most simple and crudest method. In ray tracing, the sound wave is quantised into individual rays that represent a wave front. In most cases, the rays also carry a fraction of the initial energy. Krokstad, Strøm, and Sørsdal (1968) introduced ray tracing as a method of estimating the impulse response of concert halls at various audience seating positions. Later, Schroeder (1970) used ray tracing simulations to predict the reverberation time of various enclosures, as well as compare the calculated the results against existing measures. From this study, Schroeder (1970) found discrepancies between the simulations and the measures, yet notes that ray tracing can be a valuable tool for architects and acousticians who need such measures. Later, Kulowski (1985) describes a full implementation of ray tracing in an algorithmic representation, considering what is the optimal solution for given computer system. The Kulowski (1985) method, however, does not consider 3D audio reproduction or spatialisation. This was most likely due to limitations in computer hardware of that period, highlighting the efficiency limitation that often overshadows ray tracing.

Eventually, an interactive, scalable system using ray tracing was introduced by Mueller and Ullmann (1999), and later Savioja (2000) introduced a similar system in the DIVA (Digital Interactive Virtual Acoustics) project. Over time as computer hardware has become faster, ray tracing has been combined with other reverberation methods which have been incorporated into virtual reality (VR) experiences. The RAVEN (Room Acoustics for Virtual ENvironments) software framework created by Schröder (2011) uses ray tracing in combination with the Image Source Method (see Section 3.2.2) for real-time rendering with promising results.

3.2.1.1 Typical implementation

Ray tracing is the most straightforward of all geometric approaches to implement. A high number of rays are emitted in all directions from a single source point arranged as a sphere, and are traced around the space. Rays reflect from any surface they may collide with, losing energy in the process due to surface absorption, whilst potentially being detected by a spherical receiver to form the RIR. A detected ray's current energy is registered to the RIR at the corresponding delay time. As a ray travels further away from their starting position, it loses energy due to air absorption. When the energy of a particular ray is treated as negligible, that ray is no long rendered, or 'annihilated' from the process. The entire rendering process is completed once all rays have been annihilated.

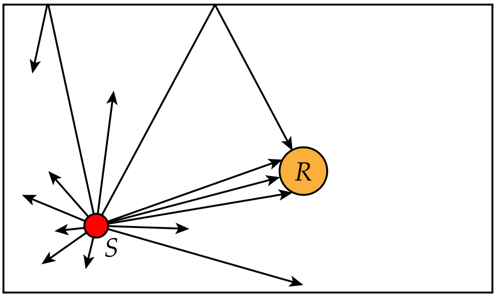


Fig. 3.4: 2-D representation of ray tracing where S is a point source, whilst R is a receiver of a given radius. The arrows represent rays being emitted from the source and reflecting off the surfaces in the virtual room.

3.2.1.2 Emission distribution

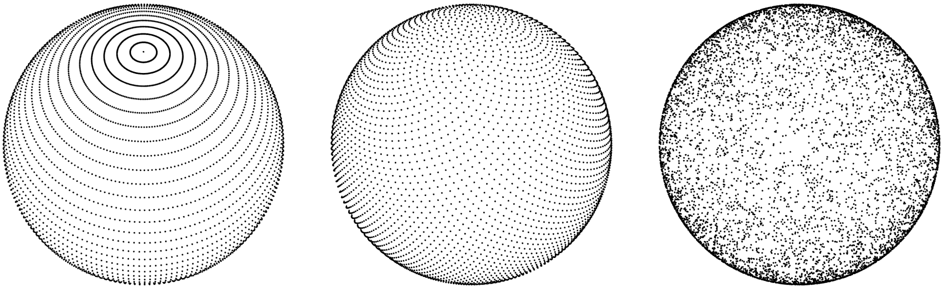


Fig. 3.5: Three examples of ray emission distributions: **Left:** Even Latitude, **Middle:** Fibonacci Lattice, and **Right:** Monte Carlo or random distribution. Notice how with the 'Even Latitude' distribution the density of rays is smaller at the equator than at the poles of the sphere, whilst in the other two methods the distribution is equal for any given area.

A very important point to consider when developing a ray tracing algorithm is the initial distribution of rays. In an ideal situation, unless the source has a particular directivity, it is otherwise assumed that the sound energy from a point source is emitted omni-directionally and is distributed equally (International Standards Organisation, 2009). Therefore, the rays should be arranged using a particular distribution pattern around a sphere that fulfils this requirement. The issue of uniformly distributing points on a sphere is widely discussed mathematical topic, with solutions often being applied to computer graphics when mapping image textures onto a curved surface without any noticeable distortions. González (2010) discusses a naïve solution to this problem where an equal number of points are mapped onto evenly spaced, latitudinal layers. This approach is inadequate as the density of points is greater towards either pole of the sphere than at its equator, see Figure 3.5.

González (2010) proposed a predefined arrangement called the 'Fibonacci Lattice'. This solution places points along a spiral that starts and ends at each pole of the sphere. The points are spaced by specifically $\pi(3-\sqrt{5})$, or 137.5°, a value known as the 'Golden Angle' denoted as ϕ. The normal, three-dimensional unit vector $\hat{V}$ of each ray in the emission sphere can be calculated using the following process. The azimuth angle θ of the ray is given by:

$$\theta = n \cdot \phi \tag{3.1}$$

Vector $\hat{V}$ is made up of three components: x, y and z. Before the x and y components of $\hat{V}$ can be calculated, the z component of the vector must be found using

$$z = 2 \cdot \frac{n}{N-1} - 1 \tag{3.2}$$

As $\hat{V}$ is a unit vector, the length should always be equal to 1, so the values for x, y and z are proportional to each other. When the sphere is viewed from either pole, x and y form a 2D vector that represents a circle with a radius r, which is inversely proportional to z, and is given by

$$r = \sqrt{1-z^2} \tag{3.3}$$

The x and y components can then be derived from r and θ with

$$x = r \cdot \cos\theta \tag{3.4}$$

$$y = r \cdot \sin\theta \tag{3.5}$$

where θ is the the angle of the vector. Thus, vector $\hat{V}$ is given by

$$\hat{V} = \langle r \cdot \cos\theta, r \cdot \sin\theta, z \rangle \tag{3.6}$$

Kulowski (1985) discusses the use of a Monte Carlo approach where rays are distributed randomly by choosing random numbers for each component of $\hat{V}$. If the chosen random number generator produces an even distribution of numbers, then the distribution of points can also be considered even. They assert that a Monte Carlo approach may solve the issue of not knowing the required number of rays prior to rendering. However, Savioja (2000) found a predefined distribution to be preferable for when rendering using fewer rays.

3.2.1.3 Modelling diffusion

One of the key advantages of ray tracing over other geometrical methods is its ability to model diffusion, yet the choice of method is a highly debated topic in ray tracing algorithm design.

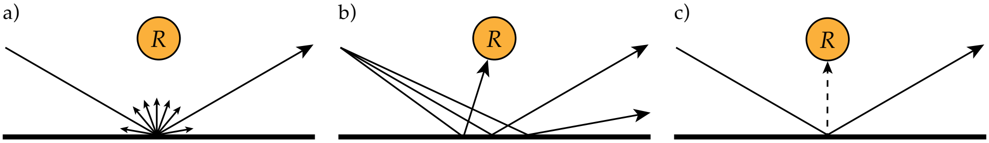


Fig. 3.6: Diffuse reflections can be modelled by **a)** splitting reflections into child rays upon reflection, or **b)** can be randomly scattered to non-specular directions. **c)** In the Diffuse Rain technique a shadow ray, shown as a dashed arrow, is randomly spawned towards the receiver, after Savioja and Svensson (2015).

Rindel (2000) discusses a method of modelling diffusion where rays are split into smaller rays upon reflection. The likelihood of this scattering is given by a diffusion coefficient. A fraction of the specular ray's energy is given to the newly spawned diffuse ray, which is also determined by this same coefficient. The diffused ray's energy is further affected by Lambert's cosine law, given in vector form by Eq. 3.7, such that as the angle θ between the incident ray and the surface normal increases, or the dot product between those two vectors decreases, the energy given to the diffuse ray decreases.

$$I_d = I_i k_d \cos\theta = I_i k_d \vec{N} \cdot \vec{R} \tag{3.7}$$

where I_d is the diffused ray's energy, I_i is the incident ray's energy, k_d is the surface diffusion coefficient, and $\vec{N}$ and $\vec{R}$ are the surface and ray normal vectors. Zeng, Christensen, and Rindel (2006) discuss another method where the ray's direction is scattered in a random direction by a given amount that also follows Lambert's law.

Schröder (2011) presents a more computationally efficient method of diffusion known as 'Diffuse Rain'. In contrast to the method proposed by Rindel (2000) where diffuse rays are spawned in random directions during surface intersection, the 'Diffuse Rain' method spawns them in the general direction of the receiver. These types of rays are known as *shadow rays*, see Fig. 3.6(c). The probability that a shadow ray will be detected is related to the distance from the point of intersection to the receiver, and the size and position of the receiver. Pelzer, Schröder, and Vorländer (2011) have found that this method reduces the required initial number of rays.

3.2.1.4 Caveats

The major caveat of ray tracing is a concept known as 'spatial aliasing', which is discussed in detail by Lehnert (1993). Spatial aliasing is a phenomenon that occurs when too few rays are used for rendering, resulting in missed reflections and a synthesised impulse response with reduced accuracy. This problem is usually caused by choosing an unsuitable initial number of rays that is not suitable for the size and geometry of the space. Essentially, large or complex spaces require a high number of rays to overcome spatial aliasing, yet as the number of rays increases, the longer it takes to render an impulse response. This means that compromises and decisions must be made by the user depending on their needs. For quick rendering they can choose to use a low number of rays at the expense of accuracy. However, if they require higher accuracy, they would choose more rays at the expense of increasing the rendering time. The time of the rendering process is naturally reduced when faster hardware is used, or if the rendering program itself is well optimised.

Another problem with ray tracing is the choice of receiver size. In order to ensure that there is a high likelihood that important reflections are captured, and that enough of the reverberant energy is captured, the receiver needs to be of a suitable size. Lehnert (1993) find that errors are also likely to occur with an improper receiver size, and thus calculates the radius of a spherical receiver under the assumption that the rays are emitted evenly from the source:

$$r = L_{max} \cdot \sqrt{\frac{2\pi}{N}} \tag{3.8}$$

where r is the desired receiver radius, L_{max} is the maximum length of a ray, and N is the initial number of rays. Zeng et al. (2003) notes that Lehnert (1993) did not take the volume of the space into consideration, and thus gives the following equation derived from a method initially proposed by Yang and Shield (2000):

$$r = \sqrt[3]{\frac{15V}{2\pi N}} \tag{3.9}$$

where V is the volume of the space. A key difference between Eqs. 3.8 and 3.9 is that the latter does take the maximum length of a ray into account. Zeng et al. (2003) note that this difference plays a key role in the accuracy of the simulation. They also found that neither equation appears to take into account the source-receiver distance, denoted as d_{SR}, and so propose the following:

$$r = k \cdot d_{SR}\sqrt{\frac{4}{N}} \tag{3.10}$$

where k is a weighting coefficient dependent on the volume of the space, and is given by:

$$k = \log_{10}(V) \tag{3.11}$$

When comparing all three methods of calculating the receiver radius, Zeng et al. (2003) found Eq. 3.10 gives the most accurate and stable simulation.

3.2.2 Image source method

The Image Source Method (ISM), sometimes known as the mirroring-method, uses a deterministic rather than a brute-force, stochastic approach to simulating reverberation. It was first introduced by Mintzer (1950) to model sound behaviour in rooms, and later implemented by Allen and Berkley (1979) to render a RIR of a simple shoebox shaped room. Borish (1984) extended the method to model arbitrarily shaped spaces. It has continued to be a popular simulation method due to its relatively simple implementation, and is often combined with other methods to overcome certain limitations, such as the 'spatial aliasing' issue that can be encountered in ray tracing when too few rays are used, resulting in missed early reflections (Lehnert, 1993). For example, Vorländer (1989) combined the ISM with ray tracing to improve the accuracy of early reflection modelling whilst leaving the late, diffused portion of the reverb to be simulated efficiently by ray tracing. More recently, Wendt et al. (2014) combined the ISM and a binauralised, algorithmic, FDN reverberator as part of an interactive, VR experience.

3.2.2.1 Implementation

The ISM is performed by mirroring the position of the original sound source through each surface to create virtual image sources in a process known as 'Image Expansion'. The distance from any image source to the receiver is equal to the distance the real sound path takes from source to receiver via the image source's reflecting surface. Higher order reflections are obtained by recursively reflecting each image source through each surface. All image sources can be stored in a tree structure called the 'Image Tree', where the root node is the original source, and each child node represents an image source and its reflecting surface, see Figure 3.7. The RIR is generated by measuring the distance between image source and the receiver to calculate the delay time of each impulse. This simplistic approach is valid for a concave, shoebox shaped room, although for more complex models and arbitrarily shaped rooms, the real path must be derived using back tracing, then tested for validity. Back tracing is performed by tracing a path from the receiver to the image source, and locating the intersection point between the path and the reflecting surface associated with that image source, see Figure 3.8. This creates a segment of the real path. If there is a parent image source, another segment is generated by repeatedly using the same process between the previous segment's surface intersection and parent image source, until the real root source has been reached. For each segment of this real path to be valid, the following criteria must be met:

- A segment must not be obstructed by surfaces that do not contribute to the

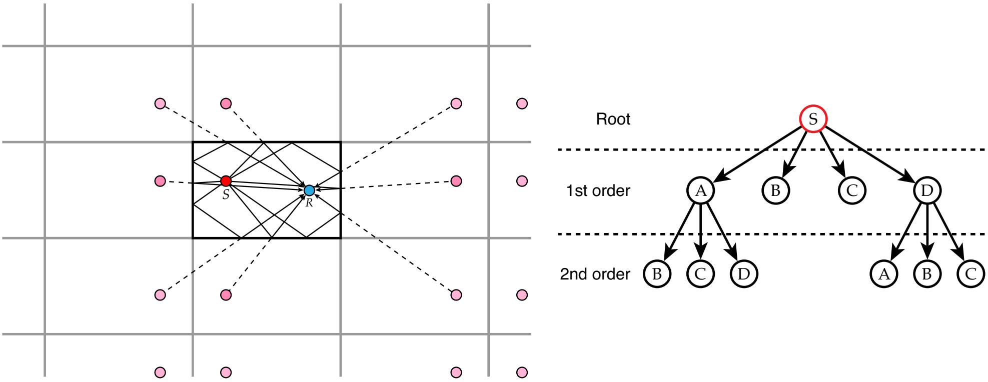


Fig. 3.7: The ISM performed on a single source and receiver in a shoe box room. **Left:** Top down view of the room, and the surrounding virtual rooms and image sources. The dashed lines represent the paths from the image sources to the receiver, and are the same length as the solid lines that represent the actual sound path. **Right:** The 'Image Tree', where at the top is the root source, followed by its child image sources and their children.

path's reflection sequence.

- The intersection point must lie within the boundaries of the associated surface.
- The image source must lie behind the associated surface.

3.2.2.2 Limitations

Firstly, due to the nature of the 'Image Expansion' process, the ISM is limited to rendering solely specular reflections. Whilst this is ideal for accurately rendering early reflections, it is unsuitable for rendering a late diffused tail. This is usually overcome by combining the ISM with another algorithm that is optimised for modelling diffusion, which will discussed later in Section 3.4.

Secondly, in its basic implementation, the number of image sources the ISM produces per reflection order increases exponentially. During the reflection of each image source, several more child image sources are spawned, thus many more recursive

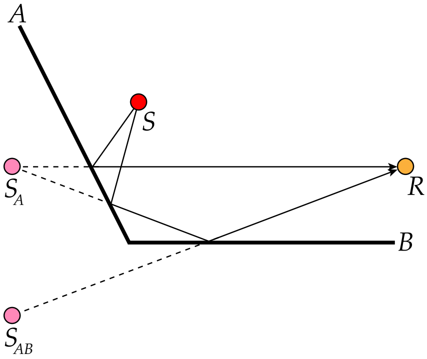


Fig. 3.8: The ISM up to second order in an arbitrary space. The image S_A produces the first order reflection, which is equivalent to the path reflected by surface A, whilst S_{AB} produces the second order which is equivalent to the path reflected by surface A, followed by B (after Vorländer (1989)).

reflections need to be performed. Thus, as the reflection order increases linearly, the number of image sources increases exponentially. The theoretical maximum number of image sources for a desired reflection order is given by:

$$S = \sum_{k=0}^{K} N(N-1)^{k-1} \tag{3.12}$$

where S is the number of images sources, K is the desired maximum reflection order, and N is the number of surfaces (or polygons). Table 3.1 demonstrates this severe exponential growth of image sources up to a reflection order of 10 in a six sided, shoebox shaped room.

Reflection Order	Number of image sources
1	6
2	36
3	186
4	936
5	4 686
6	23 436
7	117 186
8	585 936
9	2 million
10	14 million

Table 3.1: Theoretical maximum number of image sources for a given reflection order

The validity criteria discussed in the previous section implies that many of the image sources generated are in fact *invalid*, an implication that becomes a major limitation for the ISM when attempting to render high order reflections. Most of the generated images sources will be redundant because:

- They will be reflected onto the incorrect side of a surface.
- They will be reflected into a node already occupied by another image.
- Paths from image sources to a receiver may be occluded by the surfaces not part of the true reflection sequence of that path.

This problem is worsened when the ISM is implemented in a system that uses triangular polygon geometry, which is commonly used in modern 3D modelling. A rectangular polygon can be decomposed into two triangle polygons. Therefore, in a 6 sided shoebox shaped room, there are 12 triangle polygons. At a reflection order of 5, this results in a theoretical maximum of over 31 billion image sources. This imposes a major computational overhead in terms of the amount resources needed to render high order reflections using the ISM. Rapid rendering speeds would require a high performance computer with large amounts of memory. Furthermore, only a small percentage of the image sources will be valid, highlighting that the basic implementation of the ISM at high orders becomes inefficient.

3.2.3 Beam tracing

Beam tracing is an evolution of the ISM. It was initially used in optics research by Heckbert and Hanrahan (1984), and also by Dadoun, Kirkpatrick, and Walsh (1985) to solve what is known in computer graphics rendering as the 'Hidden Surface Problem', such that objects are drawn in the correct order on the screen. It was adapted to acoustics simulation by Funkhouser et al. (1998) who implemented the method for accelerated calculation of high order specular reflections, enabling for accurate real-time simulation of a virtual space. The limitation of their implementation at the time was that sound sources were fixed to a position since moving them would require recalculation. The beam tracing technique was later optimised by Laine, Siltanen, Lokki, and Savioja (2009) to enable movable sources as well render much more complex scenes.

3.2.3.1 Implementation

Beam tracing limits the number of reflection path combinations that a real time ISM has to search through by pre-calculating all the possible sound paths using pyramidal beams, and storing these paths into a beam-tree. This beam tree describes the order in which each beam was generated. This is similar to the ISM where the root of the tree is the sound source, and each child image has a corresponding child beam that is

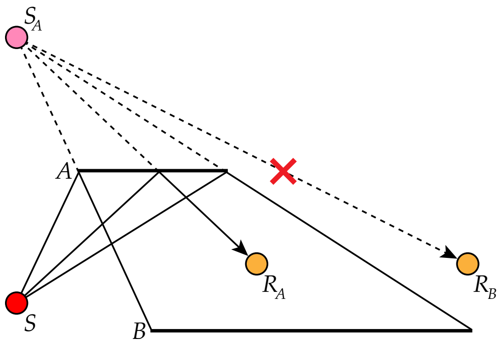


Fig. 3.9: Visual representation of first order beam tracing. A beam is spanned from the source S to the edges of surface A. An image source S_A is created, followed by a child beam spanned from A to B. If more surfaces were included in the diagram, then more child image sources and beams would be recursively created. If a receiver lies within a beam, such as R_A, then a path is back traced to S. If not, then no path can be traced, as seen with R_B.

recursively traced out into the scene, see Figure 3.9. The IR at the receiver's position is calculated by testing which beams the receiver lies within, although unlike the ISM, there are less visibility checks that have to be performed at run-time since these have already been performed during the construction of the beam tree. These checks also address the exponential growth of reflections problem in the ISM by culling reflection paths that never produce a valid reflection path during pre-calculation. The main drawback of beam tracing, however, is that it is limited to only modelling specular reflections. Furthermore, whilst it is able to render high order reflections more efficiently than the ISM, this is generally unnecessary.

3.3 Digital waveguide mesh

This section will now give a brief overview of a popular wave based method known as the Digital Waveguide Mesh (DWM). The common limitation of geometric methods is their inability to accurately model wave effects such as diffraction and interference, which have an effect on the RIR at low frequencies (Murphy, 2000). The DWM address these limitations by modelling the propagation of sound waves in an enclosed space. It is a type of finite difference time domain (FDTD) method and provides a rather accurate approach to physically modelling a reverberant space using a grid based approach. They are based on the physical modelling methods for simulating membranes and resonators (Smith, 1992), which make use of bi-directional digital delay lines, or wave guides, to simulate wave propagation through a medium. The DWM arranges these as an N-dimensional mesh to model wave propagation through a space.

Smith (1987) originally developed a digital reverberator based upon a closed network of lossless wave guides of arbitrary length. Inputs and outputs to the wave guide can be placed at any point. Networks are created by connecting the wave guides together, where the connecting point is called a 'node'. Losses such as absorption are introduced by inserting attenuators or lowpass filters in between connected wave guides. Van Duyne and Smith (1993) adapted this method of artificial reverberation by creating a 3-D mesh of discretely spaced 1-D wave guides arranged along a given dimension. The wave guides are connected using lossless scattering junctions, J, that distribute the energy between connecting wave guides, see Figure 3.10. Savioja (2000) derived the following difference equation for an N-dimensional, rectangular mesh:

$$p_k(n) = \frac{1}{N} \sum_{l=1}^{2N} p_l(n-1) - p_k(n-2) \tag{3.13}$$

where p is the sound pressure at a junction k at time step n, and l represents the neighbouring junctions.

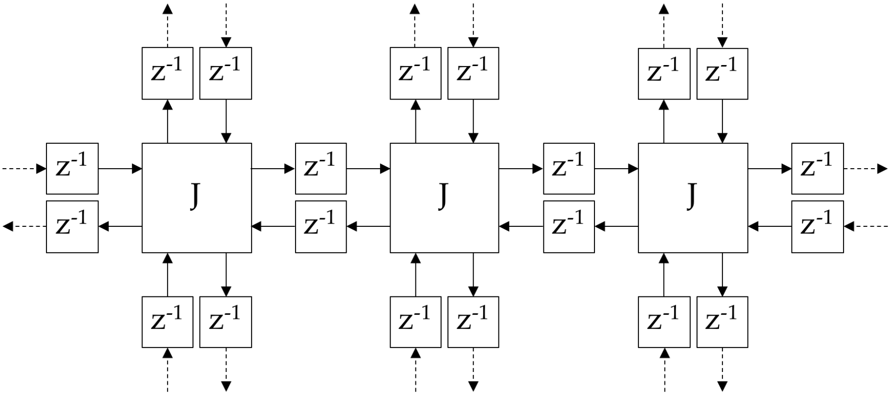


Fig. 3.10: 2-D digital waveguide mesh topology, after Savioja (2000).

3.3.1 Limitations of DWM

The DWM can only model frequencies up to a limited update frequency, thus limiting the maximum frequency response of the resulting RIR, depending upon the distance between junctions. Savioja (2000) found the maximum update frequency of an N-dimensional mesh to be:

$$f_s = \frac{c\sqrt{N}}{\Delta x} \tag{3.14}$$

where c is the speed of sound and Δx is the distance between nodes. For a 3-D mesh, the maximum frequency can be given by:

$$f_s \approx \frac{588.9}{\Delta x} \tag{3.15}$$

For an update frequency of 500 Hz, the distance between junctions must be roughly 1.2m. However, in order to propagate an audio signal with a bandwidth of 20kHz, the distance between junctions must be 0.03m apart. In a shoebox shaped room of 9m x 5m x 16m in size, to propagate the same signal a total of 24 000 nodes are required, whilst with a 500 Hz bandwidth only approximately 600 nodes are required. This demonstrates that the DWM becomes exponentially more computationally expensive as the room dimensions and complexity increase. Thus, DWMs are best implemented using parallel computing or General Purpose Graphical Processing Unit (GPGPU) computing methods to improve rendering time, as found by Savioja (2010) and Thomas (2017). Savioja (2010) in particular managed to render room acoustics with up to 7 kHz bandwidth in real-time using GPGPU computing methods, albeit for simple room geometry such as the basic ’shoe-box’ shape.

3.3.2 Scattering delay network

De Sena, Hacιhabiboğlu, Cvetković, and Smith (2015) propose the Scattering Delay Network (SDN) for efficient modelling of room acoustics. The SDN was inspired by the DWM and FDN reverberator types, where De Sena et al. (2015) observed that the FDN is in fact related the DWM in that the structure of an FDN can be viewed as a type of waveguide. It can also be viewed as an advancement of the earlier waveguide methods initially proposed by Smith (1987). An SDN simplifies the DWM using an algorithmic-like method, and uses one scattering node per surface of a virtual room. Like the DWM, each node is connected to each other using bi-directional delay lines, simulating sound waves reflecting off and propagating from each surface to neighbouring surfaces. Virtual sources and microphones are connected to the scattering nodes, and the direct path from source to microphone is simulated using a single delay line. The SDN is visualised in Figure 3.11.

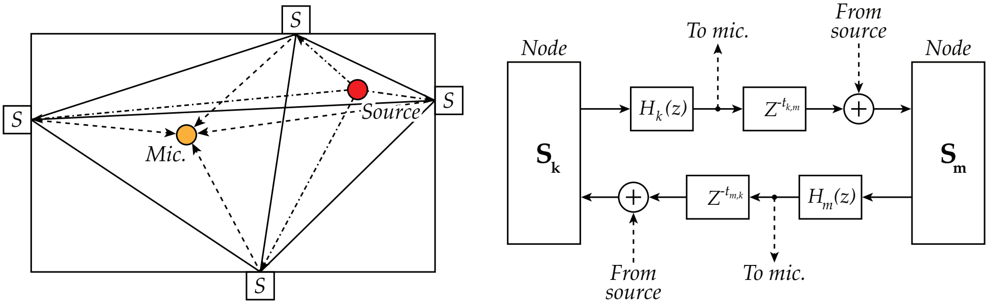


Fig. 3.11: **Left:** 2-D visualisation of the SDN reverberator simulating a simple shoebox shaped room. Solid lines represent connections between scattering nodes, marked S, whilst dashed lines represent interconnections between the nodes and source or microphone. **Right:** Algorithmic representation of an SDN showing the connections between two nodes (labelled S), the source and microphone. After De Sena et al. (2015).

De Sena et al. (2015) note that the main caveat with an SDN is that, whilst it can accurately model first-order reflections, it will only be able to approximate higher orders since the SDN is limited to calculating the paths for first-order reflections only. This introduces temporal errors in reflections of second-order and above, and is demonstrated in Figure 3.12. Furthermore, Like an algorithmic reverberator, it does not produce meta-data about each reflection such as the direction of arrival. However, since the inter-node delay times are calculated using a simple geometric method, it may be possible to extrapolate meta-data from them.

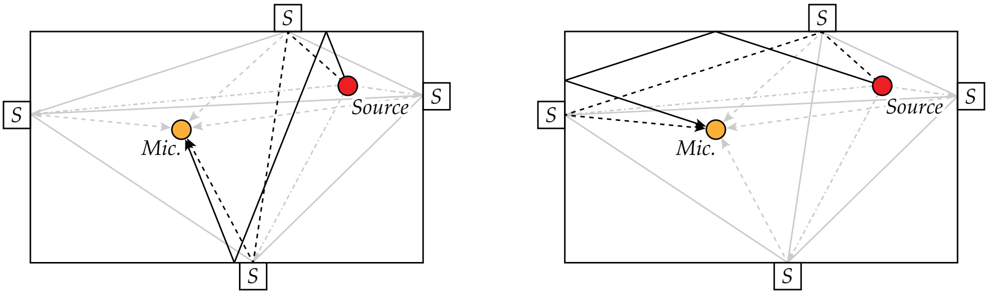


Fig. 3.12: Examples of second-order reflections rendered using an SDN. The solid black lines represent the actual reflection paths, whilst the dashed lines represent the paths approximated by the SDN. After De Sena et al. (2015).

3.4 Hybrid Methods

After much deliberation, it can be seen that each discussed method is not without limitation. Hybrid methods combine the best features of two or more algorithms in order to solve the limitations of each other. The methods can be categorised into three common types: Multi-geometric, Geometric-Algorithmic, and Geometric-DWM.

3.4.1 Multi-geometric

Multi-geometric approaches combine two geometric methods to overcome the accuracy limitations of each type. For example, as mentioned earlier, ISM cannot model diffusion or effectively model the late reverberant tail, yet ray tracing can. Vorländer (1989) implements this by combining the ISM with ray tracing, where the ISM modelled early reflections whilst ray tracing modelled the late reflections. However, Vorländer (1989) also used the ray tracing to optimise the ISM algorithm itself. The ray tracer calculates the reflection sequence for a few rays cast in random directions, discarding any duplicates. Using these now pre-determined sequences, the Image Expansion process can now efficiently calculate the exact reflection paths without requiring any extra tests to determine the visibility or validity of image sources. This of course prevents the explosion in the number of sources. The ray tracer then calculates the remaining late reflections, and the resulting RIR can then be used for auralisation. Naylor (1993) implemented a similar method to Vorländer (1989) for the commercial virtual acoustics package ODEON. This program also uses a ray tracing method to pre-determine possible reflection sequences such that the ISM can then be used to accurately model early reflections without the overhead produced by the basic implementation of the ISM, whilst the ray tracing method continues to model the late portion of the RIR.

3.4.2 Geometric-Algorithmic

Geometric-Algorithmic types combine a geometric method along with an algorithmic reverberator. This is usually to achieve accurate modelling of the early reflections and general size properties of a room model. The earliest geometric-algorithmic hybrid approach was proposed by Schroeder (1970) who modelled the early reflections using a ray tracing algorithm, and then modelled the late reflections using the Schroeder (1961) algorithmic reverberator. Moorer (1979) implements a similar technique, however, the early reflections are instead rendered with the ISM, and the algorithmic portion simulates the diffuse tail with an *improved* version of the Schroeder (1961) algorithm.

Wendt et al. (2014) combined an optimised and simplified ISM implementation with a FDN reverb (Jot & Chaigne, 1991) for use in a real-time simulation, where their method considers an empty, six sided, shoe-box shaped room. Because it is a simple geometric shape where all surfaces are visible to each other, the occlusion tests that are usually performed to valid an image path can be omitted. Further to this, the explosion in the number of images can be controlled to the point where *all* image sources and their associated paths to the receiver are valid. This is achieved by ensuring that the image sources are not reflected into duplicate nodes which would otherwise create an invalid image path, thus creating a symmetric pattern of image sources. With this method, the number of rendered sources is now given by:

$$S = \frac{4}{3}N^3 + 2N^2 + \frac{8}{3}N + 1 \tag{3.16}$$

Using this simplified implementation of the ISM, a reflection order of 10 would result in 3360 image sources, which is over 4000 times less than would be calculated using an ISM designed for arbitrary models. This simplification of the algorithm allows for quick rendering of shoe-box shaped models.

3.4.3 Geometric-DWM

Although the DWM offers the most accurate approach, as discussed earlier it is more ideal for modelling low frequency sound propagation as to keep the computational cost down, thus it can be combined with geometric methods that can render high frequency propagation. Thomas (2017) developed a hybrid algorithm that combined the DWM, the ISM and ray tracing to model both low and high frequency propagation. Thomas (2017) aimed to create an efficient and perceptually plausible algorithm.

3.5 Perceptually controllable reverb

Jot (1997) proposed a reverberation processor known as 'Spatialisateur', or Spat, that had the possibility for perceptual control over reverb parameters. Spat simulated reverberation using a similar approach to Moorer (1979) using a multi-tap delay line for early reflection simulation, however, these were fed into an FDN reverberator. Whilst real-time convolution was possible at the time, Jot (1997) opted to synthesize the early reflections with a multi-tap delay line. The early reflections were calculated using a random process and applied to the multi-tap delay line. The outputs of the delay line were then distributed between multiple output channels, as well as summed and fed into an algorithmic, FDN reverberator. Spat allows for separate, dynamic control of the early reflections and late reverberation portions, thus making it possible to control the reverb using perceptual parameters.

Jot (1999) asserts that a perceptual control method "leads to a more intuitive and effective user interface", as opposed to a physical approach. Jot (1999) defines the physical control approach as a method of the defining reverberation in terms of room dimensions, geometry, materials and source or listener positions, much like a geometric reverberation method. To Jot (1999) the physical control approach was seen as inferior, asserting that changes such as envelopment and reverberance will change with listener position, and thus are not easily predictable. Plus, with the available computer hardware at the time of writing, the physical approach would have been difficult to achieve, whereas Spat was a more efficient reverberation processor.

Jot (1999) proposes that the perceptual control method would allow perceptual attributes to be changed in a much more predictable manner, and so discusses the application of Spat for the perceptual control of the acoustical quality of a room. Here, high level controls were provided to the user to manipulate three different groups of attributes:

- Source perception: *Source presence*, *Brilliance* and *Warmth*
- Source/room interaction: *Envelopment* and *Room presence*
- Room perception: *Late reverberance*, *Heaviness* and *Liveness*

This is achieved by manipulating the energy and spectrum of different temporal groups of reflections. In the context of VR experiences such as video games that use arbitrary scene geometry data, the method proposed by Jot (1999) is perhaps unsuitable for modelling the exact behaviour of the acoustics for a given scene, yet rather produces a general sense of reverberation. Another limitation of the method is that the directional effects of particular early reflections, such as lateral reflections, are not taken into account, thus the level adjustments that the proposed control method applies can be seen as coarse and generalised. Whilst modelling a particular space would be seen as a physical control method, in theory both methods could be combined such that a geometric simulation method could model the exact RIR for a given source and listener position, and the perceptual control algorithm would post-process the RIR to allow for control over the above listed attributes. The benefit of such a system would be that the exact reflection times can be accurately modelled, yet still be perceptually controlled and optimised.

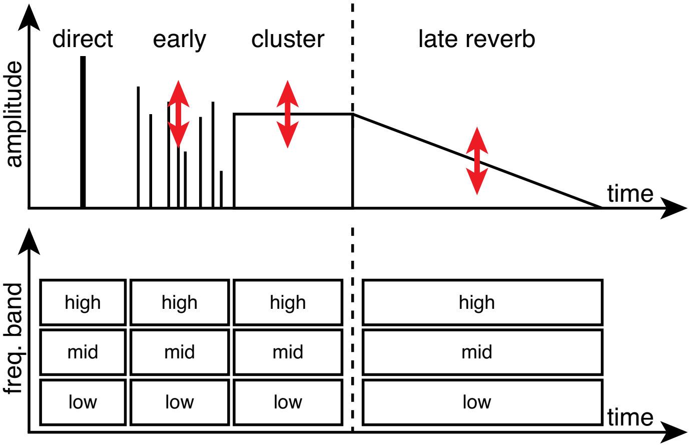


Fig. 3.13: Spatialisateur reverberator with perceptual control. The level and spectrum of an RIR is controlled in segments, after Carpentier et al. (2014).

Carpentier et al. (2014) adapted the implementation of the Jot (1999) model to create another hybrid reverberator. Their system would realise the perceptual control using a convolution reverb instead of a multi-tap delay to recreate early reflections from a measured RIR, whilst still using an FDN module for creating the late reverberation tail. The energy decay of the RIR was analysed and applied to the FDN so that the late reverberation matched that of the original RIR. The motivation behind the algorithm was to create a more efficient reverberation method, along with having perceptual control over the effect. Carpentier et al. (2014) give an example of the perceptual control of the *DirE* parameter of measured RIRs. Briefly, Carpentier et al. (2014) describes *DirE* as "the energy of the *temporally extended* direct sound energy...", controlling the intelligibility or 'presence' of the auditory source. It is given by"

$$DirE = E_{R0} + E_{R1} + E_{R2,excess} + 0.18 \cdot E_{R2,masked} \tag{3.17}$$

where E_{R0} is the estimated energy of the direct sound between 0 to 20ms, E_{R1} is the energy of early reflections arriving between 20 to 40ms, and E_{R2} is the energy of early-late cluster of reflections arriving between 40 to 100ms (Carpentier et al., 2014). The 'presence' is controlled by adjusting the gain of the reflections in each E time segment. In operation, the RIR is split temporally into the three E segments. The input signal is convolved with each segment, filtered through a three band spectral filter, then summed to form the new early reflections. The signal is mixed with FDN modelled late reverberation to form the final, perceptually controlled signal. Whilst this method arranges the RIR into temporal segments to achieve perceptual control over groups of reflections, again like the method proposed by Jot (1999), it does not take into account the direction of arrival of early reflections. Thus, the method is still limited to controlling all reflections within certain time segments regardless of their direction. Direction of arrival is seen to have a large contribution to sub-attributes of SI where ASW, for example, is greatly influenced by lateral reflections (Barron & Marshall, 1981).

As an alternative to the method proposed by Jot (1999), Pellegrini (2002) developed a perceptually motivated parametric, physical model to design virtual environments and control several attributes including source distance and room size. This was achieved by manipulating the level of single reflections that were associated with the perception of those attributes. The system proposed by Pellegrini (2002) was limited to using a rectangular 'shoe-box' shape such that the direction of the first order early reflections with a process equivalent to first order ISM. However, it could be adapted to allow for any arbitrary shaped room, and it demonstrates that perceptual controls can be applied to physical methods.

Rafii and Pardo (2009) proposed an intelligent control system for the algorithmic reverberator developed by Moorer (1979) that allows users to control the reverb using subjective perceptual descriptors. These descriptors included 'bright', 'clear', 'boomy', 'bathroom like' and 'church-like'. Up to 1024 IRs of the reverberator at a wide range of settings were measured to cover the range of the following parameters: RT60, Echo Density, Clarity, Central Time and Spectral Centroid. To create a perceptual model, using experimental data from a series of listening tests performed by subjects, the system was trained to what extent a change in each measure affected a given perceptual descriptor. In this case, the system developed by Rafii and Pardo (2009) was trained to find the strongest correlations between measures and each descriptor using linear regression, such that a control interface using those descriptors could then be created and used as an input to the AI system. This control interface was evaluated in terms of its effectiveness to manipulate each descriptor using human ratings. Rafii and Pardo (2009) found that the system was most effective and consistent at affecting clarity whilst poor at controlling more abstract terms such as 'bathroom like'. This can be understood to be a limitation in what is agreed to be a ’bathroom like’ quality in reverberation. Rafii and Pardo (2009) acknowledge that this is preliminary study and will require much deeper psychoacoustic research to improve the system. However, the presented system is limited to controlling the Moorer (1979) algorithm and its parameters, and does not analyse the effect of individual groups of reflections on spatial impression, which can only be achieved with a geometric reverberator.

3.5.1 Perceptual optimisation

What all the above systems achieve is a method of perceptual control using high level controls, yet none of them attempt to intelligently optimise and improve the acoustical quality behaviour, just simply to control it. The above systems leave the user to modify perceptual attributes to their needs. This is perhaps suitable for applications such as creative music production and video games (Jot & Trivi, 2006), where the individual perceptual parameters are independently controlled by the user and not by an automated system. However, the methods may not always achieve pleasant and subjectively preferable acoustical properties. The hybrid methods proposed by Jot (1999) and Carpentier et al. (2014) do not analyse the initial subjective properties of the acoustics, and only provide high level user control.

It was noted that the perceptual control methods manipulate the energy of entire groups of reflections, rather than focusing on individual reflections in a *microscopic* or *granular* fashion, which can be highly useful for controlling timbre when certain reflection patterns can create unpleasant tonal colouration (Halmrast, 2000). Furthermore, they do not take into account the effects of directionality of rays, which is known to have effects on the spatial impression (see Chapter 2). If direction was taken into account, then directional dependent perceptual control methods could be achieved. Whilst some directionality was taken into account by Pellegrini (2002), it only serves to pan the source and simulate source movement. From an optimisation standpoint, having finer, granular analysis and perceptual control methods than simple level and spectral adjustment could help towards the development of a system that could intelligently improve the acoustical qualities of the simulated room acoustics from a preference perspective.

3.6 Summary

This chapter reviewed the current and popular methods for achieving virtual room acoustics, investigating the implementation of each method, their advantages and disadvantages, and their suitability for this study.

Section 3.1 reviewed several algorithmic methods of digital, artificial reverberation that use a network, or algorithm, of comb-filters and all-pass filters to simulate reverb. For the artificial reverb to be plausible, the algorithm must increase the echo density of a signal to at least 1000 echoes per second. Several algorithms were investigated, and it was found that each algorithm was optimised for a particular room, be it a concert hall or an auditorium. Whilst these algorithms are capable of producing pleasing reverb suitable for music production, they alone cannot model the exact nature of a particular room model, and thus have been coupled with another algorithm for hybrid modelling, for example by Carpentier et al. (2014). They are ultimately unsuitable for this study as they do not accurately model and give access to individual reflections, or any kind of multi-dimensional metadata such as the reflection's direction of arrival, exact delay time or octave-band energy.

Section 3.3 discussed the Digital Waveguide Mesh (DWM), which is a wave based modelling method. The DWM models the wave behaviour of sound in a room model by decomposing the space into a three-dimensional mesh of equispaced elements. The spacing between adjacent elements determines the maximum frequency of the wave that can be modelled. The main advantage of the DWM, and wave-based methods in general, over geometric methods is their ability to accurately model wave phenomena such as edge diffraction, and the interaction between reflection wave fronts. They are, however, computationally expensive and are currently not ideal for real-time rendering of the entire spectrum of an RIR. Like algorithmic methods, they are also not able to give direct access to each reflection or any multi-dimensional meta-data about them.

Section 3.2 looked at geometric methods, namely ray tracing, the image source method (ISM), and beam tracing. Ray tracing renders and impulse response by firing individual particles in all directions around a space to sample each possible sound path. Whilst not ideal for accurate modelling of early reflections, it is suitable for modelling diffusion and scattering. The ISM on the other hand takes a deterministic approach to rendering, and is able to accurately model low order early reflections, although its efficiency is quickly lost when rendering higher order reflections. Beam tracing is an adaptation of the ISM and uses geometric beams to pre-calculate all the possible sound paths, regardless of the microphone positing, thus allowing for real time rendering. Whilst no geometric method is able to properly model wave behaviour, they are able to compute the exact paths that sound takes as it propagates through the model. The main advantage of geometric methods over the other discussed methods is their ability to store meta-data about each individual reflection.

Finally, an overview of the current methods of perceptual control of reverberation was given in Section 3.5. Whilst the current methods propose control over perceptual attributes such as *source presence* (Carpentier et al., 2014), localisation and room size (Pellegrini, 2002), or even provide intelligent control over reverb parameters (Rafii & Pardo, 2009), they do not analyse and perceptually optimise the subjective quality of the reverb, and thus leave it to the user to adjust the reverb. They also do not take into account the directional effects that reflections may have on perceptual attributes, namely spatial impression. This would be beneficial as much more precise control and optimisation could be applied whilst potentially not affecting other attributes for scenarios such as in virtual concert hall listening, where negative aspects such as unpleasant colouration could be suppressed without affecting the spatial impression. Further to this, except for Pellegrini (2002), none of these reviewed methods apply perceptual control to geometric methods, and only to algorithmic reverberation or pre-measured RIRs. Whilst Jot (1999) argues that perceptual control over an algorithmic reverberator is more efficient and can be perceptually plausible, the application is limited to that particular type of reverberator. The benefits of using a geometric or a hybrid method today is the increased accuracy of modelling particular room models or virtual reality scene data in, which today is becoming easier to achieve in real-time thanks to modern, high performance computer hardware (Savioja, 2010). Such an algorithm can render any given scene to a certain degree such that the resulting reverberation would perceptually match what a subject is able to see. The reflection meta-data that can be produced from geometric methods could then be used for real-time perceptual optimisation and refinement of the acoustics in the scene by manipulating aspects of reflections related to the perceptual attributes discussed in Chapter 2, namely Apparent Source Width, Listener Envelopment and Tonal Colouration.

After reviewing the three main approaches, it was decided that a hybrid geometric method was the most suitable for this study. However, at the time of development it was found that no geometric method was available that could export individual reflections to allow for analysis, perceptual optimisation and spatialisation.

Chapter 4

Custom virtual acoustics algorithm

This chapter discusses a custom virtual acoustics algorithm that was developed for the purposes of investigating the possibilities of perceptual control and optimisation of virtual room acoustics[1]. For this to be achieved, the requirements and needs must first be considered. To allow for perceptual control, the virtual acoustics algorithm must fulfil the following criteria:

1. Give access to all captured reflections.

2. Store meta-data about each captured reflection including the direction of arrival.

3. Export the captured reflections and their meta-data in a raw format.

4. Model decay due to surface and air absorption in octave bands using existing absorption coefficients, such as those given by Vorländer (1989).

First consider the algorithmic reverberators discussed in Section 3.1. They able to

[1] The source code for this program is available from `https://github.com/ValleyAudio/homr`

satisfy point 2, yet are unable to satisfy the remaining points because they are unable model the individual sound paths or accurately material absorption, but merely replicate the general behaviour of room reverberance. The DWM can also satisfy point 2, yet it also cannot meet the remaining requirements, namely points 3 and 4 which require the direction of arrival of the captured reflections to be resolved, as well as a 'granular' representation of the reflections where each individual sound ray can be captured and exported independently. Geometric reverberators are the most suitable for the task as they are able to meet all three requirements. At the time of development, there was no software package or framework available that would also meet points 1, 3 and 4, and so an algorithm was programmed from scratch. As the reverberator is aimed to eventually be used for rendering of virtual reality scenes, the algorithm's design should be focused on perceptual plausibility rather than complete accuracy.

For the sake of simplicity and to meet time-scale requirements, a custom multi-geometric algorithm, referred to as the CA from here on, that combines ray tracing and the ISM was considered to be the best possible solution. Whilst ray traced rendering of an entire room impulse response is currently too intensive for current computer hardware to achieve in real-time, it meets all the above requirements for this study.

4.1 Specification

The CA developed for the study was created using the C++ programming language. It is optimised for vector mathematical operations using the Vector Class Library (VCL) developed by Fog (2017), and features:

- Hybrid rendering using the ISM for early reflections and ray tracing for late reflections.
- Choice of either Fibonnaci Lattice or Monte Carlo emission distributions for the ray tracing portion.
- Rendering of arbitrary models that use triangular polygon meshes.
- Modelling of surface and air absorption in octave bands, defined using existing absorption coefficients, such as those provided by Vorländer (2007).
- Virtual microphones and sources with respective polar pattern and directivity simulation.
- Multi-channel rendering of a RIR as a WAV file.
- Exporting of captured rays in raw format called the RIV (Raw Impulse Vector).

The RIV stores each captured ray sequentially with the following meta-data:

- Distance travelled in metres.
- Direction of arrival vector.
- Energy per octave band.
- Reflection order.
- Reflection type (specular or diffused).

The benefits of the RIV format versus an ordinary digital audio file formatted RIR is that it provides greater access, and thus greater control over each ray that represents the RIR for a given position. The RIV allows for post processing to be applied to each independent reflection. This means that the perceived reverb could be controlled, altered and, most importantly, perceptually optimised without modification of the room model or the source and receiver positions.

4.1.1 Rendering pipeline and spatialisation

The rendering pipeline is the order of operations that the program takes to render an RIR. An overall view of the pipeline is presented in Figure 4.1. The model is loaded into the program along with a configuration file that informs the program of the positions of the source(s) and receiver(s), what reflection order the ISM is limited to, the number of rays to use during ray tracing, and the sample rate of the output WAV file.

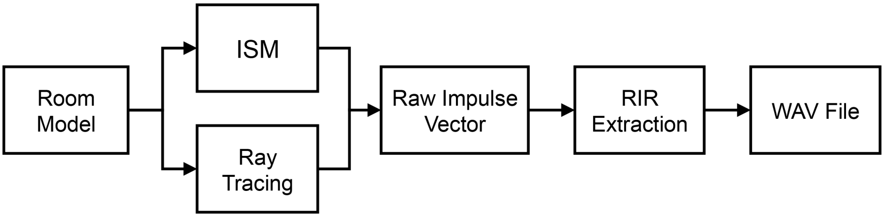


Fig. 4.1: Rendering pipeline of the custom algorithm

The RIV that is exported prior to 'RIR Extraction' can be converted to a Binaural RIR (BRIR) through a spatialisation method based upon a combination of techniques proposed by Savioja et al. (1999) and Schröder (2011). The method used here was programmed in MATLAB. Figure 4.2 is a flow chart of the method.

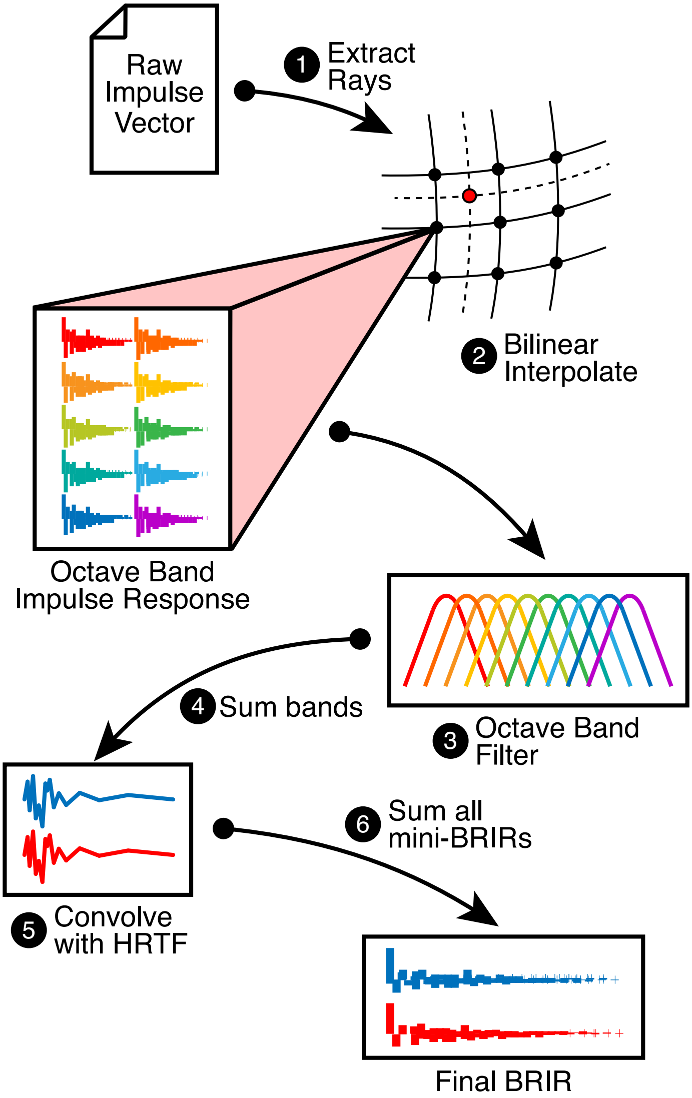


Fig. 4.2: Flowchart of the spatialisation process.

The HRTFs from the MIT KEMAR database Gardner and Martin (1995) are stored in a grid along with a spherical map of their azimuth and elevation angles. The map is composed of quadrilateral patches, where each corner of a patch is contains a HRTF location. The rays are extracted from the RIV and, by using the spherical map, are sorted into the patches using their direction angles. By applying bilinear interpolation, the octave band energy of the now sorted rays is distributed to each corner point of the patch to form directional, octave-band impulse responses. The bilinear interpolation scheme is visualised in Figure 4.3, and is given as follows:

$$\begin{aligned} p_a &= (1 - c_\phi)(1 - c_\theta) \\ p_b &= (1 - c_\phi)c_\theta \\ p_c &= c_\phi(1 - c_\theta) \\ p_d &= c_\phi c_\theta \end{aligned} \tag{4.1}$$

where p_a to p_d are the four corner points of the patch, and c_θ and c_ϕ are the azimuth and elevation interpolation coefficients. The coefficients are obtained from

$$\begin{aligned} c_\theta &= \frac{\theta \bmod \theta_{grid}}{\theta_{grid}} \\ c_\phi &= \frac{\phi \bmod \phi_{grid}}{\phi_{grid}} \end{aligned} \tag{4.2}$$

where θ and ϕ are the ray's azimuth and elevation angles, and θ_{grid} and ϕ_{grid} are the spacing angles between each point. After distribution, the octave-band of each impulse response is filtered with a corresponding bandpass filter. The filtered bands are then summed together and convolved with the corresponding HRTF to form a mini-BRIR. Finally, these mini-BRIRs are summed to form the final BRIR.

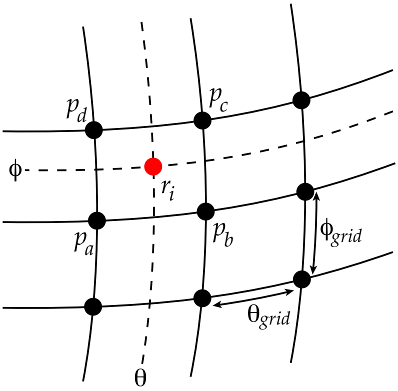


Fig. 4.3: The energy of the ray r_i at a given azimuth θ and elevation ϕ location is distributed between the four points p_a, p_b, p_c and p_d. After Savioja et al. (1999).

4.2 Comparison with ODEON

To verify that the CA was able to render an RIR with a considerable degree of accuracy, it was deemed suitable to compare the output to an existing geometric algorithm that uses the same methods as the CA, in this case ODEON 14.0 . The latter software package has been previously evaluated in a round robin study performed by Bork (2000), where several virtual acoustics programs were compared against each other both objectively and subjectively. The CA will be compared to ODEON using a similar approach used by Bork (2000). The CA is similar to ODEON as it also uses a hybrid ISM / Ray tracing approach, thus it technically has the same capabilities. For this comparison, both ODEON and the CA render BRIRs of a room model. The BRIRs will be compared in terms of the timing and amplitude of reflections, and of the decay time curves.

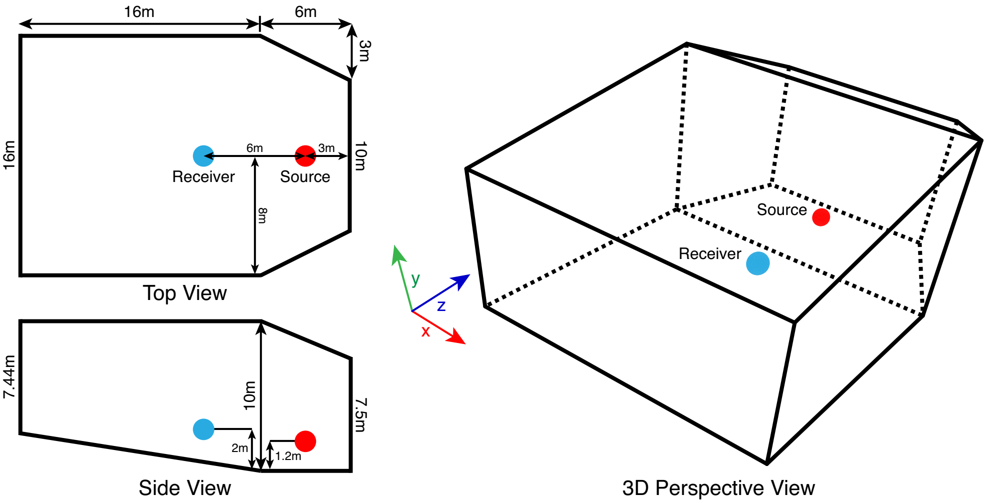


Fig. 4.4: Left: Plan views and **Right:** 3D perspective of the ODEON example room model, including source and receiver positions.

4.2.1 Model and rendering setup

The 'Example' room model that is available with the ODEON software was used for the comparison as it is exhibits a simple concert hall shape, and was converted to also work with the CA. It was ensured that each surface was of the same dimensions and used the same absorption coefficients. Figure 4.4 shows a plan and 3D view of the model. An omnidirectional source was placed centrally in the room 3m away from the stage wall and 1.2m from the floor, whilst the receiver was positioned 6m away from the source and 2m from the stage floor, and so there is a height difference of 0.8m between the two points, see Figure 4.4. To test the abilities of both programs at rendering an impulse with high precision, the 'Precision Rendering' mode of ODEON was chosen, which by default uses 160 000 rays. Therefore, the CA was set to render a response using third order ISM and ray trace using the same number of rays. Third order was chosen to ensure accurate rendering of early reflections within 80 ms when considering the distance and highest number of reflections a sound path will traverse within that time frame. It was not possible to create an impulse response with ODEON that was not binaural, therefore the spatialisation process was applied to the output RIV from the CA to create a BRIR.

4.2.2 Analysis of the resulting BRIRs

Figure 4.5 shows the average frequency spectrum of each algorithm. At a glance the spectra appear to be similar in that they both show a steady slope in reducing magnitude as frequency increases up to 10 kHz, where after that there is a prominent peak in both plots.

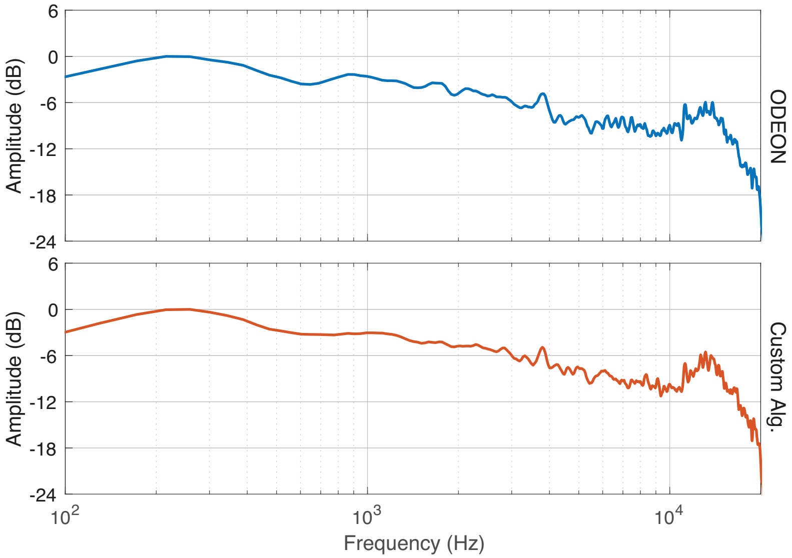


Fig. 4.5: Spectral plots of the BRIRs rendered by each algorithm, where the top was produced by ODEON, and the bottom by the CA.

To analyse the spectral differences between the to two BRIRs, the delta spectrum was obtained by subtracting the spectrum of one BRIR from the other, and is shown in Figure 4.6. It can be seen that the difference in frequency response is fairly small, where the largest difference is approximately 1.7 dB at 16 kHz. This suggests that it is unlikely that there is any noticeable difference in frequency response. As for why there are differences, they are most likely to be caused by spectral differences between the HRIR databases used during spatialisation in each algorithm, although it is not possible to directly confirm this.

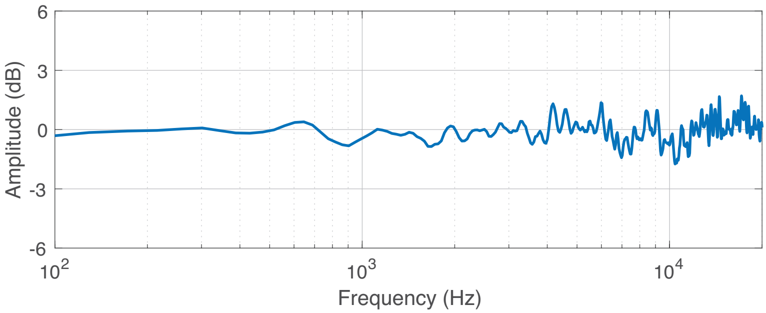


Fig. 4.6: Delta spectrum between the BRIRs produced by each algorithm.

Figures 4.7 and 4.8 are the waveforms of the BRIRs generated by the two programs. On initial observation of the first 80ms, it can be seen that the peaks of the reflections in the CA BRIR are 'sharper' and more defined than in the ODEON BRIR. This is possibly due to the different HRTF databases used during the spatialisation processes when producing the BRIRs. The first feature to verify is the pattern and decay of the early reflections, or first 80ms of the BRIRs. It can be seen that onset time of the direct sound peak is identical in both BRIRs. The next reflections that that were measured between 5 to 20ms are most likely the initial floor, ceiling and side wall reflections. The timing of these in the CA BRIR appear to match those in ODEON BRIR. Beyond 20ms it becomes increasingly difficult to determine if the reflection patterns are still matching, although prominent peaks between 30 to 60ms in the ODEON BRIR appear to be present in the CA BRIR.

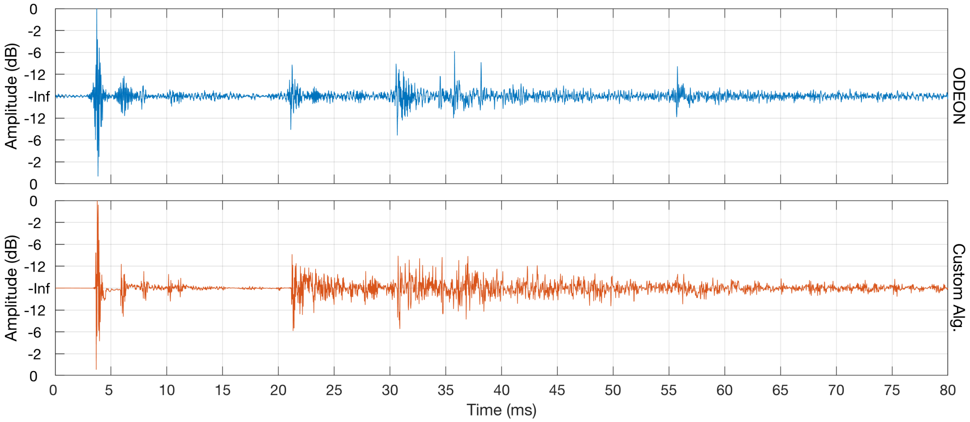


Fig. 4.7: Plot of the first 80ms of the BRIRs rendered by each algorithm, where the top was produced by ODEON, and the bottom by the CA.

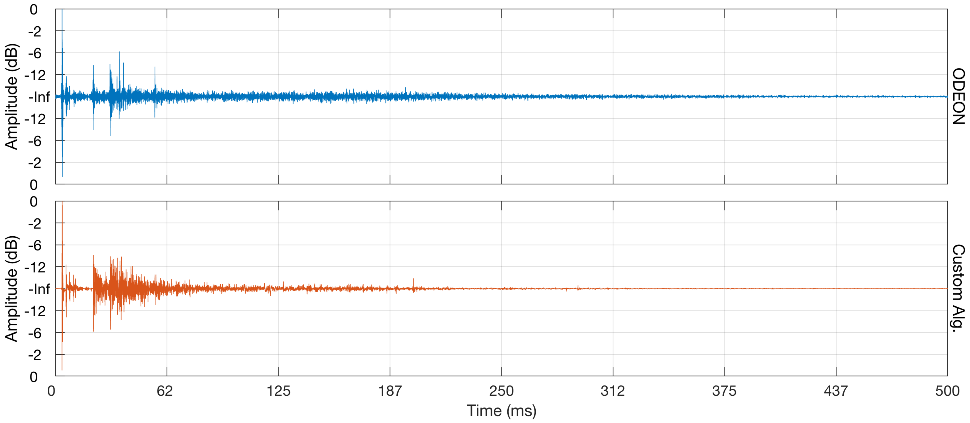


Fig. 4.8: Plot of the first 500ms of the BRIRs rendered by each algorithm, where the top was produced by ODEON, and the bottom by the CA.

However, there appears to be much denser amount reflection energy present in the CA BRIR in this time period. This is likely due to a potential difference in how the diffuse energy is attenuated in the CA, or by how this energy is accumulated during the spatialisation process discussed earlier in Section 4.1.1. The overall amplitude of the prominent and most visible peaks in the CA BRIR appear to be similar to the what is seen in the ODEON BRIR.

The first 500ms in Figure 4.8 will now be observed. Like what was seen in Figure 4.7, within the first 60ms there appears to be a larger amount reflection energy in the CA BRIR than in the ODEON. Again, it is speculated that this is due to how the diffuse energy is attenuated or accumulated. However, this early energy appears to quickly decay before 62ms. Post 62ms, the reverberant energy in the CA BRIR seems to decay faster than in the ODEON BRIR. Again, it is speculated that this is possibly due to potential differences in attenuation and absorption methods used in both programs.

To objectively determine the differences between the two algorithms, the following acoustical measures were calculated using the 'Institute of Sound Recording' (2017) MATLAB toolbox.

Parameter	CA	ODEON
$\mathrm{RT}_{60,mid}$ (s)	2.49	3.01
Early Decay Time (s)	0.71	1.78
[1-IACC_{E3}]	0.46	0.37

Table 4.1: Acoustical objective parameters measured for each algorithm.

The first noticeable difference between the two programs is that the $\mathrm{RT}_{60,mid}$ and Early Decay Time values measured from the ODEON BRIR are longer than in the CA BRIR. This verifies the observation made in Figure 4.8 where the early energy up to 80ms appears to decay quicker in the CA BRIR. This is potentially caused by a difference in how the two algorithms model energy decay due to air and surface absorption. To recap, the air absorption modelling in the CA is implemented using the method discussed by Kuttruff (2016), however, it is unclear as to how it is implemented in ODEON when studying the related literature. The predicted ASW of the CA BRIR is significantly higher than of the ODEON BRIR, suggesting that the auditory image would appear to sound wider when rendered with the CA.

4.2.3 Discussion

Whilst the CA implemented methods described in the existing literature, and in particular the literature related to ODEON such as Naylor (1993) and Rindel (2000), it is possible that these methods have been further refined and developed over time. Since the source code for ODEON is 'closed' to the public and is proprietary, it is not possible to fully determine how the methods are exactly implemented in the version of ODEON used for this comparison. However, it is important to recall here that the aim of developing the CA was not to fully replicate ODEON. The intention was to create a virtual acoustics algorithm that provided features that were not available in other programs at the time of development, with the most important feature being the ability to access all captured reflections during and post rendering, as mentioned earlier in this chapter. Finally, the two programs have different use applications. The custom algorithm is intended for music production and virtual reality, whilst ODEON is intended as an architectural measurement tool, and therefore must provide the most accurate predictions. Again, whilst one could consider total accuracy to be of utmost importance, the most accurate artificial reverberator may in fact be most unpleasant sounding, suffering from potentially poor spatial impression and distracting tonal colouration. The direct access to capture reflections provided by the CA opens up the possibility for perceptual control and optimisation that could improve the subjective qualities of the initially unpleasant reverberation.

4.3 Summary

In this chapter, a custom reverberation algorithm was developed in C++ to fulfil the requirements of this study, which was to give unprecedented access to each individually captured reflection so that perceptual control methods can be applied. The algorithm combines the ISM and ray tracing, is able to model arbitrary models, and can export the captured reflections as a monolithic file known as a 'Raw Impulse Vector' (RIV). The chapter also discusses a spatialisation process that is used to convert a RIV into a BRIR. The algorithm was compared to an existing commercial package, ODEON 14.0, through analysis of BRIRs of a simple concert hall generated by both programs. It was found that:

- The custom algorithm could accurately render the timing of early reflections up to 80ms.

- The initial early decay time was faster in the custom algorithm BRIR than in the ODEON BRIR, and the -60 dB reverb time (RT_{60}) was found to be 504ms quicker in the CA BRIR. This is most likely due to a difference in how absorption is modelled in both programs, though this is unclear and difficult to verify.

- The ASW, measured using [1-$IACC_{E3}$], was found to be higher in the CA BRIR than in the ODEON BRIR.

Whilst there is a difference between the two algorithms, the custom algorithm was not developed to be a complete recreation of ODEON, under pilot testing the custom algorithm appears to generate natural, plausible room reverb. The algorithm was also not intended model room acoustics with total accuracy, albeit it is able to model early reflections with a good level of precision; it was primarily designed to provide complete access to the reflections such that perceptual control methods can be applied to geometrical virtual acoustics.

Chapter 5

Saturation in the degree of perceived apparent source width

As discussed in Chapter 2, Barron and Marshall (1981) proposed using lateral fraction (L_f) as a measure of apparent source width (ASW). Their research found that the in the presence of a direct sound, ASW increases as the angle of an early reflection increases and becomes lateral. From this they derived the L_f measure as a function of reflection energy and angle, and further expanded the measure to be a ratio between early lateral reflections and all early reflections. The study did not directly look into the threshold point at which subjects can just notice differences in ASW between reflection angles(i.e. the just noticeable difference threshold), thus it makes the assumption that ASW continuously increases as the reflection angle increases from 0° to 90°, or decreases from 180° down to 90°. When interpreting the results produced by Barron and Marshall (1981) (see Fig. 5.1), it can be speculated that ASW may reach its maximum earlier than predicted. Overlap in the amount of perceived spatial impression for a reflection angle between 30° to 160° implies that there is no significant difference in the perceived amount between these two angles. It can be

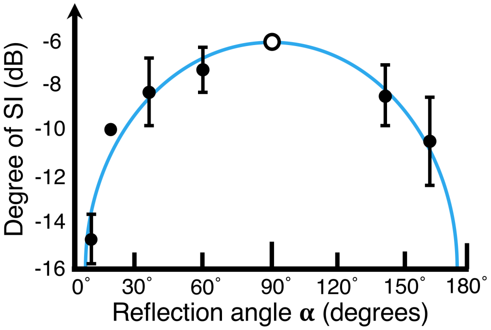


Fig. 5.1: Mean and 95% confidence intervals of subjective spatial impression versus reflection angle. The solid line is the predicted degree of SI (or ASW), after Barron and Marshall (1981).

hypothesised that subjects are likely to perceive maximum ASW when a reflection arrives between these angles. From this it may be possible to group reflections that arrive within this region as ones that are capable of producing maximum width. As L_f is a ratio of lateral reflection energy to all reflection energy, manipulating the energy of these would influence the amount of perceived ASW the most. In a music production situation, if an engineer wanted to crate a greater sense of spaciousness without affecting other attributes, they could increase the amount of energy of the lateral reflections without needing to modify the geometry of the space. In a VR scenario, if the source was difficult to localise, confusing or distracting the listener, the ASW be reduced so that the source image can be easily located. To the author's knowledge, no such study has been performed that directly attempts to look for a region of saturation, or use such a region to selectively manipulate reflection energy to affect spatial impression.

The IACC$_{E3}$ measure proposed by Hidaka et al. (1995) is also another important measure of ASW. Whilst it is not directly dependent on reflection direction, the inter-aural time-difference (ITD) increases with azimuth angle, thus decreasing the degree of IACC. Furthermore, time-varying fluctuations in the ITD have been found to also be related to the perception of ASW (Mason & Rumsey, 2001). This establishes that the direction and energy of a reflection play a vital role in the perception of ASW. Thus, it would be necessary to analyse these objective measures in relation to the hypothetical saturation region and to establish a connection so that a novel method of improving the spaciousness of a virtual concert hall can be developed.

This discussion raises the following research questions:

1. Between what horizontal angles does the hypothetical region of ASW saturation lie?
2. Is there evidence provided by the objective measures to support the existence of this region?

The experiment is split into two parts: part 1 describes a psychometric test that was performed over headphones and was designed to locate the boundaries of the region; and part 2 describes a multiple comparison verification test that was performed using an arc of loudspeakers to simulate reflections in an ITU-R BS.1116 (2015) compliant listening room.

5.1 Experiment part 1 : Finding the ASW saturation regions

5.1.1 Test methodology

Since the objective of the experiment is to find a threshold point, a threshold detection method should be used. Two types of methods were considered for this test: the 'Method of Adjustment' (MOA), and the 'Staircase Method'. MOA is an indirect scaling method where subjects have to manually adjust a parameter of a stimulus until they reach a point where it is perceived as being the same as the reference (Bech & Zacharov, 2007). The staircase method has the same objective as the MOA, however, the experimenter controls the parameter and response method, rather than the subject (Cornsweet, 1962).

5.1.1.1 Comparison of threshold detection methods

The most simple form of staircase test is the 'Yes/No' test, where subjects are asked if they can hear a difference between two stimuli, to which they respond 'Yes' or 'No' (Cornsweet, 1962). This is otherwise known as a Two-Alternative Forced-Choice, or 2AFC test. Depending on the response, the parameter is adjusted by a fixed step size in certain direction. A change in step direction is known as a *reversal*, and the sequence of values between reversals is known as a *run*. After a pre-determined number of reversals, the test is terminated, and the threshold value is calculated from the last number of runs, see Figure 5.2. García-Pérez (1998) recommends that for a reliable result, at least twenty reversals must be performed, and that the threshold value should be the mean of the last twelve runs.

The advantage of the staircase method is the standardised method of presentation, in which all subjects will be presented with the stimuli in the same fashion. Cornsweet (1962) also notes that in comparison to the MOA procedure, all staircase methods create a history of all of subject's responses, and thus it is clear to the experimenter as to how the subject was responding throughout the course of the test. Meanwhile, because MOA gives full control to the subject, Cornsweet (1962) notes that it is unclear as to how the subject settles onto the threshold value.

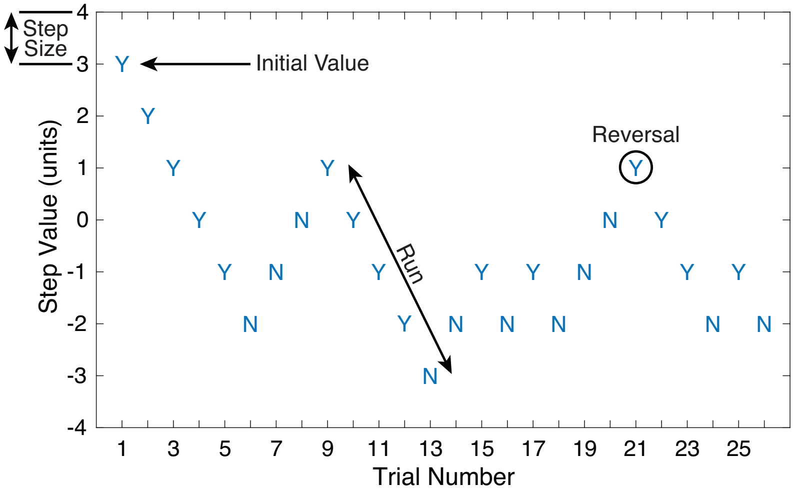


Fig. 5.2: Typical response data from a 'Yes/No' staircase test, after Levitt (1971).

However, a major drawback of the 'Yes/No' test in particular is that subjects can easily become aware of the stimuli order and the mechanism. This causes an anticipation bias where subjects are expecting a difference to be heard, which is not a major problem when at the beginning of the test where it is easy to distinguish the stimuli, however, the bias increases the amount of false responses when the differences become imperceptible. This problem was alleviated by Levitt (1971) who devised the 'Transformed Up/Down Method'. This approach reduces the amount of false responses by only reversing the step change after a certain number and pattern of responses are given. For example, rather than decreasing or increasing the step value after every single 'Yes' or 'No' response, Levitt (1971) proposed that it should only decrease after two 'Yes' responses as to doubly ensure that the subject did hear a difference. Levitt (1971) in fact proposed multiple response patterns with varying rates of false responses.

Another drawback of traditional staircase methods is the usage of a fixed step size.

Whilst a smaller step size would yield higher precision, it does not allow for rapid convergence onto the threshold value, especially if the initial step is far from the predicted threshold point. Levitt (1971) proposed that the step size should start fairly coarse, then reduce when the first reversal has been detected. The subject will converge towards the threshold region faster, thus improving the efficiency.

In contrast to the staircase method, in the MOA procedure the subject is tasked with adjusting the parameter of a stimulus until the difference between it and the reference is imperceptible. However, contrary to the staircase method, the subject has control over the parameter and response method. Cardozo (1965) notes that the main advantage of this procedure is that it is more engaging for the subject, thus enabling greater concentration. It also allows the subject to use their own stimulus order, which potentially can make the procedure more efficient than the staircase method.

There are some drawbacks to the MOA, however. Firstly, Stevens (1958), Cardozo (1965) and Ehrenstein and Ehrenstein (1999) note that because listeners are in control of the test rather than the experimenter, MOA allows for a less standard method of measurement between subjects, which can ordinarily lead to greater error in the result. Wallis and Lee (2017) also note that the use of the fixed step size hampers the efficiency and accuracy of the procedure. Therefore, they proposed an adaptive version of MOA called the ’Adaptive Method of Adjustment’, or AMOA. The AMOA uses a coarse initial step size in which subjects locate the approximate threshold point. The procedure then places the current step value one step above the threshold and reduces the step size. This process is repeated a further two times, with the final step value representing the threshold point. Whilst Wallis and Lee (2017) found that this method can reduce response error, they noted that this method has not been formally compared to established test methods, so its effectiveness is at the moment unknown.

5.1.1.2 Description of the chosen method

After considering the two most popular methods of threshold detection, and comparing their various pros and cons, a modified version of the Levitt (1971) 'Transformed Staircase' method was implemented in Max 7. This was chosen because it offers higher precision and a more reliable threshold point than the MOA procedure. The design implemented here uses a 2AFC paradigm with a 2-Down-1-Up rule. The test stimuli step position begins at either 0° or 180° depending on which angle region is being tested for. Levitt (1971), and later García-Pérez (1998) recommend using a large initial step size that is reduced after the first reversal such that there is an increased rate of convergence towards the threshold point, making the test procedure more efficient. Therefore, the initial step size in this test was 10°, and is then reduced to 5° after the first reversal to increase the efficiency in locating the threshold point. Following the 2-Down-1-Up rule, the step position would increase if the subject responded with "Yes-Yes", and would decrease if they respond with "Yes-No" or "No".

5.1.1.3 Stimuli

An anechoic sample of Danish male speech from the Bang and Olufson 'Music for Archimedes' project (Hansen & Munch, 1991) was used for this experiment. This sample has both transient and continuous characteristics as well as a constantly changing frequency spectrum. These behaviours create a more general sense of ASW rather than width perceived from a less complex source, or one that has a narrow spectrum. The speech sample was set to play and continuously loop in real-time, and was split into direct sound signal and reflection signal. The reflection angle θ was varied in 5° steps between two angles, either 0° and 90° to obtain the front boundary angle, and 90° and 180° to obtain the rear boundary angle, see Figure 5.3. The two signals were convolved via the 'multiconvolve~' object (Harker & Tremblay, 2012) for Max 7 with their corresponding diffuse-field compensated HRTF (head-related transfer function), the MIT KEMAR database (Gardner & Martin, 1995). The reflection signal was delayed by t milliseconds and attenuated by 6dB. The chosen values for t were 5ms, 10ms, 20ms and 30ms.

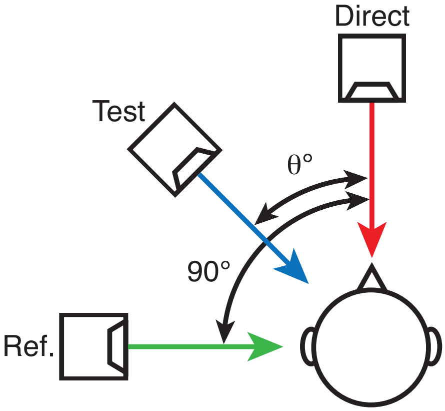
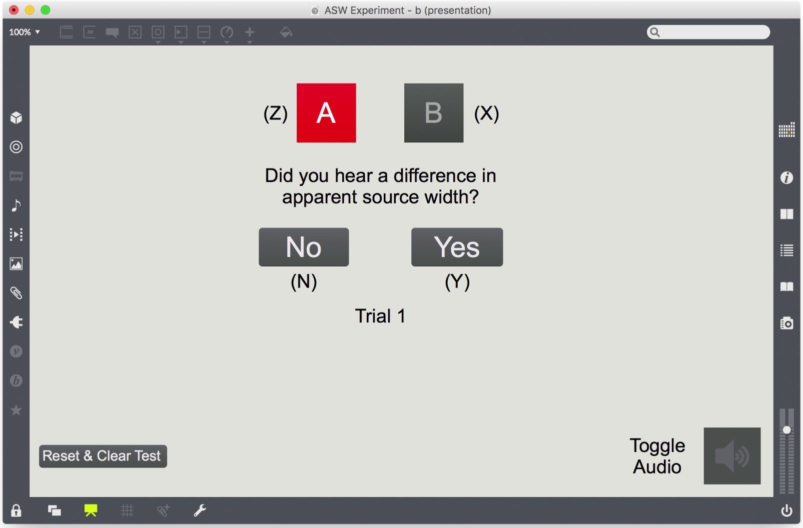


Fig. 5.3: Visual representation of the setup for the experiment. Virtual reflections are played to the listener over headphones, where they listen to a sound field using either the 'Test' or 'Reference' reflection in the presence of the 'Direct' sound.

5.1.1.4 Interface

Figure 5.4. For each trial the subjects were presented with two buttons labelled 'A' and 'B' that would freely toggle between which stimulus was being played. The subjects were instructed to listen carefully to stimuli A and B, and decide whether



Did you hear a difference in apparent source width?

Fig. 5.4: Screenshot of the staircase test interface used in the experiment. The subjects were unable to proceed onto the next trial unless they had listened to both stimuli. The reference and test button assignment is randomised on for each trial.

they can hear a difference in perceived ASW, although they can only give a response after both stimuli were played. The stimuli order was randomised for each trial and are played simultaneously, although only one is being heard at any point. The interface could be controlled using either a mouse or keyboard, where keys were mapped to both the stimuli buttons and response buttons. During both pilot and real testing, seven out of ten subjects verbally reported afterwards that the combination of keyboard control of the interface and non-sequential playback of stimuli made the test more engaging.

Prior to main testing, the subjects were presented with a practice interface during a training phase in order to familiarise them with the controls and the stimuli. This practice test did not contribute to the final results. The main test terminated after either 20 reversals or 128 trials. None of the subjects reached the 128 trial limit, and all completed each session in an average time of ten minutes.

5.1.1.5 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dBA by playing pink noise from one headphone cup, and measured using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

5.1.1.6 Subjects

For this test, a group of ten subjects consisting of music technology students and researchers at the University of Huddersfield participated. Five subjects had experience with spatial audio and critical listening, whilst the remaining subjects had mixed critical listening ability. It is recommended in ITU-R BS.1116 (2015) to employ experienced listeners that can discern fine perceptual differences, however due to a constraint in the number of experienced listeners available during the testing period, subjects with mixed listening ability had to be included in the experiment to increase the number of observations.

5.1.2 Statistical analysis of the test results

The average reflection angles that define the edges of region of maximum ASW, or ASW_{max} , were obtained by averaging the last twelve reversals of the subject response data, as recommended by García-Pérez (1998). These were grouped by either front or rear region, denoted here on as θ_F and θ_R . The results were then sorted by their corresponding delay time, t_{delay}. A Shapiro-Wilk test for normality (Shapiro & Wilk, 1965) found that all data except for θ_F at 10ms were normally distributed. This exception meant that non-parametric statistical analysis methods had to be performed. The equivalent 95% confidence interval, or notch regions (McGill, Tukey, & Larsen, 1978) of the results are presented in Figure 5.5. It is clear that there is overlap between notch edges which, according to McGill et al. (1978), indicates that there is no significant difference in values of θ_F or θ_R between each delay time with 95% confidence. A Wilcoxon signed rank test between each delay time for both regions found that there is no significant effect of delay time upon

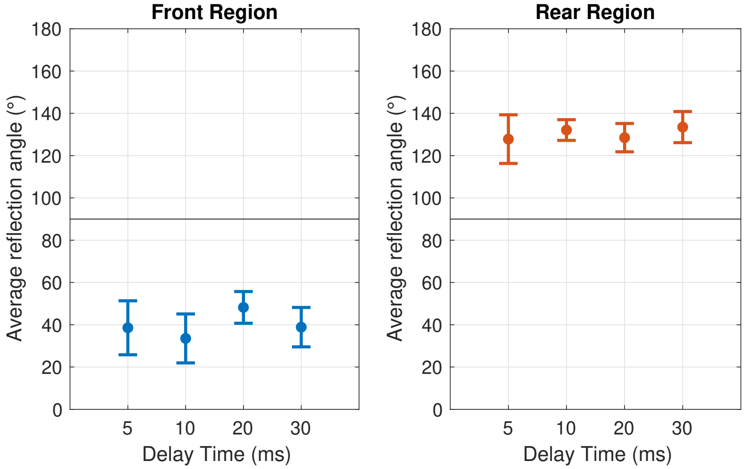


Fig. 5.5: Median values and corresponding 95% confidence interval notch edges of the ASW_{max} boundary averages of all subjects.

the values of θ_F or θ_R ($p > 0.05$). Therefore, the average reflection angle for all delay times for both regions can be calculated. It was found that $\theta_F = 38.9°$ and $\theta_R = 134.1°$, see Figure 5.6. These two values define the edges of the ASW$_{\text{max}}$ region.

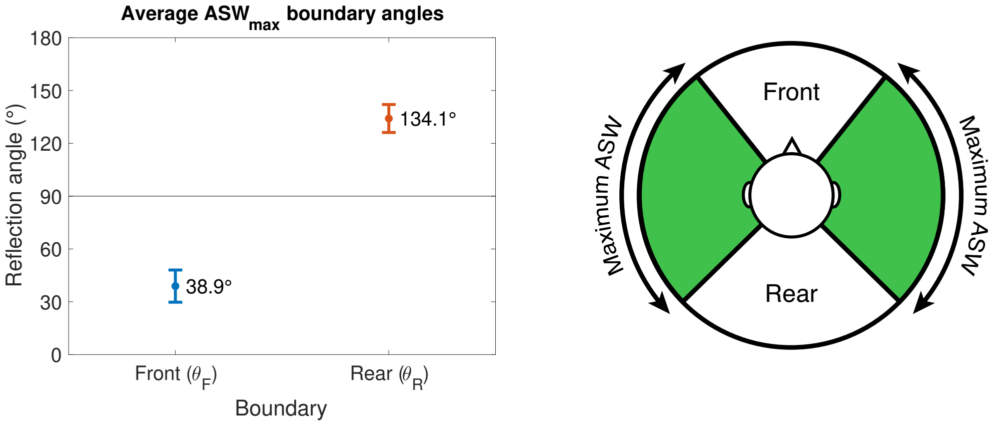


Fig. 5.6: Average boundary angles for the ASW$_{\text{max}}$ regions, and top-down visual representation of the ASW$_{\text{max}}$ regions on either side of the listener.

5.1.3 Relationship between the IACF versus reflection angle

The IACF, discussed in Section 2.1.1.2 of Chapter 2, for each artificial binaural IR were calculated within a ± 1ms lag range, and plotted versus reflection angle and lag offset in Figure 5.7. Initial examination of the IACFs showed that there is no difference between reflection delay times, as supported by the statistical analysis performed in the previous section. It can be seen in the left-hand plot in Figure 5.7 that there is a strong correlation at a lag offset of 0ms. This is of course the correlation of the direct sound between both ear signals. Since the direct sound is arriving at exactly 0°, there is no ITD, which manifests as a maximum peak in the IACF no matter the reflection angle. This can be denoted as the *'primary'* peak. On closer inspection, a *'secondary'* peak can be faintly seen splitting from the primary

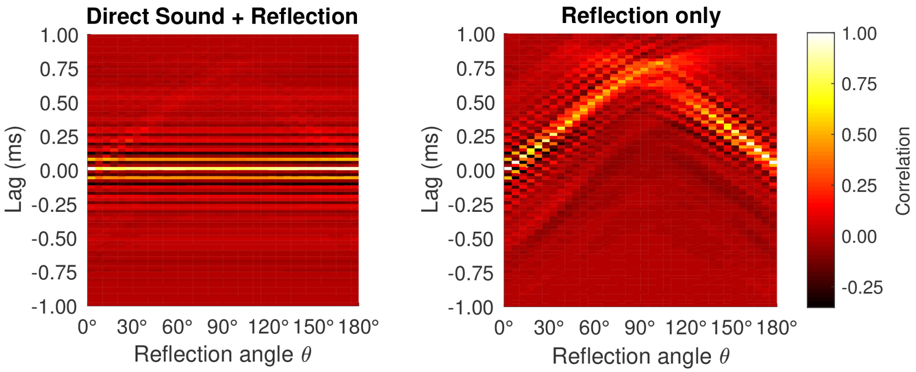


Fig. 5.7: Heat map plots of the IACF vs Reflection angle θ.

0ms, which is likely to be caused by the reflection. This is verified in the right-hand plot where the direct sound component is omitted, where in comparison to the left-hand plot the faint secondary peak matches the visible peak of the right hand plot.

The secondary peak moves away from the primary as the reflection angle increases from 0° to 90°, and moves back towards it as the reflection angle further increases to 180°. The presence of these two peaks is a possible cause for source broadening effect that occurs under these circumstances. However, it is unable to explain the saturation of the ASW between θ_F and θ_R . As ASW can be measured using the average of [1 - IACC] at 500Hz, 1kHz and 2kHz, otherwise known as the [1 - IACC_{E3}] (Hidaka et al., 1995), the relationship between IACC and reflection angle at several octave bands will now be observed.

5.1.4 Analysis of fluctuations in IACC, ITD and ILD

This section measures the IACC, ITD and ILD of the stimuli used in the test, and calculates the mean and standard deviation (SD) of time-varying fluctuations in those measures. The mean is the average of all time-varying fluctuations in a given measure, and the SD indicates the degree of variation of the fluctuations around the mean value. A word of caution, however, must be given when interpreting the SD. The SD is a value that is centred around a mean point, thus indicates how widely the measure varies around the mean and not its minimum. So whilst a low SD may initially indicate a low range of fluctuations, it does not indicate that the mean of the measure is also low.

The anechoic male speech source was octave-band filtered using a linear phase filter-bank, and then convolved with each the artificial BIRs. Prior to analysis, to simulate the neural mechanism of human hearing (Mason, Brookes, & Rumsey, 2003), the signals produced after convolution are half-wave rectified and then filtered by a first order Butterworth lowpass filter, with a cutoff frequency of 1 kHz. An octave-band filter bank rather than an 'Equivalent Rectangular Bandwidth' (ERB) bank was chosen so that the measures are compatible with existing octave-band measures such as IACC$_{E3}$. The first order slope was of the lowpass filter chosen because it allowed for analysis of the ITD-ILD trade off region around 1 kHz. IACC, ITD and ILD measurements of the signals were taken using frame-by-frame processing. Each frame was half overlapped and windowed using a Hann window. A frame length of 40ms was found to be optimal for analysis of speech signals, using recommendations provided by Mason et al. (2003). The ITD itself was derived from the offset point where the IACF is at its maximum, whilst the ILD was measured by calculating the ratio between the total energy of each windowed ear signal.

5.1.4.1 IACC

Running measurements of the IACC of continuously varying stimuli were used rather than BRIRs based on the assertion given by Mason et al. (2004) that an measurement of IACC of a BRIR can overestimate the predicted ASW, and that running measurements are more perceptually meaningful. Figure 5.8 shows the average of time-varying IACC versus reflection angle in the stimuli over several octave-bands. The vertical lines represent the boundary angles that outline the ASW$_{\text{max}}$ region. From 31 Hz to 125 Hz, the signal on average is well correlated, suggesting that there is little perceived width in this frequency range. Above 125 Hz, the range of mean IACC increases. Drawing attention to the 1, 2 and 4 kHz bands, the mean IACC appears to exhibit smaller changes in value as the reflection angle changes in the ASW$_{\text{max}}$ regions for all delay times. This suggests that the saturation in ASW could be due to the changes in mean IACC within the range of 0.6 to 0.8 falling below a JND threshold, thus subjects are unable to discern any changes between θ_F and θ_R .

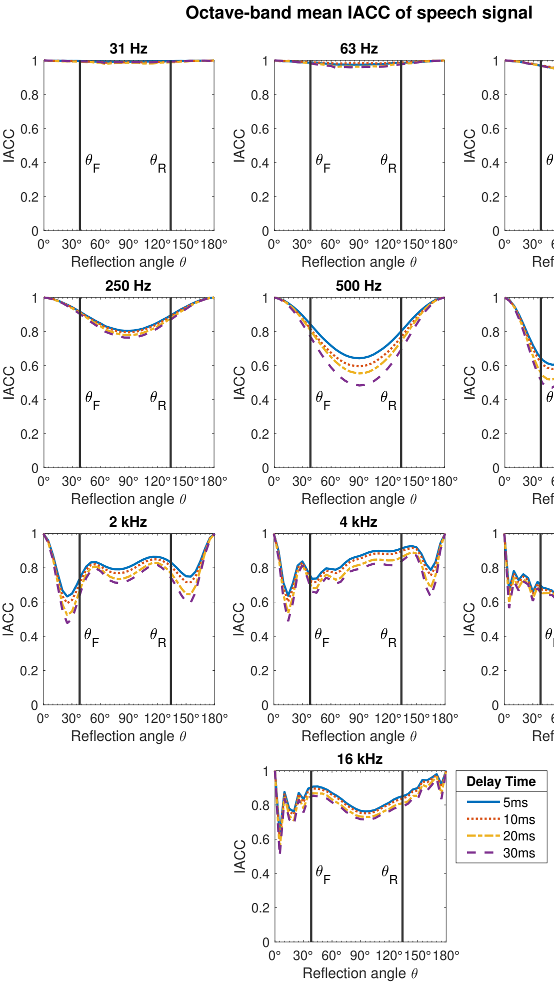


Fig. 5.8: Mean of fluctuations in IACC versus reflection angle. The vertical bars, labelled θ_F and θ_R , represent the outline of the ASW$_{\text{max}}$ region

Next, the standard deviation (SD) of the IACC fluctuations versus reflection angle were measured for each octave band, and are presented in Figure 5.9. Between 31 to 125 Hz, the range of SD is relatively small suggesting that, under the conditions set out in this test, the IACC does not fluctuate greatly in these frequency bands. However, from 250 Hz to 1 kHz, the range noticeably increases. The maximum SD increases with delay time, although the statistical results showed that this had no significant effect on the boundaries of the ASW_{max} region. Drawing attention to 1 kHz, it can be seen that the SD curve saturates or reverses at the boundary angles, with the edges of these features occurring at approximately the same angle for all delay times. This suggests that there is a possible link between SD and the position of the ASW_{max} region in this band. Above 1 kHz, whilst there appears to be some reversals and smoothing within the ASW_{max} region, these features do not line up with the boundary angles, and so it is not conclusive to whether or not there is a link between SD and ASW_{max} .

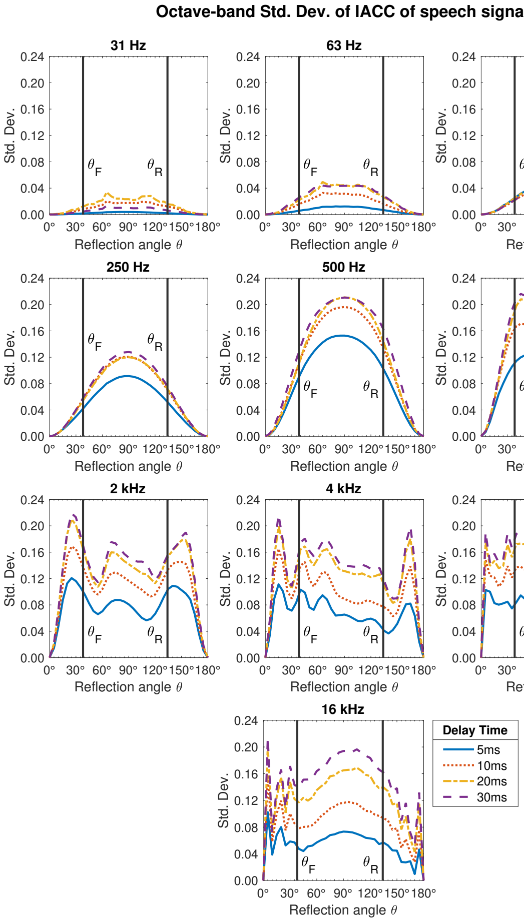


Fig. 5.9: Standard deviation of fluctuations in IACC versus reflection angle. The vertical bars, labelled θ_F and θ_R , represent the outline of the ASW$_{\text{max}}$ region

Since Okano et al. (1994) and Hidaka et al. (1995) consider the average of the IACC of BRIRs measured at 500 Hz, 1 and 2 kHz or IACC$_{E3}$, to be a good predictor for ASW, the IACC$_{E3}$ vs reflection angle will now be considered. It must be noted that they measured IACC$_{E3}$ of a static BRIR and not a time-varying signal, and therefore this measure is adapted for this analysis so that any relevant relationship can be found using continuous sources rather than BRIRs. Figure 5.10 is the mean and standard deviation of fluctuations in the IACC_{E3} plotted against reflection angle for all delay times. For the mean, it appears that the function saturates within the ASW_{max} region, suggesting that whilst ASW has a strong dependency on the IACC_{E3}, which is supported by the existing literature, it reaches maximum perceived width prematurely and does not scale to the full range of the IACC_{E3}. The SD also appears to saturate within the ASW_{max} region, and the edges of the plateau occur at approximately the same position regardless of the delay time. The seems to suggest that the ASW saturation is linked to the plateauing of the degree of variations in IACC_{E3} fluctuations.

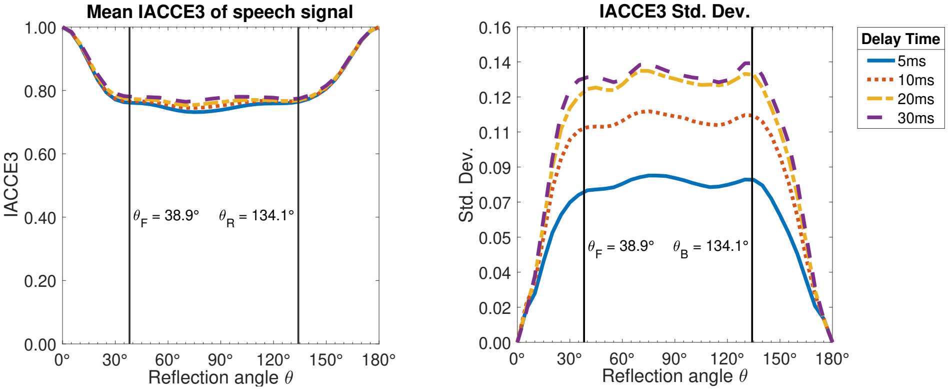


Fig. 5.10: Mean (left) and standard deviation (right) of fluctuations in IACC_{E3} versus reflection angle. The vertical bars, labelled θ_F and θ_R , represent the outline of the ASW_{max} region

5.1.4.2 ITD

Since Blauert and Lindemann (1986) has found that the SD of time-varying fluctuations in the ITD to be related to the degree of perceived auditory spaciousness and ASW, octave-band analysis of these measurements will now be considered. Figure 5.11 shows the octave-band, standard deviation of fluctuations in ITD versus reflections angle. Like the octave-band mean, there are no notable features between 31 to

500 Hz, other than the functions following a relatively smooth curve as reflection angle changes. However, at 1 and 2 kHz, there are definite reversal and plateaus. Focusing on the 1 kHz band, for all delay times the function plateaus close to the ASW$_{max}$ boundary angles, meaning that the range of the fluctuations for ITD in this band show little change within this region. This could also be potentially linked to the saturation in ASW, where between reflection angles the change in ITD between angles is below a just noticeable difference threshold.

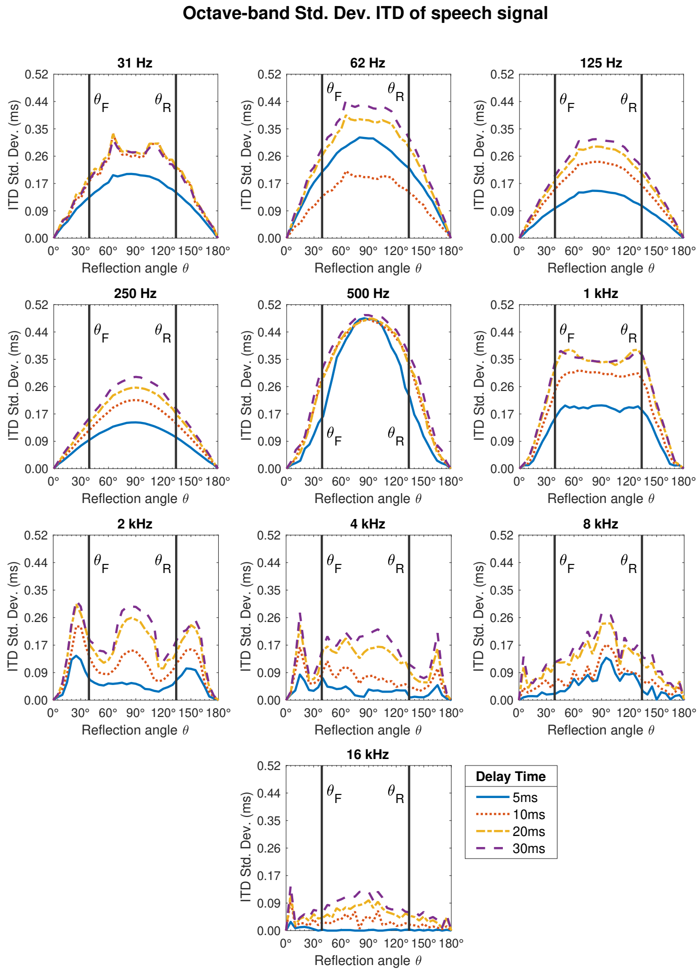


Fig. 5.11: Standard deviation of fluctuations in ITD versus reflection angle. The vertical bars, labelled θ_F and θ_R , represent the outline of the ASW$_{\text{max}}$ region

5.1.4.3 ILD

Figure 5.12 shows the SD of fluctuations in ILD versus reflection angle in the stimuli over several octave-bands, and again the vertical lines represent the ASW$_{\text{max}}$ boundary angles. Between 31 to 500 Hz, for all delay times the SD quickly yet smoothly approaches maximum, whilst not showing any major differences within the ASW_{max} region. From 1 and 4 kHz, however, the SD at low delay times plateaus approximately at the ASW_{max} boundary angles, and at higher delay times the curve become increasingly erratic. This plateauing at lower delay times could be linked to the saturation in ASW because of the changes between the range in fluctuations decreases below a potential JND threshold. However, this may not be possible at higher delay times in frequency bands about 1 kHz as these exhibit larger changes in ILD between angles at higher delay times. Therefore, these observations may potentially not be linked to the existence of the ASW_{max} region.

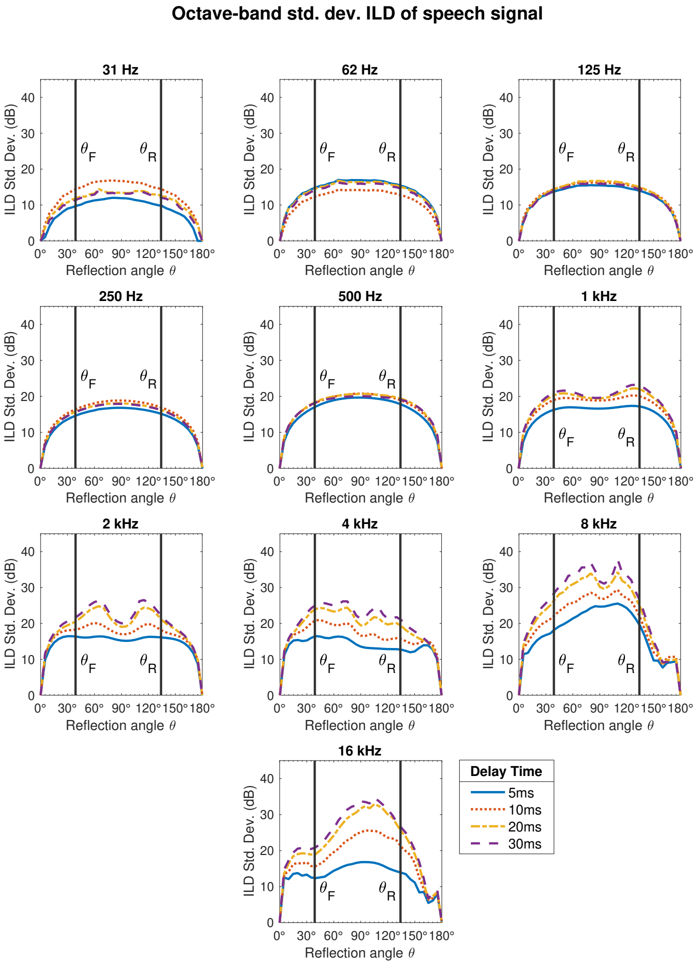


Fig. 5.12: Mean of fluctuations in ILD versus reflection angle. The vertical bars, labelled θ_F and θ_R , represent the outline of the ASW$_{\text{max}}$ region

5.1.5 Discussions

5.1.5.1 Discussion of the psychometric test results

The aim of this experiment was to verify and expand on the findings produced by Barron and Marshall (1981). Based on an interpretation of their results, see Figure

5.1, it was hypothesised that ASW reaches its maximum between 30° to 160°. The results from the psychometric test found that ASW saturates as the reflection angle arrives within an region between 38.9° and 134.1°, which lies within the hypothesis region. It was also important to test the effect of delay time upon the location of these boundary angles. Analysis found the effect to be insignificant, which was expected as Barron and Marshall (1981) also found that reflection delay time had little effect on the perceived spatial impression. Thus, the null hypothesis that delay time has no effect on location of the boundary angles can be retained. Whilst the reflection delay time used for the test was limited to 30ms, it can be assumed from the findings of Barron and Marshall (1981) that the $\mathrm{ASW_{max}}$ region can be valid for delay times up to 80ms. However, caution must be taken with this assumption as Barron and Marshall (1981) used an orchestral motif for their experiments, whilst this particular experiment used a speech source. This also implies that the region boundary locations are perhaps only valid for this particular sound source, although the hypothesis of their existence was based upon findings that used an orchestral motif. Therefore, it can be assumed that similar boundary angles exist for other sources within a certain amount of tolerance, although experimentation using said sources would be able to determine if this assumption is correct.

5.1.5.2 Possible linkage with saturation in the IACCE3

The other key aim of this experiment was to find a possible cause for the ASW to saturate within the $\mathrm{ASW_{max}}$ region through objective analysis. The most relevant objective measures for ASW are L_f, the IACF and IACC_{E3}. However, L_f was not considered as it would not be an ideal measurement for the sound fields used in this experiment, and would mathematically produce the same curve shown in

Fig. 5.1. Considering first the IACF, it was hypothesised that there would be an observable feature that would lie close to the boundary angle. However, no such feature was found. It should be noted that Mason et al. (2004) found that predicted measures of ASW from static measurements of impulse responses do not adequately match the perceived measures. Therefore, running measurements of IACC were performed.

Analysis of the time-varying, octave-band IACC measurements of stimuli used in the psychometric found that the saturations of the mean and SD of fluctuations in IACC_{E3} coincidently occur between the angles θ_F and θ_R , see Figure 5.10. It has been established by Hidaka et al. (1995) that the IACC_{E3} plays a major role in the perception of ASW, thus the findings from the objective measurements suggest that the perceptual ASW saturation occurs because differences in mean IACC_{E3} fluctuations decrease as the reflection angle increases and approaches the the ASW_{max} region. According to Pollack and Trittipoe (1959) and Klockgether and van de Par (2016), as the initial IACC decreases, the JND in IACC increases. Okano (2002) also found that the JND for $[1 - IACC_{E3}]$ is 0.065 ±0.015. This suggests that as the IACC_{E3} plateaus and the differences in the measurement fall below the JND threshold, the ASW would be become saturated. This supports the coincidence of the location of the plateau edges with the boundary angles of ASW_{max} .

5.1.5.3 Relationship between saturation and fluctuations

The participating subjects informally reported post-test to have experienced fluctuations in the ASW when listening to the stimuli, however some were able to focus on a static source image whilst simultaneously experiencing source broadening. This suggests that subjects were in fact perceiving a secondary, moving source image.

At low delay times this would manifest itself as a single, fused source image, as suggested by Litovsky, Colburn, Yost, and Guzman (1999) regarding the precedence effect. This puts forward a possible scenario where the image could be moving due to the fluctuations in ITD and ILD, and so listeners experience a fluctuation in the perceived ASW and are unable to focus on the secondary image. The frequency and range that this image moves within could be what determines the maximum magnitude of the perceived ASW. This of course is was found to be the case in previous literature such as Grantham and Wightman (1978) and Mason (2002), who found that when the rate of fluctuations increased, the source image became increasingly difficult to focus on and locate. Blauert and Lindemann (1986) also found that time-varying fluctuations in ITD are important to the perception of spaciousness, and thus their possible linkage to the saturation in ASW was considered.

The analysis of the mean and standard deviation of fluctuations in ITD and ILD over time were measured per octave band for each reflection angle. A saturation in the standard deviation of ITD in the 1 kHz band was found to lie at similar reflection angles coincident to the location of the ASW_{max} region. This supports the fact that there is a possible linkage between the behaviour of this measures and their contribution to ASW. Therefore, if the range of these fluctuations either saturates, or the degree of change between reflection angles falls below a JND threshold, there would be little noticeable change in ASW when the reflection angle lies within the ASW_{max} region. There is disagreement, however, in the literature as to which bands are most related to the perception of ASW. Barron and Marshall (1981) found that in the 1 kHz band "... the subjective impression is one of source broadening...", whilst Morimoto and Maekawa (1988) found that low frequency components below 510 Hz were much more strongly related. Hidaka et al. (1995) later suggested that the perception of ASW is best predicted using the average of the IACC measured at the 500 Hz, 1 and 2 kHz bands. The findings from this test support that the 1 kHz band was the most related to the degree of perceived ASW. The results also support the findings of Mason (2002) who showed that time-varying fluctuations in these inter-aural differences correlate with the degree of perceived ASW.

5.2 Experiment part 2 : Verification of the regions

The main limitation with the previous test described in Section 5.1 is that the reflections were simulated over headphones, rather than being mimicked by physical loudspeakers. Furthermore, the test was effectively anechoic due to the anechoic nature of the HRTFs used for that test. For practical purposes such as in control room where surface reflections are present, it was decided that the test results from Section 5.1 should be verified using physical loudspeakers that would mimic a single room reflection.

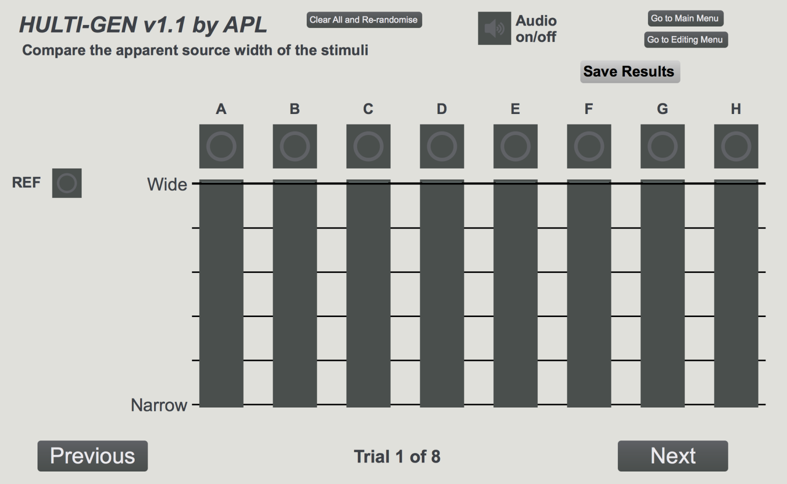
5.2.1 Experimental design

5.2.1.1 Stimuli

For this experiment, the anechoic male speech used in the previous experiment was used, in addition to an anechoic cello excerpt. This gives a total of two stimuli programme items. The addition of the cello excerpt was deemed important as to verify if the ASW$_{\mathrm{max}}$ region is directly applicable to a musical source.

5.2.1.2 Interface

The test was designed using HULTI-GEN version 1.1 developed by Gribben and Lee (2015), see Figure 5.13. Prior to main testing, the subjects were presented with a practice interface during a training phase in order to familiarise them with the controls and the stimuli.



HULTI-GEN v1.1 by APL

Compare the apparent source width of the stimuli

Fig. 5.13: The HULTI-GEN test interface used in this experiment.

Subjects were instructed listen to each stimulus including the reference, and compare and grade the relative apparent source width each of the test stimuli against the reference.

5.2.1.3 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Genelec 8040A loudspeakers. The playback level was set to a listening level of 72 dBA by playing pink noise from a single speaker, and measuring using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

With the size of the chosen room and limited number of speakers available for the experiment, it was found to be infeasible to recreate the same set of available reflection directions used in Section 5.1. A complete circle of loudspeakers spaced at 5° intervals would require 72 loudspeakers. The minimum radius of a circle of loudspeakers is given by:

$$r = \frac{WN}{2\pi} \tag{5.1}$$

where W is the width of one loudspeaker, and N is the required number of loudspeakers. Therefore, where the width of a single Genelec 8040A is 0.19m, the minimum arc radius would need to be 2.17m. This is too large to fit within the room and keep within the ITU-R BS.1116 (2015) specification regarding the minimum distance between the back of a speaker and a wall. Furthermore, whilst only a 180° arc would be used, at 5° spacing there were simply not enough speakers available to make a complete arc. However, because this experiment is not a psychometric test, and the objective is to only compare width created by a test reflection against that created by a 90° reflection, rather than between test reflections, the resolution is not of as great importance. It was decided that the number of speakers could be reduced by increasing the spacing to 10°. At this spacing, the number of speakers required for a full circle is 36, and so the minimum radius becomes 1.08m.

Since the results from Section 5.1.2 found that the ASW saturates after 38.9°, it can be assumed that artificial reflections created by loudspeakers within the ASW_{max} region would produce the same amount of ASW. Using this assumption, the number of required speakers around the 90° reference speaker can be truncated and concentrated around the boundary angle. Further to this, because the test is investigating the front and rear regions separately, the number of required speakers can be further reduced by using only a 90° arc of speakers for one region, and the subject can be rotated to either face forwards with the loudspeakers facing them from the front right, or backwards with the loudspeakers facing them from the rear left (see Figure 5.14). As the test from Section 5.1 assumes that the subject's ears are symmetrical, it is not of great important to which side of the listener the reflections arrive from.

5.2.2 Subjects

For this test, a group of ten subjects consisting of music technology students and researchers at the University of Huddersfield participated. Nine of the subjects had experience with spatial audio and critical listening, whilst the remaining subject had mixed critical listening ability.

5.2.3 Statistical analysis of the results

Figures 5.15 and 5.16 show the McGill et al. (1978) notch edges, or equivalent 95% confidence intervals of the data for perceived width versus speaker angle per delay time for each source. Focusing first on the front region, at all delay times for both sources there is an expected increase in median width as reflection angle increases

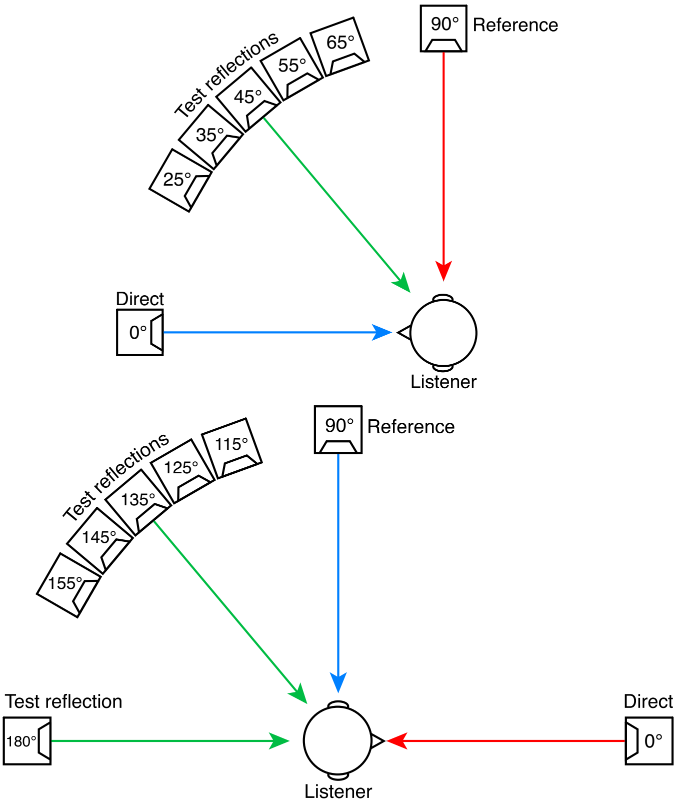


Fig. 5.14: Subject positions in the physical setup, where **top** is for testing the front region, and **bottom** is for testing the rear region.

from 0° to 90°. The lack of overlap between 0° and either 25° or 35° and above suggests that at higher reflection angles there was an expected significant difference in width. In most cases, there is a significant difference in width between the 0° and 35° reflection whilst, however, there is no significant different in width caused between the 35° and 90° reflection. In fact, for all front region cases, in all delay times except 30ms, there seems to be no significant difference in width between all angles above 35°. The trends in the median values graphs imply that there is a saturation in width begins to occur between 25° and 45°.

Focusing now on the rear region, 90° to 180°, there is an expected decrease in median width as the reflection angle approaches 180°. However, subjects are still able to perceive width at 180°, and at 5ms delay time there is no significant difference in width between a 90° or 135° reflection and 180° reflection, with the 135° reflection being the closest to θ_R . Theoretically at this angle there should be no perception of width. At higher delay times there is a significant difference between 90° and 180°, and no significant difference between 90° and 135°. Interestingly, for the Cello excerpt at 20 and 30 ms delay times, there are distinct groups that form between 90° to 135°, and 145° to 180°. Whilst there is overlap between 135° and 145° it can be seen that the perceived ASW has potentially saturated in the 90-135° region and, is lower in the 145°-180° region, with the overlap occurring near the θ_R edge of the $\mathrm{ASW}_{\mathrm{max}}$ region.However, in comparison to the results discussed in Section 5.1.2, it is not entirely clear that the perceived ASW in this experiment had saturated at the same threshold angles found in the previous experiment, or even saturated at all.

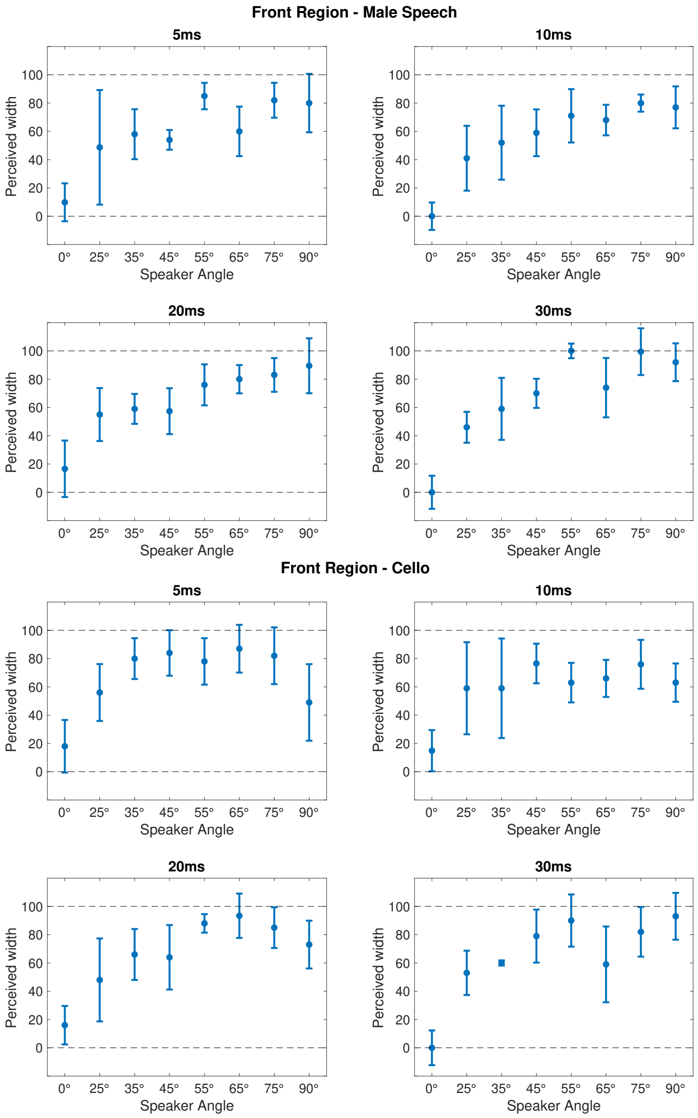


Fig. 5.15: Equivalent 95% confidence intervals of perceived width versus speaker angle per delay time. Note the non-linear step between speaker angles

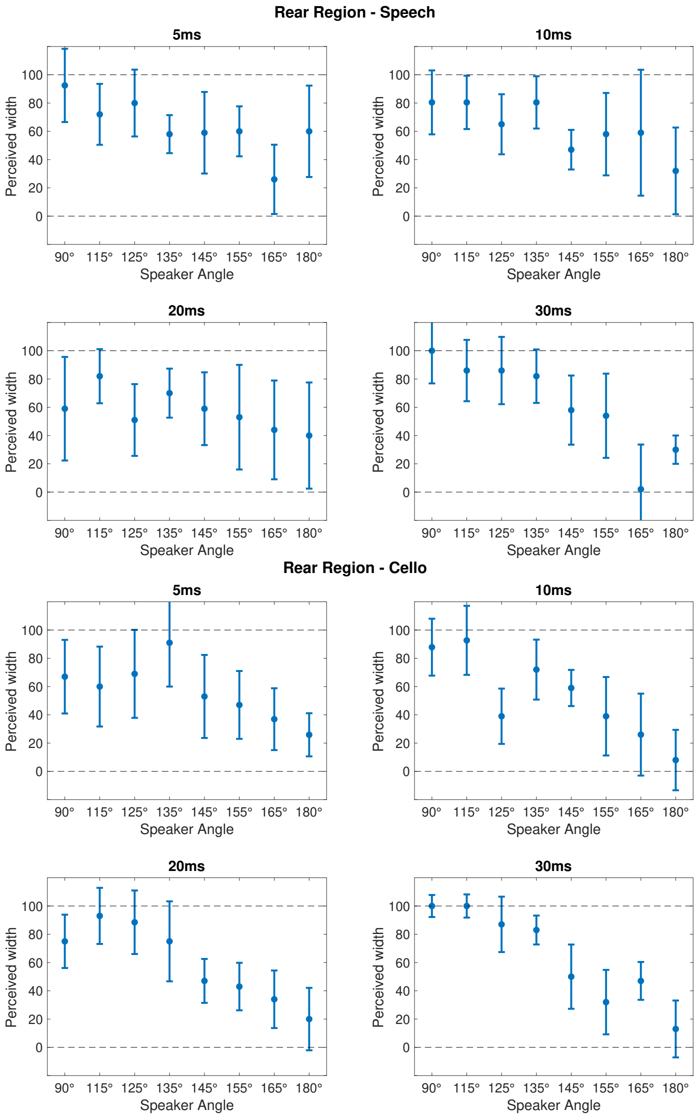


Fig. 5.16: Equivalent 95% confidence intervals of perceived width versus speaker angle per delay time. Note the non-linear step between speaker angles

Assuming that ASW had indeed saturated, these two observations suggest that subjects are perceiving larger differences in width between lower reflection angles and 90° than at larger reflection angles in the front region, especially after the angle

θ_F found in part one of this experiment where subjects appear to not hear differences at all. This then would imply that only the saturation in the front portion of the ASW$_{\text{max}}$ region appears to function in a slightly reverberant setting. However, for rear angles there appears to be very little difference in perceived width between reflection angles, although there is a trend in the median width values to decrease as the reflection angle continues to increase to 180°. It is unclear whether ASW is reliably saturating in the rear region, and as to why subjects are still perceiving width caused by a 180° reflection. Again, this observation differs from what was found in the previous experiment in Section 5.1.2. This could be caused by three factors: the lack of low anchor for little or no width; presence of room reflections causing a natural width increase; or the test method itself. If the low anchor was included, it would have perhaps reminded subjects of the minimum width. However, even if the statement is assumed to be true, it does not adequately explain why subjects were not able to properly distinguish ASW between 90° and 180° reflections. As this experiment used a multiple comparison test, rather than the threshold test used in the previous experiment, it is most likely that this type of presentation method had the greatest influence over the subjects' responses. This is discussed in greater detail in the subsequent section.

In order to further determine if ASW is saturating within the ASW$_{\text{max}}$ region in non-anechoic conditions, a Wilcoxon signed-rank test with Bonferroni adjustment applied was performed with a significance level of $\alpha = .05$. For both regions and sources, the test found that there was in fact no significant differences in width between any pair of angles ($p > .05$). This is most likely due to the conservative nature of the Bonferroni adjustment as the test disagrees with the observed significant differences between angles in the notch region plots. This causes the result of the test to appear unclear in terms of whether there was a significant difference in width within the ASW_{max} region. However, the observed saturation in median values within this region suggests that there are less noticeable differences in width between angles.

5.2.4 Discussion

The results from the statistical analysis suggest two things. First, that subjects were perceiving significantly more width when reflection angle is above 0°; secondly, and more importantly, subjects were unable to distinguish any significant differences in width not only in the ASW_{max} region, yet also at much lower angles. However, the latter also suggests that a majority of subjects are perhaps unable to distinguish significant changes in width when the stimuli is creating a fluctuating sense of width. This same fluctuation was observed in the previous experiment and could be the main factor in a subject's ability to distinguish differences in ASW. As speculated in the previous sub-section, it is possible that the response and presentation method may have had an influence on the outcome of this test. This can be seen as an unfair comparison since the staircase test from the previous test was performed in effectively anechoic conditions simulated over headphones, whilst this test used a multiple comparison method in a non-anechoic environment, and so a room effect may have had an influence. As the room is an ITU-R BS.1116 compliant room, the room effect may have had little influence. Considering the test presentation method, in the previous test the subjects simply had to respond with whether they could hear a difference in width, whilst in this test they were tasked with quantifying an amount of perceived width in comparison to a reference. Furthermore, in the previous test the subjects would only compare the width between two stimuli, whilst in this test they were able to freely switch between all conditions. These two response methods are different from each other, and so could produce different results.

Since the complex source types used in this test may have invoked a rapid fluctuation in ASW, they may have caused a difficulty in quantifying the amount of difference in width. Therefore, this test could be performed using a marker or indicator response method, such the LED strip proposed by Lee, Johnson, and Mironovs (2016), that would allow for subjects to judge the average amount of width they were able to hear. This test, however, is out of scope of this study as it was mainly concerned with verifying the findings in part one of this experiment, yet it will be revisited in the future.

5.3 Future Work and Practical Implications

The first limitation of both experiments in this chapter was the use of a single reflection and not multiple reflections, for example the arrival of a symmetric reflection arriving at the right-hand side as well as the left-hand side. This would produce perhaps a more balanced sense of ASW where the source image and any perceived width would be centred in front of the listener, which would be a slightly more realistic approach. Nevertheless, it is valid to measure lateral fraction for both *single* or *multiple* reflections as it only measures the ratio of lateral to all early energy, and does not require reflections to arrive at both sides of the listener for measurement. Therefore, using this assumption, the use of a single reflection for this experiment can be justified.

Secondly, part 1 of this test was limited to using generalised HRTFs from the MIT

KEMAR dummy head database (Gardner & Martin, 1995). The solution to this limitation would require individualised HRTFs of several subjects with an azimuthal resolution of 5° in order to match the resolution of the MIT dataset, which is challenging to obtain. The benefit of using a generalised dataset is that it gives a more practical context to the findings, as it is common to use such generalised datasets in real-world applications that call for immersive audio such Google VR SDK and YouTube 360 (Google, 2017).

The third limitation was that a small number of anechoic sources were used overall. The first experiment used only a speech signal because it contained both continuous and transient qualities together. A series of future experiments can investigate the effects of source type upon the location of the ASW_{max} region boundaries, and see if the region has a source dependency. From the observed saturations in IACC, ITD and ILD, hypothetically there may not be a source dependency, though future testing is the only way to verify this.

Finally, whilst the test was performed using a reflection arriving only on the horizontal place, Barron (1971) and later Furuya et al. (1995) found that a reflection solely in the median plane has little effect on the perceived horizontal width of a sound source. Barron and Marshall (1981) also found that with an azimuth of 90°, an elevated reflection (e.g. a ceiling reflection), does not contribute significantly to the amount of lateral energy, but to the total amount of early energy, and thus will produce a lower degree of ASW. With this in mind, it is only possible to generalise the ASW_{max} region to any non-elevated (0° elevation) reflection whose azimuth angle lies within it. Further tests in the future, however, will investigate the possibility of saturation in ASW in the median plane at various azimuth angles such that the region can expanded to be two-dimensional. The experiment was also limited to using a direct sound component arriving from directly in front of the listener at 0°. This limitation, however, is practical for concert hall setting where a listener will most likely be facing forwards towards the sound source on the stage.

The ASW_{max} region found and verified in this experiment can be used to form the "width control" parameter of the perceptual control method proposed for this study. Any reflection arriving with the ASW_{max} region will be treated as one that generates the maximum amount of perceived ASW. As predicted by lateral fraction, by altering the energy of lateral reflections within this group, it should be possible to control the amount of perceived width. In the context of perceptual optimisation, a concert hall that creates a great sense of SI is subjectively excellent (Hidaka et al., 1995). Thus, "width optimisation" could potentially improve the SI of a concert hall, albeit currently only within the domain of virtual room acoustics.

The technique can be inversely applied in order to improve the localisability, as in the ease of localisation, of a sound source. Hartmann (1983) found that early lateral reflections can reduce the localisability of a sound source. Consider a VR scenario, where users may want to be able to focus and effortlessly locate a sound source. If many lateral reflections are arriving at the user's position, rather than the user change their position, spatial filtering can be applied to reduce the lateral reflection energy and theoretically improve the localisability of a sound source. This would be great benefit if it is combined with an auto-mixing algorithm that could continuously optimise the auditory experience. It would be worth investigating the effectiveness of the "width optimisation" as part of perceptual control method scheme in the future.

5.4 Summary and conclusion

This experiment was conducted in order to find a region on the horizontal plane in which ASW is perceived to be at its maximum, such that the region can be used as part of the perceptual control method. The existence of the region was hypothesised after interpretation of results from the investigation into the relationship between lateral fraction and spatial impression performed by Barron and Marshall (1981). The experiment was split into two parts, where part one utilised a psychometric threshold detection method in order to find the angle where the perceived ASW saturates. Part one found that:

- ASW appears to saturate when a reflection arrives between 38.9° and 134.1°. These two angles form a lateral region of maximum ASW, denoted as ASW_{max}.
- The reflection delay time between 5 to 30 ms had no significant effect on the location of the region's boundary angles.
- The most likely cause for this novel finding is a plateau or saturation in the measured IACC_{E3}, where the function appears to a fairly level minimum within the ASW_{max} region.
- Another likely cause is saturation in the standard deviation of time-varying fluctuations in ITD occurring within the ASW_{max} region, most notably in the 1 kHz octave band.

Since the first part was performed over headphones using a simple, artificial sound field and HRTFs, the experiment was essentially performed in an anechoic setting. Therefore, to verify the main findings, a test using a loudspeaker arc to recreate the scenario of part one, however, in a none-anechoic yet controlled environment. A cello sample was included to further verify the findings with a musical source type. This second test found that:

- ASW does appear to reach a maximum when a reflection arrives in the previously found $\mathrm{ASW}_{\mathrm{max}}$ region, although it is unclear if it reached a saturation point at the same angles found in Part 1.

- Statistical analysis suggests that the saturation point may lie outside of the $\mathrm{ASW}_{\mathrm{max}}$ region, although this is speculated to be symptoms of using a combination of loudspeaker playback as opposed to headphones, and the test presentation method where random selection of stimuli may cause difficulty for subjects to discern small differences in width.

The novel, $\mathrm{ASW}_{\mathrm{max}}$ region found in this experiment can now be used as part of the proposed perceptual control method that will manipulate ASW.

Chapter 6

Perception of colouration in relation to the properties of a single reflection

This chapter will investigate the effects of reflection direction and level difference upon the audibility and acceptability of colouration. Recall that Halmrast (2000) proposed there may exist both ’good’ and ’bad’ types of colouration, where a good type is one that enhances the sound or does not distract from the listening experience, whilst a bad type causes a distraction and creates an unpleasant listening experience. What was not found in the reviewed literature was the effect of reflection level, relative delay time, and direction or arrival upon the perception of colouration, and how these two variables relate to the preference. In reality, reflections will arrive from various directions at differing levels over time, thus these attributes may relate to preference of colouration, be it ’good’ or bad, or alternatively ’acceptable’ or ’unacceptable’.

Whilst the study carried out by Halmrast (2000) analysed the effects of comb filtering upon RIRs at various positions in different concert halls, it did not investigate the fundamental process of how direction and level of single reflection affect the perception of colouration in terms of its audibility and acceptability. Brunner et al. (2007) later studied the effect of reflection level on the audibility of colouration, finding that listeners detected changes in colouration at an average reflection level difference of 18 dB. That study, however, was concerned with the audibility of colouration rather than its acceptability.

This discussion creates the following research questions. In the context of concert hall acoustics, for a typical -6 dB first order reflection:

1. At what angle does a reflection arriving from either in front of or behind the listener create *audible* colouration?

2. At what angle does a reflection arriving from either in front of or behind the listener create *acceptable* colouration?

3. How much reduction is required until the colouration becomes acceptable?

With respect to questions one and two, it is expected that as the reflection angle increases towards a lateral direction, the ASW will broaden (Barron & Marshall, 1981), and so it is possible that the colouration could become less distracting. As discussed in the literature review in Section 2.2, this was previously speculated by Ando (1977), although it has not yet been formally tested. Also, whilst Seki and Ito (2003) found that the audibility of colouration reduces when a reflection arrives lateral to a listener, the relationship between colouration and direction of arrival have not been properly investigated. In regard to the final question, it is expected that the severity of the colouration will reduce as the reflection level decreases, thus it can also be hypothesised that colouration with reduced severity may be regarded as more acceptable.

6.1 Experiment part 1: Elicitation of attributes

Whilst Halmrast (2000) described colouration as a change in timbre, the literature discussed in Chapter 2 establishes that it is a broad attribute that encompasses terms that describe different types of changes in the timbre in different situations. Before any investigations into the effect of reflections angle and level on tonal colouration can be performed, it is important to understand the perception of colouration produced by a single reflection in the presence of a direct sound using the methodology described below. Using an elicitation method based on the QDA (Quantitative Descriptive Analysis) (Stone & Sidel, 2004), terms specific to this scenario can be defined and used to describe the differences in timbre. The terms will be used to create a glossary to be handed to subjects during future tests that describes the nature of the colouration. This glossary should help trained subjects to focus on the types of tonal colouration they may perceive in later tests.

6.1.1 Test method

6.1.1.1 Stimuli

For this experiment, the same anechoic speech and cello stimuli used in the previous chapter were used, giving a total of two stimuli programme items. To recap the stimuli generation method utilised in Section 5.1.1.3 of chapter 5, a direct and unattenuated sound is combined with a delayed copy of the sound which is attenuated by 6 dB to simulate a reflection exhibited in a concert hall. This choice of level difference is based upon the methodology of Barron and Marshall (1981) who also used the same level difference in their investigation. The reflection is delayed by 5, 10, 20 or

30 ms. Both the direct sound and reflection are convolved with an appropriate HRTF. The direct sound was convolved with either a HRTF measured at 0° or 180°. The test was split over two sessions: one for each source type. In each session, the subject was presented with each stimulus for each delay for that particular source type, and were instructed to elicit terms regarding any audible tonal colouration. The subjects were able to listen to each stimulus as many times as they wished before moving on to the next.

For the purposes of this particular experiment the reflection was fixed to arrive from either directly front of or behind the listener. This ensured that the colouration would be perceived to be most audible, and that neither ITD and ILD caused by offset angles would not interfere with the nature of the colouration, thus controlling its perception. A reflection from behind was not considered as this produced a very similar perception of colouration, and so can be treated as a redundant condition. The result of this is two BRIRs where in both the direct sound is arriving from in front of the listener at 0°, whilst the reflection either arrives at 0° in front, or 180° behind the listener. Finally, the anechoic male speech sample used in the previous experiment was convolved with each BRIR to produce the stimuli.

6.1.1.2 Subjects

Four subjects, excluding the author, took part in this experiment. It was of utmost importance that the subjects had a high critical listening ability, as suggested by ITU-R BS.1116 (International Telecommunication Union, 2015). All four subjects had experience with critical evaluation of spatial audio, and one in particular had previously performed experiments regarding tonal colouration. It is also worth noting that the subjects had an awareness that the reflection angle was being altered, although this did not appear to affect the result of this investigation. They were therefore carefully instructed to focus on only the timbral aspects and overlook the fact that the reflection angle is changing.

6.1.1.3 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dB $\mathrm{LA_{eq}}$ by playing pink noise from one headphone cup, and measuring the average level using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

6.1.2 Elicited terms

After the test, the occurrences of terms and phrases were counted, and are presented in Table 6.1.

Elicited Terms
Metallic (17), Roughness (9), Clarity(8), Muddy(6), Width (6), Fullness (5), Comb Filtered (4), Natural (3), Separation (3), Thin (3), Hollow (2), Tube-like (2), Echo (1), Externalised (1), Glassy (1), Image Split (1), Narrow (1), Phasey (1), Resonant (1), Robotic (1)

Table 6.1: Occurrences of the elicited terms for the perceived colouration

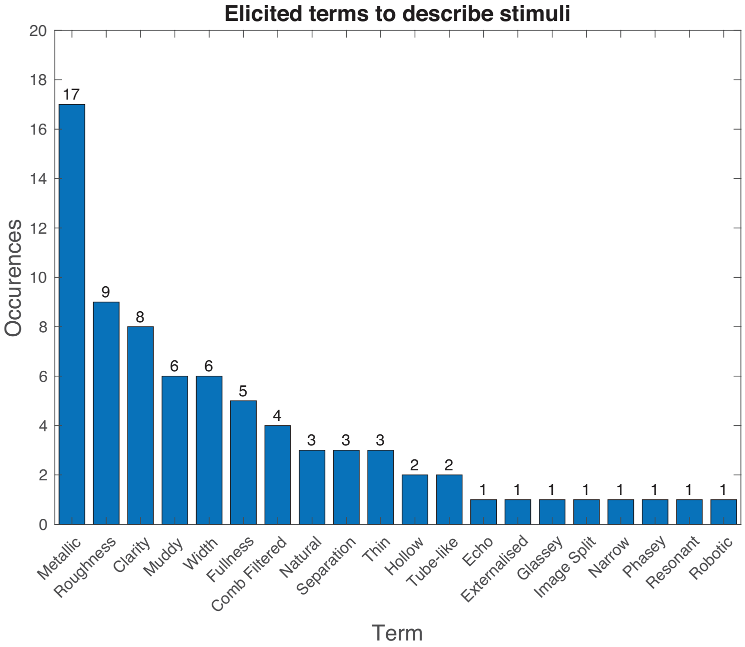


Fig. 6.1: Occurrences of the elicited terms for the perceived colouration

6.1.3 Group discussion on elicited terms

The four subjects later participated in a group discussion. The objective of the discussion was to decide upon which are the most salient terms that were being perceived at different delay times. The occurrences of each term were withheld from the subjects. Each subject was given a pair of headphones, the same type used in the elicitation test, connected via a four-way headphone amplifier so that the stimuli could be simultaneously presented to the subjects if required. The author chaired the discussion, yet would not interject or influence the discussion as such not to create bias, and would merely be there to progress the discussion forward. The discussion took two hours to complete. The following table shows the terms grouped by their corresponding delay time:

Delay Time (ms)	Elicited Terms
5	Metallic, Harsh, Thin, Roughness, Phasiness
10	Chorusing, Phasiness, Thicker
20	Roughness, Fullness, Scratchiness, Sharpness
30	Modulation, Echo, Resonance, Ambience, Roughness

Table 6.2: Elicited terms for each delay time

After the terms were grouped by delay time, definitions of each term were created by the group. This was done so that the terms could be used as examples of tonal colouration, and as a glossary for the subsequent tests in this experiment.

Attribute	Definition
Metallic	'Robotic', or as if the sound is 'inside a metal tube'. Can also be described as a 'scraping on metal' quality
Resonant Pitch	A noticeable resonant, constant pitch that is present in the sound
Roughness	Distorted, modulated quality in the high frequency band
Scratchy / Glassy	Sharp and piercing quality in the high frequency bands
Fullness	Representation of both low and high frequency regions
Natural	Sound does not appear to be artificial, and it feels realistic and acceptable
Clarity	Amount of definition in the high frequency band. Opposite of muddiness
Muddiness	Lack of definition in the high frequency band. Opposite of clarity

Table 6.3: Glossary of terms used to describe 'tonal colouration'.

6.2 Experiment part 2: Effect of reflection angle

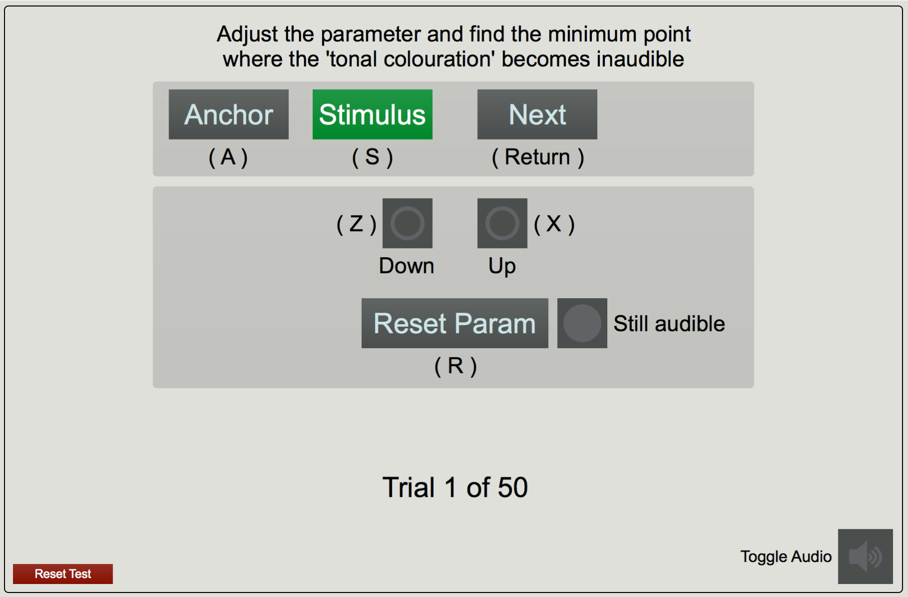
To test if the direction of arrival of a reflection had an effect on the acceptability and audibility thresholds of colouration, a psychometric test was performed to obtain two threshold angles. This test follows the same procedure as the test described in Section 5.1, where a reflection's direction of arrival was changed and subjects had to locate the angle that the tonal colouration becomes acceptable / inaudible. From this, the goal of the test was to define regions of tonal colouration unacceptability and audibility, where the boundaries of this region depends on the listener's own tolerance of colouration, and what they deem acceptable.

6.2.1 Stimuli

This experiment used the same stimuli created for the first test in the experiment performed in Chapter 5, where an anechoic sample of Danish male speech from the Bang and Olufson 'Music for Archimedes' project (Hansen & Munch, 1991) was used. The sample was set to play and continuously loop in real-time, and was split into direct sound signal and reflection signal. The reflection angle was varied between two angles, either 0° and 90° to obtain the front boundary angle, and 90° and 180° to obtain the rear boundary angle. The two signals were convolved with their corresponding diffuse-field compensated HRTF (head-related transfer function), the KEMAR database (Gardner & Martin, 1995). The reflection signal was delayed by five different values of t milliseconds and attenuated by 6 dB. The chosen values for t were 2.5, 5, 10, 20 and 30 milliseconds. This creates ten possible conditions.

6.2.2 Test method

Initially, the traditional adaptive staircase method (Levitt, 1971) used from the previous chapter was considered. However, informal testing using this method found that it was difficult to easily distinguish changes in colouration, and that it became tiring to find any possible threshold points using a staircase method. Alternatively, the 'Method of Adjustment' (MOA), was also considered for this experiment as it is a much more interactive than the staircase method and, suggested by Cardozo (1965), it is suitable for stimuli in which there may be more than one attribute that can change, which is the case here because ASW will change as well as colouration. However, the reasoning put forward by Cardozo (1965) is unverified. An expanded discussion on threshold detection test methods can be found in Chapter 5, Section 5.1.1.1. For the test, subjects were presented with the interface shown in Figure 6.2.



Adjust the parameter and find the minimum point
where the 'tonal colouration' becomes inaudible

Trial 1 of 50

Fig. 6.2: The interface used in this test, created in Max 7.

For each trial, subjects were asked to listen carefully to the tonal colouration of the stimulus and adjust a hidden parameter (in this case the reflection angle) up or down until it becomes acceptable or inaudible, depending on the session, then proceed onto the next trial. The nature of the tonal colouration that they may perceive during the test was described to them using the glossary created in the elicitation test. A familiarisation stage was performed prior to actual testing in order for subjects to get used to the different types of tonal colouration at each delay time. The main test was conducted over two sessions containing ten trials for acceptability and audibility. Due to the relatively low number of subjects, five repetitions of each trial were performed.

To ensure subjects find a threshold point, subjects were instructed to explore the full range of reflection angles, then settle onto a point where they are sure the tonal colouration has become acceptable / inaudible. Subjects were also provided with an anchor where the reflection angle is at 0° or 180° depending on which region is being tested, in order to remind themselves what the tonal colouration is supposed to sound like. They were advised to not directly compare the stimulus to the anchor in terms of relative acceptability / audibility, and only to seldom use it as a reminder and only focus on when the tonal colouration of the test stimulus has become acceptable / inaudible. Finally, subjects can mark the tonal colouration within the trial as being "Still Unacceptable" / "Still Audible" if are unable to find a suitable point after exploring the full range of reflection angles. In this scenario, they were instructed to position the parameter where they felt the attribute was the closest to becoming acceptable / inaudible. Subjects were strongly advised to only use this as a last resort because could invite inaccurate results, particularly if a subject is not experienced. It was found that these 'flags' were rarely used throughout the experiment.

6.2.3 Apparatus

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dB $\mathrm{LA}_{\mathrm{eq}}$ by playing pink noise from one headphone cup, and measuring using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

6.2.4 Subjects

For this test, a group of eight subjects consisting of music technology students and researchers at the University of Huddersfield participated. Four subjects have experience with spatial audio and critical listening, whilst the remaining subjects have mixed critical listening ability and, because they took part in the elicitation test, were already familiar with tonal colouration. It is recommended in ITU-R BS.1116 (2015) to employ experienced listeners that can discern fine perceptual differences, however due to a constraint in the number of experienced listeners available during the testing period, subjects with mixed listening ability had to be included in the experiment to increase the number of observations.

6.2.5 Results

Prior to any statistical analysis, a Shapiro-Wilk test for normality was performed to deduce the suitability of performing parametric tests upon the data (Shapiro & Wilk, 1965). It was found that the data for a majority of conditions did not have a normal distribution. This meant that non-parametric tests were chosen for statistical analysis.

6.2.5.1 Effect of delay time on audibility threshold angle

Figure 6.3 shows the median audibility threshold angles and the notch edges, or equivalent 95% confidence interval (McGill et al., 1978), per delay time. Here colouration is deemed audible if the reflection angle is either below the threshold angle in the 'Front' region, or above the threshold angle in the 'Rear' region. A Friedman test with Bonferroni adjustment (α = .05) found that for both the front and rear regions, reflection delay time had a significant effect on the audibility threshold angle ($p < 0.05$). To further investigate which delay time pairs were significantly different, a Wilcoxon signed-rank test was performed.

For the front region, the test found that there was significant difference in audibility threshold angle between 2.5 and 30 ms, yet no significant difference between all other delay time pairs. Whilst the Friedman test found that there was a significant effect, the Wilcoxon test shows that there is only a significant difference between the two extreme pairs ($p < .05$), yet not between other pairs ($p > .05$). This agrees with the 'Front' sub-plot in Fig. 6.3 where there is overlap between non-significantly different pairs of notch edges. Whilst the null hypothesis can be retained, it is unclear if delay time has a significant effect on the audibility threshold angle.

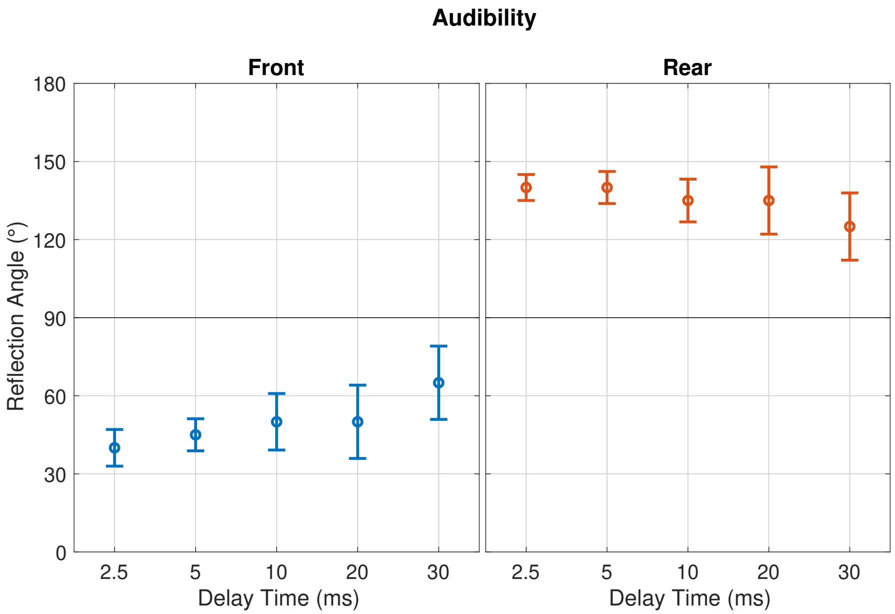


Fig. 6.3: Median audibility threshold angles per delay time, along with notch edges, or equivalent 95% confidence interval.

For the rear region, the test found that there was a significant difference in threshold angle between 2.5 and 30ms, and 5 and 30 ms. However, this does not agree comparing with the 'Rear' sub-plot in Fig. 6.3 where this overlap between all notch edges which implies that there is no significant difference between the pairs ($p >$ .05). Whilst there could be Type-I error in that particular sub-plot, the test found that there was no significant difference between other pairs. Like the front region, whilst the null hypothesis can be retained, it is unclear if delay time has a significant effect on the audibility threshold angle.

Upon further observation of Fig. 6.3, in the front region there appears to be a slight positive trend in median values as the reflection angle increase. At 2.5 ms the median is 40°, whilst at 30 ms the median is 65°, which is a 25° difference. In the rear region, there is also a slight negative trend in median values as the reflection angle continues

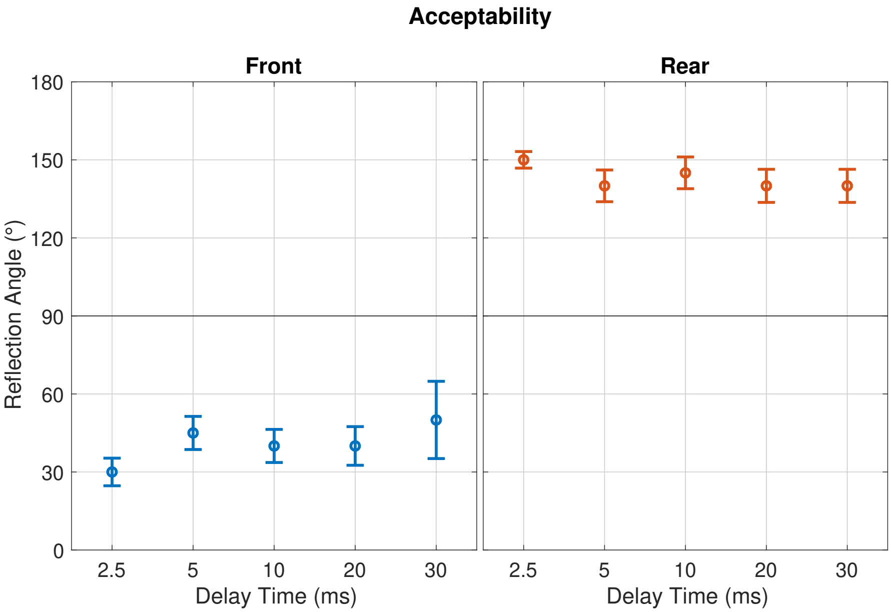


Fig. 6.4: Median audibility threshold angles per delay time, along with notch edges, or equivalent 95% confidence interval.

to increase from 90° to 180°. Here at 2.5 ms the median value is 140°, whilst at 30 ms it is 125°, a 15° difference. It could suggest that delay time is having a small effect on the audibility threshold angle, however, the statistical analysis did find that it had no significant effect.

6.2.5.2 Effect of delay time on acceptability threshold angle

Figure 6.4 shows the median audibility threshold angles and the notch edges, or equivalent 95% confidence interval (McGill et al., 1978), per delay time. Here colouration is deemed ’Un-acceptable’ if the reflection angle is either below the threshold angle in the ’Front’ region, or above the threshold angle in the ’Rear’ region. A Friedman test (α = .05) found that for both the front and rear regions, reflection delay time had a significant effect on the audibility threshold angle (p <

0.05). To further investigate which delay time pairs were significantly different, a Wilcoxon signed rank test was performed. Bonferroni adjustment was applied to the p values.

For the front region, the test found that there was a significant difference in threshold angle between 2.5 and 5ms, yet no significant differences between all other pairs ($p >$ .05). Whilst the Friedman test found that there was a significant effect, the Wilcoxon test shows that there is only a significant difference between 2.5 and 5ms ($p <$ .05). This agrees with the 'Front' sub-plot in Fig. 6.4 where there is overlap between non-significantly different pairs of notch edges. Thus, the null hypothesis must be retained.

For the rear region, the test found that there was also a significant difference in threshold angle between 2.5 and 5ms. This agrees with the 'Front' sub-plot in Fig. 6.4 where there is overlap between non-significantly different pairs of notch edges. Like the front region, the null hypothesis must be retained. It can then be concluded that the delay time had no significant effect on acceptability threshold angle in either region. Upon further observation of Fig. 6.4, there does not appear to be any notable trends. However, in comparison it should be noted that the median acceptability thresholds are lower than those for audibility in the front region, and higher in the rear region, see Table 6.4.

Median Threshold Angles (°)						
		Delay Time (ms)				
Region	Attribute	2.5	5	10	20	30
Front	Audibility	40	45	50	50	65
	Un-acceptability	30	45	40	40	50
Rear	Audibility	140	140	135	135	125
	Un-acceptability	150	140	145	140	140

Table 6.4: Median threshold angles (°) per delay time (ms) for both regions and attributes.

6.2.6 Objective analysis

Figures 6.5 to 6.9 show the left and right channel spectra of the BRIRs used in the experiment at several reflections angles between 0° and 180°, and at each delay time. The initial observation that was noted was the expected narrowing of the distance between the comb-filter teeth as the delay time increases between each figure. However, upon further inspection, the most notable finding was how the depth of the teeth in the left channel in frequencies above approximately 1 kHz seem to reduce as the reflection angle increase from 0° to 45°, or decreases from 180° to 135°. This is observed at all delay times.

This reduction is possibly due to the ILD between the left and right channels increasing as the reflection direction moves towards right of the listener. It suggests that subjects were perceiving less audible or distracting comb-filtering as a whole as the reflection direction approaches 90°. Also, because the reflection direction transitions from being frontal to lateral, the expected increase in width may lead to a more natural comb-filtering effect, much like in a real room. However, because this reduction appears to be only observed above 1 kHz, it may be that a reduction in high-frequency colouration contributes to effects such as ′roughness′ or ′sharpness′.

Further testing would be required to see if this is the case.

With regard to the direct effects of reflection direction upon the perception of colouration, it was found by Ando (1977) that a lateral reflection arriving at approximately 55° created a more preferable listening experience. It was speculated that this due to the enhanced ASW and reduced colouration, although the linkage to a reduction in perceived colouration was not further investigated. This analysis and results from this experiment appear to support this speculation, as it can be seen that there is a reduction in comb-filtering in one of the ear channels around the obtained threshold angles.

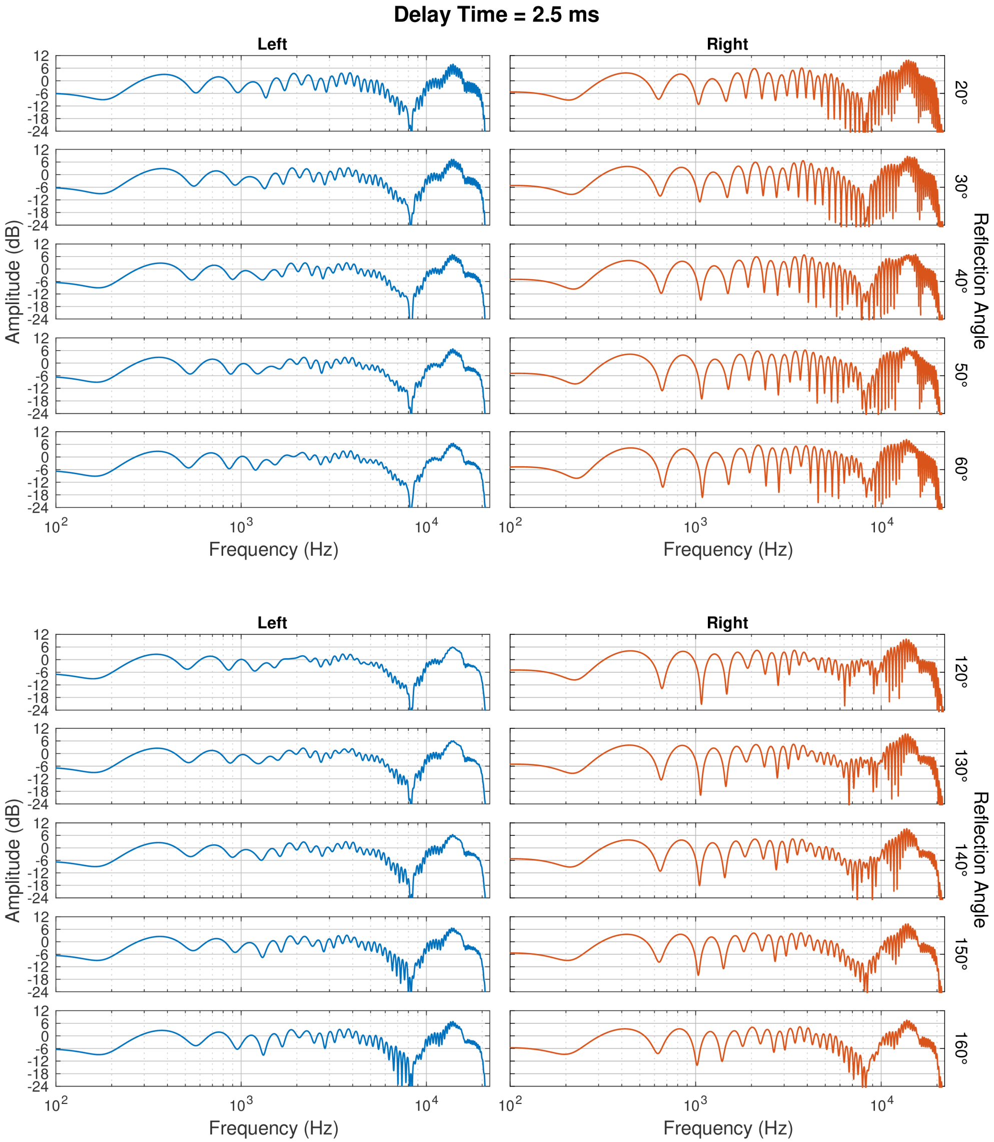


Fig. 6.5: Left and right channel spectra of BRIRs at several reflection angles around the threshold angles, at a delay time of **2.5 ms**.

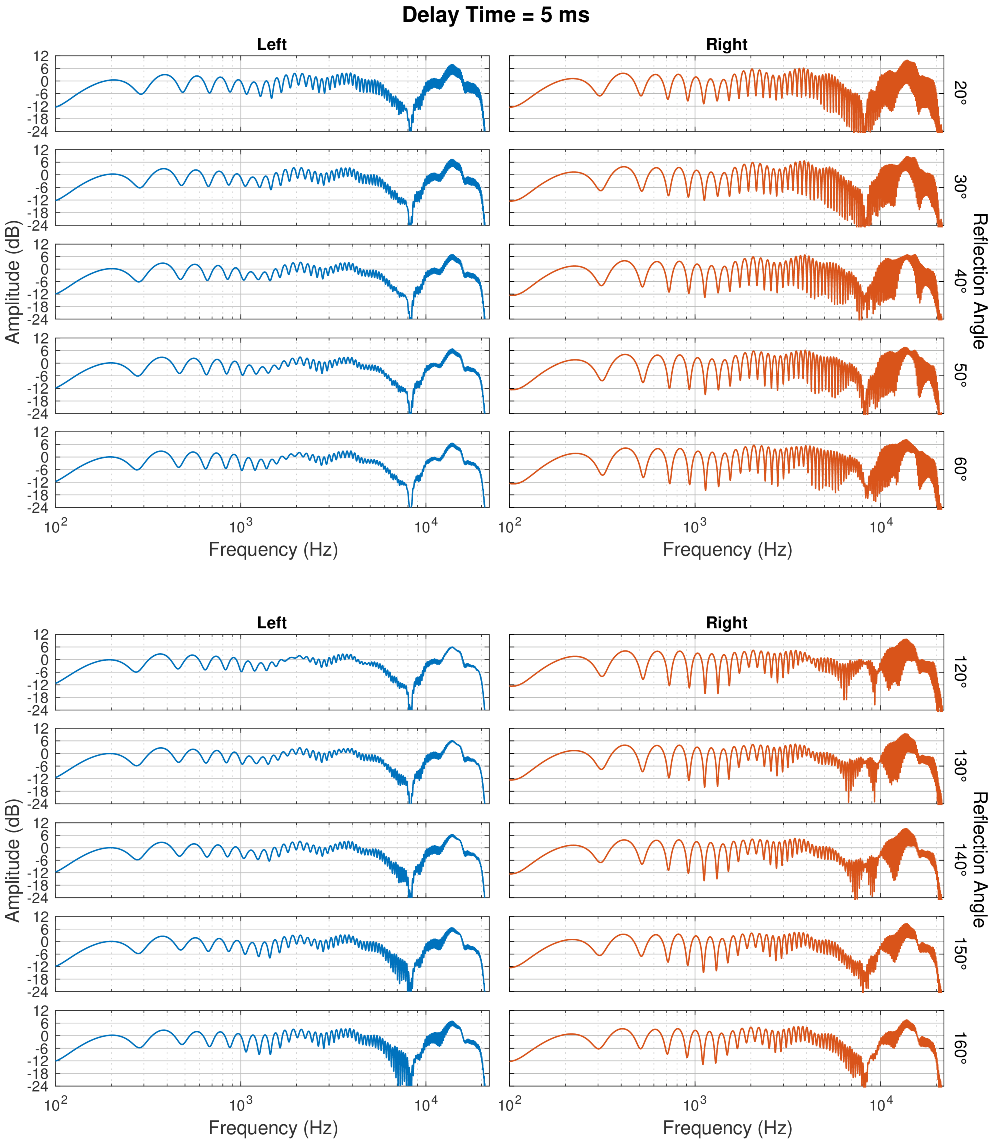


Fig. 6.6: Left and right channel spectra of BRIRs at several reflection angles around the threshold angles, at a delay time of **5 ms**.

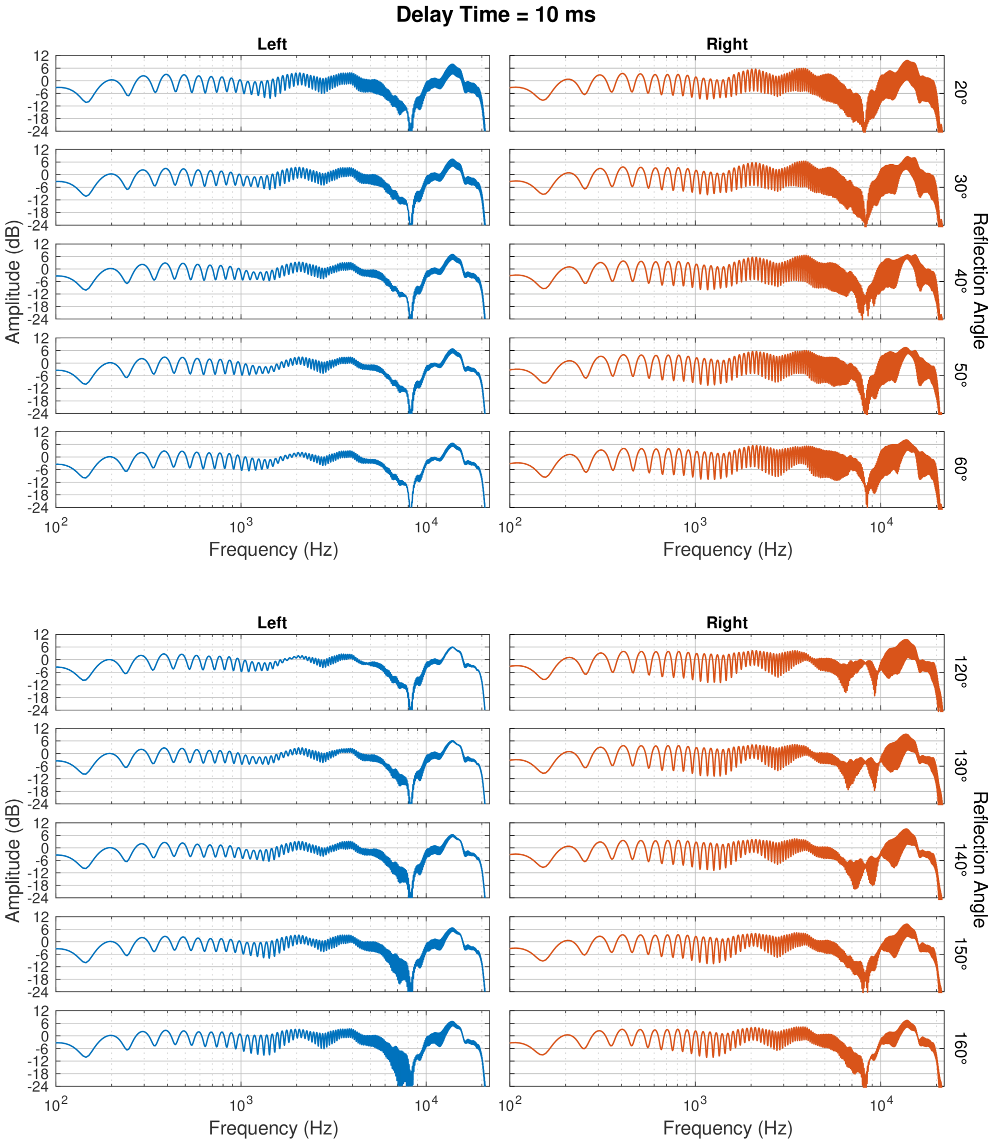


Fig. 6.7: Left and right channel spectra of BRIRs at several reflection angles around the threshold angles, at a delay time of **10 ms**.

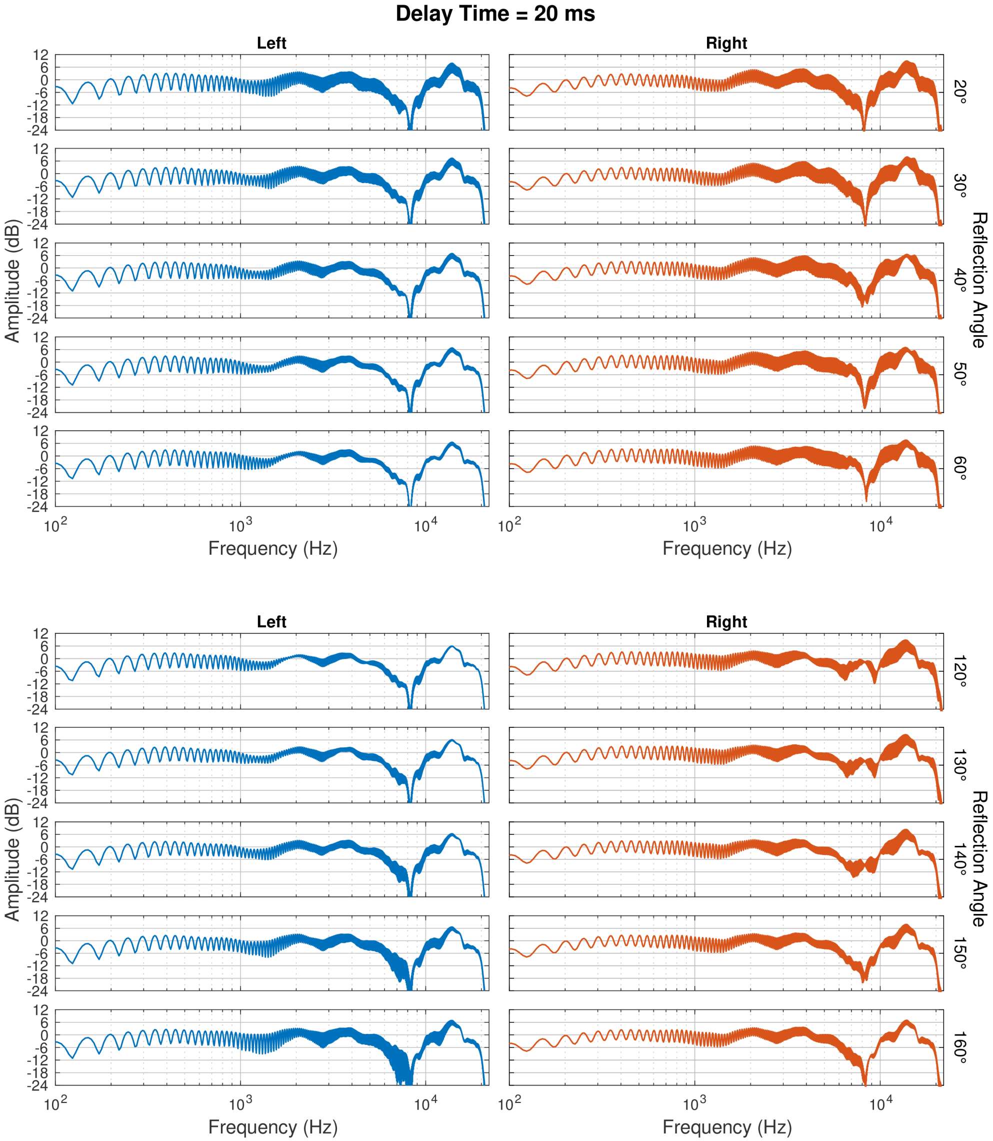


Fig. 6.8: Left and right channel spectra of BRIRs at several reflection angles around the threshold angles, at a delay time of **20 ms**.

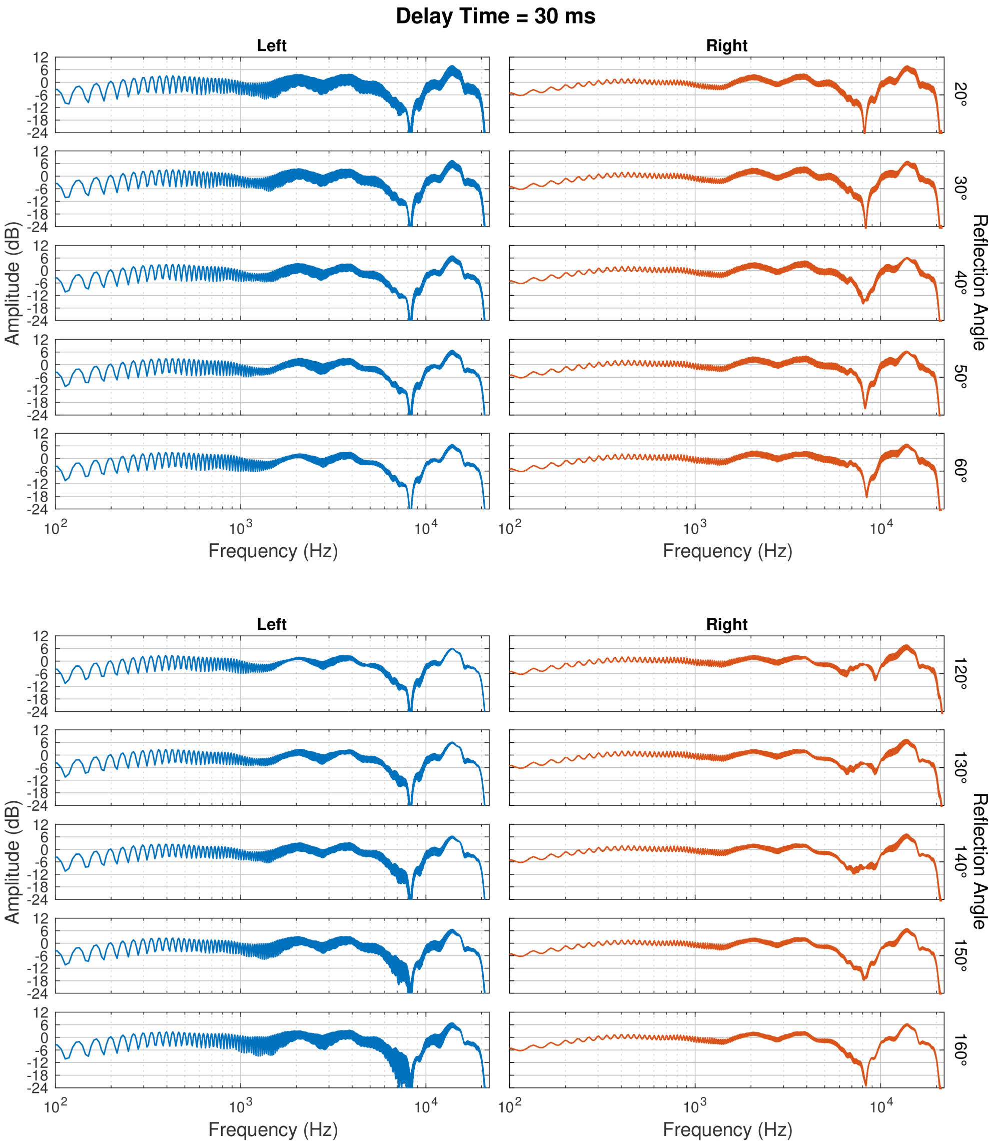


Fig. 6.9: Left and right channel spectra of BRIRs at several reflection angles around the threshold angles, at a delay time of **30 ms**.

6.2.7 Regions of colouration audibility and unacceptability

Since there was no significant effect of delay time upon threshold angle for both attributes, it was decided that the median angle for each region for each attribute could be calculated, see Table 6.5. The corresponding 95% confidence interval error plots for each threshold angle are presented in Figure 6.10.

Attribute	Front	Rear
Audibility	50	135
Un-acceptability	41	143

Table 6.5: Median threshold angles (°)

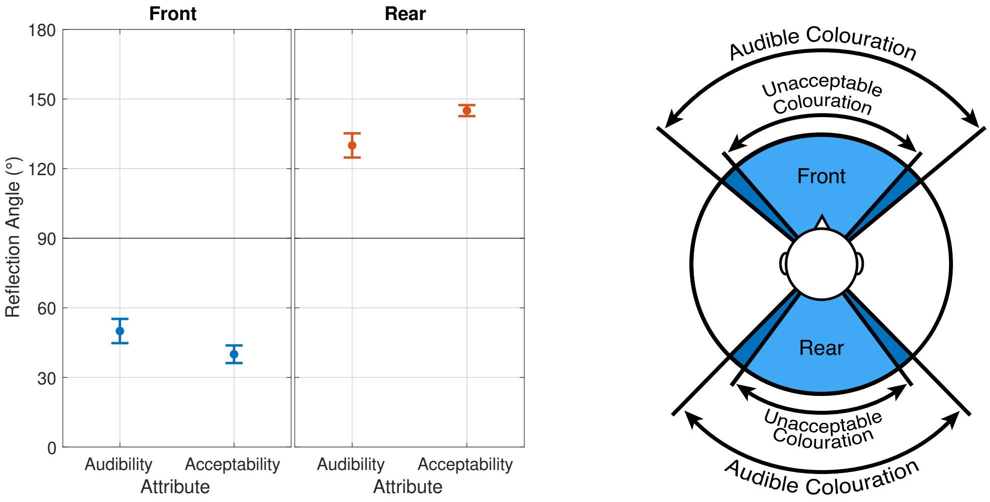


Fig. 6.10: a) Left: Equivalent 95% confidence interval notch regions for each attribute and median reflection angle. **b)** Right: Top-down view of the regions defined by the boundary angles.

From observing Figure 6.10a, it can clearly be seen that there is no overlap between the notches for each attribute in each area. Since the region of un-acceptability lies within the audibility regions, it suggests that perhaps when subjects cross the edge of the un-acceptable region, they can still perceive colouration, yet find it to be acceptable. To verify whether this is the case, a Wilcoxon signed-rank test was performed between the attributes for each area. The test found that there is a very significant difference between the edges for each attribute ($p < .01$). This necessitates the need for a set of regions to be independently defined for each attribute.

Like the ASW_{max} region discussed in Chapter 5, regions of unacceptable and audible tonal colouration can be created in front of and behind the listener. The region of audible colouration lies within 0° to 50° in front of the listener, and 135° to 180° behind the listener. Similarly, the regions of unacceptable tonal colouration lie within 0° to 41° in front, and 143° and 180° behind. These boundaries define two regions where a single reflection arriving between 2.5 to 30 ms and within either region would create subjectively unacceptable tonal colouration, see Figure 6.10b. Inversely, whilst similar reflections arriving outside of the regions may still create audible colouration, the degree would be regarded as subjectively acceptable. Since one of the goals of this study is to create a method of controlling the tonal colouration of reverb to make it more acceptable, the region of unacceptability will only be used for the final part of this experiment, and by extension as part of the control method.

Interestingly, the boundaries of this region lie adjacent to the boundaries of the ASW_{max} region. This potentially means that when grouping the reflections based on their angle of arrival, there may be an 'un-sortable' region or gap lying in between these the ASW_{max} and unacceptability regions, as in these reflections create acceptable colouration yet do not cause the greatest amount ASW. Whilst it is not possible to directly test the two data sets of the first experiment performed in Chapter 5 to this to see if there is any significant difference between the boundary edges, because of their close proximity the 'grey area' between these two regions, for simplicity they will be considered to be part of the 'un-acceptable colouration' region.

6.3 Experiment part 3: Effect of reflection level

Now that the regions of unacceptable colouration has been defined, the next task was to determine the absolute acceptability thresholds depending on reflection level. It is hypothesised that the reflection level will affect the acceptability of the tonal colouration based on the audibility threshold of colouration (Brunner et al., 2007). If the colouration is inaudible, it would not be perceived so would be regarded as totally acceptable. However, in any echoic space, the reflections will interact with the direct sound and create audible colouration, yet it could be regarded as pleasant or acceptable. Therefore, is is likely that there is a threshold level at which the colouration just becomes unacceptable. The colouration could be made more acceptable by simply reducing the reflection level. This forms the following research question:

In relation to the direct sound, by how much does a -6 dB reflection need to be reduced in order to produce acceptable colouration?

6.3.1 Test method

Since it was assumed that only the colouration attribute would change with reflection level, the same transformed staircase test developed by Levitt (1971) was used as this allowed for the threshold to located with higher precision than the 'Method of Adjustment'.

6.3.2 Stimuli

The process used to generate the stimuli in this test was similar to the process used in the previous section. However, rather than altering the reflection direction, it is fixed to two angles, 20° and 40°, and the level difference is instead changed. The level ranges between -48dB and -6dB, and is changed initially in 3dB steps, then 1.5dB steps after the first reversal. The 48dB level difference is well above the average difference of 18dB found by Brunner et al. (2007).

6.3.3 Interface

The interface used for the test in Section 5.1 was adapted to create an absolute threshold test, see Figure 6.11. For each trial subjects are presented with a button that would trigger playback of the stimulus. The subjects were instructed to listen carefully to the stimulus, and determine if the perceived colouration is acceptable. This was performed without a reference stimulus so that an absolute threshold rather than a relative threshold could be found. Again, the interface could either be controlled using a mouse or a computer keyboard, where keys were mapped to both the stimuli buttons and response buttons. The test terminates after either 20 reversals or 128 trials. Pilot testing revealed that subjects can potentially become stuck back at the initial reflection level, finding this colouration in this condition to be acceptable. In this case the step direction may be unable to reverse, meaning the number of required reversals would not be met, and so the subject would otherwise have to observe all 128 trials. If this event would occur, the test modified to automatically terminate after 10 'Yes' responses when the step is stuck at this minimum value. This approach was chosen on basis that subjects who felt this condition to be acceptable

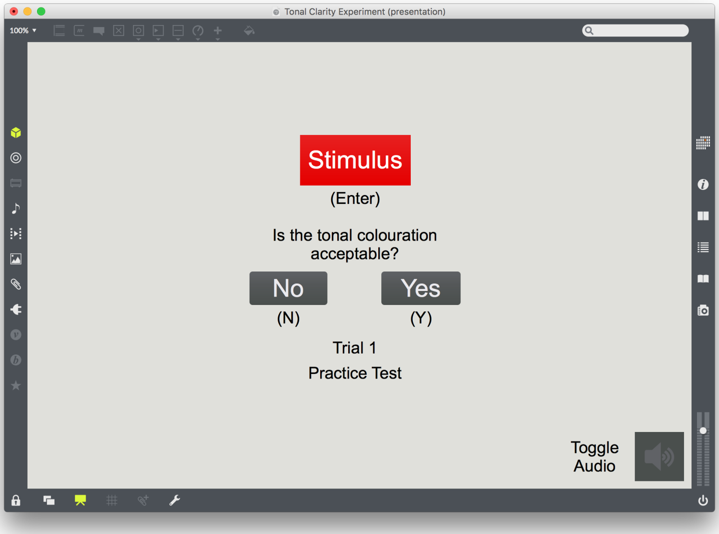

Tonal Clarity Experiment (presentation)
100%
Stimulus
(Enter)
Is the tonal colouration acceptable?
No
(N)
Yes
(Y)
Trial 1
Practice Test
Toggle Audio

Fig. 6.11: The absolute threshold test interface, developed in Max 7.

would be likely to repeatedly give the same response, yet it gives subjects a chance to potentially reduce the gain within an adequate number trials. The trial counter for this condition is reset to 0 when the gain value becomes unstuck. All subjects completed 20 reversals in each test session with an average time of ten minutes. Whilst this suggests that the 'fail-safe' was unnecessary, the pilot testing still revealed that it was possible, and so this feature was still implemented in the final experiment design.

Prior to main testing, subjects performed two training trials in order for them to be familiarised with the test interface and the colouration present in the stimuli.

6.3.3.1 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dB LA_{eq} by playing pink noise from one headphone cup, and measuring using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

6.3.3.2 Subjects

For this test, a group of eleven subjects consisting of music technology students and researchers at the University of Huddersfield participated. Six subjects have experience with spatial audio and critical listening, whilst the remaining five subjects have mixed critical listening ability. It is recommended in ITU-R BS.1116 (2015) to employ experienced listeners that can discern fine perceptual differences. However, due to a constraint in the number of experienced listeners available during the testing period, subjects with mixed listening ability had to be included in the experiment to increase the number of observations.

6.3.4 Results

Figure 6.12 shows the median values of the mean level reduction per delay time. Prior to any statistical analysis, a Shapiro-Wilk test for normality was performed to

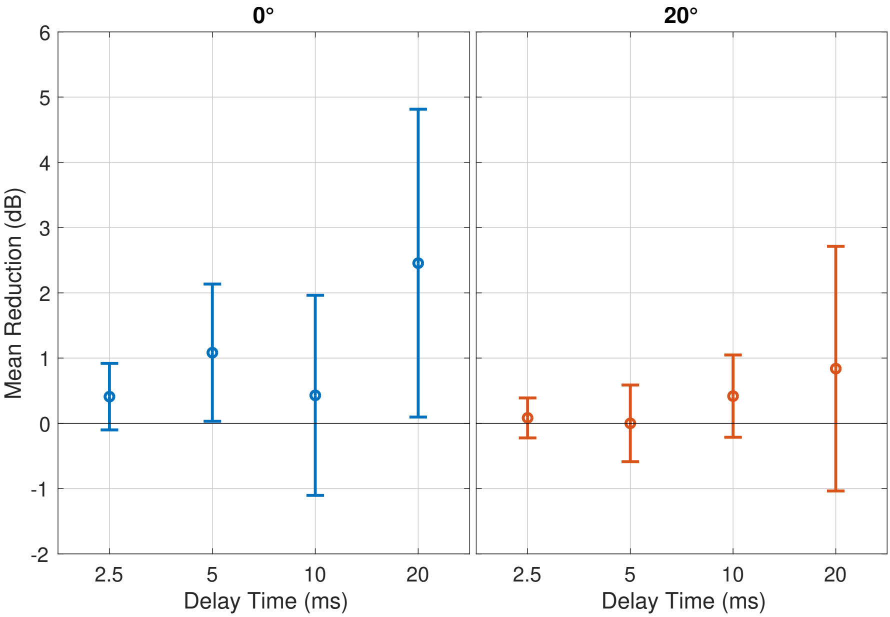


Fig. 6.12: Median values of the mean level reduction applied to the reflection at different delay times, in order to make the colouration acceptable.

deduce the suitability of performing parametric tests upon the data. It was found that the data for a majority of conditions did not have a normal distribution. This meant that non-parametric tests were chosen for statistical analysis.

6.3.4.1 Effect of delay time on the mean level reduction

To test if there is any effect of delay time upon the mean reduction level for either region, a Friedman test was performed in MATLAB with Bonferroni adjustment and a significance value of $\alpha = .05$. The test found that there was no effect of delay time on the mean reduction level ($p > .05$). This is supported by the overlap between notch edges in both sub-plots of Fig. 6.12. Whilst there is no significant difference between delay time pairs, with a 0° reflection angle there appears to be a much a larger mean level reduction of 2.5 dB at 20 ms. At this delay time, there should be

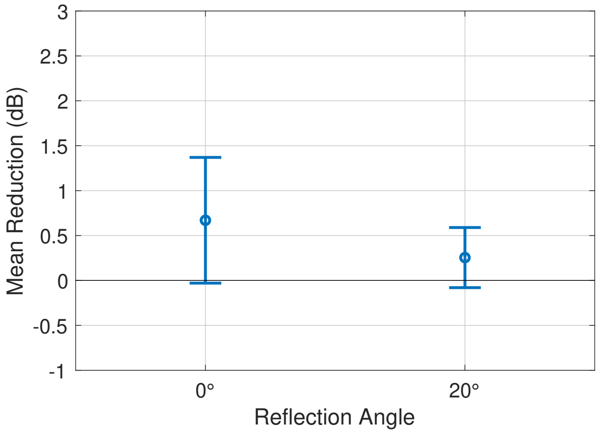


Fig. 6.13: Median values of the combined mean level reduction for all delay times applied to the reflection.

a noticeable pitch centred at 50 Hz, which produces a metallic, pitched quality of colouration which may require more level reduction to be acceptable. At 20° at the same delay time, the median average level reduction is much smaller at just under 1 dB which is a relatively small reduction.

6.3.4.2 Effect of reflection angle on the mean level reduction

Since there is no significant difference in level reduction between delay times, the mean reduction values can be combined and grouped by the two reflection angles. The median values of the mean level reduction for each reflection angle are plotted in Figure 6.13. Another Friedman test (α = .05) performed in MATLAB with Bonferroni adjustment found that there is no significant difference in level reduction between the two reflection angles (p > .05). This is supported by the overlap between notch edges in both sub-plots of Fig. 6.13. In general, the median value is closer to 0 dB at 20° than at 0°, however the overall median reduction at both reflection angles is relatively small, being under 1 dB. Whilst the reduction levels are small, at 0°, potentially more reduction was required than at 20°, yet this is still unclear.

6.4 Discussion

Part one of the experiment established what types of tonal colouration would be perceived when a single, delayed reflection is mixed with the original direct sound.

Part two of the experiment discussed in Section 6.2 found that the perceived colouration became acceptable when the reflection arrived between 41° to 143°. Whilst these values are the average threshold angles, since the HRTF dataset used in the experiment has an azimuth resolution of 5° at 0° elevation, the values must be rounded down 40° and 140°. Between these two angles the colouration, whilst still likely to be audible, is acceptable. It is also worth noting that because the reflection angle is increasing, according to lateral fraction (Barron & Marshall, 1981) the ASW would also increase. It could be that the colouration becomes less noticeable as the auditory image fluctuates in width and position, as found in Chapter 5. This could cause the colouration effect to become less distracting as listeners begin to perceive an increase in width. This is made plausible due to acceptability threshold angles being adjacent to the ASW_{max} boundary angles. It can then be hypothesised that the colouration gradually becomes less distracting as the reflection angle approaches the ASW_{max} boundaries.

The analysis of part three of the experiment has found that at both reflection angles, 0° and 20°, that little level reduction was required to make the perceived tonal colouration become acceptable. When compared to the results in part two, this result was unexpected. In part two, a reflection was deemed unacceptable at 0°, yet required little further reduction in part three. This could have occurred for either of or both of the following two reasons: the lack of an anechoic reference does meant that subjects had nothing to compare the colouration of the stimulus against, and so the 0 dB reduction scenario may have already been acceptable; the subjects adapted to the colouration during the test such that they eventually deemed the 0 dB reduction.

When investigating each individual subject's responses, the latter reason appears to the case, particularly for the 0° scenario. A trend begins to emerge where five out of the eleven subjects find the initial colouration to be unacceptable and apply a degree of reflection level reduction. Yet, as the test progresses, they appear to become conditioned to the colouration and the amount of level reduction drifts back towards the initial value. In some cases, subjects drift completely back to the initial value meaning that their internal reference for what they consider an unacceptable degree of colouration may change depending on the amount of time that they are exposed to the phenomenon. Thus, it is possible that what these subjects considered unacceptable required a level difference that was perhaps beyond the initial step value of the experiment. These psychological drifts in subject responses during prolonged testing have been discussed in general by Zielinski, Rumsey, and Bech (2008).

6.4.1 Experimental limitations

Whilst the findings from this experiment are novel, there are a number of limitations in the methodology.

Firstly, the test performed in Section 6.2 is valid only for a -6 dB reflection level. This level, however, was deemed realistic and mimics the absorption of a first order reflection in a concert hall, where the first order wall reflection often has significantly less energy than the direct sound. It was also chosen to be consistent with the reflection level used in the experiment performed in Chapter 5.

Next, the experiment may be subject to anticipation bias from a few subjects who had an awareness that the reflection angle would be manipulated in subsequent tests. They were not, however, aware of the parameters and mechanics of the threshold tests, and so it is unclear if their awareness had any effect. Whilst they could have been excluded from the test, it was decided to include them as they are regarded as experienced critical listeners, and without their participation the number of subjects would have been too low for this experiment. Investigation of the test results performed by the author found that the responses provided by the four biased subjects were similar to unbiased subjects, thus the possibility of an anticipation bias affecting the experiment was deemed inconclusive.

Thirdly, for the final test performed in Section 6.3, it is possible that using a -6 dB limit for the reflection level, as well as using -6 dB for the initial level, produced what could be already acceptable colouration and may not have required much more reduction. Again, as discussed in the first point, it was thought to have not caused a limitation as it was assumed that a -6 dB would mimic what is exhibited in a concert hall. This test did not anticipate what the actual reflection level relative to the direct sound level should be for the colouration to be perceived as unacceptable. Since subjects were able to identify the colouration as being unacceptable in the previous parts of the experiment, where an angular region of acceptable colouration was identified in Section 6.2, it was further assumed that the initial -6 dB level did cause unacceptable colouration. Further investigations using a lower level difference between the direct sound and the reflection shall be performed in the future to overcome this limitation.

6.5 Conclusion

This chapter has investigated the effects of level and angle of a single reflection on the audibility and acceptability of tonal colouration on an anechoic speech source. The experiments conducted were as to expand the work performed by Halmrast (2000), Seki and Ito (2003) and Brunner et al. (2007) to further understand what would be considered as subjectively unacceptable colouration using a psychophysical testing method.

Part one was designed to establish the tonal colouration that may be perceived with the anechoic speech source, and to decided on a set of terms that could be used in the subsequent tests. Part two used the Method Of Adjustment, or MOA, psychometric test to establish what the effect of reflection angle on audibility and acceptability, and to find threshold angles for both attributes as to define areas of ′unacceptable colouration′.Finally, part three then investigated the effect of reflection level on acceptability, where the objective was to understand how much level reduction needs to be applied in order to make the colouration become subjectively acceptable. This was performed using a modified version of the Levitt (1971) transformed staircase test to find the absolute threshold level, and then derive how much reduction was to be applied to make the colouration acceptable. The key findings are as follows:
Experiment part 1:

- The most common term that elicited was Metallic, occurring 17 times.

- Delay time appeared to affect the characteristic of the tonal colouration, where at low delay times the sound was perceived as 'harsh', 'metallic' and 'phasey', whilst at higher times there was perceived 'Fullness' and 'Modulation'.

- For almost all delay times, the common perceived quality is 'Roughness'.

Experiment part 2:

- When an early reflection attenuated by -6 dB arrives between 50° and 135°, the colouration is perceived as inaudible. Thus, these two angles define a region of inaudible colouration.

- When the same reflection arrives either ±41° or ±143°, the perceived colouration may be unacceptable, thus defining regions of unacceptable colouration.

- Whilst the audible and unacceptable regions are coincident, there is a significant difference between their boundary angles, such that there are transition areas where the colouration may be audible yet acceptable.

- Delay time had no significant effect on the position of the boundary angles of any region.

- The regions of unacceptable colouration are adjacent to the ASW_{max} region defined in Chapter 5. Therefore, there could be a relationship between ASW and the acceptability of tonal colouration.

Experiment part 3:

- For a -6 dB reflection, at both 0° and 20° very little level reduction was required to make the colouration become acceptable.

- Whilst there was no significant effect of delay time on the reduction at either reflection angle, for 20ms delay time at 0° the median value of mean reduction was higher than at lower delay times.

- The reflection angle did not have a significant effect on the mean level reduction. However, the median value at 0° was higher, which potentially means more reduction is required for frontal reflections rather than lateral reflections.

In conclusion, the findings of the experiment provide a novel insight into how the audibility and acceptability of colouration change with reflection angle and level. The regions of unacceptable colouration in front of and behind the listener can be used as part of another perceptual control method that can be used to reduce the level of reflections that cause unacceptable colouration to improve the tonal colouration of a virtual space.

Chapter 7

Application of the perceptual control methods: Part 1

The previous chapters have researched and identified two regions in the horizontal plane that can be used to group reflections as those having the greatest influence on either ASW or tonal colouration. Both chapters performed fundamental experiments using a single reflection. Chapter 5 identified a region of maximum ASW, known as ASW_{max}, in which any single reflection arriving within this region produce the maximum amount of perceived ASW. After interpreting the findings of Barron and Marshall (1981), adjusting the level of reflections within this region as a group is hypothesised as having the greatest influence on ASW. Chapter 6 identified regions of unacceptable tonal colouration in front of and behind the listener. It was also found that reducing the level of these reflections caused the subjective acceptability of the colouration to improve.

This chapter will now demonstrate how those regions can be used as part of a perceptual control method where early reflections can be selectively level adjusted in order to control either ASW or colouration. The objective of this chapter is to assess the effectiveness of the perceptual control methods at manipulating either attribute independently when applied to virtual room acoustics. The investigation is split into three parts: the first will briefly discuss the proposed method of perceptual control that incorporates the regions identified in Chapters 5 and 6. The second part will investigate the perceptual effects of the proposed methods using an elicitation method. This test was of particular importance since it was initially unclear as to what aspect of tonal colouration the listeners should focus on. Part three then assesses the effectiveness of the method on various 'raw impulse vectors' (RIVs) by measuring the degree of difference they had on their associated attribute. RIVs of two virtual rooms, a typical concert hall and a large room, were rendered by the custom, artificial reverberation algorithm discussed in Chapter 3.

An important point to reiterate in this investigation is that the proposed methods could be taken from being simple, perceptual attribute controls and be developed to be part of an optimisation method. It could identify reflections that contribute to what is possibly a positive or negative attribute, and manipulating them in order to improve the overall listening experience. However, it should achieve this regardless of the room and source type, and more importantly it should do so by not affecting other attributes, otherwise it may lead to negative artefacts or side effects. Thus, for the perceptual control methods to be considered effective, they should ultimately solely affect their corresponding attribute.

From this discussion, the following research questions were asked:

1. *What are subjective effects of the proposed perceptual control methods?*

2. *To what degree will manipulating the gain of reflections within the ASW_{max} region*

affect ASW?

3. *To what degree will manipulating the gain of reflections within the unacceptable colouration region affect colouration?*

7.1 Proposed control method

As discussed in Chapter 3, the RIV file that the custom algorithm produces allows for perceptual control to be applied before the rendering of the BRIR. First, early reflections that arrive within 80ms of the direct sound are extracted. Next, depending on the attribute that is being manipulated, reflections that fall within a respective region are further isolated. The broadband energy of these reflections can then be manipulated. The direct sound in both types of methods is unaffected. Figure 7.1 shows a visual representation of the regions that will be used to select captured rays from an example RIV. Barron and Marshall (1981) found that the perceived ASW is dependent on the level of the lateral reflections, therefore, the ASW can be controlled by adjusting the level of extracted reflections that arrive within the ASW_{max} regions.

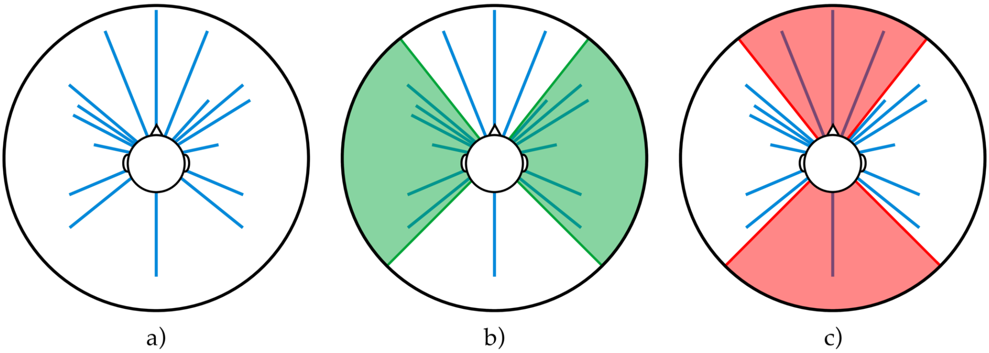


Fig. 7.1: Top-down, visual representation of early rays extracted from an example RIV **a)** with two regions being used to select rays for manipulation, where in **b)** the lateral highlighted areas represent the ASW_{max} region, and in **c)** the front-back areas represent the regions of 'unacceptable colouration.

Likewise, as demonstrated in the previous chapter, the colouration can be improved or controlled by reducing the level of the reflections that fall within the ’unacceptable colouration’ regions. This method of gain control can now be applied to RIVs of virtual rooms generated by the custom algorithm for the main experiments.

7.2 Experiment part 1: Formal elicitation of attributes

Building upon the first research question, implementing and grading the control method directly without knowledge about all the possible attributes it may affect would not be ideal. The previous experiments in this study were performed using a single reflection, yet in this chapter the perceptual control is being applied to a reverberant sound field. Previous literature has identified that a change in a given attribute can be interlinked with changes in other attributes, for example Wallis (2017) found that the inclusion of a delayed signal from an elevated loudspeaker can affect fullness, loudness, source distance, and clarity simultaneously.

Further to this, as discussed in the previous chapter, tonal colouration is a broad attribute and may be ambiguous to listeners. It cannot be graded directly because listeners need to be able to focus on one particular aspect or sub-attribute of colouration. If they did not, then they may perceive different aspects of the colouration, and so it would be very unclear if the colouration control method was operating, and to what degree. Subjects should focus on the most salient sub-attribute of colouration. Therefore, before grading the optimisation methods, a formal elicitation test must be performed to identify what attributes the level adjustment may affect, and to select the most salient such that the degree to which they are affected can be identified.

The objective of this part of the experiment was to elicit attributes that describe the perceptual effects of different types of spatial filtering in virtual rooms and to the group them into common sets of attributes.

7.2.1 Experimental design

This test will be based upon a QDA (Quantitative Descriptive Analysis) method proposed by Stone and Sidel (2004). Bech and Zacharov (2007) states that this method consists of six test phases:

- Present subjects with stimuli and ask them to elicit attributes.
- Remove or group duplicate attributes.
- Further discuss attributes.
- Introduce test stimuli to activate attributes at a wide range of intensities.
- Introduce test stimuli with smaller differences.
- Perform actual testing where each elicited attribute is graded for each stimulus.

Whilst the benefit of QDA is its ability to produce a complete description of a stimulus (Stone & Sidel, 2004), it is a time consuming method for a multitude of reasons. Stone and Sidel (2004) state that the there should be around four to five training sessions, each lasting a duration of 90 minutes. To limit the duration of the entire process, Wallis (2017), whose study mainly focused on the perceptual effects of vertical inter-channel crosstalk, implemented a modified version of the QDA method. This version was based upon the approach taken by Lee (2006) whose approach presented the subjects with a list of potential attributes taken from previous studies related to a perceptual effect. Subjects graded the audibility of changes in each attribute. This data would then contribute to the main grading experiment. However, what differed from the established QDA method, where grouping analysis was performed by all subjects in a group discussion, is that grouping was instead performed by the experimenter through self interpretation and informal discussion with subjects. The most salient of attributes were then used for the grading experiment. This reduced the process down to two experiments: elicitation then grading. Wallis (2017) argues that the presentation of a list potential attributes to subjects prior to any testing may create a bias, such that the attributes graded were based on the subject's own interpretation of the attribute and its description. Therefore, Wallis (2017) decided to remove the list of potential attributes during elicitation.

The method proposed by Wallis (2017) can be broken into two parts: a two stage elicitation test followed by an audibility grading experiment. The two stages or phases of the elicitation test are a free elicitation phase followed by a group discussion. The free elicitation phase consists of presenting the subjects with two stimuli with a difference in effect between them. In a test performed by Wallis (2017), one stimulus has no vertical inter-channel crosstalk, whilst the other has the crosstalk applied at maximum. Subjects are tasked with eliciting any audible differences and terms between the two stimuli onto a blank sheet. The group discussion phase, the terms are presented all at once to the subjects. The task of the group is to fit terms into an attribute group. If no term fits into any group, then a new attribute group is created along with its own description. Francombe (2014), whose research focused on the perceptual evaluation of audio-on-audio interference, suggests splitting the group discussion into three sessions, each with a specific task:

- End-point definition.
- Attribute description.
- Attribute list combination

The total discussion time takes approximately five hours to complete which, as Wallis (2017) notes is difficult to achieve as this discussion requires 10 to 12 subjects who must be present for all stages. Wallis (2017) thus chose to use only one group session that focused on the attribute description creation.

The next stage is the audibility grading experiment. The objective of this experiment is to determine what are the most salient effects by considering the audibility of each attribute. The most salient attributes are then used in a later grading experiment. During the test performed by Wallis (2017), pairs of stimuli, one with and the other without vertical inter-channel crosstalk, were presented to the subjects. They then graded the audibility of a given attribute (e.g. loudness) between the two stimuli. The two most salient attributes, in this case ’vertical image spread’ and ’fullness’, were chosen to be the focus of two independent grading tests.

7.2.1.1 Test method

With the above considerations, it was decided that a modified and simplified version of the QDA method used by Wallis (2017) would be conducted for this study. The test is split into two parts:

1. Elicitation of attributes

2. Sorting and grouping of attributes

These two parts would be followed by the main test in Section 7.3. For the sake of efficiency and to reduce the entire experiment duration, it was decided that an audibility grading test was to not be performed. This is due to a number of reasons:

- The availability of subjects at the time of experimentation was limited.
- Pilot testing, as well as the assumptions regarding the relationship between reflection level and lateral fraction, found that source widening was very likely to be a salient attribute and thus would not need to graded in terms of audibility

As an alternative, the number of occurrences of the elicited terms will be used to indicate which attributes are the most salient. To ensure terms are not unnecessarily duplicated, which would bias the possible saliency of an attribute, subjects were advised to only use that term once per trial where possible.

Since the main focus was to elicit spatial and tonal differences, a list of contextual attributes for the subjects to consider was provided:

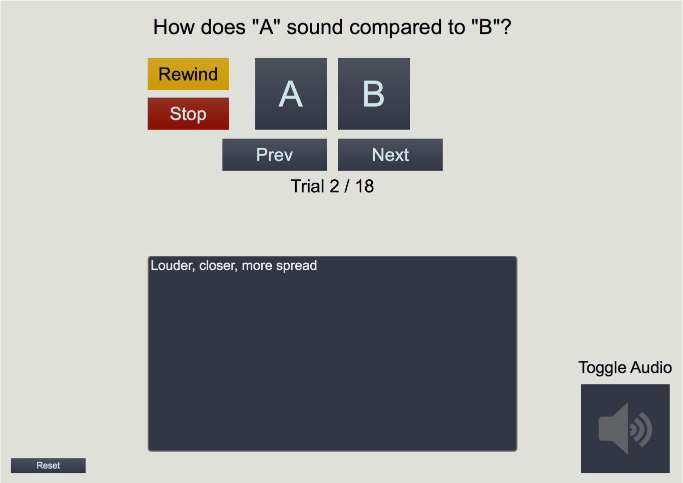
Contextual attributes
Spatial Impression
Width
Spread
Envelopment
Colouration
Loudness

Table 7.1: List of contextual attributes for subjects to refer to during the experiment

To avoid bias, descriptions of the contextual attributes were omitted in order for subjects to deduce their own definitions and terms such that they would elicit them. The subjects were not limited to providing only terms within these contexts.

7.2.1.2 Interface

The first part of the test was designed to acquire a set of elicited terms that described the spatial and tonal qualities of perceptual filtering. Figure 7.2 shows the interface used for this test.



How does "A" sound compared to "B"?

Trial 2 / 18

Louder, closer, more spread

Fig. 7.2: Interface used for the elicitation test designed in Max 7

For each trial, the subjects were presented with two stimuli, each with one of the following conditions: selected reflection levels are boosted by 6dB, which is the maximum amount used throughout this and subsequent tests; and unprocessed. The allocation of conditions to buttons A and B was randomised per trial. Subjects were instructed to carefully listen to each stimulus and elicit any perceptual differences in A compared to B. Terms were elicited into a text box on the listening test interface.

7.2.1.3 Stimuli

Several BRIRs of two simple yet distinctly different models, a large room and a concert hall, were created using the custom artificial reverb algorithm (see Figure 7.3). The rooms were designed with the following specifications:

Attribute	Concert Hall	Large 'Shoebox' Room
Dimensions (WxHxD)	16m x 10m x 22m	9m x 5m x 16m
Wall Material	Plasterboard	Plasterboard
Floor Material	Empty chairs	Wood
Ceiling Material	Plasterboard	Plasterboard
Source XYZ Coords (m)	(0.0, 1.75, 19.0)	(0.0, 1.75, 14.0)

Table 7.2: Virtual room specifications

The source in both rooms had an omni-directional directivity pattern. An omni-directional receiver was placed at 3m, 6m and 12m from the source at a height of 2m from the floor. The receiver in both rooms was slightly offset by 1cm to right to create a subtle de-coherence between the left and right channels in the early reflections during the BRIR synthesis stage, whilst also maintaining a central source image. The concert hall model is based upon the 'Example' model provided with ODEON 14.0, whilst the large room is based upon the Wendt et al. (2014) laboratory model with the depth extended to accommodate the chosen source-receiver distances. RIVs of each model at each distance were rendered using the custom algorithm described in Chapter 4. The RIVs were generated using both third order ISM for early reflections, and ray tracing for the late portion using 50 000 rays and a receiver radius of 0.25m.

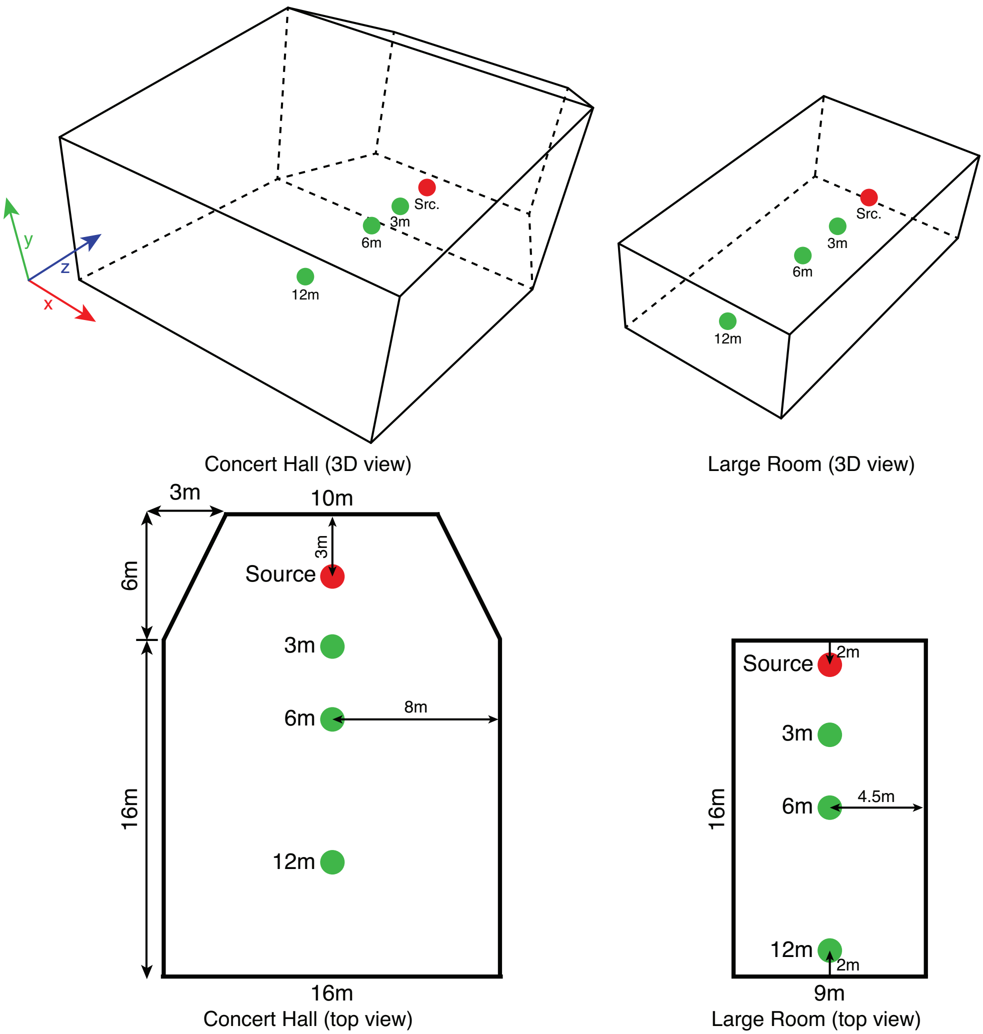


Fig. 7.3: Wire-frame diagrams of the two room models. **Left:** The example concert hall model available with the ODEON acoustics tool. **Right:** A large 'shoebox' shaped room.

To apply the proposed perceptual control method discussed in Section 7.1 and synthesise BRIRs, the spatialisation process discussed in Chapter 4 Section 4.1.1 was utilised, however, after rays are extracted from the RIV prior to BRIR rendering, rays are sorted using the appropriate region, level adjusted by up to ±6 dB in 3 dB steps, then re-injected into the RIV, after which the spatialisation method proceeds. Monophonic, anechoic male speech, classical guitar and orchestral excerpts are convolved with the BRIRs to produce the final stimuli. It is important to note that prior to convolution with the anechoic sources, no normalisation of the RIVs or BRIRs was performed. Since all RIVs were rendered using the same starting energy, or a digital level of 1.0, it is assumed the same amount of energy for all conditions was transmitted into the models during rendering. Therefore, no loudness matching was applied between stimuli so that effect of distance and room geometry upon reflection energy between renders was relative. The binaural room impulse responses that form the basis of the stimuli are presented in Figures 7.4 to 7.7. The frequency response of the BRIRs are presented in Figure 7.8.

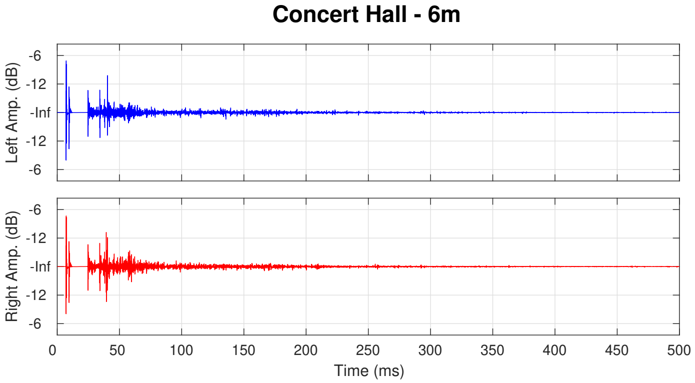


Fig. 7.4: BRIR of the Concert Hall at a source-receiver distance of 6m

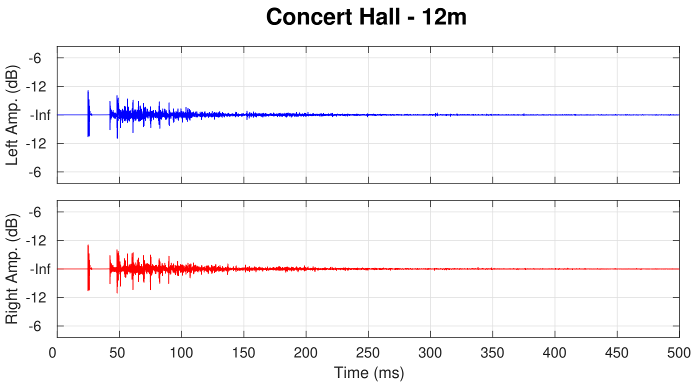


Fig. 7.5: BRIR of the Concert Hall at a source-receiver distance of 12m

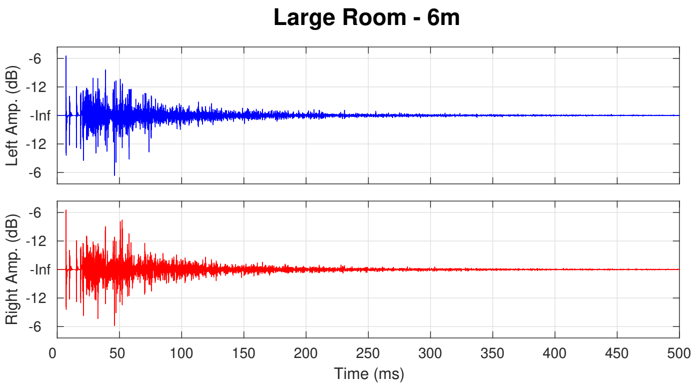


Fig. 7.6: BRIR of the Large Room at a source-receiver distance of 6m

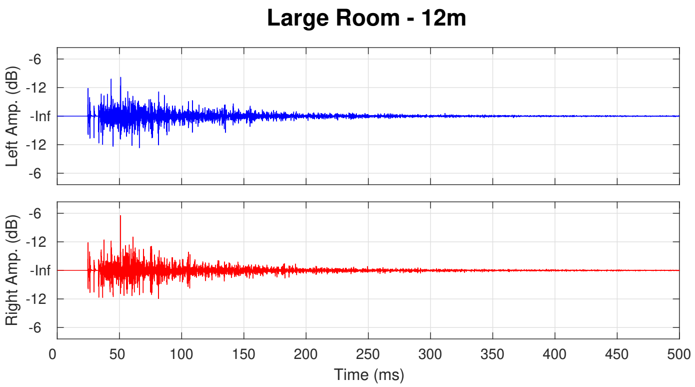


Fig. 7.7: BRIR of the Large Room at a source-receiver distance of 12m

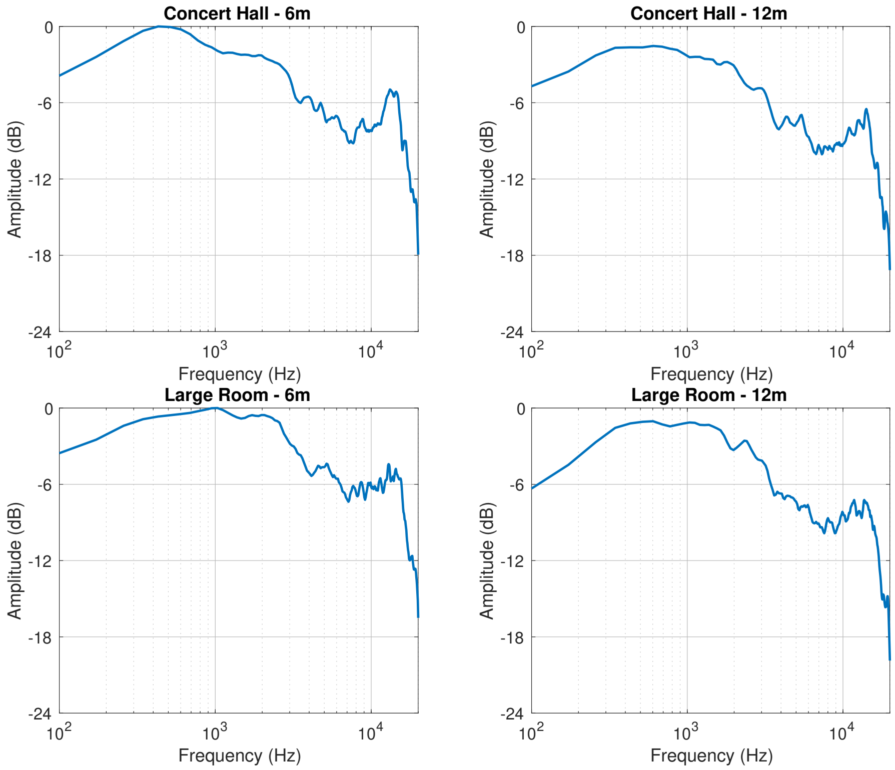


Fig. 7.8: Frequency responses of the BRIRs of the Concert Hall (top) and Large Room (bottom)

Boosting or attenuating the energy of different reflection groups should have an effect on the frequency response of the BRIR, which may have an audible difference on the colouration of the stimuli. Thus, the difference or delta spectrum of first 80ms, or early reflection region, of the BRIRs with either -6 or +6 dB gain for either region was calculated and is presented in Figures 7.9 and 7.10.

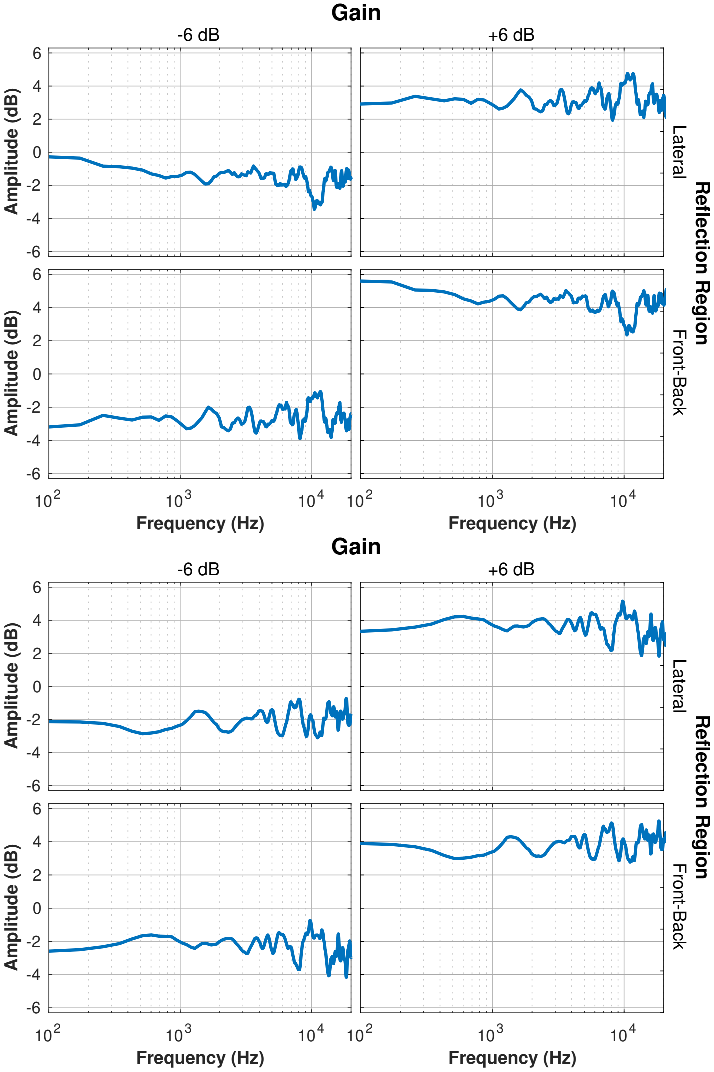


Fig. 7.9: Spectral difference plots of the first 80ms of the BRIRs of the Concert Hall at 6m (top) and 12m (bottom) with either +6 or -6 dB gain applied to each reflection group.

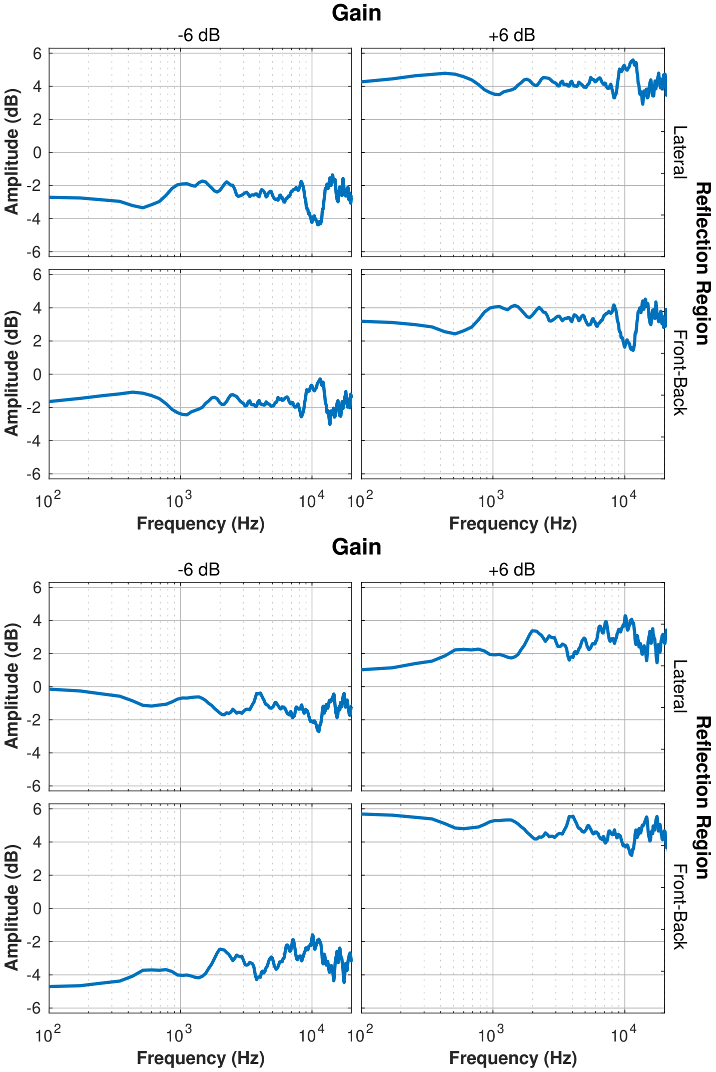


Fig. 7.10: Spectral difference plots of the first 80ms of the BRIRs of the Large Room at 6m (top) and 12m (bottom) with either +6 or -6 dB gain applied to each reflection group.

7.2.1.4 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dB LA$_{\text{eq}}$ by playing pink noise from one headphone cup, and measuring using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

7.2.1.5 Subjects

For this test, critical listeners with experience in spatial audio were required, and therefore four members of the Applied Psychoacoustics Laboratory at the University of Huddersfield were selected to participate. At the time, all reported to have no known hearing impediments. One test member was already familiar with the perception of colouration, whilst all of them in general had critical experience with spatial audio.

7.2.2 Group discussion on attributes

Terms were extracted from the results files, stripping any unnecessary words and padding before counting unique occurrences of each term. However, it was deemed important to check the context that the terms were used in, and thus the files were manually checked to extract any phrases that would potentially alter the context.

Following the test on a different day, the subjects were invited to participate in a focus discussion to group and sort the elicited terms into attributes. The discussion was held in the same room as the elicitation test, and took approximately two hours to complete. The author was present, yet only served as the chair for the discussion. The terms were presented altogether to the subjects, and they were tasked with deciding on a appropriate attributes in which to group the terms by. As the subjects were experienced listeners and have participated in similar elicitation tests of this nature prior to this particular test, they were already familiar with the terms and their definition, which of course helped facilitate the discussion.

Attribute	Occurrences	Elicited Terms
Horizontal Spread	150	Wider (50), Enveloped (28), Narrow (20), Bigger (10), Enveloping (8), Spatial Impression (8), Diffuse (7), Smaller(4), Width (4), Spatial (2), 'Horizontal Spread' (2) 'Bigger Source' (1), Fused(1), Larger (1), 'Source more frontal' (1), Reverberant (1), Spatially Bigger (1), Spacious (1)
Source Distance	60	Close (41), 'Further Away' (8), Distant (7), Further (3), Close (1)
Loudness	53	Louder (41), Quieter (12)
Fullness	44	Fuller (30), Thinner (7), Boomy (6), Bassy (1)
Clarity	37	Duller (11), Muffled (9), Clearer (7), Muddier (7), Presence (2), Boxy (1)
Brightness	21	Brighter (21)
Phasiness	16	Phasey (6), Modulated (4), Metallic (3), Harsher (3)
Vertical Source Spread (VSS)	15	Spread (9), Spacious (3), Envelopment (2), Smaller (1)
Externalisation	9	Externalised (9)
Roughness	7	Rough (7)
Timbral Naturalness	5	Natural (5)
Dynamic Range	3	'Dynamic Range' (3)
Echoic	2	Echoic(1), 'Stronger Reflections' (1)
Vertical Environmental Spread (VES)	1	'Spatially Bigger' (1), Vertical Spread (1)

Table 7.3: Attributes, occurrences and breakdown of elicited terms with their individual occurrences.

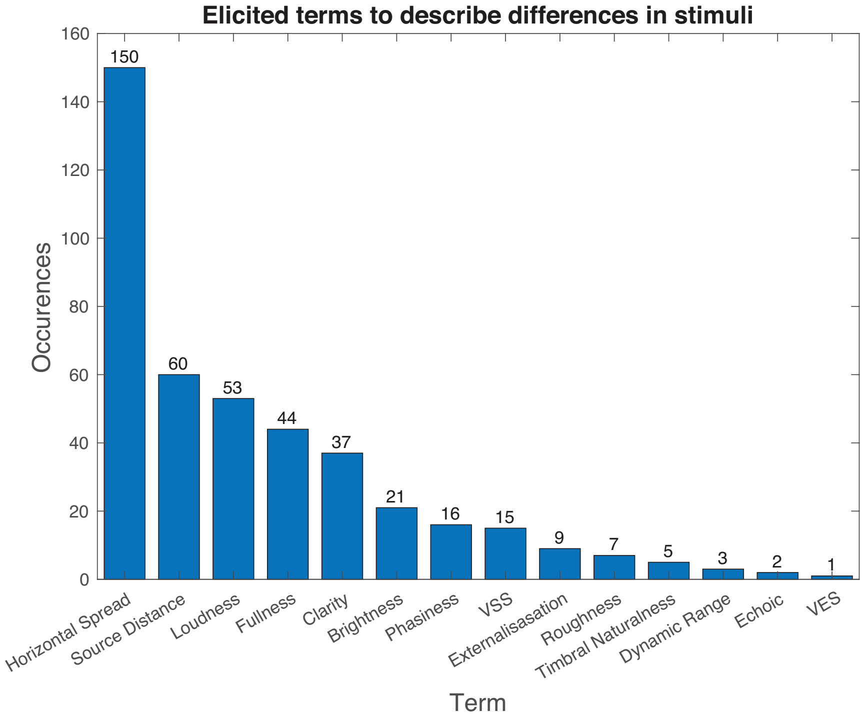


Fig. 7.11: Occurrences of the elicited terms for the perceived differences in colouration

7.2.3 Discussion

Table 7.3 shows that 'Horizontal Spread' was the most commonly occurring attribute, suggesting that differences in width between the BRIRs with and without gain adjustment applied were occurring, where the most commonly elicited terms for this difference are 'Wider', 'Enveloped' and 'Narrow'. This suggests that horizontal spread could be synonymous with ASW since these terms are often elicited for ASW, for example in Lee (2006) or Kaplanis, Bech, Jensen, and van Waterschoot (2014). Notably, 'Enveloped' is directly related to the LEV sub-paradigm of spatial impression, yet it is being used here along with 'Wider' which suggests that the two terms are linked. However, it was undetermined whether the term was either referring to source envelopment or environmental envelopment, thus caution must be given when using this term (Rumsey, 2002).

Interestingly, terms related to ’Source Distance’ and ’Loudness’ were also frequently elicited, which strongly suggests whilst affecting the horizontal spread, the gain adjustments were having also effects on the perceived loudness and source distance. The term ’louder’ by itself was in fact elicited by a similar number of occurrences as ’wider’, which suggests that perhaps there is some form of link between the two, as suggested by Lee (2013). Whilst this of course appears to not be an ideal effect, the extent of the effect can only be highlighted using a grading test[1].

The next four terms in order of occurrence are ’Fullness’, ’Clarity’, ’Brightness’ and ’Phasiness’, all of which can be regarded as timbral attributes. ’Phasiness’ is of particular interest because the terms for this attribute are associated with the perceptual effects of comb-filtering (Halmrast, 2000), as well as the effects discussed in the previous chapter. Here, ’Phasiness’ can be defined as:

"An apparent metallic, rough, comb-filter type characteristic present in the stimulus"

Thus, the colouration control method which utilises the ’unacceptable colouration’ regions can be regarded as a control for ’Phasiness’.

[1] The effects of the perceptual control method on ’Loudness’ and ’Distance’ are investigated later in Chapter 8.

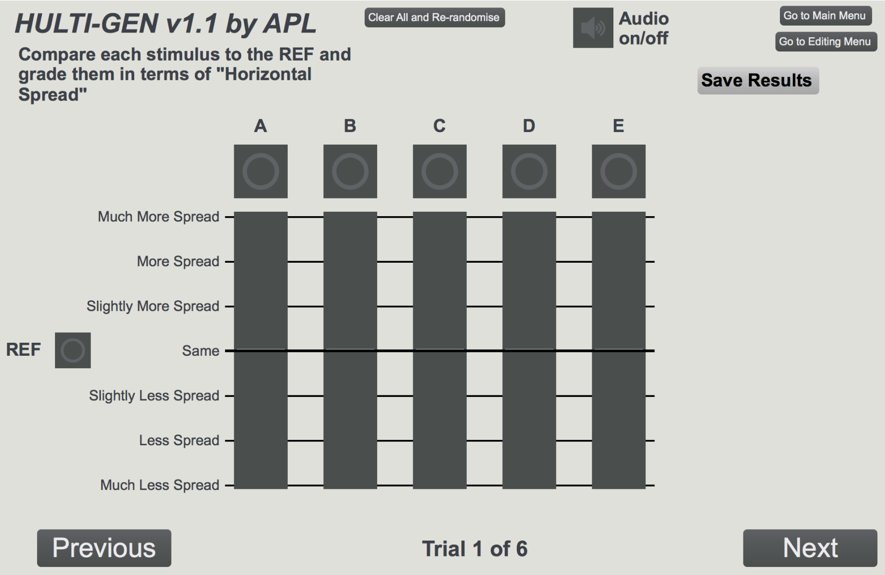
7.3 Experiment part 2: Application of the perceptual control method

7.3.1 Experimental design

The objective of this test was to grade the degree that each attribute is perceived for each level of reflection attenuation / boosting against an unprocessed stimulus.

7.3.1.1 Test method

A multiple comparison test was designed using HULTI-GEN version 1.1 developed by Gribben and Lee (2015). For each trial, subjects were presented with the reference stimulus and five test stimuli, see Figure 7.12.



HULTI-GEN v1.1 by APL

Compare each stimulus to the REF and grade them in terms of "Horizontal Spread"

Fig. 7.12: The HULTI-GEN test interface used for this experiment.

They were asked to compare each stimulus against the reference and grade any audible changes for a given attribute on a continuous scale with semantic labels. The scale ranged between -50 to 50, with -50 being 'Much less', -33 as 'Less',-16 as 'Slightly less', 0 as 'About the same', 16 as 'Slightly more', 66 as 'More' and 50 as 'Much more'. Whilst Bech and Zacharov (2007) recommend the use of anchors at each end of the scale in order to minimise the 'end-point effects' (where subjects are reluctant to use the full range of the scale due to a potential expectancy of extreme stimuli), it was found to be problematic for the main test as the -6dB and +6dB were the extreme conditions, and therefore the anchors would have been identical to two of the chosen stimuli. Also, because there is no certainty about the degree of perceptual difference between stimuli depending on the amount of gain applied, the semantic labels were provided instead of audible anchors.

7.3.1.2 Stimuli

The 6m and 12m stimuli generated in Section 7.2.1.3 were used in this test. The 3m distance was omitted in order to reduce the number of test sessions. However, five levels of adjustment ranging from -6dB to +6dB of gain was applied to either the lateral reflections or the front-back reflections in 3dB steps. Prior to testing, the subjects undertook a training session where they were familiarised with the test interface and the types of stimuli that they would be presented with during the main test. Clear instruction of the type of attributes they would be focusing on was given at the beginning each session.

7.3.1.3 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dB LA_{eq} by playing pink noise from one headphone cup, and measuring using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

7.3.1.4 Subjects

For this test, a group of twelve subjects consisting of music technology students and researchers at the University of Huddersfield participated. Four subjects have experience with spatial audio and critical listening, whilst the remaining subjects have mixed critical listening ability. It is recommended in ITU-R BS.1116 (2015) to employ experienced listeners that can discern fine perceptual differences, however due to a constraint in the number of experienced listeners available during the testing period, subjects with mixed listening ability had to be included in the experiment to increase the number of observations.

7.3.2 Statistical analysis of the results

Prior to any statistical analysis, a Shapiro-Wilk test for normality was performed to deduce the suitability of performing parametric tests upon the data (Shapiro &

Wilk, 1965). It was found that the data for a majority of conditions did not have a normal distribution. This meant that non-parametric tests were chosen for statistical analysis.

7.3.2.1 Overall effects of the perceptual control methods

Figures 7.13 to 7.16 show the median values and notch edges of the responses to the effects of each perceptual control method upon the perceived ASW and phasiness. The notch edges are a non-parametric equivalent to the commonly used 95% confidence interval (McGill et al., 1978). If there is no overlap between two conditions, there is a significant difference between a pair.

Focusing first on the effect of the lateral control method on ASW, upon inspection of Fig. 7.13 it can be seen that in most cases a reduction in lateral reflection energy had a slight to zero effect on the perceived ASW (i.e. there was no perceived difference between with and without reduction). However, for an increase in energy, it can be seen that median values for most cases are between 16 and 33 on the grading scale, which is equivalent to between "Slightly Wider" and "Wider". This implies that the increase in reflection energy created an increased perception of ASW. It also appears the lateral control method was more effective on the orchestra source, especially in the concert hall 6m condition where there is a noticeable linear trend in the median values, and relatively short error bars.

On the other hand, when observing Fig. 7.14, the reduction in front-back energy also had slight to zero effect on the perceived ASW. However, an increase had some effect on the ASW, again with certain cases, the median values lie between 16 to 33, or "Slightly Wider" and "Wider", on the grading scale. This suggests that at times the front-back control method did have an effect on the perceived ASW, albeit a slightly lesser observable effect when compared to the effects of the lateral control method. Interestingly, for the male speech source in the large room 12m condition, there is an observable negative trend in the median values, which suggests that an increase in front-back energy had the complete opposite effect. Considering that front-back energy is being increased as opposed to lateral, according to L_f (Barron & Marshall, 1981) this increase should have caused ASW to decrease[2].

Shifting focus now to phasiness, 7.15 shows the effects of the lateral perceptual control method on phasiness. In the concert hall at both distances, for all sources the notch edges overlap which suggests that the lateral control method may have had no significant effect on phasiness. However, in the large room, at 6m the guitar shows some significant difference in phasiness as the lateral reflection energy is increased by 6 dB, albeit only lying on the "Slightly less" category of the grading scale. At 12m, when the energy is increased by 6 dB there is a slight yet significant increase for the guitar, yet a slight and also significant reduction for orchestra. Overall, there appears to be some slight negative trends in median values, yet no observable major differences, which is important as this means that the lateral control method most probably satisfies the requirement of not affecting another attribute, namely colouration.

Finally, the subjective results of the front-back control method in Fig. 7.16 are similar pattern to the results of the lateral control method shown in Fig. 7.15. Overall, except for the guitar in the large room at 6m, there is no observable significant effect of front-back gain adjustment on perceived phasiness. This suggests that method is probably ineffective at controlling this type of tonal colouration, which

[2] Objective analysis is performed in Chapter 8.

is understandable as the method used is based on the experiments in the previous chapter were performed using a single reflection, not multiple reflections. Another observable feature of Fig. 7.16 is the slight negative trend in median values in certain cases, for example in the guitar and orchestra in the large room at 12m. However, there no observable significant differences, and that the range of median values lie close to 0 or "About same", suggesting that the control method did not have any noticeable effect on phasiness.

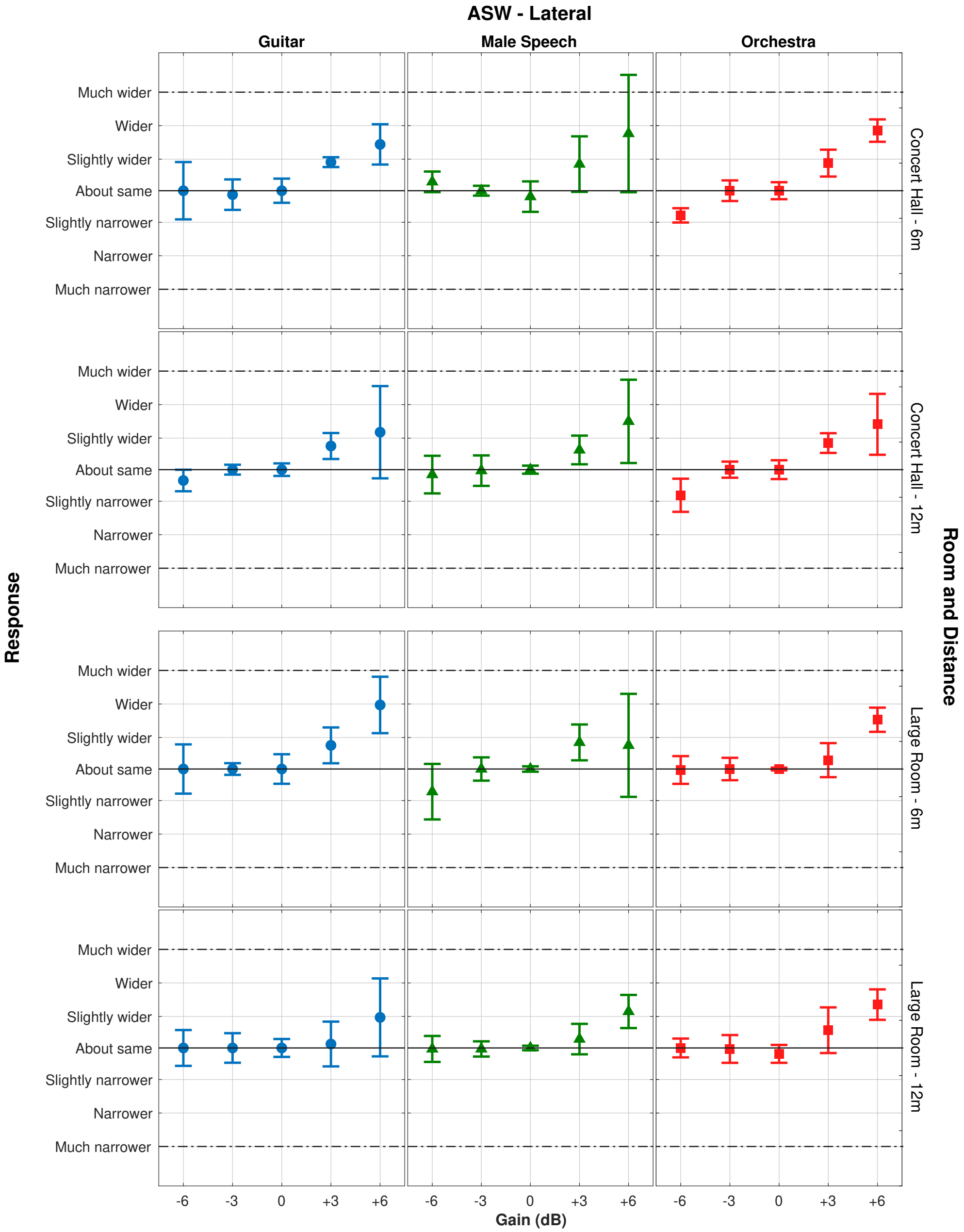


Fig. 7.13: Median values and 95% confidence interval notch edges of the subject responses to the effects of lateral level adjustment upon perceived ASW

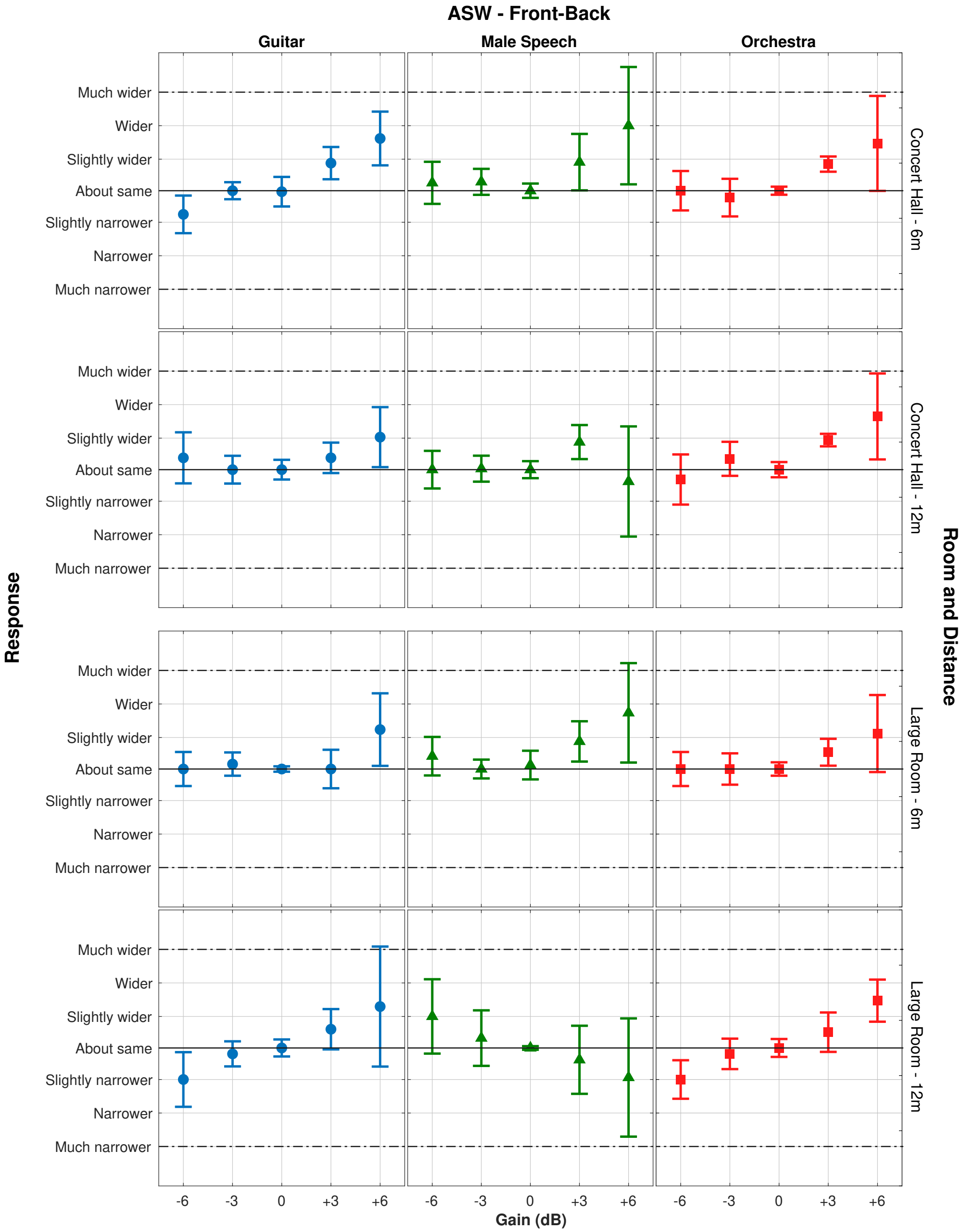


Fig. 7.14: Median values and 95% confidence interval notch edges of the subject responses to the effects of front-back level adjustment upon perceived ASW

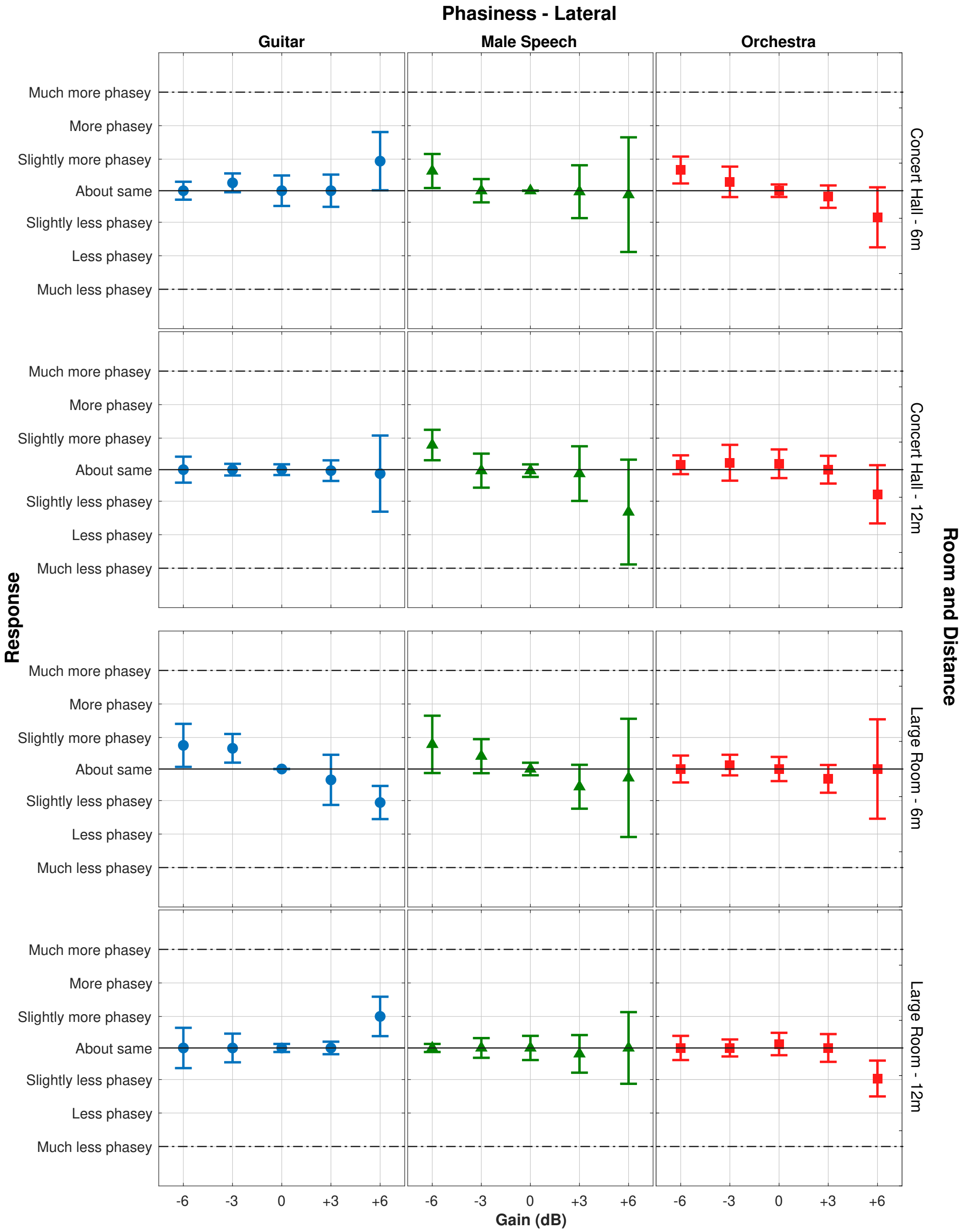


Fig. 7.15: Median values and 95% confidence interval notch edges of the subject responses to the effects of lateral level adjustment upon perceived phasiness

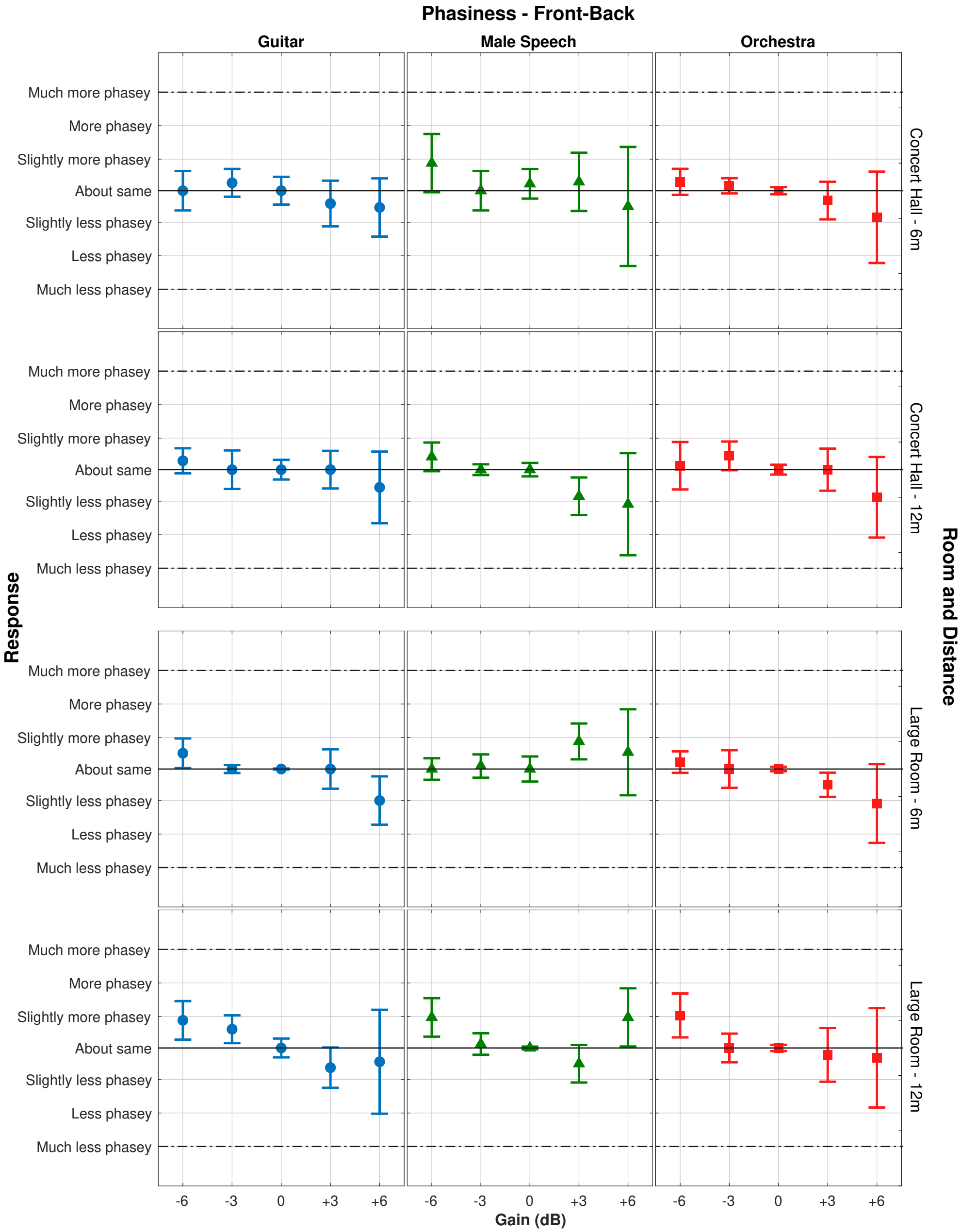


Fig. 7.16: Median values and 95% confidence interval notch edges of the subject responses to the effects of front-back level adjustment upon perceived phasiness

To verify the observations made, and to determine if there are any significant effects of each perceptual control method upon each attribute, a series of Friedman tests were performed on the experimental data using the Statistics and Machine Learning Toolbox^{TM} in MATLAB. The results for the effects on ASW and phasiness are presented in Table 7.3.2.1.

				Source		
Attribute	Method	Room	Distance	Guitar	Speech	Orchestra
ASW	Lateral	Concert Hall	6m	**.001****	.555	**.001****
			12m	.119	.119	**.006****
		Large Room	6m	**.047***	.333	**.013***
			12m	.599	**.006****	**.001****
	Front-Back	Concert Hall	6m	**.017***	.577	.297
			12m	.362	.583	.119
		Large Room	6m	.162	.126	.614
			12m	.278	.774	**.003****
Phasiness	Lateral	Concert Hall	6m	.604	.706	**.009****
			12m	.757	.251	.359
		Large Room	6m	**.029***	.155	.334
			12m	.127	.695	**.040***
	Front-Back	Concert Hall	6m	.087	.559	.329
			12m	.444	.263	.278
		Large Room	6m	.114	.804	.131
			12m	.492	**.044***	.106

Table 7.4: p values for Friedman test performed on subjective data for effects of lateral and front-back gain on the attributes. * - Significant ($p < .050$), ** - Very significant ($p < .010$).

The Friedman test reveals that the lateral level adjustment had a significant effect on the ASW of: the guitar at 6m in both rooms, the speech only in the large room at 12m, and the orchestra in all cases. Front-back gain adjustment on the other hand only had a significant effect on the ASW of the guitar in the concert hall at 6m, and of the orchestra in the large room at 12m. This suggests that overall, the lateral level adjustment had a significant effect on ASW in more cases than front-back level adjustment, thus appears to be more effective at controlling it. Plus it shows that the front-back level adjustment did not significantly affect the ASW often. The Friedman test results for ASW are also in agreement with Figures 7.13 and 7.14, where notch pairs for the significant conditions are not overlapping, except in Fig. 7.14 for the orchestra in the concert hall at 12m where there is no overlap between the 0 and +3 dB. However, the other notch edges are overlapping so this is perhaps a Type-I error produced by the plot.

For phasiness, the test showed that the lateral level adjustment had a significant effect on the perceived phasiness of the guitar in the large room at 6m, and of the orchestra at 6m in the concert hall, at 12m in the large room. This agrees with Figure 7.15 where certain pairs of notches for these significant cases are not overlapping. In comparison, the test showed that in all cases except for male speech in the large room at 12m, the front-back gain adjustment had no significant effect ($p > .05$). This of course signifies that this particular method of perceptual control failed its intended task of controlling phasiness.

7.3.2.2 Effect of the perceptual control methods on ASW

To determine if there are any significant differences between pairs of gain values, a Wilcoxon signed-rank test with Bonferroni adjustment applied was performed with a significance level of $\alpha = .05$. Furthermore, the non-parametric effect size (Fritz, Morris, & Richler, 2012) was also calculated to judge the degree of effect the control methods had on ASW. Since the main objective was to find out if the methods caused a difference to ASW in comparison to the 0 dB, or unaltered reference, only the p and r values between $\pm g$ and 0 dB pairs will be considered.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Gain (dB)	*p*	*r*	*p*	*r*	*p*	*r*
Concert Hall	6m	–6	1.000	-.06	1.000	.13	.537	-.40
		–3	.684	-.37	1.000	-.09	1.000	-.23
		+3	.327	.43	1.000	.14	1.000	.28
		+6	.332	.43	1.000	.22	.161	.48
	12m	–6	1.000	-.15	1.000	-.23	.127	-.49
		–3	1.000	-.01	1.000	-.31	1.000	-.19
		+3	.674	.38	1.000	.31	.649	.38
		+6	1.000	.22	1.000	.26	1.000	.34
Large Room	6m	–6	1.000	-.20	1.000	-.27	1.000	-.35
		–3	.859	-.35	1.000	-.05	1.000	.05
		+3	1.000	.06	1.000	.29	1.000	.19
		+6	1.000	.34	1.000	.24	.454	.41
	12m	–6	1.000	.04	1.000	-.19	1.000	.09
		–3	1.000	.03	.586	-.39	1.000	.09
		+3	1.000	.18	1.000	.11	1.000	.28
		+6	1.000	.22	.938	.34	**.034***	.56

Table 7.5: *p* values for Wilcoxon sign-rank of results for lateral level adjustment on ASW. * - Significant ($p < .050$), ** - Very significant ($p < .010$).

The test found that between 0 dB and each gain value, there was only a significant difference in ASW between 0 and 6 dB for the orchestra in the large room at 12m. Whilst this agrees with the lack of overlap between the notches for these values in that condition, the remaining results do not agree with most other observations where, in almost all cases, there is no overlap between this pair of values. Furthermore, whilst it is also possible that the Bonferroni adjustment is too conservative, the results show that even a ±6 dB change in gain did not have a significant effect on ASW. However, the *r* values show that there was at least some effect of gain adjustment.

For the guitar, it can be seen that the gain adjustment had a larger effect on ASW in both rooms at 6m than at 12m, where the *r* values range from -.37 to .43 in the concert hall, and -.35 to .34 in the large room, which is within the medium effect size category (Coolican, 2009). At 12m in both rooms this range is less. The opposite is observed for speech where the effect size range is higher at 12m than at 6m in both rooms, again with the respective r values at 12m being $\pm$.31 in the concert hall, and -.39 to .34 in the large room, both of which are all in the medium effect size category. Finally, for orchestra it appears that the effect size range is considerable both rooms and distances, and mostly larger than the ranges for the other two sources. In the concert hall, the range is -.40 to .48 at 6m, and -.49 to .34 at 12m. At 6m in particular, +6 dB had an effect size of .48, and similarly at 12m -6 dB the effect size was -.49, both of which are close to the 'large' category. In the large room, whilst the range was not as high, at 12m the +6 dB gain had an effect size of .56 which is definitely a large effect size.

What these values suggest is that the lateral perceptual control method, whilst it did not have a significant effect, it had a medium effect which is of course reflected by the positive trends in the median values in Fig. 7.13. Furthermore, the effect size range for orchestra is greater than that of the other two sources, which suggests the control was more effective for this particular source. Again, this can be observed in Fig. 7.13 where the notch ranges are smaller for orchestra, especially in the concert hall at 6m.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Gain (dB)	*p*	*r*	*p*	*r*	*p*	*r*
Concert Hall	6m	–6	1.000	-.09	1.000	.16	1.000	-.01
		–3	1.000	.28	1.000	.26	1.000	-.05
		+3	.103	.51	1.000	.19	.127	.49
		+6	.200	.46	1.000	.22	1.000	.26
	12m	–6	1.000	.06	1.000	.19	1.000	.09
		–3	1.000	.02	1.000	.22	1.000	.33
		+3	1.000	.05	.610	.39	.059	.54
		+6	.723	.37	1.000	.02	1.000	.30
Large Room	6m	–6	1.000	.16	1.000	-.16	1.000	-.11
		–3	1.000	.26	.488	-.41	1.000	-.03
		+3	1.000	-.10	1.000	-.02	1.000	.18
		+6	.322	.43	1.000	.26	1.000	.18
	12m	–6	1.000	-.31	1.000	.26	.059	-.54
		–3	1.000	-.24	1.000	.15	.898	-.35
		+3	1.000	.13	1.000	-.04	1.000	.22
		+6	1.000	.17	1.000	-.07	.820	.36

Table 7.6: *p* values for Wilcoxon sign-rank of results for front-back level adjustment on ASW. * - Significant ($p < .050$), ** - Very significant ($p < .010$).

For the front-back perceptual control, the Wilcoxon signed-rank test found that between 0 and up to ±6 dB gain in all cases, the method had no significant effect on perceived ASW. In almost all cases, this does in fact agree with Fig. 7.14 where the notch edges for all gain values overlap. However, there are a few instances where there is no overlap, notably in the concert hall at 6m where for speech +6 dB does not overlap 0 dB, and for orchestra where +3 dB also does not overlap 0 dB, although the +6 dB condition for orchestra does overlap 0 dB so the 0 and +3 dB case is potentially anomalous. This is also exhibited at 12m in the concert hall. The only instance of disagreement is for the orchestra in the large room at 12m where there is a definite lack of overlap between 0 dB and both ±6 dB pair of conditions, whilst the test found no significant difference for either pair ($p > .05$). Again, this could potentially be an anomaly in the plot. These results suggest that the front-back perceptual control had no significant effect on ASW for all three sources. Since the test was mostly in agreement, the effect size values were not further considered.

7.3.2.3 Effect of the perceptual control methods on phasiness

To find out if there are any significant differences between pairs of gain values in the subjective data for phasiness, a Wilcoxon signed-rank test with Bonferroni adjustment applied was performed with a significance level of $\alpha = .05$. Again, the non-parametric effect size r was also calculated as to determine the degree of effect the control methods had on phasiness. As mentioned in the last section, the main objective was to find out if the methods caused a difference to phasiness in comparison to the 0 dB, or unaltered reference. Thus, only the p and r values between all 0 dB and $\pm g$ pairs will be considered.

The test found that with the lateral perceptual control applied, in all cases there were no significant differences in perceived phasiness between 0 dB and all gain values. This largely agrees with the overlapping notch edges between these pairs, see Fig. 7.15. However, the only instance where the test does not agree is for the guitar in the large room at 6m, where there is a lack of overlap between -3 and 0 dB, and 0 and +6 dB. Again, when studying the p values in Table 7.7, it is possible that the Bonferroni adjustment was too conservative. The respective effect size r values for these pairs are .47 and -.42, which is a moderate to large effect size. This suggests that perhaps the lateral gain adjustment did have a noticeable effect in this scenario, albeit the median values do not cross the "Slightly..." categories on the grading scale.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Gain (dB)	*p*	*r*	*p*	*r*	*p*	*r*
Concert Hall	6m	–6	1.000	-.24	1.000	.32	.195	.47
		–3	1.000	-.05	1.000	.19	.879	.35
		+3	1.000	-.17	1.000	.06	1.000	-.16
		+6	1.000	.09	1.000	.05	1.000	-.24
	12m	–6	1.000	.03	.469	.41	1.000	.01
		–3	1.000	.03	1.000	.13	1.000	.23
		+3	1.000	-.27	1.000	.15	1.000	-.22
		+6	1.000	-.09	1.000	-.05	1.000	-.23
Large Room	6m	–6	.684	.37	.264	.45	1.000	.05
		–3	.195	.47	.234	.46	1.000	.11
		+3	1.000	-.17	1.000	-.19	1.000	-.24
		+6	.400	-.42	1.000	-.14	1.000	-.20
	12m	–6	1.000	-.04	1.000	-.22	1.000	.06
		–3	1.000	-.25	1.000	-.11	1.000	-.18
		+3	.313	-.45	1.000	-.30	1.000	-.28
		+6	1.000	.17	1.000	-.12	.098	-.51

Table 7.7: *p* values for Wilcoxon sign-rank of results for lateral level adjustment on phasiness. * - Significant ($p < .050$), ** - Very significant ($p < .010$).

Finally, the test found that with the front-back perceptual control applied, in all cases there were no significant differences in perceived phasiness between 0 dB and all gain values. This also mostly agrees with Fig. 7.16, except for the guitar in the large room at 6m where the +6 dB notch region does not overlap the 0 dB notch region, and for male speech in the same room at 12m where -6 dB does not overlap 0 dB. Again, whilst perhaps the Bonferroni adjustment applied to the *p* values is too conservative, due to the relatively low number of subject numbers it is likely that there is no significant effect that is otherwise suggested by the plot, which is possibly erroneous. The respective *r* values, however, for these two cases are -.34 and .51, which are moderate and large effect sizes. The test revealed that, overall, the front-back perceptual control method did not have a significant on phasiness.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Gain (dB)	*p*	*r*	*p*	*r*	*p*	*r*
Concert Hall	6m	–6	1.000	.06	1.000	.31	1.000	.19
		–3	1.000	.24	1.000	-.06	1.000	.24
		+3	1.000	-.21	1.000	-.16	1.000	-.32
		+6	1.000	-.30	1.000	-.19	1.000	-.20
	12m	–6	1.000	.07	1.000	.28	1.000	.28
		–3	1.000	.00	1.000	.15	.352	.42
		+3	1.000	-.12	1.000	-.26	1.000	.10
		+6	1.000	-.29	1.000	-.18	1.000	-.22
Large Room	6m	–6	1.000	.21	1.000	.01	1.000	.17
		–3	1.000	-.17	1.000	.09	1.000	.07
		+3	1.000	-.09	1.000	.14	.273	-.45
		+6	.957	-.34	1.000	-.02	1.000	-.25
	12m	–6	1.000	.23	.098	.51	.176	.47
		–3	1.000	.19	.430	.42	.234	.47
		+3	1.000	-.28	1.000	-.01	1.000	.07
		+6	1.000	-.14	1.000	.31	1.000	-.05

Table 7.8: *p* values for Wilcoxon sign-rank of results for front-back level adjustment on phasiness. * - Significant ($p < .050$), ** - Very significant ($p < .010$).

7.3.3 Discussion

The statistical analysis has shown that the perceptual control methods have mixed effectiveness over their corresponding attributes. The lateral perceptual control had, overall, a mostly significant effect on the perceived ASW, especially for the guitar at 6m in both rooms, and for the orchestra in all room and distance conditions. However, whilst the objective of the lateral control method was to either increase or decrease the amount of ASW, the Wilcoxon signed rank test showed that there were no significant differences in ASW, although the effect size showed otherwise. The positive trends in the median values of the subject results shown in Figure

7.13 suggest that a 6 dB increase in lateral reflection energy does appear to have a potentially noticeable effect on the perceived ASW, yet a 6 dB reduction does not. There are two possible reasons for this. The first is that the lateral reflection energy was initially quite low, such that reducing it further did not cause a noticeable difference, although this does not appear to be the case for the orchestra source which shows that a reduction had some effect in the concert hall. Another possible reason for the reduction not having a significant effect is that there are reflections that lie close to, yet not within, the $\mathrm{ASW}_{\mathrm{max}}$ regions that the control method acts upon. This is likely due the operation of L_f, whereby reflections just outside of the $\mathrm{ASW}_{\mathrm{max}}$ region may still cause some degree of source broadening.

As for the effects of the front-back perceptual control method on perceived ASW, the analysis suggests that the method had no significant effect. This was further supported by the Wilcoxon signed rank test. The possible reason for this is that, in contrast to what was hypothesised in the previous paragraph, the front-back reflections have less influence on ASW than lateral reflections present in the $\mathrm{ASW}_{\mathrm{max}}$ region. However, there is an noticeable increase in the median values for the +6 dB conditions (see Figure 7.14), along with a large increase in notch width. This suggests that for some subjects, the increase in front-back reflection energy may have caused a noticeable increase in ASW. Since this is due to front-back reflections, it is possible that a perceived change in width was due to another factor such as a change in loudness or distance, which were frequently occurring attributes in the elicitation test. To verify this, the effects of the perceptual control method on perceived loudness and distance should be observed using the same multiple comparison grading test that was used in this experiment. By considering the occurrence of the elicited attributes, it is hypothesised that there may be some effect.

For phasiness, the analysis showed that both the lateral and front-back perceptual control methods had no significant effect. Whilst this showed that the lateral control method did not affect the colouration, however, it does reveal that the front-back control method was not effective at manipulating phasiness. This is probably due to the method being too simple for controlling this attribute, as it was based on results from the experiment performed in Chapter 6, which focused on the effects of a single reflection and not a reverberant sound field or multiple reflections. The ineffectiveness of the method is also likely to be compounded the subjective nature of phasiness, where some subjects graded in opposite directions to other subjects (i.e. some graded "more", whilst others graded "less"). Whilst subjects were familiarised with the attribute prior to testing, it was unknown if they could definitely perceive the effect. The test relied on a subject's ability to perceive a difference in phasiness, and if some subjects were unable to, then they would not grade any large amount of difference. The subjective nature of this attribute is reflected in Fig. 7.16 where the notch regions become larger as the gain is changed, yet the median values never exceed ±16, or "Slightly..." category.

7.3.4 Limitations

What this experiment does not allow for is a direct comparison between the two perceptual control methods. It can be seen that the lateral method did have an effect on ASW, and that the front-back method also had some lesser effect. However, this raises the question of *'how much less?'*, which could be answered by using a similar comparison test. The test should present both perceptual control methods within the same trial, allowing for a direct comparison of the two methods in terms of effectiveness on ASW, and phasiness for that matter.

Another limitation is the use of only two virtual rooms and receiver positions. More rooms would allow for a greater variety of spaces for which to test the performance of the perceptual controls on because, after all, the ultimate goal is for the perceptual optimisation to be *agnostic* of the room and be a generic method. It would also be potentially worthwhile to investigate the performance using a variety of different surface materials. The only drawback of this would be the sheer increase in the number of test sessions and trials, which in turn would require a long test period which was unfeasible for this study. Therefore, these points can be revisited in a future study.

7.4 Conclusion

This chapter focused on the application of perceptual control methods on ASW and tonal colouration. Section 7.1 discussed the two proposed methods that controlled the amount of energy in two regions, ASW$_{\text{max}}$ and 'Unacceptable Colouration', which were discussed in Chapters 5 and 6. Early reflections, or rays, that arrive within either region would be extracted from the RIV file produced by the custom, virtual acoustics algorithm. The gain of these grouped rays would have their energy boosted or attenuated by up to 6 dB.

The first part of the experiment was a modified QDA type elicitation test. It focused on extracting terms to describe the audible differences that could be perceived when the control methods were applied to RIVs. Two different rooms were used for the entirety of this experiment: a concert hall and a large room. In each room, with a fixed source position, RIVs were rendered with a source-receiver distance of 3m, 6m and 12m. The perceptual control methods were applied with gain adjustments of up to ±6 dB in 3 dB steps, spatialised to create BRIRs which were then convolved with three anechoic sources with varying characteristics. Subjects elicited the differences between the 0 dB condition and +6 dB condition. Analysis of the terms, along with a group discussion, found that the four most salient attributes were *'Horizontal Spread'*, *'Loudness'*, *'Distance'* and *'Fullness'*. The attribute that matched tonal colouration the closest was *'Phasiness'*. The changes in horizontal spread, which was assumed to be equivalent to ASW, were expected. However, because the test was limited in not in assessing the audibility of the terms, the degree of change in ASW and Phasiness, or the other salient attributes for that matter, is unclear.

The second part of the experiment was a multiple comparison listening test. Using the same stimuli as the elicitation test, whilst omitting the 3m distance, this test investigated the effectiveness of the perceptual control methods on manipulating ASW and phasiness. It was found that lateral level control had a significant effect on ASW for certain situations, yet not at all on phasiness. The method was effective on the guitar source at 6m in both rooms, on speech at 12m in both rooms, and on the orchestra in all scenarios. Despite the analysis not finding a significant difference in ASW between 0 dB and all other degrees of gain, the positive trend in median values and the calculated effect sizes suggest that there was a moderate to large effect on ASW. However, it appeared that the lateral control method was more effective at increasing ASW, rather than reducing it. It was also found that the front-back control method had some moderate yet similar effects on ASW too.

The lateral control method had no significant on the perceived phasiness, although it is hypothesised that some subjects may have perceived a difference, albeit a debatable one. The same was observed for the front-back control method which also had no significant effect. This then highlights that the lateral control method is not likely to inadvertently affect tonal colouration, however, it also shows that the front-back control method was not effective at manipulating the colouration.

This experiment found that both the lateral and front-back perceptual control methods had an effect on ASW, as well as potentially having an effect on loudness and distance. Thus, the effects of both methods on perceived loudness and distance must be considered in order to determine how well they can affect their corresponding attribute without affecting others. This experiment also did not allow for direct comparison between the two methods as they were not compared together during the second part of the experiment. Therefore, an investigation that compares the two methods directly must be considered.

7.5 Summary

This chapter investigated the effects of two perceptual control methods on ASW and tonal colouration. An experiment was performed to determine the most salient perceptual effects, then investigate the effectiveness of each method on controlling ASW and phasiness. This investigation has found the following:

- The most commonly elicited effects of level adjustment are 'Horizontal Spread', 'Loudness', 'Distance' and 'Fullness'. This suggests that the methods affected not only ASW, they may have also affected loudness and distance perception.

- When lateral gain adjustment is applied, it causes a significant effect on ASW

- Lateral gain adjustment did not have a significant effect on *'phasiness'*.

- When front-back level adjustment is applied, it also has some significant effect on ASW.

- Front-back level adjustment also did not have a significant effect on *'phasiness'*.

Chapter 8

Application of the perceptual control methods: Part 2

The experiments performed in Chapter 7 investigated how the two types of control methods affected the ASW and 'phasiness' directly. However, they did not allow for a direct comparison between the methods. Such a comparison would reveal the effectiveness of each type of spatial filtering and whether they can control one attribute whilst leaving another unchanged. According to the previous experiments, it was hypothesised that modifying the gain of the lateral reflections that fall within the $\mathrm{ASW}_{\mathrm{max}}$ region should change the ASW without affecting the acceptability of colouration. Inversely, modifying the gain of the reflections that fall within the front-back region should affect the phasiness, albeit insignificantly as seen from the experiment in Chapter 7, and should not greatly affect ASW. Furthermore, whilst modifying the level of the front-back reflections may affect ASW, it should not do so as greatly since, according to Barron and Marshall (1981), these reflections should have less influence due to their lower angle of arrival. However, it was in fact found that ASW appeared to change when an increase in level was also applied to front-back reflections. The lateral reflections should influence the perception of width far more greatly than front-back reflections, as increasing the lateral reflection energy would increase L_f , whilst increasing the front-back reflection energy would reduce it. Thus, by extension, this would reduce the perceived ASW.

It was found in the elicitation test performed in Chapter 7 that the attributes 'loudness' and 'distance' had a relatively high occurrence, suggesting that the changes in reflection energy had a potential effect on those. Hypothetically, an increase in the reflection energy would cause a perceptual increase in loudness. Ideally, from a perceptual control viewpoint, it is desirable that no attribute other than ASW or phasiness should change, yet the elicitation test suggests that changes in perceived loudness and distance could be resultant artefacts or manifestations caused by the changes in reflection level. For this to be investigated, tests that observe the effect of manipulating the level of the two reflection groups upon loudness and distance were included in this experiment. Ideally, if the differences in those attributes is insignificant, the control methods can be considered successful.

From this, the following research questions were established:

1. *Does a particular type of control method effectively alter their corresponding attribute more than the other type?*

2. *Can each control method affect only their associated attribute without affecting others?*

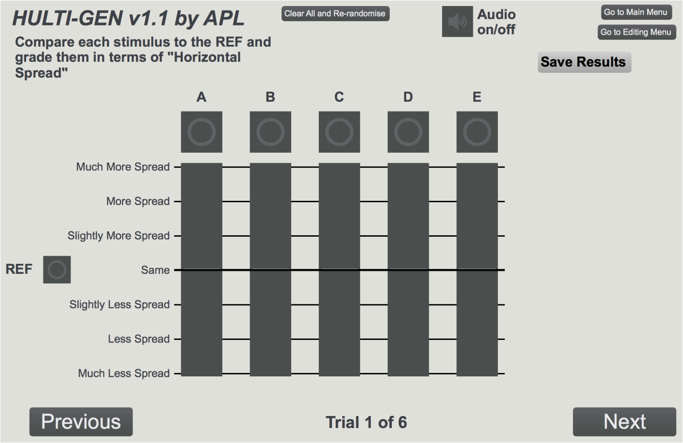
3. *To what degree, if any, does each method affect loudness and distance perception?*

The third question is of most importance as this will determine the effectiveness of each perceptual control method in their ability to affect their associated attribute with little effect on others.

8.1 Experimental design

8.1.1 Test method

The same HULTI-GEN interface used in the second part of the previous experiment discussed in Chapter 7 is used here. The response method and labelling scheme was also kept identical.



Compare each stimulus to the REF and grade them in terms of "Horizontal Spread"

Fig. 8.1: The HULTI-GEN test interface used in the experiment

8.1.2 Stimuli

The same pool of stimuli that was created in Chapter 7 were used in this experiment. It was decided that a 6dB difference in gain would reveal more obvious changes in an attribute than a 3dB difference. Thus, only the -6 and +6 dB gain conditions were used rather than the intermediate values. This kept the number of stimuli to five per trial as before. Prior to testing, the subjects undertook a training session where they were familiarised with the test interface and the types of stimuli they would be presented with during the main test. Clear instruction of the type of attributes they would be focusing on was given at the beginning each session.

8.1.3 Apparatus and equipment

The stimuli were played to the subjects using a Merging Horus audio interface at a sample rate of 44.1 kHz, and Sennheiser HD650 Blue Stage headphones. The playback level was set to a comfortable listening level of 68 dB $\mathrm{LA}_{\mathrm{eq}}$ by playing pink noise from one headphone cup, and measuring the average level using a Casella CEL-450 loudness level meter. The test took place in the ITU-R BS.1116 (2015) compliant critical listening room at the University of Huddersfield. Loudspeakers, equipment and other permanent fixtures in the room were hidden from view using curtains to reduce the likelihood of visual bias affecting the test results.

8.1.4 Subjects

A group of ten subjects consisting of researchers and music technology students took part in this experiment. Four subjects had experience with critical listening and spatial audio, and had participated in the previous experiments. The remaining subjects were found to have mixed listening ability. It was decided to include subjects with a mixed listening ability as this would give the experiment an applied context rather than a fundamental one. All subjects reported to have normal hearing with no known impediments at the time.

8.2 Statistical analysis

Prior to any statistical analysis, a Shapiro-Wilk test for normality was performed to deduce the suitability of performing parametric tests upon the data (Shapiro & Wilk, 1965). It was found that the data for a majority of conditions did not have a normal distribution. This meant that non-parametric tests were chosen for statistical analysis.

8.2.1 Effect of control method and level manipulation

Figures 8.2 to 8.5 show the median responses for ASW, Loudness, Distance and Phasiness. The medians have been presented with notch edges, representing the equivalent 95% confidence interval (McGill et al., 1978). An initial glance at these plots showed that manipulation of reflection energy had some observable effect on each attribute.

			Source		
Attribute	Room	Distance	Guitar	Speech	Orchestra
ASW	Concert Hall	6m	.481	1.000	.265
		12m	**.022***	1.000	**.024***
	Large Room	6m	**.001****	.159	**.008****
		12m	.821	.987	1.000
Loudness	Concert Hall	6m	**< .001****	**< .001****	**< .001****
		12m	**< .001****	**< .001****	**< .001****
	Large Room	6m	**< .001****	**< .001****	**< .001****
		12m	**< .001****	**< .001****	**< .001****
Distance	Concert Hall	6m	**< .001****	**< .001****	**< .001****
		12m	**< .001****	**< .001****	**< .001****
	Large Room	6m	**< .001****	**< .001****	**< .001****
		12m	**< .001****	**< .001****	**< .001****
Phasiness	Concert Hall	6m	1.000	1.000	1.000
		12m	1.000	1.000	1.000
	Large Room	6m	1.000	1.000	1.000
		12m	1.000	1.000	1.000

Table 8.1: Friedman test p values. * - Significant ($p < .050$), ** - Very significant ($p < .010$).

Friedman tests were performed in MATLAB for each combination of attribute, room, source-listener distance and source type. Due to the relatively small number of subjects, Bonferroni adjustment was applied to the obtained p values, presented in Table 8.1. The significant value was $\alpha = .05$. For each attribute, the null hypothesis is that the gain adjustment does not have a significant effect on the perception of that attribute.

For ASW, it was found that the level adjustment of either reflection region caused some significant effect on the perceived width for: Guitar in the Concert Hall at 12m and the Large Room at 6m; and Orchestra in the Concert Hall at 12m the Large Room at 6m. For all other combinations of conditions, the level adjustment had no significant effect on ASW. This largely disagrees with the Fig. 8.2 where it there is no overlap between the notch edges of the reference, 'None' condition and all other conditions in several scenarios. For the Loudness and Distance attributes, it was found that for all conditions, level adjustment over both reflection regions caused a significant effect on both attributes. This agrees with notch edges plotted in Figures 8.3 and 8.4. Finally, for 'Phasiness' it was found that neither type of spatial filtering had any significant effect on Phasiness. This strongly implies that the perceptual processing had no significant effect on 'Phasiness'.

These initial observations show that the control techniques may not have operated as expected, such that they potentially had a greater effect on attributes other than their intended target. In order to further determine which particular methods on each reflection region causes the differences found in the attributes, Wilcoxon signed rank tests were performed in MATLAB with a significance value of $\alpha = .05$. Alongside the tests, the non-parametric effect size r values were also calculated.

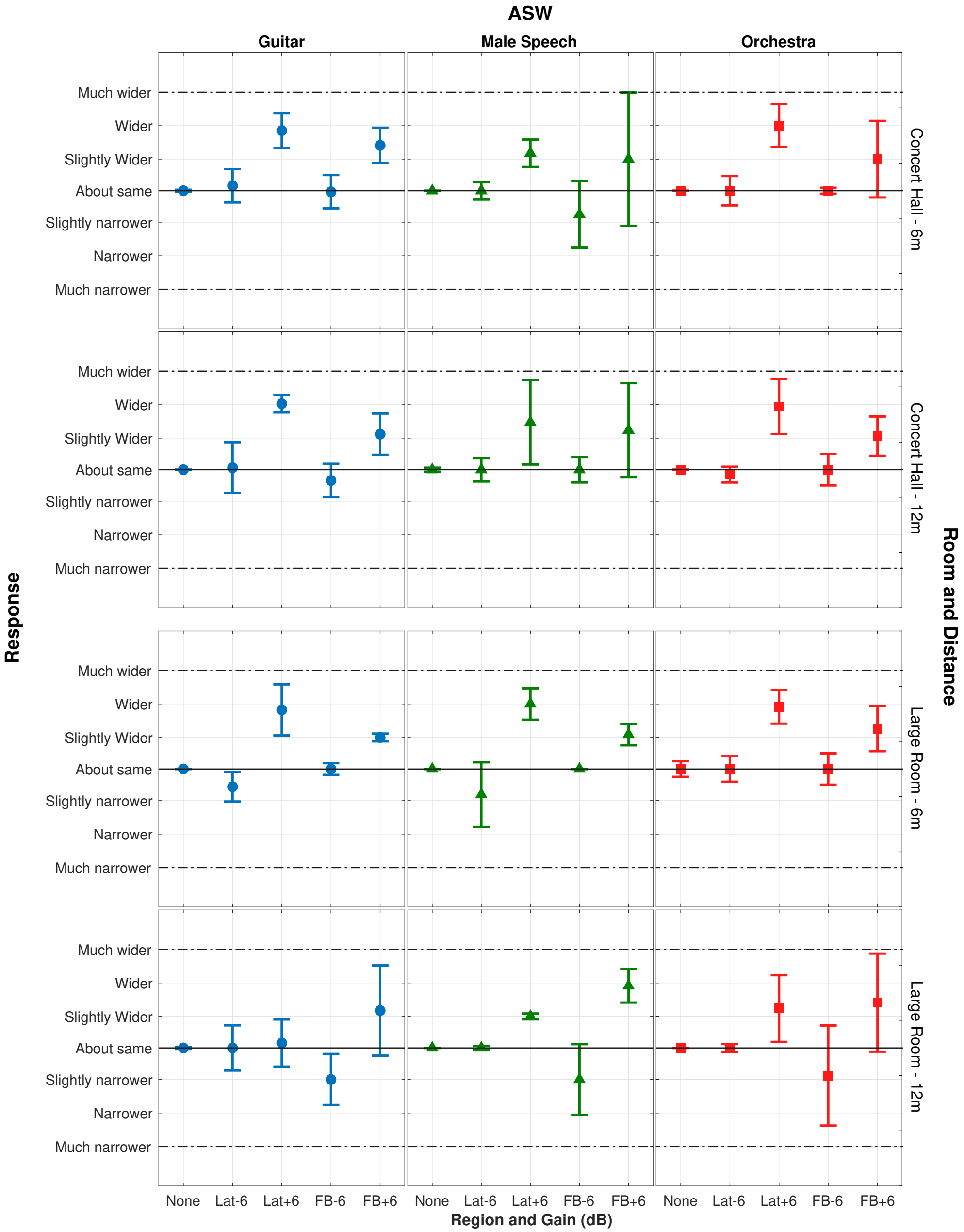


Fig. 8.2: Notch regions of subject responses for ASW

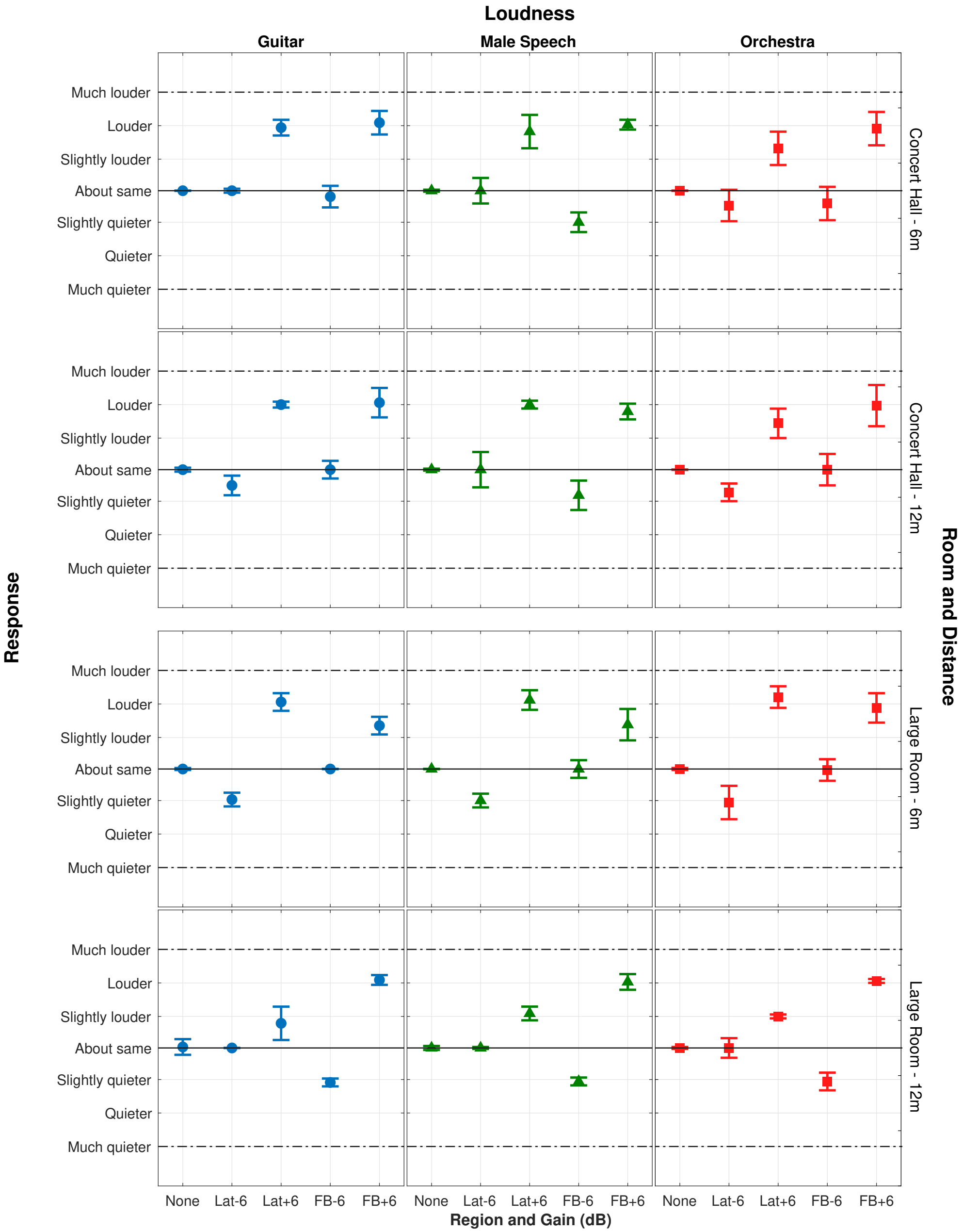


Fig. 8.3: Notch regions of subject responses for Loudness

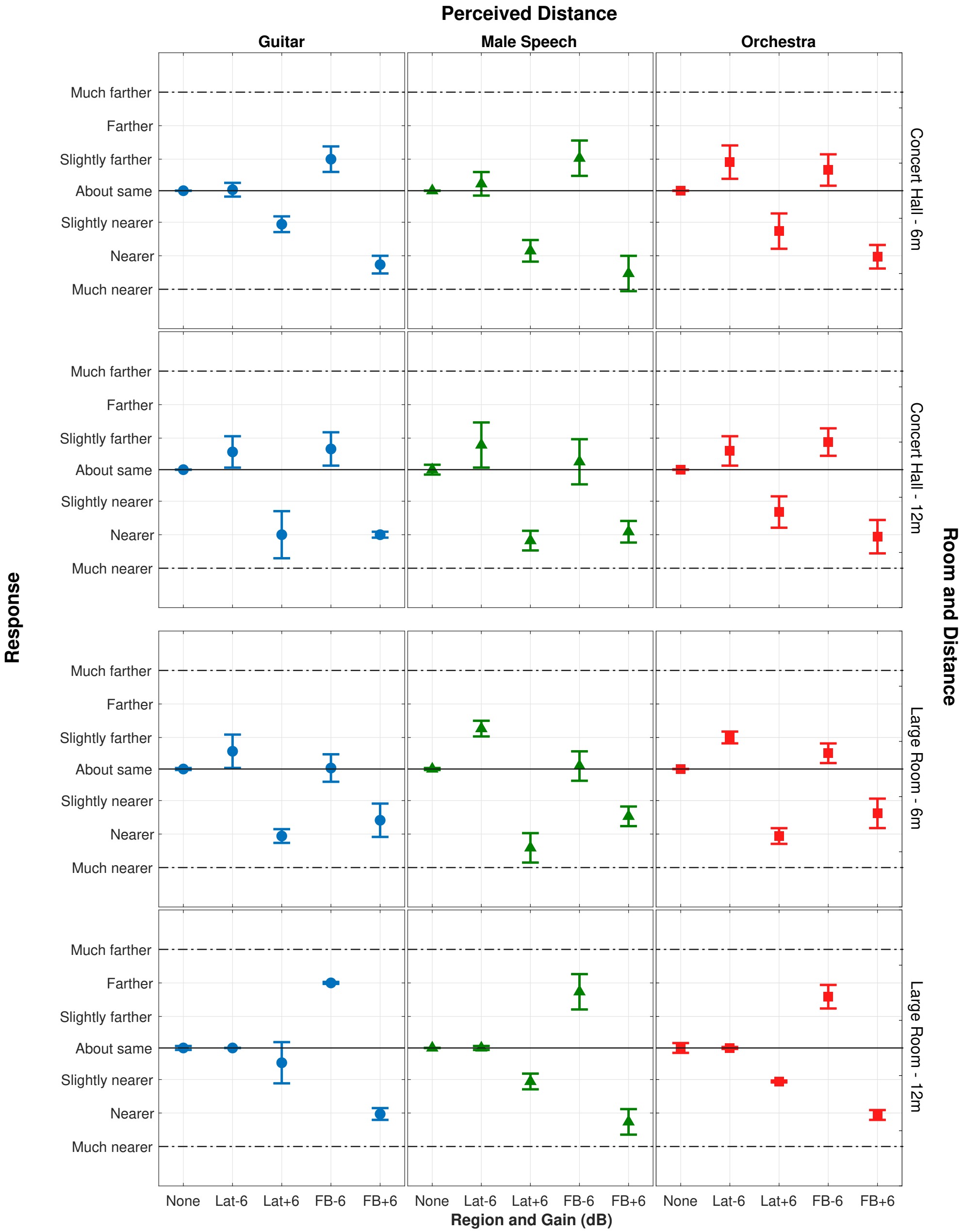


Fig. 8.4: Notch regions of subject responses for Perceived Distance

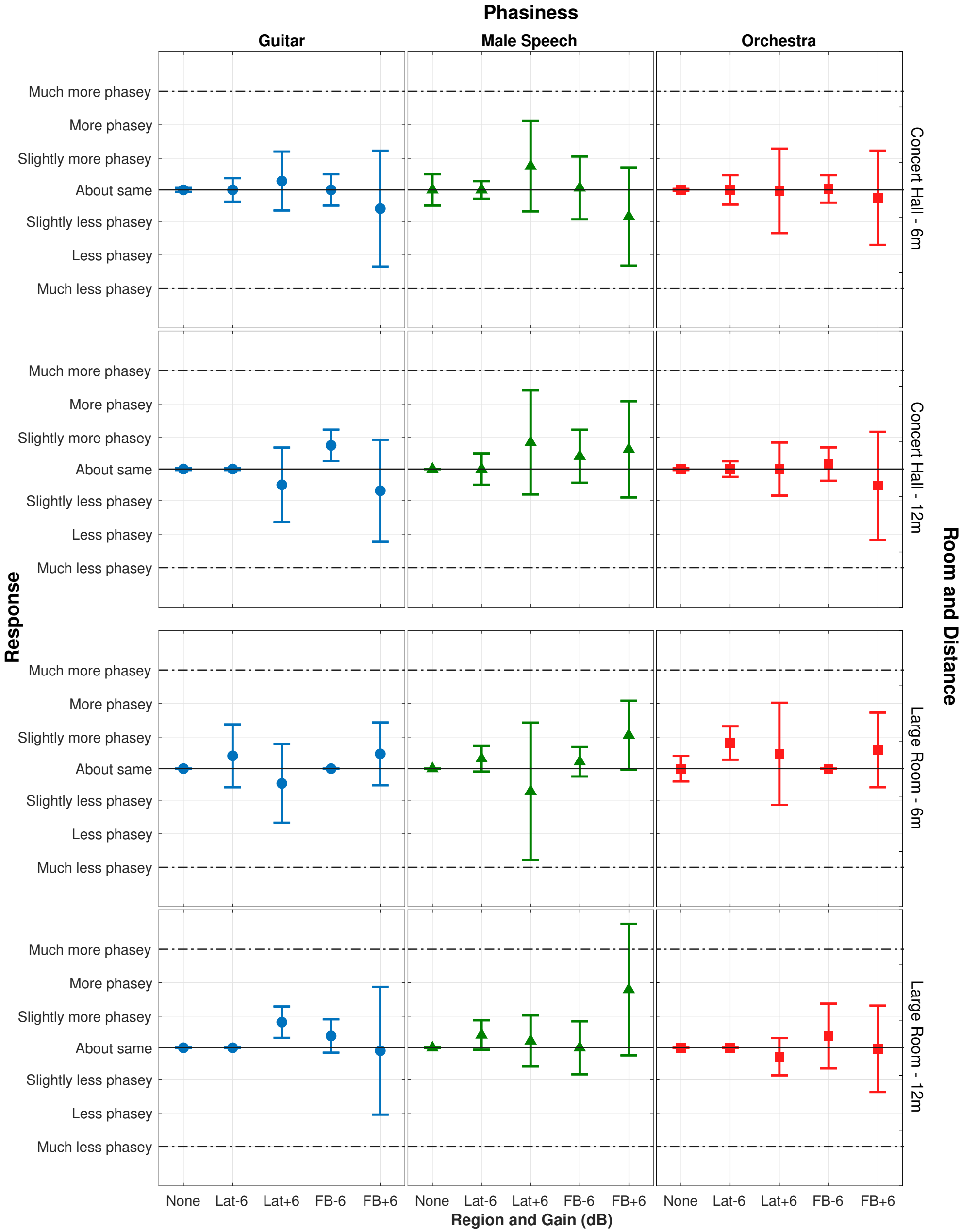


Fig. 8.5: Notch regions of subject responses for Phasiness.

8.2.1.1 ASW

Table 8.2 shows the Bonferroni adjusted p and effect size r values for the Wilcoxon signed rank test between the reference 'None' condition and each reflection group and gain condition.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Condition	p	r	p	r	p	r
Concert Hall	6m	Lat -6	1.000	.08	1.000	-.02	1.000	.08
		Lat +6	.449	.45	1.000	.33	.176	.53
		FB -6	1.000	-.33	1.000	.03	1.000	.08
		FB +6	1.000	.29	1.000	.24	1.000	.13
	12m	Lat -6	1.000	.11	1.000	-.13	1.000	-.34
		Lat +6	.449	.46	1.000	.37	.234	.50
		FB -6	1.000	-.29	1.000	-.08	1.000	-.17
		FB +6	.547	.43	1.000	.13	.742	.41
Large Room	6m	Lat -6	1.000	-.34	1.000	-.06	1.000	-.12
		Lat +6	.137	.54	.195	.51	.059	.58
		FB -6	1.000	-.21	1.000	.16	1.000	.27
		FB +6	.391	.47	.215	.50	.195	.52
	12m	Lat -6	1.000	-.19	1.000	.33	1.000	.00
		Lat +6	.781	.41	.371	.47	.410	.46
		FB -6	1.000	-.29	1.000	-.06	1.000	.08
		FB +6	1.000	.17	.156	.52	1.000	.30

Table 8.2: Bonferroni adjusted p and effect size r values between 'None' and each reflection group and gain value.

For all three sources in both rooms and at both source-receiver distances, the test found no significant difference in perceived ASW between the 'None' condition and all other conditions ($p > .05$). This mostly agrees with the Friedman test, except for the cases where the Friedman test found a significant effect on guitar and orchestra for Concert Hall 12m, and Large Room 6m. This is likely due to there being significant differences between other pairs of conditions, however, the main focus here is between 'None' and each condition. Whilst in most cases an increase in reflection energy had a medium to large effect size, with *r* values lying between .30 and .50, it did not have a significant effect on the perceived ASW, thus the null hypothesis must be retained.

When comparing the responses between the large room at 6m and 12m, there is a notable pattern that occurs between the median values for Lat+6, FB-6 and FB+6. At the 6m position, the median values for Lat+6 for all sources are higher than those for FB+6, and lie on the 'Wider' scale category. However, at the 12m position, they are lower than the median values for FB+6, and lie between the 'Same' and 'Slightly Wider' scale categories. These observations suggest that at 6m, there may have been potentially more lateral reflections available within the $\mathrm{ASW}_{\mathrm{max}}$ region to boost the level of, thus creating a more noticeable increase in perceived width. When at 12m, there may have been fewer reflections available to boost, thus not creating as much a dramatic difference in perceived width.

8.2.1.2 Loudness

Table 8.3 shows the Bonferroni adjusted *p* and effect size *r* values for the Wilcoxon signed-rank test between the reference 'None' condition and each reflection group and gain condition. For all three sources, the test found that in almost all cases an increase in either lateral or front-back reflection energy had a significant effect on perceived loudness. The only excluded case was in the large room at 12m where an increase in lateral energy did not have a significant effect. The effect size for all significant differences in loudness are above .60, signifying that the increase in energy of either group of reflections had a large effect. Thus, the null hypothesis can be rejected.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Condition	*p*	*r*	*p*	*r*	*p*	*r*
Concert Hall	6m	Lat -6	1.000	-.09	.938	-.40	.078	-.57
		Lat +6	**.020***	.63	**.039***	.60	**.020***	.63
		FB -6	.625	-.45	**.039***	-.60	.078	-.56
		FB +6	**.020***	.63	**.020***	.63	**.020***	.63
	12m	Lat -6	.078	-.56	.156	-.53	.273	-.49
		Lat +6	**.020***	.63	**.020***	.63	**.020***	.63
		FB -6	1.000	-.33	**.039***	-.60	1.000	-.28
		FB +6	**.020***	.63	**.020***	.63	**.020***	.63
Large Room	6m	Lat -6	.078	-.57	**.039***	-.60	.078	-.57
		Lat +6	**.020***	.63	**.020***	.63	**.020***	.63
		FB -6	1.000	-.33	.625	-.44	.156	-.53
		FB +6	**.020***	.63	**.020***	.63	**.020***	.63
	12m	Lat -6	1.000	-.41	1.000	-.30	.625	-.45
		Lat +6	1.000	.33	**.020***	.63	**.020***	.63
		FB -6	**.020***	-.63	**.039***	-.60	**.039***	-.60
		FB +6	**.020***	.63	**.020***	.63	**.020***	.63

Table 8.3: Bonferroni adjusted *p* and effect size *r* values produced by a Wilcoxon signed-rank test. The test compared the effect of gain adjustment upon the perceived loudness. * - Significant ($p < .05$), ** - Very Significant ($p < .01$)

On the contrary, the test found only in the large room at 12m did a reduction in lateral reflection energy have a significant and large negative effect on the perceived loudness, yet in all other cases it had no significant effect. It was also found that a reduction in front-back reflection energy only had a significant effect on perceived loudness for speech in the concert hall at both distances, and for all three sources in the large room at 12m only. In these cases, the effect size *r* was also negative and large. However, for other cases, the null hypothesis must be retained.

As noted previously with ASW in 8.2.1.1, when comparing responses for Lat+6 in the large room between 6m and 12m, the median values for Lat+6 are much higher at 6m than at 12m. For all sources, the median response at 6m lies on the 'Louder' category of the scale, whilst at 12m the median lies at 'Slightly Louder'. As suggested before, it is possible at 12m that there is a lack of lateral reflections available to manipulate. Whilst the difference is still significant, it is possible that an increase of lateral reflection energy at this distance could have been less noticeable.

Comparing the results for loudness against the results for ASW, with a reduction in energy, the response values for guitar for both ASW and Loudness lie close to zero. When there is an increase in energy, there is both an increase in ASW and Loudness. Whilst for speech, the increase in both lateral and front-back energy had medium effects on perceived width, there are not as large as the effects they had on guitar. However, both the median values of ASW and Loudness increase with an increase in energy for the speech source. This also applies to the orchestra, yet the effect of both types of energy boosting are larger. This trend between median values is also visible at 12m in the concert hall. In the large room at 6m, it can be seen that the median values for Lat+6 for both ASW and Loudness are larger than FB+6. However at 12m, the opposite occurs where the median values for FB+6 for both ASW and loudness are larger than Lat+6.

It is clear that an increase in both lateral and front-back reflection energy increased the perceived loudness to significant degree. Since the aim of neither control method was to affect the perceived loudness, this difference can be considered a negative artefact, such that an increase in perceived loudness may effect other attributes in an unknown manner. Since the aim of increasing the level of lateral reflections was to enhance the ASW in a controlled manner, it should solely perform that task, yet the analysis suggests that it is also affecting loudness.

8.2.1.3 Perceived distance

Table 8.4 shows the Bonferroni adjusted *p* and effect size *r* values for the Wilcoxon signed rank test between the reference 'None' condition and each reflection group and gain condition. Excluding guitar in the large room at 12m, for all other sources and cases an increase in either lateral or front-back reflection energy had a significant effect on the perceived distance. For all significant pairs, the effect size *r* was below -.50 and therefore indicates that the reduction in perceived source distance was large. This agrees with the lack of overlap between the corresponding 'None' and each +6 condition.

On the other hand, there are a few cases where a reduction in reflection energy had a significant effect on perceived distance of the speech source only. For a reduction in lateral energy, there was a significant effect in the concert hall at 12m and in the large room at 6m (p = .020). A reduction in front-back energy had a significant effect in the large room at 12m only (p = .039). All of these significances had an effect size above .50 indication that the effect was large, thus the reduction in energy caused a significant increase in perceived distance.

			Source					
			Guitar		Speech		Orchestra	
Room	Distance	Condition	*p*	*r*	*p*	*r*	*p*	*r*
Concert Hall	6m	Lat -6	1.000	.21	1.000	.20	0.312	.50
		Lat +6	**.020***	-.63	**.020***	-.63	**.020***	-.63
		FB -6	**.039***	.60	.117	.56	1.000	.34
		FB +6	**.020***	-.63	**.020***	-.63	**.020***	-.63
	12m	Lat -6	.977	.39	.020	.63	.156	.53
		Lat +6	**.020***	-.63	**.020***	-.63	**.039***	-.60
		FB -6	.078	.56	.078	.56	.117	.54
		FB +6	**.020***	-.63	**.020***	-.63	**.020***	-.63
Large Room	6m	Lat -6	.508	.44	.020	.63	.430	.45
		Lat +6	**.020***	-.63	**.020***	-.63	**.020***	-.63
		FB -6	1.000	.11	1.000	.34	1.000	.27
		FB +6	**.020***	-.63	**.020***	-.63	**.020***	-.63
	12m	Lat -6	1.000	-.15	1.000	.11	.625	.17
		Lat +6	1.000	-.33	**.020***	-.63	**.020***	-.60
		FB -6	**.020***	.60	**.039***	.63	**.039***	.60
		FB +6	**.020***	-.63	**.020***	-.63	**.020***	-.63

Table 8.4: Bonferroni adjusted p and effect size r values produced by a Wilcoxon signed-rank test. The test compared the effect of gain adjustment upon the perceived distance. * - Significant ($p < .05$), ** - Very Significant ($p < .01$)

It is interesting to note that when studying Fig. 8.4, in the concert hall at both distances, the median values each FB+6 case lie around -33, or the 'Nearer' category of the grading scale. However, the median values for each Lat+6 case in the same room are lower, or much nearer, at 12m than at 6m. At different distances, in this particular space the distribution of lateral reflections changes is greater than that of front-back reflections. The distribution of lateral reflections at 12m appeared to have a greater effect on the difference in perceived distance than at 6m. For front-back reflections, between 6m and 12m there was probably little difference in distribution, thus the extent of the difference remained the same.

In the large room, however, the median values for FB+6 are closer to -16, or 'Slightly Nearer', at 6m than at 12m where they lie just below -33, or 'Nearer'. This means that the difference distance was perceived to be greater at 12m than at 6m. Furthermore, at 6m the median values for Lat+6 are around -33, whilst at 12m they lie between 0 and -16. This is opposite to what was observed with an increase in front-back energy. This is also contrary to what was observed in the concert hall.

Finally, the median values of Lat-6 show that a reduction in lateral reflection energy had a small effect on the perceived distance. Whilst this is partially true for FB-6, it is only in the large room at 12m did a reduction in front-back reflections cause a significant increase in perceived distance for all sources. Again, this is likely due to distribution of the reflections for this particular space at this source-receiver distance.

To summarise, as found previously with Loudness in Section 8.2.1.2, the effect of increasing both lateral and front-back reflection has a considerable effect on the perceived source-listener distance. This can be seen as another potentially negative side effect of increasing early reflection. Whilst the aim was to either control ASW or Phasiness independently, the manipulation of reflection energy was not supposed to affect the perceived source-listener distance, highlighting a weakness in the design of this control technique.

8.2.1.4 Phasiness

For Phasiness, the Wilcoxon test found that for all conditions in both rooms and both source-listener distances, there was no significant effect of either control technique upon the perceived phasiness ($p > .05$), thus it was felt unnecessary to provide a table of p or r values. The test agrees with almost all of the notch edge overlaps between the gain adjustment conditions and the ’None’ condition, except for FB-6 for guitar in the concert hall at 12m, and Lat+6 also for guitar in the large room at 12m. With the exception of the guitar in the large room at 12m, for all other conditions the median values do not exceed either the ’Slightly More’ or ’Slightly Less’ categories of the scale. They in fact lie close to the ’About Same’ category.

The reduction of front-back reflection energy (FB-6) was designed to reduce the amount of perceived phasiness. However, as seen in Figure 8.5, this method had no significant effect on the perceived phasiness, thus demonstrating that it was not effective at reducing the phasiness. It is worth noting though that the notch edges of the FB-6 cases in speech did widen, suggesting that some subjects did potentially perceive a difference. This also does suggest that the effect of reducing front-back energy has on the perceived phasiness is subjective.

It can also be seen that the notch edges for Lat+6 and FB+6 in some cases did widen considerably. Whilst there no significant difference, for some subjects the increase in reflection energy may have been fairly noticeable, yet with the given experimental data this is difficult to conclude. In future work, it would be worth interviewing subjects and perform a large scale elicitation test as to understand what effects increasing the reflection energy have on colouration, if any.

8.2.2 Interaction between ASW, loudness and distance

When comparing between Figures 8.2 to 8.4, there appears to be an interaction between ASW, loudness and perceived distance when there is a change in reflection energy. For example, when observing the responses given in the ’Large Room - 6m’ condition, for all sources when the energy of either lateral or front-back reflections is increased, there is a significant increase in ASW and decrease in perceived distance when there is an increase in loudness. This suggests that boosting the lateral reflection energy to increase ASW may be simply due to the increase in loudness. A change in loudness is known to also cause a change in perceived source distance (Stevens & Guirao, 1962), which can be observed when comparing Figures 8.3 and 8.4.

To determine if there is any interaction between ASW, loudness and perceived distance, a 'Spearman's rho' correlation analysis was performed on the combined subjective scores for each room, source-receiver distance and source combination results using MATLAB.

Correlations between ASW and Loudness

		Source		
Room	Distance	Guitar	Speech	Orchestra
Concert Hall	6m	.359*	.390**	.346*
	12m	.610**	.290*	.633**
Large Room	6m	.666**	.524**	.592**
	12m	.327*	.468**	.275

Table 8.5: Spearman's rho correlation analysis performed on the subjective scores for ASW and Loudness for room, distance and source combination. * - significant, ** - very significant.

Table 8.5 shows that at in the concert hall for 12m, for the guitar and orchestra there is a strong correlation between ASW and Loudness, and in the large room at 6m, there is a moderate to strong correlation for all three sources. There is also a moderate correlation for speech at 12m in the same room. This suggests that, potentially, ASW increases due to an increase in loudness caused by the increase in reflection energy, regardless of the region of reflections that are boosted.

Correlations between ASW and perceived distance				
		Source		
Room	Distance	Guitar	Speech	Orchestra
Concert Hall	6m	**-.453****	**-.354***	-.278
	12m	**-.526****	-.246	**-.444****
Large Room	6m	**-.651****	**-.531****	**-.589****
	12m	**-.311***	**-.471****	-.261

Table 8.6: Spearman's rho correlation analysis performed on the subjective scores for ASW and Distance for room, distance and source combination. * - significant, ** - very significant.

The correlation values presented in Table 8.6 exhibit a similar pattern to the values presented in Table 8.5. This suggests that there the increase in ASW may also be associated with a decrease in perceived distance (i.e. the source sounds closer, thus appears to feel wider).

Correlations between loudness and perceived distance				
		Source		
Room	Distance	Guitar	Speech	Orchestra
Concert Hall	6m	**-.803****	**-.844****	**-.808****
	12m	**-.706****	**-.825****	**-.725****
Large Room	6m	**-.852****	**-.879****	**-.871****
	12m	**-.746****	**-.918****	**-.874****

Table 8.7: Spearman's rho correlation analysis performed on the subjective scores for Loudness and Distance for room, distance and source combination. * - significant, ** - very significant.

Table 8.7 shows that in all conditions for all sources, there is a strong negative correlation between loudness and perceived source distance. This strongly suggests that the changes in loudness due to the manipulation of reflection energy affect the perceived distance. When comparing all three tables, a strong trend can be observed between ASW, loudness and perceived distance simultaneously. Whilst there could be an influence of room type and physical source-receiver distance on these attributes, the above tables do not show any clear connection.

8.3 Objective analysis

The previous section showed that in certain cases there is possibly some strong interaction between ASW, loudness and perceived distance. This observation implies that ASW could be dependent on the sound strength factor G which, as discussed in Chapter 2 is a ratio of the sound energy emitted by an omni directional sound source measured at an arbitrary distance to the same sound energy measured at 10m in a free field. Sound strength can be regarded as being closely related to subjective 'loudness' (Beranek, 2011), and is given by:

$$G = 10 \log_{10} \frac{\int_0^\infty p^2(t)dt}{\int_0^\infty p_{10}^2(t)dt} \tag{8.1}$$

where p is an RIR measured at an arbitrary distance from the sound source, and p_{10} is a free field measurement at 10m from the same source.

Beranek (2011) found that early sound strength combined with BQI (Binaural Quality Index), otherwise known as [1 - IACC$_{\text{E3}}$] proposed by (Hidaka et al., 1995), to be useful in calculating DSB (Degree of Source Broadening), a measure that can be given by

$$DSB = 31 \cdot BQI + \frac{5}{3}(G_E) \tag{8.2}$$

where BQI is equivalent to [1 - IACC$_{\text{E3}}$] and G_E is the sound strength measured for the first initial 80ms from the initial onset of an impulse response.

Beranek (2011) concluded that G is an important factor to consider when planning new concert halls or evaluating existing ones. Lee (2013) later found that the perceived ASW was best predicted using G_E as the decrease in G_E closely followed the linear decrease in perceived ASW as the distance from source to listener doubled. Since G_E is related to both source loudness and ASW, as found from comparison between Figures 8.2 and 8.3 it is possible that the apparent loudness of the source influences the perceived ASW. The objective parameters LF, ASW, G_E and DSB are presented in the following sub-sections.

8.3.1 Lateral Fraction

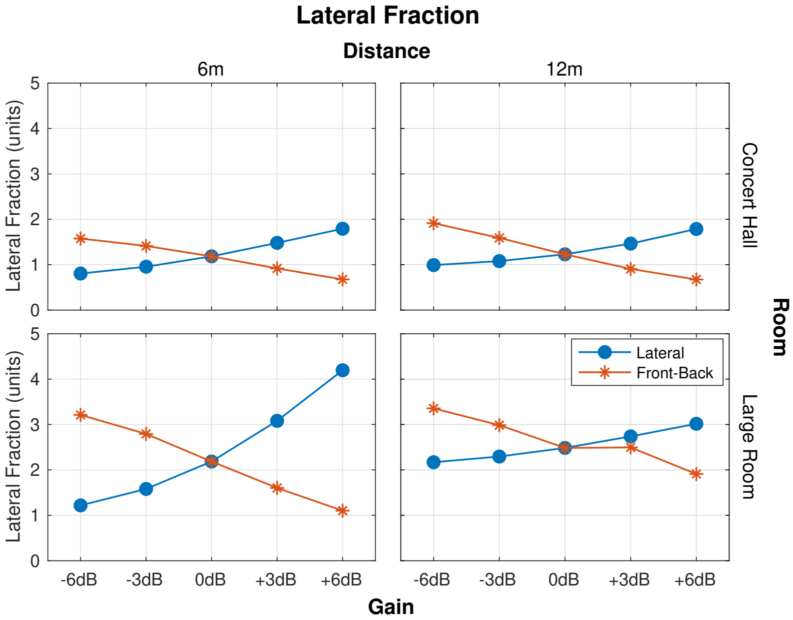


Fig. 8.6: Lateral fraction versus reflection energy gain per room per distance. Note: The lateral fraction parameter does not have any particular standard unit.

Observing first lateral fraction L_f, it can be seen that an increase in lateral reflection energy from -6 dB through to +6 dB gain will result in a fairly non-linear increase in $\mathrm{L_f}$, albeit in most cases it is small difference of 2 dB. The greatest difference in L_f is observed in the large room at 6m where there is a 3 dB difference in the measure between -6 dB and +6 dB lateral reflection gain. On the other hand, it is also seen that an increase in front-back energy from -6 dB through to +6 dB gain will result in a linear decrease in L_f. The difference here is also small, being as little as 1 dB, with the greatest difference occurring again in the large room at 6m where there is a difference of 2 dB in L_f between -6 dB and +6 dB reflection gain. Both of these observations were expected as L_f is a ratio of lateral to all early reflection energy. As the lateral reflection energy is increased during the energy adjustment process, there is a greater ratio of lateral energy to all reflection energy, and so L_f is expected to increase (Barron & Marshall, 1981). Inversely, as the front-back reflection energy is increased, the ratio of lateral to all early reflection energy decreases, and therefore L_f will decrease. However, in comparison with Fig. 8.2, this measurement does not match up with the observed increase in ASW. It was initially expected that only an increase in lateral reflection energy and not front-back reflection energy would result in an increase in perceived ASW, yet it is observed that the increases in front-back also yields a similar result.

8.3.2 Binaural Quality Index

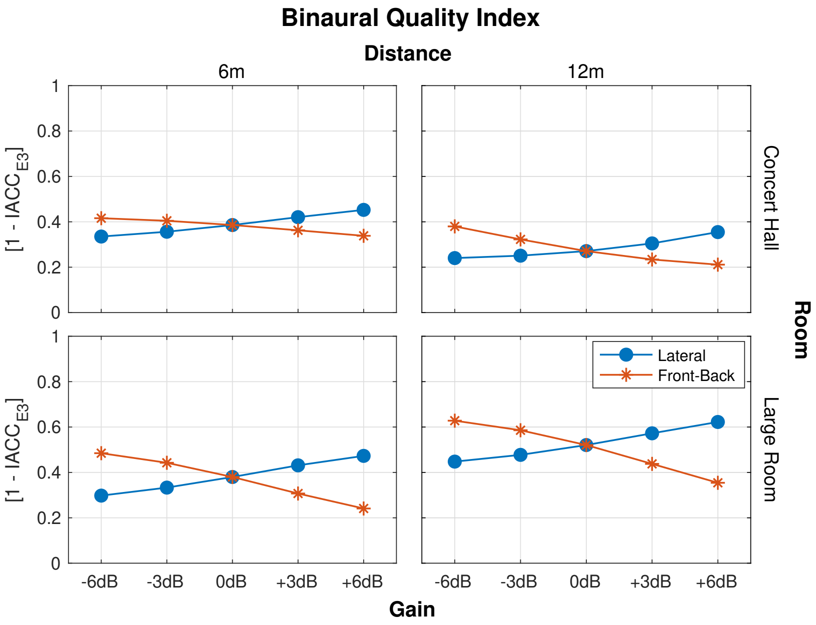


Fig. 8.7: Binaural Quality Index versus reflection energy gain per room per distance

The more widely established measure of ASW is [1-IACCE_3], or the binaural quality index (BQI) presented in Fig. 8.7. In all rooms at both source-listener distances, the increase in gain applied to the lateral reflection energy from -6 dB to +6 dB causes a slight increase in BQI, whilst when applied to front-reflection energy it causes a reduction in BQI. Whilst in all cases it appears to be subtle, the difference could be over the JND threshold of 0.065 ± 0.015, established by Okano (2002). The threshold, however, was measured for BQI values in the range of 0.4 to 0.8, and therefore is only mostly valid within that range.

Initial BQI		
	Distance	
Room	6m	12m
Concert Hall	0.39	0.27
Large Room	0.38	0.52

It is only at 12m in the large room that the initial BQI value is within the reference range for the JND threshold. However, at 6m in both rooms, it is only just below 0.4. Since the JND threshold at 0.4 is 0.065, it is implied that a reduction of up to 0.065 below 0.4, equating to 0.335 would be noticed. Since BQI is a prevalent measure for ASW, an analysis into the differences in BQI was carried out.

In the concert hall at 6m, a -6 dB reduction of lateral reflection energy results in a difference in BQI of 0.051. The is below the JND threshold and only just within the ±0.015 tolerance. Since this reduction had no significant effect on the perceived ASW, it can be regarded as an unnoticeable difference. A +6dB increase yields a difference of 0.067, which is above the threshold, and matches the observed subjective result. At 12m, the reduction in lateral reflection energy yields a difference of 0.031, which is also below the difference threshold, thus matching the subjective result, where a reduction had no significant effect on ASW. An increase on the other hand created a difference of 0.084, which is above the threshold and thus matches the subjective result.

In the large room at 6m, the reduction produces a difference of 0.082. Whilst this is above the JND threshold, it does not agree with the subjective result where in this condition the reduction had no significant difference. An increase created a difference of 0.093, which is above the threshold and matches the subjective result. At 12m, the -6dB reduction created a difference of 0.073, yet again this does not match with the same subjective result, whilst a +6dB increase created a difference of 0.102. Whilst this is considered a relatively large difference, the statistical analysis of the subjective results found that the increase did not always create a significant increase in perceived ASW. It is possible to deduce from these observations that the slight increases in BQI is related to the increase in ASW when the lateral reflection energy is increased. However, it does not explain as to why in the large room at 12m does an increase in said energy not significantly increase the perceived ASW.

Upon observation of the gain applied to front-back reflection energy and is effect on BQI, it can be seen that there is a slight negative correlation where a reduction of -6dB increases the BQI, whilst an increase of +6dB reduces the BQI. This is similar to what is observed with lateral fraction, yet like L_f, it does not match the observed subjective result where in some cases an increase in front-back reflection energy was causing the perceived ASW to increase. Although in some cases the effect was not significant, it was measured to be large. From this, it can be concluded that the BQI measure alone cannot adequately explain as to why the increase in front-back reflections was causing an increase in width.

8.3.3 Early Sound Strength

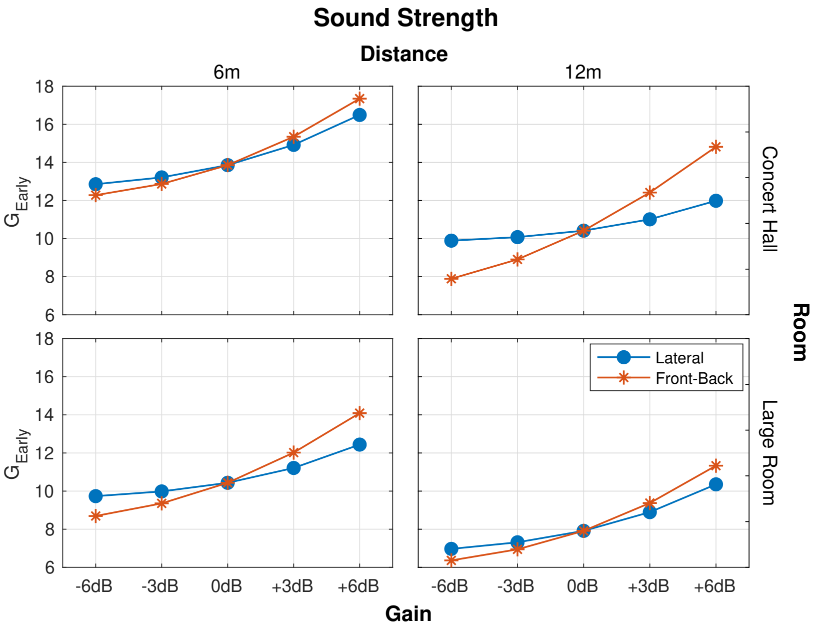


Fig. 8.8: Early Sound Strength versus reflection energy gain per room per distance

The early sound strength, G_E, is plotted against applied level adjustment in Figure 8.8. On initial observation, for either reflection group, it can be seen that the sound strength increases almost linearly with an increase in gain from -6 dB to +6 dB. As G_E is considered as a subjective measure for perceived loudness, it can be compared to the subjective responses presented in Fig. 8.3. Okano (2002) states that the JND threshold for G_E is 0.5 ± 0.2 dB.

In the concert hall at 6m, a 6 dB reduction of lateral reflection energy resulted in a 1 dB reduction in G_E, whilst at 12m the reduction is 0.52 dB. An 6 dB increase in lateral reflection energy at 6m increased G_E by 2.63 dB, and at 12m it increased it by 1.57 dB. In the large room at 6m, a 6 dB reduction in lateral reflection energy resulted in a 0.70 dB reduction in G_E, and at 12m results in a 0.95 dB reduction. On the other hand, a 6 dB increase in reflection energy at 6m yields a 2.01 dB increase in G_E, whilst at 12m it yields a 2.44 dB increase. All of the observed differences in G_E are above the JND threshold, thus when compared to the subject results for loudness in Fig. 8.3, G_E appears to correlate with the change in perceived loudness caused by a change in lateral reflection energy.

Focusing now on front-back reflections, in the concert hall at 6m a 6 dB reduction in energy results in a 1.58 dB reduction in G_E, whilst at 12m it results in a 2.52 dB reduction. In the large room at 6m, a reduction in energy results in a 1.74 dB reduction, whilst at 12m it results in a 1.55 dB reduction. A 6 dB increase in front-back reflections causes a 3.49 dB increase in G_E, and at 12m it causes a 4.40 dB increase. In the large room, a 6m an increase in energy results in a 3.66 dB increase in G_E, and at 12m it causes a 3.42 dB increase. All of the observed differences in G_E are above the JND threshold, thus when compared to the subject results for loudness in Fig. 8.3, G_E appears to correlate with the change perceived loudness resulting from a change in front-back reflection energy.

When the differences in G_E are compared to those caused by changes in lateral reflection energy, it is clear that the changes of front-back have a greater range of values and differences in G_E. This comparison could be linked to the median responses presented in Fig. 8.3. In the concert hall, the median values of perceived loudness for FB+6 are higher than those of Lat+6. Incidentally, at 6m the increase in front-back reflection energy created an increase in G_E that is 0.86 dB higher than when lateral reflection energy is increased. At 12m, the difference is 2.8 dB. However, in the large room at 6m, whilst the median values of FB+6 are lower than those of Lat+6, the increase in front-back reflection energy still produces a larger difference of 1.65 dB in G_E than with an increase in lateral reflection energy. The observation shows that whilst G_E appears to be linked to loudness, it does not explain why the median values for FB+6 are lower than Lat+6 at 6m in the large room since the prediction does not match the subjective results.

8.3.4 Degree of Source Broadening

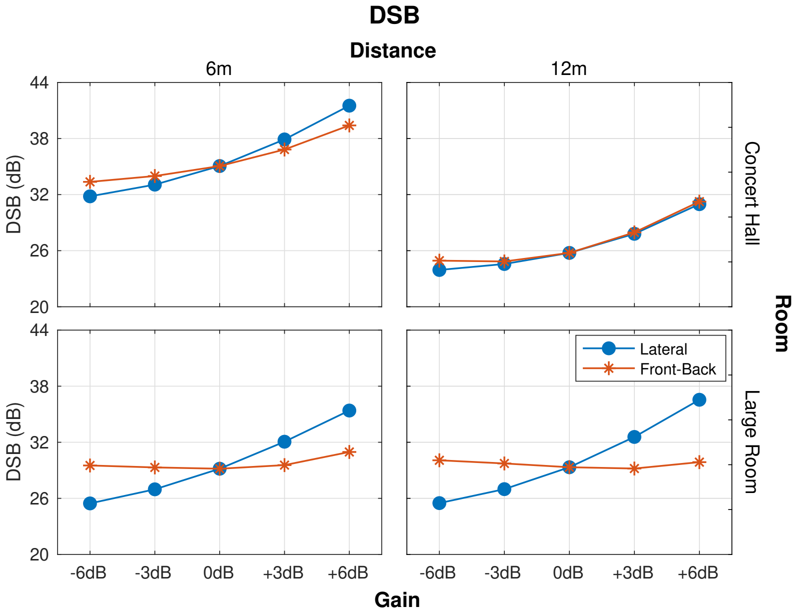


Fig. 8.9: Degree of Source Broadening versus reflection energy gain per room per distance

Degree of Source Broadening, DSB, proposed by Beranek (2011) is a combination of the BQI and G_E. Like the previous measures, it was calculated for each room, source-listener distance and reflection group, see Fig. 8.9. In all cases, DSB follows an increasing, monotonic curve as gain from -6 dB to +6 dB is applied to lateral reflections. However, as the same gain adjustment is applied to front-back reflections, DSB only follows the same curve in the concert hall, whilst in the large room it remains relatively flat between both extremes of the adjustment.

A major point of interest is that in the concert hall, there is an expected decrease in initial DSB at 0 dB gain as the distance is double. This is expected as the G_E parameter decreases as the distance decreases. Yet, in the large room, the initial DSB at the same gain value remains almost the same between the two distances. When comparing the plots for BQI and G_E, Figs. 8.7 and 8.8, it can be seen that as the distance doubles, the initial G_E decreases whilst the initial BQI increases. When the two parameters are then combined to calculate DSB, the ratio between them at the two distances remains near constant. This also explains as to why the DSB curve for the front-back reflections remains flat: the corresponding BQI and G_E functions 'cancel out'.

When compared to ASW presented in Fig. 8.2, in both rooms, DSB follows the increase in lateral reflection energy, except at 12m in the large room where DSB does not match the observed subject result. It is possible that the reduced sound strength at this distance means that the increase +6 dB in lateral reflections is not enough to create a significant increase in ASW. It could be that the initial lateral reflection energy is very low at this distance in this particular room, thus the change in ASW may not have been as noticeable.

8.4 Discussion

The statistical and objective analysis gives an important insight into the effectiveness of the perceptual control methods. Focusing first on the effects of level adjustment in the ASW_{max} region, it appears that increasing reflection energy in this region, in some cases, significantly increased the perceived ASW. This was the desired outcome when boosting the energy in the ASW_{max} region, satisfying the first research question in that this method was effective at altering ASW. However, a reduction in ASW_{max} region energy did not effectively reduce the ASW, so the control method was only partially effective in that it could enhance ASW, yet not reduce it. This is possibly due to reflections still being present in the front-back region causing ASW to be at a *minimum*, such that even with lateral reflections being suppressed there is still a minimum amount of ASW due to the presence of unsuppressed front-back reflections.

Shifting focus to the effects of manipulating the energy in the front-back region. It was found that manipulation of the front-back reflection energy had no significant effect on phasiness. Thus, this dissatisfies the first research question in that it was *not* effective at altering phasiness. Furthermore, whilst it was not expected for boosting the energy in the front-back region to have a significant effect on ASW, the observations made during statistical analysis show otherwise. This of course dissatisfies the second research question in which this method was only supposed to affect phasiness and not ASW.

Regarding the third research question, *'To what degree does each method affect loudness and distance perception'*, the statistical and objective analysis found a fairly strong interaction between ASW and either loudness or perceived distance, and a very strong interaction between loudness and perceived distance themselves. The third research question is itself an extension of the second research question, thus because it appeared that there was a fairly strong interaction between ASW and either loudness or perceived distance, both perceptual control methods dissatisfy the second question as they also affect other attributes. Therefore, the only method that appeared to partially operate with its intended effect was the increase in $\mathrm{ASW}_{\mathrm{max}}$ region energy, yet with the drawback of inadvertently affecting loudness and perceived distance.

8.4.1 On the effect of loudness on ASW and perceived distance

It is worth noting that subjects were quoted to have found that, in some cases, if the source was perceptually louder than in the reference condition, it was then perceived to be closer and therefore graded as being wider or more horizontally spread. Inversely, if the source was perceived to be quieter than in the reference condition, it was felt to be farther away and so was perceived to be narrower. Naturally, if one was to visualise the auditory source, if it is close to the listener it would, in reality, appear to be larger and louder. Drawing this observation in parallel to the results of this experiment, whilst there were no visuals during the test, it is possible that subjects are mentally visualising the auditory source as being wider or more horizontally spread when there is an increase in perceived loudness. This is understandable when considering the fact that subjects were instructed to directly focus on the perceived, horizontal width of the source image, see Figure 8.10.

This suggests that perhaps the description of ASW being *'the perceived auditory width of the source'* is too general, or that it can be misunderstood. The subjects' descriptions show that perhaps at times the environmental width increased, whilst at other times the changes in actual perceived auditory width of the source changed due to changes in loudness. This further suggests that perhaps two distinguishing sub-paradigms of width are required, one which describes the apparent source width in terms of its perceived size, and on that described the perceived environmental width. This was initially suggested by Rumsey (2002) to reduce confusion in the types of width and envelopment.

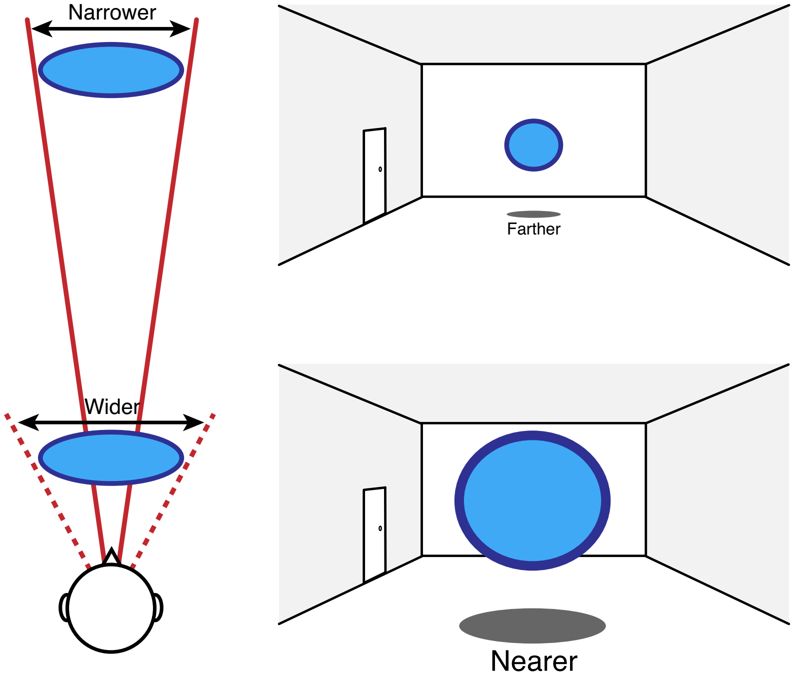


Fig. 8.10: Visualisation of the link between width, distance and loudness. An object far away would be seen as smaller and quieter. However, if it was nearer, it would appear to be larger and louder.

As shown in the objective analysis, parameters L_f, BQI and DSB do not explain as to why the increase in front-back energy caused a noticeable increase in ASW. These measures show that ASW should not change, or even decrease. As noted by Lee (2013), the G_E parameter appears to be a better indicator for measuring ASW. However, as it is an energy parameter, not a cross-correlation parameter, it is measuring perceived loudness. Therefore, the increase in G_E leads to an increase in loudness. Since it is combined in the DSB measure (Beranek, 2011), it can be considered an indicator of width. Therefore, in combination with the findings in this study, loudness must somehow influence ASW. Secondly, there is research performed by Stevens and Guirao (1962) into the relationship between loudness and perceived distance, which was later reviewed by Zahorik, Brungart, and Bronkhorst (2005), who found that loudness and distance are inversely related. From these speculations, G_E could be an indicator of perceived distance, thus could link ASW, loudness and perceived distance together, and better predict the perceived size of the source image.

8.5 Conclusion

Since the alteration of front-back energy had unexpectedly appeared to have an effect on ASW in the previous chapter, the main objective of this experiment was to assess the effectiveness of each perceptual control technique in their ability to enhance or suppress only their corresponding attribute. The manipulation of early lateral reflection energy should have only affected ASW, and the manipulation of early front-back reflection energy should have only affected phasiness. However, the statistical analysis performed in Section 8.2 found that only an increase in lateral energy caused an increase in ASW that was room and distance dependent, and that a reduction on energy had no effect. It was also found that manipulation of front-back energy had no significant effect on phasiness, and that an increase also had an effect on ASW.

As the previous chapter also found that the manipulation of the energy in both regions potentially had an effect on perceived source loudness and distance, the secondary objective of this experiment was to observe any possible effects the perceptual control may have on those attributes, and to what degree. The same statistical analysis found that increasing either early lateral or front-back reflection energy had a significant effect on loudness and distance. This of course signifies that the perceptual control techniques had failed to affect only their corresponding attributes. However, correlation analysis found that there was a distinct interaction between ASW, loudness and distance, suggesting that auditory width may have been increased due to the increased loudness, which in turn reduced the perceived source distance.

It was also found that the objective predictors L_f, BQI and DSB did not adequately match the perceived change in ASW due to the increase in front-back reflection energy. They predicted that in some cases ASW should have remained the same, or in fact reduced. Since the increase in loudness and reduction in perceived distance, it was found that early sound strength G_E parameter better matched the observed effect. Therefore, it is speculated that ASW increased due to an increase in loudness, which subjects translated to a reduction in distance, and thus naturally felt that source image had become wider. This of course is experienced in everyday scenarios where objects will appear visually and audibly larger and louder when closer.

To summarise, the main conclusions from this experiment are:

- In certain room conditions, the boosting of lateral reflections had some effect on ASW.
- Manipulation of front-back reflections does not reduce perceived phasiness, and manipulation of Lateral reflections appear to have no significant effect either.
- Increasing the energy of either lateral and front-back reflections unintentionally

affects loudness and perceived distance.

- The effectiveness of the control method does not greatly depend on source type.

- The degree of effects on loudness and distance is room dependent.

8.6 Future work

The points discussed in the previous section demonstrate that the current spatial control techniques, whilst novel, are limited and thus requires future work to be properly realised. To reiterate, the ultimate aim with the perceptual control techniques under test here are to enhance or diminish perceptual attributes without affecting others. Whilst the experiments performed here found that both techniques did not meet that requirement, they had an effect. The boosting of lateral reflection energy had caused some significant increase in perceived width, yet the reduction of front-back reflections had no significant reduction in phasiness. Thus, the techniques need to be refined through further research and experimentation.

It is known from analysis that there is a strong interaction between ASW, Loudness and Distance, and they could be linked by the G_E parameter, which is an energy parameter. The current algorithm assumes that simple manipulation of lateral ray energy, a principle based upon the operation of lateral fraction (Barron & Marshall, 1981) and the results from Chapter 5, should have been more than adequate for controlling ASW. On the other hand, the objective analysis in this experiment found that the link between G_E and ASW, and the correlation between ASW and loudness, suggest that loudness of early reflections can have an effect on ASW. This falls in line with the predictions made by the L_f parameter. This leads to a hypothetical improvement to the ASW control technique: to enhance ASW without affecting loudness, the average energy of the early reflections after the modification process should be equal or at least near to the energy prior to modification.

The current research presents the first few steps towards a perceptual optimisation technique that will analyse the reflections in a raw impulse vector (RIV) and, whilst still being agnostic of the room type or source-listener distance, decide whether increasing lateral reflection energy, front-back reflection energy, or other parameters, will result in a more subjectively optimal listening experience for the listener. This type of algorithm will need deeper analysis techniques that are able to account for several parameters, for example neural networks and machine learning, or even simpler, a better understanding of the psychoacoustic parameters and their underlying functionality.

Chapter 9

Summary and Conclusions

9.1 Chapter 1

Chapter 1 introduces the reader to the primary aim of thesis which was to develop and apply perceptual control methods to virtual acoustics. This was believed to be achievable by taking advantage of reflection meta-data that is otherwise normally inaccessible in then existing virtual acoustics programs. Such methods would be highly beneficial to VR experiences where the spatial impression of the scene could be improved by enhancing positive effects whilst suppressing the negative. Although it was found that perceptual control methods for artificial reverberation already existed, it was realised that these did apply them to geometrical virtual acoustics which are currently being used in VR experiences today. The primary research questions set out by the introduction were:

1. What reflection meta-data relates to the perception of ASW and tonal colouration?

2. Can the meta-data be used to sort and group reflections by how much they will influence each attribute?

3. What perceptually motivated methods should be used to manipulate these grouped reflections in order to control each attribute?

4. How effective are the control methods at manipulating these attributes?

5. Is it possible to control them independently?

9.2 Chapter 2

In Chapter 2, the spatial Impression (SI) paradigm, and perception of tonal colouration is introduced and reviewed. The properties of SI and its two sub-paradigms 'Apparent Source Width' (ASW) and 'Listener Envelopment' (LEV), as well as the predictors and measures for these two sub-paradigms were introduced. Several measures for the two sub-paradigms were reviewed. For ASW, the measures include: lateral fraction (L_f) which depends on the intensity and direction of reflections; the inter-aural cross-correlation coefficient (IACC) which analyses the similarity between ear signals; and time-varying fluctuations in inter-aural time and level differences, or ITD and ILD. LEV is generally agreed to be best predicted using the late lateral reflection measure ($\mathrm{LF}_{80}^{\infty}$). When reviewing the effects of reflection direction on ASW in particular, interpretation of the results from research regarding L_f suggest that there may be a region on the horizontal plane where the perceived ASW saturates before the reflection has reached 90°. This region would define an area of maximum ASW. In combination with the understood effects of the reflection energy on ASW, the energy of reflections arriving in this hypothetical region should have the greatest influence on the ASW. Thus, ASW could be controlled by manipulating the energy of only these reflections. From this review, it was hypothesised that:

- A region may exist on the horizontal plane that defines reflections arriving within it as those that cause the maximum amount of perceived ASW. The energy of these reflections would have the greatest influence of the degree of ASW.

- Whilst delay time is understood as not having a significant effect on the degree of ASW, it is possible that it could affect the boundary angles of the hypothetical region.

The second part of the chapter reviewed the perception tonal colouration, which describes the effects that reflections, and inherently the geometry of a space, have on the timbre of an audio source. The nature of this effect is dependent on the delay time between the direct sound and the reflection, and for a single reflection this usually results in the ′comb filter′ effect. It can also result in other effects which can be described as ′boxiness′, ′metallicness′ or ′fullness′. Whilst tonal colouration was initially understood to be a negative effect and a distraction, it was later understood that the colouration effects could be pleasant and would help support or enhance the audio source, which is understood to be important for orchestral instruments. More recent literature has begun to investigate the effects of reflection direction on the perception of colouration, although it is still a poorly understood area of research. However, the effects of reflection direction on a subject′s preference of SI are more understood. For the proposed perceptual control method, it is vital to understand the effects of reflection direction on perceived colourations, and whether a certain directional region causes colouration that is more pleasant than other directions, such that reflections can be sorted by whether they produce *acceptable* or *un-acceptable* colouration. From this, it was hypothesised that:

- Since tonal colouration will inherently be perceived in any reverberant space, and that listeners appear to have a preference on SI depending on reflection direction, it could be possible that reflection direction affects the listener's preference of the colouration.

- The tonal colouration caused by a lateral reflection could be more preferable than that caused by a frontal reflection, which in turn could define a region of unacceptable colouration.

- Due to effect of time delay on the nature of the tonal colouration, it is possible that the acceptability of the colouration may change with delay time, thus possibly affecting the angle at which the colouration is acceptable.

9.3 Chapter 3

The first part of Chapter 3 reviewed several methods and approaches to simulating room acoustics and reverberation, their advantages and disadvantages regarding real-time and accurate modelling room acoustics, and their application in perceptually controllable reverb. The three types of methods that were considered are 'Algorithmic','Wave based' and 'Geometric'. Algorithmic reverberators use a network of digital delays and filters to model reverb, and due to their efficiency they are widely used in music production to add *life* and *realism* to a song. However, algorithmic reverberators on their own cannot accurately model the exact reflection times and decay of a particular space or geometric model, and are usually combined with other methods in order to achieve accurate simulation. Wave based methods, and specifically the ’Digital Waveguide Mesh’ (DWM) can accurately model sound propagation through a geometric model, as well as simulate wave phenomena such as cancellation and diffraction. However, they are limited to modelling relatively low frequency waves due to the high amount of computer resources required, especially for real-time usage.

Finally, geometric methods were considered, and these are split into three common methods: Ray Tracing, the Image Source Method (ISM) and Beam Tracing. These methods model the individual paths of reflections, and each method is suitable to either modelling early or late reflections. Whilst they are unable to model wave phenomena, if carefully programmed and are well optimised for fast rendering, they can dynamically model the acoustics of a space in real-time. The biggest advantage that geometric methods have over the other methods is that they allow access and control over independent reflections, which is vital to perceptually controllable reverb that requires this level of control. The review also discusses so-called ’Hybrid Methods’ that combine two types of reverberation methods to cover the drawbacks of each method. One example would be the combination of the ISM and ray tracing, where the ISM can accurately model early reflections, whilst ray tracing models the diffused reverb tail. Hybrid methods also allow for efficient real-time usage, which is important for applications such as video games and virtual reality.

Finally, the chapter reviewed perceptual control methods for reverberation. These provide high level control over several perceptual attributes such as clarity, envelopment, brilliance and warmth. Some methods approach this by controlling the level and spectrum of specific temporal groups of reflections, whilst others use statistical models to provide perceptual control by dynamically controlling size, decay and damping. When reviewing the existing literature, it was found that these controlled groups of reflections based solely on their arrival time and did not take into account reflection direction. Therefore, this would not allow control over attributes such as ASW, which depends on the reflection direction. This limitation was partly due to usage of algorithmic reverberation methods, rather than geometric methods that would otherwise allow for direction of arrival to be taken into consideration. To summarise, it was found that:

- For perceptual control over ASW and tonal colouration, it was decided that the most suitable artificial reverb method for this project was a geometric method.
- Current, commercially available geometric methods would only produce an impulse response in the form of an audio file, and do not commonly give access to the captured reflections along with their meta data, which is required for the proposed perceptual control method.
- A custom, geometric reverberator was to be developed that would give access to each rendered reflection so that the perceptual control methods could be applied.

9.4 Chapter 4

Chapter 4 discusses the development of a custom geometric, artificial reverberation algorithm. The custom algorithm, denoted as CA, was written in the C++ programming language, and fulfils the reflection access requirements discussed in chapter 3. It combines both the ISM and ray tracing methods to allow for accurate modelling of the early reflections, as well as late reflection and diffusion modelling. The algorithm is also able to model octave-band surface and air absorption. The CA produces both a room impulse response (RIR), as well as a novel file format known as a 'Raw Impulse Vector' (RIV). This file stores each captured sound ray including, its delay time, direction of arrival, octave band energy and surface history. The chapter also discusses a spatialisation process that can convert the RIV into a binaural RIR (BRIR), and the perceptual control methods that can be applied to reflections extracted from the RIV during spatialisation. To verify the validity the CA, it was compared against the existing commercial software package ODEON 14. A BRIR of a simple concert hall model was rendered by both programs, which was then both visually and objectively compared. It was found that:

- The custom algorithm could accurately render the timing of early reflections up to 80ms.

- The initial early decay time was faster in the custom algorithm BRIR than in the ODEON BRIR, and the -60 dB reverb time (RT_{60}) was found to be 504ms quicker in the CA BRIR. This is most likely due to a difference in how absorption is modelled in both programs, though this is unclear and difficult to verify.

- The ASW, measured using [1-$IACC_{E3}$], was found to be higher in the CA BRIR than in the ODEON BRIR.

Whilst there is a difference between the two algorithms, the CA was not developed to be a complete recreation of ODEON, and does appear to generate perceptually plausible reverberation. The purpose of the CA was not to model room acoustics with total accuracy, albeit it is able to model early reflections we a good level of precision; it is designed to provide more *open* access to the reflections such that perceptual control methods can be applied to geometric acoustics.

9.5 Chapter 5

Chapter 5 builds upon the hypothesis brought forward in Chapter 2 where it was speculated that the degree of perceived ASW saturates before a lateral reflection reaches 90°, and that a region of maximum ASW may exist between 30°to 160°. This region, denoted as $\mathrm{ASW}_{\mathrm{max}}$, and its corresponding boundary angles are hypothetically not time dependent. To determine if this region exists, a threshold test, the 'Transformed Staircase Method', was performed over headphones. In the presence of a direct sound, the angle of a delayed, -6 dB reflection was varied between 0°to 90°, or from 180°to 90°, creating a varying comparison sound field. The direct sound and reflection were convolved with corresponding HRTFs. Using a male speech sample as an audio source, subjects were asked if they could perceive a difference in ASW between the comparison sound field and a reference sound field which consisted of the fixed, 90°reflection. The test would locate the average reflection angle where subjects would perceive a just noticeable difference in ASW. Statistical and objective analysis found that:

- A delay time between 5 to 30 ms has no significant effect on the front or rear average reflection angles.

- The average front and rear reflection angles were calculated to be 38.9°and 134.1°, thus defining a region of maximum ASW, or $\mathrm{ASW}_{\mathrm{max}}$, between these two angles. Therefore, any reflection arriving within this region would produce maximum ASW.

- The most likely cause for this novel finding is a plateau in the measured IACC$_{E3}$, where the function appears to reach fairly level minimum approximately between the two reflections.

- It is also likely caused by a saturation in the standard deviation of fluctuations in the ITD and ILD in the 500 Hz and 1 kHz bands, which also occur within the ASW$_{\text{max}}$ region.

Since these findings are limited to binaural playback over headphones, it was decided to verify the findings by simulating similar sound fields using loudspeakers. An arc of loudspeakers was created on the horizontal plane to mimic the sound fields used in the previous section. A multiple comparison test was used, where each stimulus was a sound field consisting of the direct sound from a speaker in front of or behind the listener, and a reflection arriving from a random speaker in the arc. Subjects were tasked with grading the perceived degree of ASW caused by each stimulus. This time, both a speech and cello sample were used. The verification test found that:

- Subjects were unable to distinguish differences in width at angles outside of the ASW$_{\text{max}}$ region.

- The presentation method, or sequential versus random selection of stimuli may have influenced the ability for subjects to distinguish these differences.

9.6 Chapter 6

Chapter 6 investigated the effect of reflection direction and level upon the audibility and acceptability of tonal colouration. The objective of the chapter was to locate the reflection azimuth angles where the colouration becomes acceptable. These angles would define horizontal regions of unacceptable colouration, one in front of and one behind the listener. A reflection arriving within these regions would cause unacceptable colouration. The investigation was split into three parts: informal elicitation, and two threshold tests.

An elicitation test was designed to clarify the types of colouration and timbral effects a single reflection has in the presence of a direct sound. A small set of experienced subjects listened to a speech sample in various single reflection sound fields at several azimuth angles and delay times between 2.5 to 30 ms. Whilst it was already well understood in previous literature what timbral effects a single reflection has, for the purposes of this study it was worth documenting the effects for the playback method and parameters used for this experiment. The elicitation test found that:

- The most common term elicited was 'Metallic', occurring a total of 17 times.
- The delay time of the reflection, as expected, affected the characteristic of the tonal colouration, where at low delay times the sound was perceived as *'Harsh'*, *'Metallic'* and *'Phasey'*, and higher delay times there was a perceived *'Fullness'* and *'Modulation'*.
- For almost all delay times, the most common perceived quality was *'Roughness'*.

To find the independent regions of audible and unacceptable colouration, a 'Method of Adjustment' test was performed. For each reflection delay time, the direction was adjusted from its initial starting position either directly front of or behind the listener until they found the point where the colouration becomes inaudible / acceptable. The test found that:

- Colouration is audible when an early reflection arrives between ±50°in front of the listener, and ±135°behind the listener.

- The same colouration is perceived as unacceptable when the reflection arrives between ±41°in front, or ±143°behind the listener.

- Whilst the audible and unacceptable regions are coincident, there is a significant difference between their boundary angles, such that there are transition areas where the colouration may be audible, yet is acceptable.

- Delay time had no significant effect on the position of boundary angles of any region.

- The regions of unacceptable colouration are adjacent to the ASW_{max} region defined in Chapter 5. Therefore, there could be a relationship between ASW and the acceptability of tonal colouration.

- Spectral analysis of the sound field showed a reduction in the comb-filtering effects in the left ear signal as the reflection angle became lateral and entered the acceptable colouration region.

Once the regions were established, it was hypothesised that a reduction in reflection level within the unacceptable colouration region may make it become acceptable. To find how much a -6 dB reflection would need to be reduced by to make the colouration acceptable, a 'Transformed Staircase' test was performed. At two fixed reflection angles, 0°and 20°in front of the listener, and four delay times between 2.5 to 20ms, the reflection level was manipulated until the subject felt that the colouration was acceptable. It must be noted that because the colouration in the previous threshold detection test was perceived to be unacceptable both in front and behind the listener, it was assumed that the same amount reduction would be required both in front of and behind the listener. Thus, for the sake of efficiency, it was decided that only the front region would be tested. The test found that:

- For a -6 dB reflection, at both 0°and 20°very little level reduction was required to make the colouration become acceptable.

- Whilst there was no significant effect of delay time on the reduction at either reflection angle, for 20ms delay time at 0°the median value of mean reduction was higher than at lower delay times.

- The reflection angle did not have a significant effect on the mean level reduction. However, the median value at 0°was higher, which potentially means more reduction is required for frontal reflections rather than lateral reflections.

It must be stressed that the conclusions presented for this chapter are valid for speech signals, thus the effect of source type upon the characteristics of the colouration, the location of the un-acceptable region boundaries and the amount of level reduction is not fully understood. After the regions of unacceptable colouration had been found, they could then be used to sort reflections extracted from a raw impulse vector as those that cause unacceptable colouration, such that their level can be reduced to improved the perceived colouration.

9.7 Chapter 7

After establishing the regions of maximum ASW and unacceptable colouration, Chapter 7 investigated how those regions can be used to perceptually control those attributes. Section 7.1 proposed that the degree of ASW could be controlled by manipulating the level of the reflections within the ASW$_{max}$ region, based upon the operation of 'Lateral Fraction' (Barron & Marshall, 1981) where ASW is linked to the ratio of lateral early energy to total energy. Likewise, as found in the previous chapter, it was also proposed that the tonal colouration can be controlled by manipulating the level in the unacceptable tonal colouration regions. Both methods adjust the level by up to ±6 dB in 3 dB steps. The main point of adjusting reflections in directional groups was to avoid affecting other attributes. It was assumed that because the ASW$_{max}$ and unacceptable colouration regions are not coincident, adjusting the level of the reflections in each region would only influence their intended attribute. Thus, Chapter 7 investigated the effectiveness of the level adjustment in each set of regions upon affecting their intended attribute whilst simultaneously not affecting the other.

Since the test in Chapter 6 investigated tonal colouration using only a single reflection, the nature of any perceived colouration both with and without the changes in reflection level was unclear. Therefore, prior to testing the effectiveness, a formal elicitation test was performed. The elicitation test was a modified version of the 'Qualitative Descriptive Analysis'. The test was modified to hasten the elicitation process by shortening the group discussion stages and removing the grading process, whereby the most salient attributes would be chosen based on their occurrence. Subjects listened to two stimuli at a time: a reference without any perceptual control, and one with level adjustment applied to a region by either + or - 6 dB level adjustment. The elicitation test found that:

- The most commonly occurring difference due to level adjustment was in 'Horizontal Spread', suggesting that it affects ASW.
- The next most commonly occurring differences were in 'Loudness', 'Distance' and 'Fullness', suggesting that the level adjustment is having a strong influence on the loudness and source distance perception.
- The most commonly occurring difference that was related to unacceptable colouration was in 'Phasiness'.

Now that the possible differences were understood, as well as the type of difference in colouration, the next step was to investigate the effect of the level adjustment of each region on ASW and 'Phasiness'. A multiple comparison test was performed in the same virtual rooms and source-receiver distances used in the elicitation test, where subjects compared the perceived ASW or phasiness of a stimulus with level adjustment against an unmodified reference stimulus. The test found that:

- Gain adjustment of reflections arriving in the $\mathrm{ASW}_{\mathrm{max}}$, or lateral region, in some cases has a significant effect on ASW.
- The same adjustment did not have a significant effect on phasiness.
- Gain adjustment on reflections arriving in the unacceptable colouration, or front-back region also had some significant effect on ASW depending on the source, room and source-receiver distance.

- The same adjustment also had no significant effect on phasiness.

9.8 Chapter 8

The test performed in Chapter 7 was limited by the fact that it was not possible to directly compare the effectiveness of each method against each other. This could only be done by including level adjustment of both regions in each trial. It was also found in Chapter 7 that the level adjustment method affected the perceived loudness and source distance, thus it was hypothesised that there is possible linkage between ASW, loudness and perceived source distance. The same multiple comparison test was performed, although onlt the reference and extreme ±6 dB adjustment conditions were used. The test found that:

- In certain room conditions, the increase in reflection level in the $\mathrm{ASW}_{\mathrm{max}}$ region had some effect on ASW.

- Manipulation of either $\mathrm{ASW}_{\mathrm{max}}$ or front-back reflection level had no significant effect on perceived phasiness.

- Increasing the level of either $\mathrm{ASW}_{\mathrm{max}}$ or front-back reflections unintentionally affects perceived loudness and source distance.

- There is an apparent three-way linkage between ASW, loudness and distance, where an increase in ASW and reduction in distance could be caused by the increase in loudness.

- Early sound strength, G_E, appears to be the best predictor for ASW in all cases.

- The effectiveness of either control method does not appear to greatly depend on source type.
- The degree of the effects on perceived loudness and source distance appears to be room dependent.

9.9 Limitations and further work

The main objective of this study was to develop a perceptual control method for the purposes of improving and optimising the spatial impression and tonal colouration of a virtual concert hall independently. It investigated a novel, perceptually motivated method of controlling ASW and tonal colouration by manipulating the level of early reflections arriving within specific, horizontal regions. For ASW, this meant controlling the level of a group of lateral reflections, which in turn would have the greatest influence on the perceived ASW. Likewise, for colouration, to improve the perceived acceptability, this meant reducing the level of a group of front-back reflections that would otherwise produce unacceptable and unpleasant colouration. These regions were obtained in experiments performed in Chapters 5 and 6.

However, there are limitations that will need to be addressed in future studies and experiments. The first and largest limitations are that the region locations were obtained in tests that used only a single reflection and a speech source. This of course may have had unseen effects on the results of the subsequent chapters which applied the perceptual control methods to multiple reflections. These limitations may have been the main causes as to why the control methods did not work as expected. To reiterate the findings from Chapters 7 and 8, whilst ASW expectedly increased by boosting the level of lateral reflections, it was also *unexpectedly* increased by boosting the level of front-back reflections. Also, the colouration was not significantly affected by altering the levels of either region. This could mean that boundaries of the regions are not in their most optimal locations because they do not fully consider the effects and behaviour of multiple reflections. Furthermore, whilst the statistical analysis performed in Chapters 7 and 8 suggested that there was no source dependency, the lack of dependency cannot be concluded because the regions were obtained using only a speech source, and that no test was performed to formally investigate the effects of source type upon the effectiveness of the control methods. Future tests that will investigate the potential source type dependency of the regions using a wider variety of sources, including different speech types, musical sources and both broadband and band-limited noise.

What was also not considered in Chapters 7 and 8 was the preservation of total early energy during the level adjustment. What may have caused the increase in loudness and decrease in perceived source distance was the increase in total early energy, as predicted by G_E. It is possible that ASW could be enhanced without an increase in loudness by increasing the level in the ASW_{max} region whilst simultaneously decreasing the level outside of the region as a method of preserving the original amount of early energy. This method of *energy preservation* can be explored using future experiments similar to those performed in this study. Whilst this improvement could be applied to colouration control, because the method had no significant effect on the attribute, it is worth pursuing more fundamental research such as addressing the single reflection limitation discussed in the previous paragraph.

All tests, except for the verification test in Chapter 5, were limited to playback over headphones. This was imposed due to fundamental nature of the experiments which required binaural presentation, and (simulated) anechoic conditions. Whilst it may have been possible to conduct the experiments using discrete speakers for simulating a single reflection, there was a limited number of available loudspeakers, and thus by extension as limited number of virtual sources. Furthermore, there was a limited volume of space to accommodate them in the ITU-R.1116 critical listening room at the University of Huddersfield. As discussed in Chapter 5 Section 5.2, the available loudspeaker model (Genelec 8040A) would have reduced the resolution of inter-reflection angle. Furthermore, the critical listening space itself was not anechoic, which was vital for the fundamental experiments as it the sound fields could be controlled.

The apparent linkage between ASW, loudness and source distance discussed in Chapter 8, which of course will need to be explored further. It can be speculated that the increase in loudness, for example, was due to an increase in total energy. However, the observed correlation between ASW and loudness in this study, as well as in previous literature, suggests that the two are strongly linked. This was not considered in this study, thus the findings from the experiments in Chapters 7 and 8 could affect the implementation of the perceptual control methods proposed in this study.

Lastly, the custom virtual acoustics program developed for this study was not tested for 3D audio speaker setups such as 9.1, Auro 3D and Dolby Atmos, and has been used for mainly rendering BRIRs. Future studies will look into: how the program can be used for developing a 3D audio virtual acoustics program; developing the program further to accurately model wave phenomenon; using the research found during the development to create a real-time reverberator for use in a VR experience, and ultimately develop an intelligent perceptual optimisation system.

Appendix A

Investigation into the perceptual effects of image source method order[1]

A.1 Abstract

This engineering brief explores the perceived effects and characteristics of impulse responses (IRs) generated using a custom, hybrid, geometric reverb algorithm. The algorithm makes use of a well known Image Source Method (ISM) and Ray Tracing methods. ISM is used to render the early reflections to a specified order whilst ray tracing renders the remaining reflections. IRs rendered at varying ISM orders appear to exhibit differences in perceptual characteristics, particularly in the early portion. To understand these characteristics, an elicitation test based was devised in order to acquire terms for the different characteristics. These terms were grouped in order to provide attributes for future grading tests.

[1] Johnson, D., Lee, H. (2016a). Investigation into the perceptual effects of image source method order. In *Audio engineering society convention 140.*

A.2 Introduction

Virtual acoustics software is often used for architectural design and more recently, virtual or augmented reality. Software packages often use geometric methods for modelling the propagation of sound waves, where commonly used are the Image Source Method (ISM) (Allen & Berkley, 1979) and Ray Tracing (Krokstad et al., 1968). ISM can only model specular reflections, whilst ray tracing can model both specular and diffuse reflections. It could been seen at first glance, that ray tracing would be a preferable method however, the major difference between the two algorithms is their detection method. ISM is deterministic and uses a point receiver, so all valid reflections will be detected, however ray tracing is stochastic and requires the use of a receiver sphere, which if not optimal, may lead to systematic errors such as missed rays or detection of invalid reflections (Laine et al., 2009).

However, ISM can become computationally expensive at increasing orders, and so it is desirable to combine it with another algorithm to improve efficiency, for example Wendt et al. (2014) combine ISM with a filter delay network to model a shoebox room, and Aspöck, Pelzer, Wefers, and Vorländer (2014) combine ISM with ray tracing to model arbitrarily shaped models. The ISM order is usually dependent on the desired range of early reflections. Concert hall acoustic research generally suggests that the early lateral reflections, or first 50-80ms of an impulse response (IR) are an important factor in the perception of spatial impression (Bradley & Soulodre, 1995b), whereas late reverberation contributes to the perception of listener envelopment (LEV). However, there are some research questions that have not yet been formally answered:

1. As ISM order increases, what is the perceived effect of the increased amount of specular energy?
2. At what point should ISM stop rendering?

As the first step towards addressing the questions, an elicitation test was conducted in such a way that subjects freely described the perceived qualities of various IRs generated with varying ISM orders. These terms were then grouped into a common set of attributes through a group discussion.

A.3 Virtual acoustics software

For the purposes of concert hall acoustics and psychoacoustics research, a 3D virtual acoustics rendering tool was developed in C++. This tool is able to render the acoustics of a virtual three-dimensional room model and produce a multichannel impulse response.

A.3.1 Rendering algorithm

The tool implements a hybrid system consisting of two sub-algorithms: the Image Source Method (ISM), and Ray Tracing[1]. The ISM is based on the work developed by Allen and Berkley (1979), Vorländer (1989), Lehmann and Johansson (2008), Wendt et al. (2014) and Aspöck et al. (2014), and the ray tracing sub-algorithm is based on work developed by Krokstad et al. (1968), Kulowski (1985) and Elorza (2005). To briefly explain the algorithms, firstly the ISM computes the impulse response by the

[1] The program developed for the main study is a much improved version of program discussed in this engineering brief. The main improvements are in how surface and air absorption are modelled, and the method of diffusion

means of recursive image expansion to a specified order. The impulse response is formed by computing the distance from each virtual image source to the receiver. ISM is limited to rendering specular reflections only. Ray tracing emits rays in all directions from the source. These are reflected indefinitely until their energy is negligible. The impulse response is formed by detecting if any rays cross a receiver sphere. It is possible to render diffuse reflections by randomising the ray direction upon reflection. This is based upon the vector scattering method described by Rindel (2000).

The rendering process involves calculating the early to intermediate reflections with ISM. All the possible images to a specified order are calculated, then using a process of back-tracing and visibility testing, each reflection path is validated. Once this has been completed, the remaining reflections are calculated using ray tracing. During ray tracing, rays that would be calculated using ISM are exempt from detection until the reflection order overtakes the ISM order. As each ray is 'detected' by a receiver, it is cached into a two-dimensional array. Each row of the array belongs to a receiver in the model. This enables any post-processing and analyses to be performed upon a raw impulse response without the need for re-rendering the entire impulse response from scratch.

A.3.2 Impulse response forming

A discrete multichannel IR is formed by extracting the delay times of each ray in reference to the speed of sound. The energy is response is formed by multiplying the amplitude by the corresponding receiver's polar pattern, and forming octave-band channels for each energy band. These are then filtered using linear phase, Butterworth band-pass filters and summed to form the final impulse response. This is done for each receiver.

A.3.3 ISM order and receiver size

A set of factors to consider is the approach ISM and ray tracing use in order to render room acoustics. ISM is deterministic, in which it is guaranteed to find all valid reflections. Ray tracing is stochastic, whereby it relies on probability that rays will cross a receiver sphere. Systematic errors may occur if the ray count and receiver radius are not optimal. The receiver size can be determined using equation A.1, described by Jiang and Qiu (2003):

$$r = \sqrt[3]{\frac{15V}{2\pi N}} \tag{A.1}$$

where r is the receiver radius, V is the room volume and N is the ray count.

This implies that it is likely that the density of reflections may increase as ISM order increases, up to the point where ray tracing will continue the rendering process. Therefore, this may have an effect on the perceptual characteristics of the IR.

A.4 Experimental design

In order to understand what the perceived effect of ISM order is, a free elicitation test based upon the QDA method (Stone & Sidel, 2004) was devised. The experiment was split into two phases: an individual testing phase, followed by a group discussion phase. The listening tests were conducted in the ITU-R BS.1116-compliant listening room at the University of Huddersfield.

A.4.1 Room model

The room model constructed for the experiment is a typical concert hall sized room with dimensions of 16.18m x 10.0m x 26.18m (W x H x D). All surfaces in the model have an absorption coefficient of 0.01 across the entire frequency spectrum. The reason for this is to minimise the effect absorption may have upon experiments. A single sound source is positioned at (8.09, 2.0, 23.0). Two omni-directional receivers configured as a 0.5m spaced AB pair are positioned at 3m, 6m and 12m from the receiver, and are also raised to a height of 2m.

A.4.2 Stimuli

For the experiment, three anechoic recordings from the Bang and Olufsen Archimedes project CD (Hansen & Munch, 1991) were used: trumpet, spoken dialogue, and conga excerpts. These recordings were chosen in order to examine the effects that different characteristics such as continuous notes, transients, and complex harmonic content may exhibit when processed with the artificial reverb. A total of thirty six IRs were generated. These can be categorised into three groups depending on the source receiver distance: 3m, 6m and 12m. Each IR in each group was generated using a different ISM order from 1 to 6, however with a constant number of 75 000 rays, with a calculated receiver radius of 0.6m. The IRs are interleaved into eighteen, stereo, 44.1kHz WAV files. Each sound source is then convolved with each IR to produce fifty four stimuli in total. Figure A.1 shows an example of the IRs produced at ISM order 1 and ISM order 6 at 3m distance. The density from 100ms is higher in the IR rendered at ISM order 6 than ISM order 1. This difference may be caused by systematic errors in an IR rendered at order 1, essentially solely with ray tracing.

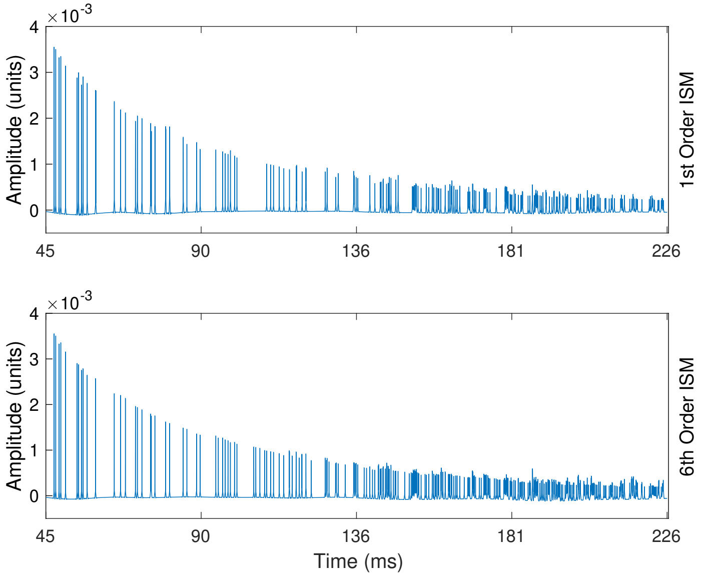


Fig. A.1: RIRs rendered using ISM order 1 and 6 between 45-226 ms

A.4.3 Subjects

The experiment requires experienced critical listeners. Five members of the Applied Psychoacoustics Laboratory of the University of Huddersfield participated in the elicitation test and group discussion. All subjects had experience in spatial audio evaluation.

A.4.4 Elicitation test

Phase one was a listening test consisting of nine trials, and in each trial there are six audio samples. The test interface was designed in Cycling 74's Max 7. The order of the samples was randomised per trial. In each trial, subjects are asked to play the six samples, and then describe and grade the apparent effects they were able to perceive by typing into a text box. The samples are looped and play in synchronisation so it is possible to switch between samples freely. After each test the data is analysed and any elicited terms are extracted and added to a spreadsheet.

A.4.5 Group discussion

The objective of the group discussion is to debate the meanings behind each elicited terms and under what spatial or timbral attribute to group them as. After a short period of time after the end of phase one, the subjects were asked to attend a group discussion in the same listening room. The terms gathered from phase one were displayed to the subjects. The objective was explained to the subjects, and over a period of two hours the terms from phase one were interpreted, discussed, and then grouped into categories that were deemed suitable the by entire group.

A.5 Results and discussions

Attribute	Description	Elicited Terms
Environment image spread (51)	Environment related component. Horizontal / vertical spread.	Boxy (11), Diffuseness (1), Echoey (8), Frontal (1), Larger (5), 'Opening up' (2), Reverberant (4), Room size (5), Short reverb (1), Smaller (3), Spaciousness (8), Spatial impression (2)
Roughness (38)	How rough / smooth after initial transient	Comb filtered (1), Natural (14), Phasey (4), Rattley (2), Roughness (11), Smooth (2), Unnatural (4)
Distance (24)	The impression of source-receiver distance	Closer (9), Distance (9), Nearer (2), Presence (1), Proximity (1), Wetness (2)
Clarity (13)	Perception of intelligibility	Clarity (3), Clear (2), Intelligibility (1), Mid resonance (1), Muddy (6)
Thin / Full (6)	Amount of low frequency energy	Fullness (1), Low frequency energy (2), Thin (2), Weight (1)
Dull / Bright (15)	Amount of high frequency energy	Bright (5), Dark (1), Dull (6), High frequency energy (2), High end (1)
Vertical Shift (4)	Perceived vertical shift in image	Localisation shift (1), Vertical image shift (3)
Loudness (15)	The apparent loudness of the source	Amplitude (2), Direct sound increase (3), Louder (3), Loudness (2), Lower mix of reverb (1), Presence (1), Stronger direct sound (1), Wetness (2)
Resonant (4)	Emphasis on certain frequencies	Nasal (2), Resonant (2)
Echo Perception (5)	The perception of additional echoes	Echo (2), Echoey (3)
Source Focus (3)	Apparent separation of source from reverb	Source focus (3)

Table A.1: Attributes, occurrences and breakdown of elicited terms with their individual occurrences.

The terms elicited from phase one were grouped into the categories extracted from phase two, as shown in Table A.1. The table shows how listeners are able to describe each attribute using many different terms. Although this was not a preference test, it was also interesting to note how terms such as 'rough' and 'muddy' were described as being a negative terms. On the other hand, terms such as 'natural', 'smooth' and 'bright' were described as positive terms. This gives rise that the attributes are bipolar. The attributes with the largest occurrence are 'Environmental Image Spread', 'Roughness', 'Distance' and 'Loudness'.

The high occurrence of the 'Environmental Image Spread' attribute implies that the ISM order has an effect over the perceived environmental properties of the room. It could be for example that a low ISM order leads to a 'boxy' and 'small' characteristic whilst a higher order leads to an apparent 'opening up', increased 'Diffuseness' or an increased, 'Echoey' quality. The increasing density of specular energy at higher orders could potentially be playing an important role. Another highly occurring attribute is 'Roughness', and the most elicited terms in this attribute are 'rough' and 'natural'. Again, this suggests that the ISM order has an effect upon the apparent timbral quality and plausibility of the IR. It may be the case that sub-optimal ISM orders will result in a 'rough', unnatural sounding reverb, whilst higher orders will lead to a more 'natural' or 'smooth' reverb.

The 'Distance' attribute is likely related to differing distances the IRs have been rendered at, however it may be working in conjunction with the ISM order in which higher orders could be adding a sense of depth to the perceived reverb. This attribute may also be working in conjunction with the 'Loudness' attribute in which the source-listener distance is merely affecting the apparent loudness of direct sound against the reverb. 'Clarity', 'Thin / Full' and 'Dull / Bright' appear are timbral attributes. The ISM order here may be contributing to the overall intelligibility in which an increase in order may lead to an increase in intelligibility, however absorption may also be a contributing factor. The next stage in the investigation would be to firstly look into what extent the attributes obtained from the group discussion are perceived. This will be performed through the use of grading tests in which listeners will grade the strength in which they perceive the attribute. Secondly, the effects of the ISM order should be investigated using objective measures in order to compare analytical results with the results from the grading tests. These measures will include ICCC, C_{80}, D_{50} and G.

It is worth noting that the terms elicited in this engineering brief can be seen as being applicable only to the room model described earlier. Therefore a future stage would investigate further what effect the dimensions and geometry of a room has on the attributes. If an objective measure can be formed, it would be useful in choosing an optimum ISM order for a perceptually plausible IR. As the experiment was performed using only a stereophonic setup, other attributes such as ASW and LEV cannot be truly investigated, therefore future work will include performing ISM order experiments in multichannel 3D and binaural setups. This future work may lead to an attribute grading that may differ from stereophonic experiments.

A.6 Summary

The aim of this engineering brief was to gain insight into the perceptual effects of artificial IRs rendered using differing ISM orders. To obtain perceptual attributes associated with the perceived effects, a free-elicitation method was used to obtain terms. The terms were then categorised into attributes that would describe the type of effect that was being perceived. Analysis of the results show that the most elicited terms belonged to the attributes ‘Environmental Image Spread’, ‘Roughness’, ‘Distance’ and ‘Loudness’. These results imply that ISM order may have connection to the degree of perception of these attributes. This can be investigated further through the use of grading tests and objective measures.

Appendix B

Instructions for subjects per experiment

Section	Experiment	Instructions
5.1	Finding the ASW saturation region	Subjects were instructed to listen to each stimulus, A and B, and decide whether or not they could here a difference in ASW. Per trial, they were asked: *Did you hear a difference in apparent source width?*
5.2	Verification of the regions	Subjects were instructed listen to each stimulus including the reference, and compare and grade the relative apparent source width each of the test stimuli against the reference. Per trial, the interface instructed subjects to: *Compare the apparent source width of the stimuli.*
6.1	Elicitation of attributes associated with tonal colouration	In each session, the subject was presented with each stimulus for each delay for that particular source type, and were instructed to elicit terms regarding any audible tonal colouration. The subjects were able to listen to each stimulus as many times as they wished before moving on to the next.
6.2	Effect of reflection angle on tonal colouration	Subjects were asked to listen carefully to the tonal colouration of the stimulus and adjust a hidden parameter until it becomes acceptable or inaudible, depending on the session, then proceed onto the next trial. The interface instructed subjects to: *Adjust the parameter and find the minimum point where the 'tonal colouration' becomes inaudible / acceptable.*

6.3	Effect of reflection level on tonal colouration	Subjects were instructed to listen carefully to the stimulus, and determine if the perceived colouration is acceptable. The test interface asked subjects: *Is the tonal colouration acceptable.*
7.2	Formal elicitation of attributes regarding differences between perceptual control methods	Subjects were presented with two stimuli, each with one of the following conditions: selected reflection levels are boosted by 6dB, which is the maximum amount used throughout this and subsequent tests; and unprocessed. They were then asked to carefully listen to each stimulus and elicit any perceptual differences in A compared to B. Terms were elicited into a text box on the listening test interface.
7.3	Application of the perceptual control method	Subjects were asked to compare each stimulus against the reference and grade any audible changes for a given attribute on a continuous scale with semantic labels. The test interface instructed subjects to: *Compare each stimulus to the REF and grade them in terms of "Horizontal Spread" / "Phasiness".*
8.1	Application of the perceptual control methods: Part 2	The instructions were identical to the previous experiment, except subjects now also compared the stimuli in terms of 'Loudness' and 'Perceived distance' in addition to 'Horizontal Spread' and 'Phasiness'.

Table B.1: Breakdown of instructions provided to subjects during each listneing experiment

References

Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, *65*(4), 943-950. doi: doi: 10.1121/1.382599

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

Aspöck, L., Pelzer, S., Wefers, F., & Vorländer, M. (2014). A real-time auralization plugins for architectural design and education. In *Proceedings of the EAA joint symposium on auralization and ambisonics.*

Barron, M. (1971). The subjective effects of first reflections in concert halls - the need for lateral reflections. *Journal of Sound and Vibration*, *15*(4), 475–494.

Barron, M., & Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *Journal of Sound and Vibration*, *77*(2), 211–232.

Bech, S. (1995). Timbral aspects of reproduced sound in small rooms. I. *The Journal of the Acoustical Society of America*, *97*(3), 1717–1726.

Bech, S. (1996). Timbral aspects of reproduced sound in small rooms. II. *The Journal of the Acoustical Society of America*, *99*(6), 3539–3550.

Bech, S., & Zacharov, N. (2007). *Perceptual audio evaluation-theory, method and application*. Chichester, United Kingdom: John Wiley & Sons.

Beranek, L. (1992). Concert hall acoustics—1992. *The Journal of the Acoustical Society of America*, *92*(1), 1–39.

Beranek, L. (1996). Acoustics and musical qualities. *The Journal of the Acoustical Society of America*, *99*(5), 2647–2652.

Beranek, L. (2011). The sound strength parameter G and its importance in evaluating and planning the acoustics of halls for music. *The Journal of the Acoustical Society of America*, *129*(5), 3020–3026.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. Cambridge, United States: MIT press.

Blauert, J., & Lindemann, W. (1986). Auditory spaciousness: Some further psychoacoustic analyses. *The Journal of the Acoustical Society of America*, *80*(2), 533–542.

Borish, J. (1984). Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, *75*(6), 1827–1836.

Bork, I. (2000). A comparison of room simulation software-the 2nd round robin on room acoustical computer simulation. *Acta Acustica united with Acustica*, *86*(6), 943–956.

Bradley, J. S., Reich, R., & Norcross, S. (2000). On the combined effects of early-and late-arriving sound on spatial impression in concert halls. *The Journal of the Acoustical Society of America*, *108*(2), 651–661.

Bradley, J. S., & Soulodre, G. A. (1995a). The influence of late arriving energy on spatial impression. *The Journal of the Acoustical Society of America*, *97*(4), 2264–2271.

Bradley, J. S., & Soulodre, G. A. (1995b). Objective measures of listener envelopment. *The Journal of the Acoustical Society of America*, *98*(5), 2590–2597.

Brunner, S., Maempel, H.-J., & Weinzierl, S. (2007). On the audibility of comb filter distortions. In *Audio engineering society convention 122.*

Cardozo, B. (1965). Adjusting the method of adjustment: Sd vs dl. *The Journal of the Acoustical Society of America*, *37*(5), 786–792.

Carpentier, T., Noisternig, M., & Warufsel, O. (2014). Hybrid reverberation processor with perceptual control. In *17th international conference on digital audio effects.*

Coleman, P., Franck, A., Jackson, P. J. B., Hughes, R. J., Remaggi, L., & Melchoir, F. (2017). Object-based reverberation for spatial audio. *Journal of the Audio Engineering Society*, *65*(1/2), 66–77.

Coolican, H. (2009). *Research methods and statistics in psychology*. London, United Kingdom: Hodder.

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology*, *75*(3), 485–491.

Dadoun, N., Kirkpatrick, D. G., & Walsh, J. P. (1985). The geometry of beam tracing. In *Proceedings of the first annual symposium on computational geometry* (pp. 55–61).

De Sena, E., Hacıhabiboğlu, H., Cvetković, Z., & Smith, J. O. (2015). Efficient synthesis of room acoustics via scattering delay networks. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, *23*(9), 1478–1492.

de Vries, D., Hulsebos, E., & Baan, J. (2001). Spatial fluctuations in measures for spaciousness. *The Journal of the Acoustical Society of America*, *110*(2), 947–954.

Ehrenstein, W. H., & Ehrenstein, A. (1999). Psychophysical methods. In *Modern techniques in neuroscience research* (pp. 1211–1241). Springer.

Elorza, D. O. (2005). *Room acoustics modeling using the raytracing method: implementation and evaluation* (Unpublished doctoral dissertation). University of Turku.

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio engineering society convention 108.*

Fog, A. (2017). VCL, C++ vector class library [Computer software manual]. Retrieved from http://www.agner.org/optimize/vectorclass.pdf

Francombe, J. (2014). *Perceptual evaluation of audio-on-audio interference in a personal* *sound zone system* (PhD Thesis). University of Surrey (United Kingdom).

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General*, *141*(1), 2.

Funkhouser, T., Carlbom, I., Elko, G., Pingali, G., Sondhi, M., & West, J. (1998). A beam tracing approach to acoustic modeling for interactive virtual environments. In *Proceedings of the 25th annual conference on computer graphics and interactive techniques* (pp. 21–32).

Furuya, H., Fujimoto, K., Ji, C. Y., & Higa, N. (2001). Arrival direction of late sound and listener envelopment. *Applied Acoustics*, *62*(2), 125–136.

Furuya, H., Fujimoto, K., Takeshima, Y., & Nakamura, H. (1995). Effect of early reflections from upside on auditory envelopment. *Journal of the Acoustical Society of Japan (E)*, *16*(2), 97–104.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision research*, *38*(12), 1861–1881.

Gardner, W. G. (1992). *The virtual acoustic room* (PhD Thesis). Massachusetts Institute of Technology.

Gardner, W. G., & Martin, K. D. (1995). HRTF measurements of a KEMAR. *The Journal of the Acoustical Society of America*, *97*(6), 3907–3908.

George, S., Zielinski, S., Rumsey, F., Jackson, P., Conetta, R., Dewhirst, M., . . . Bech, S. (2010). Development and validation of an unintrusive model for predicting the sensation of envelopment arising from surround sound recordings. *Journal of the Audio Engineering Society*, *58*(12), 1013–1031.

González, Á. (2010). Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, *42*(1), 49–64.

Google. (2017). *Spatial audio resources.* Retrieved 2018-07-2, from https://github

.com/google/spatial-media/tree/master/spatial-audio

Grantham, D. W., & Wightman, F. L. (1978). Detectability of varying interaural temporal differences. *Journal of the Acoustical Society of America*, *63*(2), 511–523.

Gribben, C., & Lee, H. (2015). Toward the development of a universal listening test interface generator in max. In *Audio engineering society convention 138.*

Halmrast, T. (2000). Orchestral timbre: Comb-filter coloration from reflections. *Journal of Sound and Vibration*, *232*(1), 53–69.

Halmrast, T. (2001). Sound coloration from (very) early reflections. *Journal of the Acoustical Society of America*, *109*(5), 2303.

Hansen, V., & Munch, G. (1991). Making recordings for simulation tests in the archimedes project. *Journal of the Audio Engineering Society*, *39*(10), 768–774.

Harker, A., & Tremblay, P. A. (2012). The hisstools impulse response toolbox: Convolution for the masses. In *Proceedings of the international computer music conference* (pp. 148–155).

Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, *74*(5), 1380–1391.

Heckbert, P. S., & Hanrahan, P. (1984). Beam tracing polygonal objects. *ACM SIGGRAPH Computer Graphics*, *18*(3), 119–127.

Hidaka, T., Beranek, L. L., & Okano, T. (1995). Interaural cross-correlation, lateral fraction, and low-and high-frequency sound levels as measures of acoustical quality in concert halls. *The Journal of the Acoustical Society of America*, *98*(2), 988–1007.

Institute of Sound Recording. (2017). *IoSR matlab toolbox.* Retrieved 2018-09-1, from https://github.com/IoSR-Surrey/MatlabToolbox

International Standards Organisation. (2009). 3382-1: 2009. *Measurement of room acoustic parameters–Part 1.*

International Telecommunication Union. (2015). BS.1116. *Methods for the subjective assessment of small impairments in audio systems.*

Jiang, Z., & Qiu, X. (2003). Receiving radius determination in ray-tracing sound prediction of rectangular enclosure. *Journal of Sound and Vibration*, *301*(1), 391–399.

Johnson, D., & Lee, H. (2016a). Investigation into the perceptual effects of image source method order. In *Audio engineering society convention 140.*

Johnson, D., & Lee, H. (2016b). Taking advantage of geometric acoustics modeling using metadata. In *Interactive audio systems symposium 2016.*

Johnson, D., & Lee, H. (2017). Just noticeable difference in apparent source width depending on the direction of a single reflection. In *Audio engineering society convention 142.*

Johnson, D., & Lee, H. (2018). Perceptually optimised virtual acoustics. In *Proceedings of the 4th workshop on intelligent music production.*

Jot, J.-M. (1997). *Efficient models for reverberation and distance rendering in computer music and virtual audio reality* [Technical Report].

Jot, J.-M. (1999). Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia systems*, *7*(1), 55–69.

Jot, J.-M., & Chaigne, A. (1991). Digital delay networks for designing artificial reverberators. In *Audio engineering society convention 90.*

Jot, J.-M., & Trivi, J.-M. (2006). Scene description model and rendering engine for interactive virtual acoustics. In *Audio engineering society convention 120.*

Kaplanis, N., Bech, S., Jensen, S. H., & van Waterschoot, T. (2014). Perception of reverberation in small rooms: a literature study. In *Audio engineering society conference: 55th international conference: Spatial audio.*

Keet, W. d. V. (1968). The influence of early lateral reflections on the spatial impression. In *Proceedings of the 6th international congress on acoustics* (pp. E53–E56).

Klockgether, S., & van de Par, S. (2016). Just noticeable differences of spatial cues in echoic and anechoic acoustical environments. *The Journal of the Acoustical Society of America*, *140*(4), EL352–EL357.

Krokstad, A., Strøm, S., & Sørsdal, S. (1968). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, *8*(1), 118–125.

Kulowski, A. (1985). Algorithmic representation of the ray tracing technique. *Applied Acoustics*, *18*(6), 449–469.

Kuttruff, H. (2016). *Room acoustics* (6th ed.). CRC Press.

Laine, S., Siltanen, S., Lokki, T., & Savioja, L. (2009). Accelerated beam tracing algorithm. *Applied Acoustics*, *70*(1), 172–181.

Lee, H. (2006). *Effects of interchannel crosstalk in multichannel microphone technique* (PhD Thesis). University of Surrey (United Kingdom).

Lee, H. (2013). Apparent source width and listener envelopment in relation to source-listener distance. In *Audio engineering society conference: 52nd international conference: Sound field control-engineering and perception.*

Lee, H., Johnson, D., & Mironovs, M. (2016). A new response method for auditory localization and spread test. In *Audio engineering society convention 140.*

Lehmann, E. A., & Johansson, A. M. (2008). Prediction of energy decay in room impulse responses simulated with the image-source model. *The Journal of the*

Acoustical society of America, *124*(1), 269–277.

Lehnert, H. (1993). Systematic errors of the ray-tracing algorithm. *Applied Acoustics*, *38*(2/4), 207–221.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467–477.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, *106*(4), 1633–1654.

Mason, R. (2002). *Elicitation and measurement of auditory spatial attributes in reproduced sound* (PhD Thesis). University of Surrey.

Mason, R., Brookes, T., & Rumsey, F. (2003). Creation and verification of a controlled experimental stimulus for investigating selected perceived spatial attributes. In *Audio engineering society convention 114.*

Mason, R., Brookes, T., & Rumsey, F. (2004). Development of the interaural cross-correlation coefficient into a more complete auditory width prediction model. In *International congress on acoustics 18* (pp. 2453–2456).

Mason, R., & Rumsey, F. (2001). Interaural time difference fluctuations: their measurment, subjective perceptual effect, and application in sound reproduction. In *AES 19th international conference.*

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, *32*(1), 12–16.

Mintzer, D. (1950). Transient sounds in rooms. *The Journal of the Acoustical Society of America*, *22*(3), 341–352.

Moorer, J. A. (1979). About this reverberation business. *Computer music journal*, 13–28.

Morimoto, M., & Iida, K. (1995). A practical evaluation method of auditory source width in concert halls. *Journal of the Acoustical Society of Japan (E)*, *16*(2), 59–69.

Morimoto, M., & Iida, K. (1998). Effects of front/back energy ratios of early and late reflections on listener envelopment. *The Journal of the Acoustical Society of America*, *103*(5), 2748–2748.

Morimoto, M., Iida, K., & Sakagami, K. (2001). The role of reflections from behind the listener in spatial impression. *Applied Acoustics*, *62*(2), 109–124.

Morimoto, M., & Maekawa, Z. (1988). Effects of low frequency components on auditory spaciousness. *ACTA Acustica united with Acustica*, *66*(4), 190–196.

Morimoto, M., & Pösselt, C. (1989). Contribution of reverberation to auditory spaciousness in concert halls. *Journal of the Acoustical Society of Japan (E)*, *10*(2), 87–92.

Mueller, W., & Ullmann, F. (1999). A scalable system for 3D audio ray tracing. In *Multimedia computing and systems, 1999. IEEE international conference on* (Vol. 2, pp. 819–823).

Murphy, D. T. (2000). *Digital waveguide mesh topologies in room acoustics modelling* (PhD Thesis). The University of York.

Naylor, G. M. (1993). ODEON—another hybrid room acoustical model. *Applied Acoustics*, *38*(2-4), 131–143.

Okano, T. (2002). Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections. *The Journal of the Acoustical Society of America*, *111*(1), 217–229.

Okano, T., Beranek, L., & Hidaka, T. (1998). Relations among interaural cross-correlation coefficient (IACC_E), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. *The Journal of the Acoustical Society of America*, *104*(1), 255–265.

Okano, T., Hidaka, T., & Beranek, L. L. (1994). Relations between the apparent source width (asw) of the sound field in a concert hall and its sound pressure level at low frequencies (gl), and its inter-aural cross correlation coefficient (iacc). *The Journal of the Acoustical Society of America*, *96*(5), 3268–3268.

Pellegrini, R. (2002). Perception-based design of virtual rooms for sound reproduction. In *Audio engineering society conference: 22nd international conference: Virtual, synthetic, and entertainment audio.*

Pelzer, S., Schröder, D., & Vorländer, M. (2011). The number of necessary rays in geometrically based simulations using the diffuse rain technique. In *Fortschritte der akustik–daga, düsseldorf.*

Pollack, I., & Trittipoe, W. (1959). Binaural listening and interaural noise cross correlation. *The Journal of the Acoustical Society of America*, *31*(9), 1250–1252.

Rafii, Z., & Pardo, B. (2009). Learning to control a reverberator using subjective perceptual descriptors. In *International society for music information retrieval conference 10* (pp. 285–290).

Reichardt, W., & Schmidt, W. (1967). Die wahrnehmbarkeit der veränderung von schallfeldparametern bei der darbietung von musik. *Acustica*, *18*(5), 274–282.

Rindel, J. H. (2000). The use of computer modeling in room acoustics. *Journal of vibroengineering*, *3*(4), 41–72.

Robotham, T. (2016). *The effects of a vertical reflection on the relationship between listener preference and timbral and spatial attributes* (Masters Thesis). University of Huddersfield.

Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: Terminology, meaning and a scene-based paradigm. *Journal Audio Engineering Society*, *50*(9), 651–666.

Salomons, A. M. (1995). *Coloration and binaural decoloration of sound due to reflections* (PhD Thesis). Delft University.

Savioja, L. (2000). *Modeling techniques for virtual acoustics* (PhD Thesis). Helsinki University of Technology.

Savioja, L. (2010). Real-time 3d finite-difference time-domain simulation of low-and mid-frequency room acoustics. In *Internationl conference on digital audio effects 13* (Vol. 1, p. 75).

Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, R. (1999). Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, *47*(9), 675–705.

Savioja, L., & Svensson, U. P. (2015). Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*, *138*(2), 708–730.

Schröder, D. (2011). *Physically based real-time auralization of interactive virtual environments* (PhD Thesis). RWTH Aachen University.

Schroeder, M. R. (1961). Natural sounding artificial reverberation. In *Audio engineering society convention 13.*

Schroeder, M. R. (1970). Digital simulation of sound transmission in reverberant spaces. *The Journal of the Acoustical Society of America*, *47*(2A), 424–431.

Schroeder, M. R., & Logan, B. F. (1960). "Colorless" artificial reverberation. *The Journal of the Acoustical Society of America*, *32*(11), 1520–1520.

Seki, Y., & Ito, K. (2003). Coloration perception depending on sound direction. *IEEE transactions on speech and audio processing*, *11*(6), 817–825.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Smith, J. O. (1987). *Musical applications of digital waveguides.* CCRMA, Stanford University.

Smith, J. O. (1992). Physical modeling using digital waveguides. *Computer music journal*, *16*(4), 74–91.

Stevens, S. S. (1958). Problems and methods of psychophysics. *Psychological Bulletin*, *55*(4), 177–196.

Stevens, S. S., & Guirao, M. (1962). Loudness, reciprocality, and partition scales. *The Journal of the Acoustical Society of America*, *34*(9B), 1466–1471.

Stone, H., & Sidel, J. L. (2004). *Sensory evaluation practices* (3rd ed.). Cambridge, United States: Academic Press.

Tervo, S., Pätynen, J., Kuusinen, A., & Lokki, T. (2013). Spatial decomposition method for room impulse responses. *Journal of the Audio Engineering Society*, *61*(1/2), 17–28.

Thomas, M. R. (2017). *Wayverb: A graphical tool for hybrid room acoustics simulation* (Masters Thesis, University of Huddersfield). Retrieved from https://reuk.github.io/wayverb/

Välimäki, V., Parker, J. D., Savioja, L., Smith, J. O., & Abel, J. S. (2012). Fifty years of artificial reverberation. *IEEE Audio, Speech and Language Processing*, *20*(5), 1421–1448.

Van Duyne, S. A., & Smith, J. O. (1993). Physical modeling with the 2-d digital waveguide mesh. In *Proceedings of the international computer music conference* (pp. 40–47).

Vorländer, M. (1989). Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *The Journal of the Acoustical Society of America*, *86*(1), 172–178.

Vorländer, M. (2007). *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media.

Wakuda, A., Furuya, H., Fujimoto, K., Isogai, K., & Anai, K. (2003). Effects of arrival direction of late sound on listener envelopment. *Acoustical Science and Technology*, *24*(4), 179–185.

Wallis, R. (2017). *The analysis of frequency dependent vertical localisation thresholds and the perceptual effects of vertical interchannel crosstalk* (PhD Thesis, University of Huddersfield). Retrieved from http://eprints.hud.ac.uk/id/eprint/34350/

Wallis, R., & Lee, H. (2017). The reduction of interchannel crosstalk: the analysis of localization thresholds for natural sound sources. *Applied Sciences*, *7*(3), 278–298.

Wendt, T., van de Par, S., & Ewert, S. D. (2014). Perceptual and room acoustical evaluation of a computational efficient binaural room impulse response simulation method. *10.14279/depositonce-4103*.

Yang, L., & Shield, B. (2000). Development of a ray tracing computer model for the prediction of the sound field in long enclosures. *Journal of Sound and Vibration*, *229*(1), 133–146.

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *ACTA Acustica united with Acustica*, *91*(3), 409–420.

Zeng, X., Chen, K., & Sun, J. (2003). On the accuracy of the ray-tracing algorithms based on various sound receiver models. *Applied Acoustics*, *64*(4), 433–441.

Zeng, X., Christensen, C. L., & Rindel, J. H. (2006). Practical methods to define scattering coefficients in a room acoustics computer model. *Applied Acoustics*, *67*(8), 771–786.

Zielinski, S., Rumsey, F., & Bech, S. (2008). On some biases encountered in modern audio quality listening tests- a review. *Journal of the Audio Engineering Society*, *56*(6), 427–452.