INTRODUCTION

1.1 Project Overview

The project described in this paper aims to extend the functionality of a currently implemented system called the Virtual Singing Studio situated in the listening space of the Audio Lab at the University of York. The system allows a user to hear themselves as though they are present in another acoustic environment. This project looks into extending this functionality to allow the user to move themselves around a digitally modelled version of Hendrix Hall, one of the lecture halls on the Universities campus by generation a large grid of synthetic room impulse responses using the room acoustic simulation software Odeon.

The aims of the project are as follows:

- Extend the Virtual Singing Studio to allow a user to select any position the want within the given virtual space

- Allow the user to feel as though they are freely moving around the space

- Investigate the plausibility of the produced system

- Investigate the perception of mobility given different Room Impulse Response densities provided.

1.2 Project Motivation

1.3 Background

The following section covers material that forms the basis of the system produced as a result of this project.

1.3.1 Virtual Acoustic Environments

Virtual acoustics has been previously described [**?**] as follows:

> *"Virtual acoustics is a general term for*
> *the modelling of acoustical phenomena*
> *and systems with the aid of a computer"*

By this definition, a **VAE!** (**VAE!**) can be thought of as an environment (such as a room) for which the acoustical phenomena have been either recreated or synthesised. To produce a **VAE!**, prior knowledge regarding the room which is to be acoustically recreated must be known; how do all audible frequencies propagate around the room for a set sound source location and receiver location?

This information can be gathered by taking a **RIR!** (**RIR!**) and used to recreate the acoustics of a room for the set sound source and receiver location.

1.3.2   Room Impulse Responses

In order to reproduce the acoustical phenomena of a room, an **RIR!** must be obtained. This is done by exciting all audible frequencies within the room by using a sound source such as a loudspeaker, and recording the result using a receiver microphone.

There are a number of techniques used for exciting all audible frequencies. These include an **impulse** (such as a starter pistol) or an **exponentially swept sine** for which a sine wave is exponentially increased in frequency over a fixed period of time. Using a starter pistol means that no post processing is required as all the frequencies are excited at the same time and the impulse recorded at the receiver position can be used for convolution with an audio source. Using an exponentially swept sine requires post processing in order to time align all of the frequency dependent room reflections, thus producing an impulse response through the use of a deconvolution algorithm, however this method produces a greater signal to noise ratio thus is the desired method [**?**].

Though using an omni-directional microphone to record an **RIR!** is a set standard [**?**], it is also possible to record **RIR!**'s using techniques such as Ambisonics to capture a three dimensional sound field.

1.3.3   Ambisonics

Ambisonics is a technique used to encode and decode three dimensional spatial audio information using just four audio channels. A three dimensional sound field can be recorded using a microphone known as a Soundfield microphone shown on the left in figure 2. These microphones contain four coincident capsules, one of which is an omni-directional capsule (W) and the rest of which are figure of 8 capsules used to record sound in the X (front and back), Y (left and right) and Z (up and down) direction illustrated on the right in figure 2. By combining these signals in what is known as a B-Format audio file, the sound field surrounding the microphone can be captured. By using a system specific Ambisonic decoder, the soundfield can be accurately reconstructed by replaying the B-Format file over a spherical loudspeaker array of arbitrary size.

Sections/Background/images/soundField.jpg Sections/Background/images/soundFieldPolar.

Figure 1: **Left**: Picture of a Soundfield microphone with coincident capsules exposed **Right**: Soundfield microphone polar pattern [**?**]

1.4   Software Overview

1.4.1   ODEON

A method for measuring **RIR!**'s has been discussed, however there is also a way to synthesis **RIR!**'s by using room acoustic simulation software such as Odeon [**?**].

Odeon was designed to provide reliable predictions of room acoustics by using a hybrid of two geometric acoustic models, Ray-tracing and the **ISM!** (**ISM!**) to synthesis **RIR!**'s. These methods model how sound would propagate around a room as though it were a straight line (ray). This inherently neglects wave phenomena such as phase and diffraction, properties that are negligible at high frequencies, however they are fundamental in describing low frequency wave behaviour [**?**]. Therefore geometrical methods are not accurate at modelling sound propagation for low frequency waves.

The reason for Odeon using a hybrid of the two method comes from the inherent problems encountered in each method.

1.4.1.1   Ray-Tracing

The ray-tracing method imitates a sound source by emitting a large number of particles in various directions from a single point [**?**]. These particles are then traced around the room, losing energy each time the particle encounters a surface according to the absorption coefficient assigned to that surface. The angle at which the particle is then reflected is determined by the scattering coefficient assigned to the surface of contact, ranging from a specular reflection to a completely random reflection [**?**]. For a specific receiver position, an area around said point is defined in which rays are collected and used to calculate the results.

Ray-tracing does not provide a completely accurate result as it is a risk that some rays may not pass close enough to the receiver to contribute to the final result. The outcome of the ray-tracing method is a statistical result rather than a complete one.

1.4.1.2   Image Source Method

The **ISM!** can be used to find all possible specular reflections paths from the sound source to the receiver position. This is done by representing each specular reflection from a surface as a secondary source known as an "image source" [**?**], illustrated in figure 1. The advantage of the **ISM!** is that each surface can be modelled by an image source which provides a great deal of data regarding the contribution to the received sound making this method more accurate than ray-tracing, however as each image source then emits more rays there is a possibility to produce a great deal more image sources. This means that the number of image sources grows exponentially for each new image source than is produced making this method much more computationally expensive than the ray-tracing method. This problem can be avoided by setting a *reflection order* which determines how many times a ray can reflect off a surface before calculations are stopped, thus preventing the creation of more image source though the calculates for the prediction of the rooms acoustics will be incomplete. This obviously then makes the **ISM!** less accurate with a smaller reflection order.

Once all image sources have been calculated, a visibility check is run. This checks to see whether the image sources that have been produced can be seen

by the receiver, therefore checking whether that image source needs to be used or not. If the image source is not in sight of the receiver, it therefore does not need to be used for calculations. If the receiver position is moved, the visibility check is run again. Overall, this saves time as the image sources do not need to be recalculated.

Sections/Background/images/ISM.png

Figure 2: Illustration of the image source method (by the author)

1.4.1.3 Hybrid Method

Ideally, the **ISM!** would be used to calculate all sound rays due to its accurate results, however due to the computational limitations, Odeon uses a hybrid method to provide a reasonable compromise between calculation time and accuracy.

The hybrid method first uses the **ISM!** to calculate a number of image sources up until a specified reflection order determined by a **TO!** (**TO!**). For example, if the **TO!** = 2, the image source method will allow a ray to reflect twice which will produce a number of image sources, then it will switch to using the ray tracing method to calculate a statistical model of how the rest of the rays might propagate. This gives the user the choice between computation time and accuracy.

1.4.1.4 Scattering

Quote from Odeon Manual:

> *"In short, each time ODEON detects an image source (IMS), an inner loop of (scatter) rays (not visualised in the 3D Investigate Rays display) is started, taking care of the scattered sound which is reflected from this image source /surface"*

Reword that shit

1.4.1.5 Room Modelling

In order to predict the acoustics of a space, Odeon requires a geometry file in the form of a .par file. This file contains information regarding the room dimensions, object dimensions and positions. This can be produced within

4

Odeon itself by using the built in 'Extrusion Modeller' which allows a user to use a script like language to describe the rooms geometry. It is also possible to use 3rd party applications such as Google SketchUp [?], a software which is described in section GSU.

1.4.1.6   Material Selection

Selecting the appropriate absorption coefficients of the surfaces within a **VAE!** is crucial to its accuracy. Odeon provides a material list with common materials that can be assigned to the surfaces of a model read from a geometry file. This material list can be extended by creating new materials and assigning absorption coefficients to the appropriate frequency bands.

1.4.2   Google SketchUp

Google SketchUp [?] is an easy to use 3-D modelling software that can be used to produce room models. Unlike Odeon extrusion modeller, it allows the user the draw surfaces with a mouse, easily duplicate structures and provides simple measuring tools and markers to enable accurate modelling. Plug-ins such as SU2Odeon ("SketchUp to Odeon") [?] enables the user to convert the model into a .par file for Odeon to use as a geometry file.

1.4.3   Max/MSP

Max/MSP (Max) is a visual programming language that can be used to easily route and manipulate audio signals through the use of simple objects and patches [?]. Max allows a number of external devices to be easily connected to the software such as smart phones and motion sensors through the use of plugins.

<div align="center">

[More on Max]

</div>

1.4.3.1   Spat

Spat, developed at the research institute IRCAM [?] is designed for real-time spatialisation of sound signals in Max without having to touch any code. It provides an simple way to convolve B-format audio signals with impulse responses as well as decode the B-format signals for any given number of speakers in any arrangement by simply providing the number of speakers and angles at which they are placed.

1.5   Project Description

1.5.1   The Virtual Singing Studio

The **VSS!** (**VSS!**) is a loudspeaker based room acoustics simulator used as a tool for analysing the correlation between room acoustic characteristics and vocal performance parameters as part of Dr Jude Breretons PhD Thesis [?]. It

<div align="center">

5

</div>

is comprised of a head-mounted microphone used to capture a real-time audio input from a singer, a software patch that convolves the audio signal with a number of Ambisonic B-Format **RIR!**'s and finally a spherical array of 16 loudspeakers for which the convolved audio signal is decoded and fed to. In addition, a head-tracking device (an Oculus Rift [**?**]) is also used to track which direction the user is facing in the virtual space. A flow diagram of the system can is shown in figure **??**.

[**Add VSS system diagram here**]

The **VAE!** used in the **VSS!** was initially the National Centre for Early Music, a space in York frequently used for musical performance. Using a Soundfield microphone, 16 Ambisonic B-Format **RIR!**'s were captured. Four positions were chosen and four **RIR!**'s facing in each direction (front, left, back, right) were recorded in each. These four directional **RIR!**'s are used to approximate the room acoustic phenomena that would occur if the user were projecting in that direction. This is done by using the data from the head-tracking device to amplitude pan the convolved signals before sending them to the spherical speaker array. Though this is a crude method for achieving such a state, it provides enough variation between the two directions to be considered convincing.

1.5.2   Project Motivation

The **VSS!** addressed a problem faced when trying to research how musicians preform in different acoustic environments: having to travel to each performance space with musicians and researchers. This also indirectly provides a solution for performers wanting to rehearse is spaces that are often inaccessible and would otherwise be expensive to book and travel to. By obtaining an **RIR!** of the desired location, the only time travelling will be necessary is to initially obtain said **RIR!**. However, one limitation of using the **VSS!** is the restriction of position. If a performer wanted to try and sing at another point in the room, an **RIR!** would have to be taken in that position too. This could be done initially, taking a range of **RIR!**'s in a number of positions, however it cannot be guaranteed that all positions desirable to the performer will be available. Therefore, this project aims to implement a system that will allow the user of the **VSS!** to feel as though thy can move around a **VAE!** freely.

Previous projects have addressed mobility within **VAE!**'s before, such as [**?**]. The paper initially looks at two methods for providing mobility:

1. **Direct room impulse response rendering**: By producing a large grid of **RIR!**'s within the desired space, it is possible to allow a user to select a position in the room by interpolation between the two nearest neighbouring **RIR!**'s. This method does not provide truly accurate room acoustics for the selected location and also suffers from the fact that a large number of **RIR!**'s are required, thus the requirement for a large amount of storage space.

2. **Parametric room impulse response rendering**: By actually synthe-sising **RIR!**'s in real time for the given position of the user in the **VAE!**, it is possible to provide an accurate **RIR!** for the given location. This method avoids the need for a large amount of storage space and a system to retrieve the correct files, however, as will be seen through the rest of this report, post-processing of the obtained RIR's would not be possible with real time rendering.

As the intention was to build upon the existing **VSS!** system, it was decided that the direct room impulse response rendering method would be used, as this method would allow for the modification of the existing patch used to real-time convolution.

As the intention was to build upon the existing **VSS!** system which requires pre-recorded (and post-processed) **RIR!**'s, it was decided that the direct room impulse response rendering method was to be used and the effects of the issues related with the method studied.

1.5.3   Project Objectives

Given the project aims and the background provided, the following is a list of objectives that were set in order to complete this project:

1. Find an appropriate room to be modelled as a **VAE!**

2. Digitally model the room using either Odeons extrusion modelling or 3rd part application

3. Import room model into Odeon to finish material selection and **RIR!** set-tings

4. Produce a grid of **RIR!**'s that can be used to interpolate between to sim-ulate user positions

5. Record real **RIR!**'s in locations that can be compared to synthetic **RIR!**'s

6. Extend upon existing software patch to accommodate new functionality

7. Preform user tests:

   - Test #1: Does the perception of distance change when using real or synthetic **RIR!**'s?

7

- Test #2: How far does the user have to move in the given **VAE!** before they notice they have moved

- Test #3: How many **RIR!**'s are required for interpolation for the user to feel they are moving around the space freely

In order to test the perceptual differences when using synthetic RIRs (Test #1), real RIRs of the same space had to be taken.

Previous test in the VSS have required 'Plausibility' test, where the user must evaluate the **VAE!** produced by the VSS without reference to the real venue. This is usually because testing a virtual environment against a real one (authenticity tests) requires travel which can be expensive and difficult. Therefore, by running test based on 'plausibility' a sense of how convincing the virtual room is can be obtained. In the case of other virtual reality systems where the virtual environment does not exist in the real world, only these types of test can be run. Though plausibility test are acceptable, an authenticity test gives more objective results by comparing the simulated **VAE!** to the real thing. Therefore impulse responses of Hendrix Hall were taking, in the same format as the synthesised RIR's.

1.5.4   Project Overview

The following outlines the steps that were required to produce a system that allowed a user to feel as though they are freely moving around a **VAE!** and the tasks necessary to investigate the plausibility of said system, thus completing the project aims and objectives set

1.5.5   Room Choice

The following room features were kept in mind when searching for a room to use as part of this project:

Size:	Large enough to be used as a singing space
Simplicity:	Simple enough architecture to be able to model with the time available
Accessibility:	The room had to be easily accessible to take multiple measurements to make blue prints and take **RIR!** measurements

The room chosen was Hendrix Hall, a large lecture theatre on the University of York campus. The room contains retractable seating leaving the large space in the centre open and unobstructed. The room is architecturally simple being almost perfectly rectangular with the occasional wall indent. With it being located on the universities campus it can be booked for any time of the week meaning it can be accessed when necessary and for free.

1.5.6   User Test Planning

In order to perform the final user tests (test #1 and test #2), minimum distance
between **RIR!**'s had to be decided.