
BACKGROUND

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 room acoustic simulation software Odeon.

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

1.1 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.

1.2 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 Synthetic Room Impulse Responses in 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.

1.3.1 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.

1.3.2 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 ???. 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 1: Illustration of the image source method (by the author)

1.3.3 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 of 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 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.

1.5 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 vo-



Figure 2: **Left:** Picture of a Soundfield microphone with coincident capsules exposed **Right:** Soundfield microphone polar pattern [?]

cal performance parameters as part of Dr Jude Brereton's PhD Thesis [?]. It 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 be shown in figure ??.

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 four directions (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.

1.6 Project Motivation

The **VSS!** addressed a problem faced when trying to research how musicians perform 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 in 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 can not be guaranteed that all locations desirable to the performer will be available.

Therefore, this project aims to implement a system that will allow the user of the **VSS!** to 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: *direct room impulse response rendering* and *parametric room impulse response rendering*. Direct room impulse response rendering consists of obtaining a set number of **RIR!**'s in a grid and interpolation between them to synthesise the user's position.

This method suffered from the fact that a large number of **RIR!**'s are required, thus the requirement for a large amount of storage space.

Parametric room impulse response rendering actually synthesises **RIR!**'s in real time for the given position of the user in the **VAE!**, making them more accurate 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 is required, which would not be possible with real time rendering.

1.7 Similar Projects

1.7.0.1 Geometrical Methods