Creating a visual model of audio quality in rooms/venues

By

Joseph Davison

A report on project work carried out for the degree of

BSC (Hons) Music Technology and Audio systems

1st Supervisor Ian Gibson

2nd Supervisor XXXX XXXXXX

xx/xx/22

# Abstract

Understanding room acoustics intuitively is quite hard for humans to do especially as you change environment variables. This project intends to investigate methods of visualising audio quality in a dynamic environment to grant a sound technician a greater understanding of the room and help them evaluate different possible setups.

My approach finding a solution has two main segments:

• Firstly, I will conduct research into which audio characteristics I should incorporate into an index by which to judge audio quality at any discrete point. This could be qualities like RT60, DRR or deviation of the frequency spectrum from the source point (constructive/destructive interference, standing waves etc.).

• Once I have created an algorithm which gives an index score for any point in a room, I will build a software tool to visualise this within a 3D computer model of the room. This will require further research into 3D audio simulation methods and other software challenges like using GPU acceleration and other optimisations as the larger and more complex the model gets the longer it will take to run (linear – O(n)), however, there might be a way to reduce calculations and interpolate.

^first person

Contents

[Abstract 2](#_Toc121094428)

[Introduction 4](#_Toc121094429)

[Background Research 5](#_Toc121094430)

[Defining an Index Score 5](#_Toc121094431)

[Spatial model 7](#_Toc121094432)

[Plan of Work 11](#_Toc121094433)

[Creating an Algorithm 11](#_Toc121094434)

[Design ideas and Optimisations 11](#_Toc121094435)

[Prototype Code Examples 11](#_Toc121094436)

[Spatial model 12](#_Toc121094437)

[Visualisation of the Index Score 13](#_Toc121094438)

[Deployment 14](#_Toc121094439)

[Design ideas 14](#_Toc121094440)

[Prototype Code Examples 14](#_Toc121094441)

[References 15](#_Toc121094442)

[Appendix 16](#_Toc121094443)

# Introduction

To include the background to and nature of the problem, aims and objectives.

Summary of the report structure.

Research methodologies, data collection methodologies, qualitative, quatitative

Important to really explain the idea of a public access website and how that requires low compute time due to queues

The result should be a system that is both fast and efficient, which can accurately measure and display the audio quality of a room in real-time.

The UI should be intuitive and easy to use, so that any engineer, novice or experienced, can use it quickly and effectively.

The system should also be scalable, so that it can be deployed across multiple venues with varying complexities.

# Background Research

## Defining an Index Score

The point of this project is to create a visual model of audio quality throughout a room and to do that we must first define a formula by which to judge audio quality at each given point.

Firstly, there is a lot to be gained from looking at other similar ideas like the speech intelligibility index (SII), which is a simple method for predicting specifically how intelligible speech is within a given space. The two main inputs for this index are audibility which is a signal to noise function for a frequency band and a frequency importance function which gives a weighting to each band for how important that specific band is to the interpretation of the speech, as outlined in (ANSI/ASA, 1997). Obviously, this index is very heavily based on speech and how different frequency bands are more important than others for comprehension, unfortunately, isn’t true with music as the totality of the spectrum is important, the general idea is you want the listener to receive a signal where the whole spectrum is as close as possible to the output plus possible desired room reverb, it could be argued that perhaps a greater focus might be wanted for typical instrument bands, however, that will still cover a very large area. SII is clearly built for speech, but, the basic outline of a sum of a combination of parameters, normalised to a 0 – 1 scale, can easily be carried forward into the creation of the formula.

Secondly, industry standards in room acoustics should be considered to approach this problem, for example, in their paper *A Universal Deep Room Acoustics Estimator,* P. S. Lopez, P. Callens & M. Cernak (2021) use Reverberation Time (RT60), Direct to reverberant ratio (DRR), Clarity (C50 & C80), and Speech Transmission Index (STI) as they are “basic acoustic parameters that characterize the environment well”, all of these parameters would be useful in giving insight into a room’s behaviour however, again, fewer calculations are better due to the focus on run time, therefore only RT60, DRR, and C80 will be used. The use of these standards should hopefully mean that it is intuitive to understand for engineers who would already be familiar with such ideas, however, it is also important that an untrained person can get a good idea of quality, so simplicity is key.

Finally, the frequency domain should be considered. Because two sound sources are being used, even if all walls were 100% absorbent, there will be constructive and destructive interference between waves, alternatively, if the walls were highly reflective it would then also be important to be able to model phenomena like standing waves. one possible way to include this could be calculating the correlation between the frequency domain of the receiver and both left and right sources or perhaps left and right receivers separately with corresponding sources, however, the total calculations necessary should be kept to a minimum for this algorithm due to the importance of low computation time.

Although possibly subject to change due to computation time optimisations, the basic formula will look something like this:

Y =

## Spatial model

To perform the calculations from the section above obviously a 3D model of a room is needed, many solutions to this problem exist with a large range of complexity, however, with complexity comes both computation cost as well as implementation time, because of this, certain higher-level models will have to be left out of consideration for the final system.

One method is described in *Parametric Directional Coding for Precomputed Sound Propagation,* N. Raghuvanshi & J. Snyder (2018). In this paper Raghuvanshi and Snyder outline a system for precomputing large virtual spaces so that a binaural head related impulse response (HRIR) can later be calculated in real time as a listening position is arbitrarily moved around anywhere in the space. The main point of their paper was to create a system for realistic audio in video games (Microsoft, n.d.) meaning the precomputation time is not a problem due to encoder only needing to be “baked” once per level and then can be shipped out with the game, in their results Raghuvanshi and Snyder state on a single 8-core machine the pre-calculations can be performed in between 20 minutes and 5-6 hours depending on input parameters. While these calculations can be run in parallel allowing more powerful modern multi-core chips to complete the same calculations in less time, this is still unreasonable for a system which is being designed for general access where each new user might add hours of computation time to the queue.

The paper *A Universal Deep Room Acoustics Estimator*, Lopez, Callens & Cernak (2021), while not specifically about propagation modelling, is still worth mentioning again. This paper proposes a method for using a neural network to predict signal to noise ratio (SNR) along with 5 room acoustic parameters (RT60, C50, C80, DRR, and STI) from an input reverberant and noisy speech signal. Their results showed that the model could follow the ideal estimation reasonably close with some variance. This brings up an important question of what role artificial intelligence could play in the implementation of the model, a similar system which trains using 3D maps of a room and room impulse responses (RIR) could be quite successful in predicting reverb given a specific input sample, however, rather than finding impulse responses (IR) in a diffuse room, the point of this research is to create a system that can model the frequency spectrum at every given point in a room, this means that if real world data were to be used it would have to be hundreds of measurements and the exact locations of those measurements across multiple hundreds of venues which would be difficult to gather individually, in their paper Lopez, Callens & Cernak do reveal a few open source data sets used for their training, however, as stated above, this system would need a slightly different set which cannot be found open source or otherwise. One alternate method of training could be to use a system like the one outlined in Raghuvanshi & Snyder (2018) to pre-compute large spaces, take virtual measurements within that space and then feed that resulting data set into a neural network, however, implementation time for that method might exceed the bounds of this project and simply from a user experience stand point very few people will have a perfect 3D map complete with complicated absorption coefficients for whichever venue they might want to model.

Another method is described in *A Simple Method for Calculating the Distribution of Sound Pressure Levels within an Enclosure,* B. M. Gibbs & D. K. Jones (1972). This paper outlines a method for modelling sound propagation by imagining a 3D array of tessellated cuboids surrounding a real room with a sound source and a receiver where each adjacent box mirrors its surrounding boxes therefore representing what a wavefront would see as it reflects around the room, however, the wavefront is moving outward into the imaginary rooms instead of reflecting. This allows for simple wave calculations to be used to model the propagation as if the source is in completely free space moving radially outwards, If you calculate the distance between each of the reflected imaginary points and the source you can gather delay time and if you calculate which walls the wave has passed through to get to a point then you can figure out a chain of absorption coefficients to multiply the final signal by.

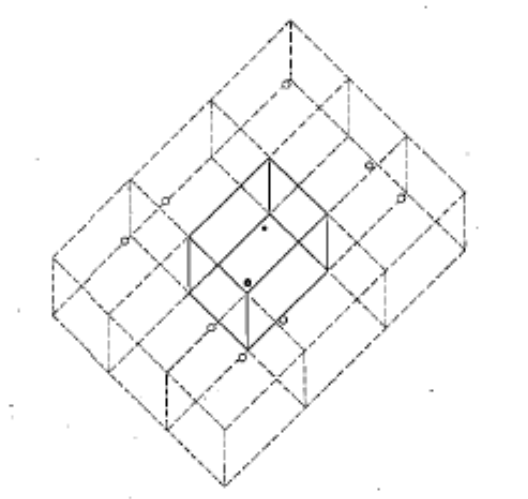


Figure : Visualisation of the reflected rooms. The smaller dot is the source, the larger dot is the receiver, and the blank circles are the imaginary reflected points.

This method is much simpler to implement and should be able to run much faster than the other methods discussed in this paper while giving fairly accurate results for a given space, however, no room is a perfect cuboid and rooms will not only contain static objects within it, which might scatter or absorb audio, but also dynamic objects, such as humans. J. S. Bradley (1996) shows there is a simple linear relationship between the absorption coefficient of a crowd and the area of ground covered, this means that should a user be able to predict attendance the system could incorporate the effect of crowds upon the reverberation time, unfortunately crowds also create their own noise which to simulate an additional sound source could possibly be included in the centre of the room emitting some form of white noise like “crowd signal” however including this would greatly increase run time by possibly half if an additional source needs to added to the original two. Furthermore, crowds will grow and diminish through time, even during a show, meaning they are ultimately impossible to predict. Additionally, the further the shape of the users venue strays from the ideal cuboid, the further the model will stray from reality, meaning there is a possibility for users to receive wildly different results from what they see in real life and some might not even be able to approximate their venue to a cube in the first place, for example large concert halls which might have irregular walls or rounded edges, this will cause the pool of possible users of the system to be decreased.

Possibly include research from the modern usage of Gibbs, Jones method – they discuss air absorption and uneven absorption coefficients

The method outlined in Gibbs & Jones (1972) is clearly not perfect but what it looses in accuracy it does make up for with simplicity and computation time which is a far more important quality in this system, therefore this method has been chosen to be the basis upon which work will be done to add functionality and refine it into a final system, a plan for this work will be discussed in the next chapter.

Representing the room in this manner allows for a very simple equation to be used to find the resulting function at each imaginary point by finding the length from the source and therefore delay time TAU.

Xn(t) = (ALPHA1 \* ALPHA2 \* …) \* INVSQ LOSS \* F(t - TAU)

Where ALPHA is the absorption coefficient of each wall the wave passes through, INVSQ LOSS is the magnitude lost to dispersion, t is time, X(t) is function at an imaginary point with n being the point number, and F(t - TAU) is the source function delayed by time gathered from distance to source.

Then summing each of these functions gives the resulting real function at that specific point

Y(t) = SUM 0-n Xn(t)

you can then perform a Fourier transformation on to get the frequency domain at regular intervals.

yHAT(f) = INTEGRAL Y(t) dt

if you do this for

## Visualisation of the Index Score

Intro

When creating a visual model, the main aim is to make it as intuitive as possible, so the end user must spend as little time to decipher the results as possible. A quick look at other visualisation methods across the music industry shows that, while there is a lot of divergence, there are a few specific patterns that often show up.

EQ

One of these is the idea that the visualisation will dynamically move along with the music, like how a digital EQ might show a spectrogram of the output, this is important as music by its nature is somewhat ephemeral, you cannot understand the music if you experience it all at once, therefore, the visualisation must develop alongside the music in the same way the acoustics of the room will develop.



Figure : Example of an EQ plugin with a spectrogram.

Colour

Another common pattern in music visualisation and design in general is the use of colour to convey information about something, one example of the use of colour is in level monitors where quieter signals are displayed as green, louder signals become more yellow/orange, and once the signal begins clipping it will display as red. The use of colour to explain the state of something is quite a fundamental part of visual design and so its inclusion in this system is undisputable and the visualiser will most likely use a similar colour layout to the aforementioned level monitors, a 0 or bad value will be displayed as red, low to medium values will be displayed as orange and medium to good values will be displayed as green.

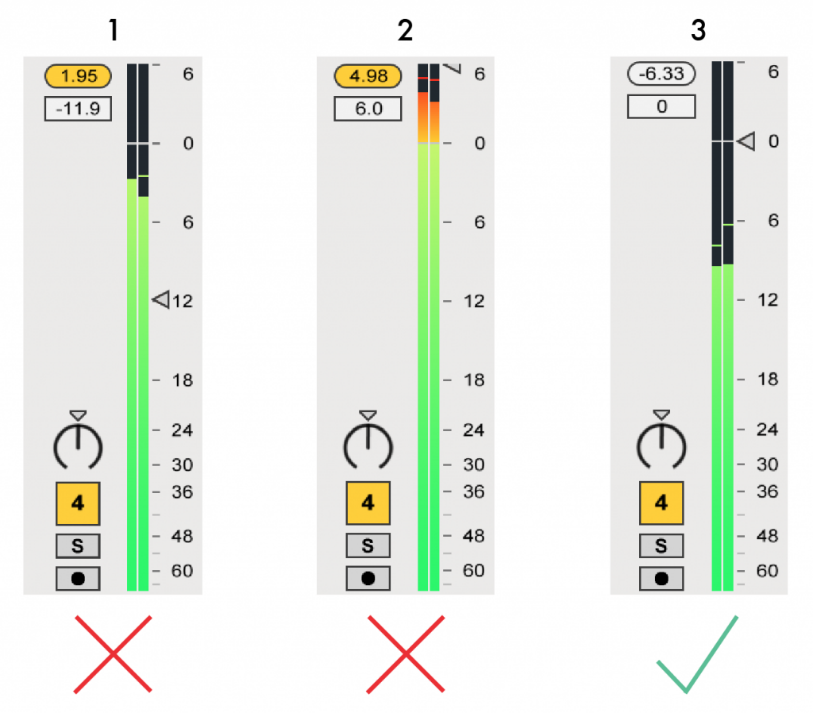


Figure : Colour showing loudness in level monitors in Ableton, (University of Washington 2022)

Interactivity

Another common attribute of visualisation systems is interactivity.

An interactive/draggable map allows the user to see all aspects of the room that could be missed if it was simply static, it could also allow them to focus in on one particular area.

conc.

The point of including these design patterns is to try to make the system blend in with other visualisation systems hopefully providing a sense of familiarity to a user, improving both the initial acclimatisation as well as the overall efficiency of displaying room characteristics.

Ultimately, the visualisation should be successful if these conditions are met:

* It is dynamic and can change along with audio.
* It is intuitive and simple to understand by use of typical design choices i.e., colour.
* It allows the user full comprehension of the 3D space.

# Plan of Work

## Creating an Algorithm

### Design ideas and Optimisations

How 3D space is dealt with. Separation of getTimes() from getIndex() and how total 3D model might not be completed. Other design complications. Showing independent extracurricular research is in the rubric.

Oop vs functional. GPU acceleration. How optimisation links with end user, perhaps they would be in the field and need a quick response, could that be possible - web deployment and scalability (nginx massive concurrency). Big O notation.

The main consideration for this algorithm is compute time

### Prototype Code Examples

Can perhaps include references to a public git hub repository.

In addition to the research, a few simple early implementations of this algorithm were created to guide the literature review.

Full examples can be found in the appendix (Appendix no.)

## Visualisation

webGL, interactive/draggable, transparency, playback with audio that you uploaded

side on/ top down, cross-sectional, interactive? Playback file compression / file type of playback file.

In order to render the visualiser in-browser the open-source library webGL will be used.

## Deployment

### Design ideas

Web page, easy access for a sound tech on the job, can “order” process time before hand and once you have the file to play back you can play it back any time with the playback functionality on the website

### Prototype Code Examples

Can perhaps include references to a public git hub repository.

Again to supplement research an early website design was created to know what the methods of deployment of this system might look like, however, unlike the algorithm prototypes, the below code is likely very close to the final product.

Full examples can be found in the appendix (Appendix no.)

# References

ANSI/ASA. (2020). *Methods For Calculation of The Speech Intelligibility Index* (Standard No. S3.5-1997 (R2020)). Retrieved from https://webstore.ansi.org/standards/asa/ansiasas31997r2020

P. S. Lopez, P. Callens & M. Cernak (2021). A Universal Deep Room Acoustics Estimator*. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. https://doi.org/10.48550/arXiv.2109.14436

N. Raghuvanshi & J. Snyder (2018).Parametric Directional Coding for Precomputed Sound Propagation. *ACM Transactions on Graphics*, 37(4), 1-14, Article 108. https://doi.org/10.1145/3197517.3201339

Microsoft. (n.d.) *Project Triton*. Retrieved December 2, 2022, from https://www.microsoft.com/en-us/research/project/project-triton/

*Gibbs, B.M. & Jones, D.K. (1972).* A Simple Image Method for Calculating the Distribution of Sound Pressure Levels within an Enclosure*. Acta Acustica united with Acustica,* 26(1), 24-32(9).

Bradley, J.S. (1996). The sound absorption of occupied auditorium seating. *Journal of the Acoustical Society of America*, 99, 990-995. https://doi.org/10.1121/1.414562

University of Washington. (2022). *Ableton Live Online Curriculum*. Retrieved December 5, 2022, from https://itconnect.uw.edu/tools-services-support/teaching learning/workshops/online-tutorials/digital-audio-online-curriculum/ableton-live-online curriculum/

# Appendix

Algo prototype code examples.

Website prototype code examples.