**INSTITUTO TECNOLOGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE**

Imagen que contiene Logotipo

Descripción generada automáticamente

**Data Science Master Program**

**REPORT #1:**

**Problem statement, contextualization, and specific and general objectives**

**Directors:** Dr. Jorge Arturo Pardiñas Mir, Dr. Riemann Ruiz Cruz

**Presented by:** Eng. Adrian Ramos Perez

adrian.ramos@iteso.mx

March 22nd, 2022

CONTENT TABLE

[I. Problem statement 3](#_Toc98949893)

[II. Contextualization 4](#_Toc98949894)

[1. Acoustic Effects on a reduced space 4](#_Toc98949895)

[Single point vs multi-point 6](#_Toc98949896)

[2. Mathematical model of the space acoustic effects of audio signal and how it can be treated 7](#_Toc98949897)

[Signal in Discrete Time 8](#_Toc98949898)

[System’s Impulse Response 8](#_Toc98949899)

[Z-Transform and Complex Plane 9](#_Toc98949900)

[System’s Transfer Function 9](#_Toc98949901)

[Inverse Impulse Response of a System 10](#_Toc98949902)

[Classic Approach 10](#_Toc98949903)

[3. Previous works on the theme 10](#_Toc98949904)

[Minimum Phase vs Mixed Phase 11](#_Toc98949905)

[Single point and multi-point approaches 11](#_Toc98949906)

[Room response and its perception 12](#_Toc98949907)

[Invertibility of the Room Response 12](#_Toc98949908)

[Non-linear techniques 12](#_Toc98949909)

[Neural Networks 13](#_Toc98949910)

[III. General objective 13](#_Toc98949911)

[IV. Specific objectives 14](#_Toc98949912)

[V. Bibliography 14](#_Toc98949913)

# Problem statement

In any enclosed environment or space, as a room, when we listen sounds, they got altered by the reproduction system and the space characteristics like its dimensions and the materials it is made of.

For example, when there is a conversation going on in a room, the signals of the voices are affected by nearby reflecting walls, as the sound travels in many directions and not only to the receiver, bouncing off the walls [1]. This “room effect” is the convolution in time domain of the speech signal with the room impulse response. The room effect is often perceived as an echo or reverb, and it can be undesirable depending on the dimensions of the room and the amount of the effect. One option to mitigate this is to pass the signal to a filter which inverts the room impulse response.

So, there are physical spaces that require special acoustic treatment which usually is very complex and expensive to get a desired acoustic, which is the case of music production studios (in specific mastering and mixing rooms), meeting rooms, cinemas, etc.

# Contextualization

## Acoustic Effects on a reduced space

When an enclosed space or room is small, there are several conditions that affect sound propagation in it. The smaller the room, the more problems it will has caused by early reflections [4]. When a sound is reproduced, sound waves propagate through air and more mediums available inside the enclosed space, causing it to interact with multiple phenomena.

The incident wave gets physically affected when it reaches enclosing surfaces. Physical phenomena that occur to the original wave are:

* Reflection
* Diffraction
* Refraction
* Diffusion
* Absorption
* Reverberation

These effects are shown in the figure 1:

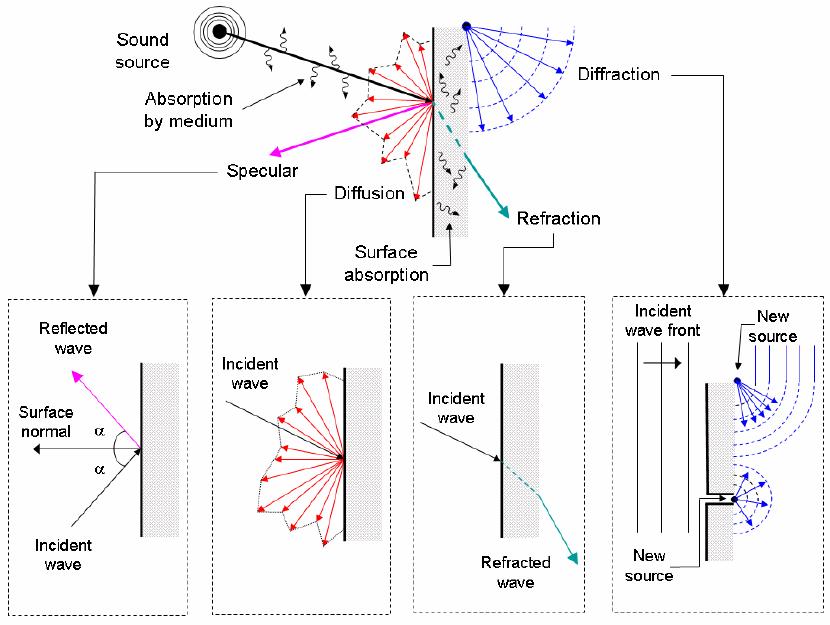


Figure 1. Reflection of waves on a wall

From these effects, small rooms (approx. less than1500 ft3) start having undesired acoustic properties and problems concerning sounds that are reproduced inside the enclosed environment.

Depending on the frequential content of a signal it will propagate in the room reaching different point depending directly on the wavelength, lower frequencies have a longer wavelength comparing to middle and higher frequencies, and thus travel and reach far points from the point of emission.

A typical placement of a stereo reproduction system in a room is shown in figure 2:

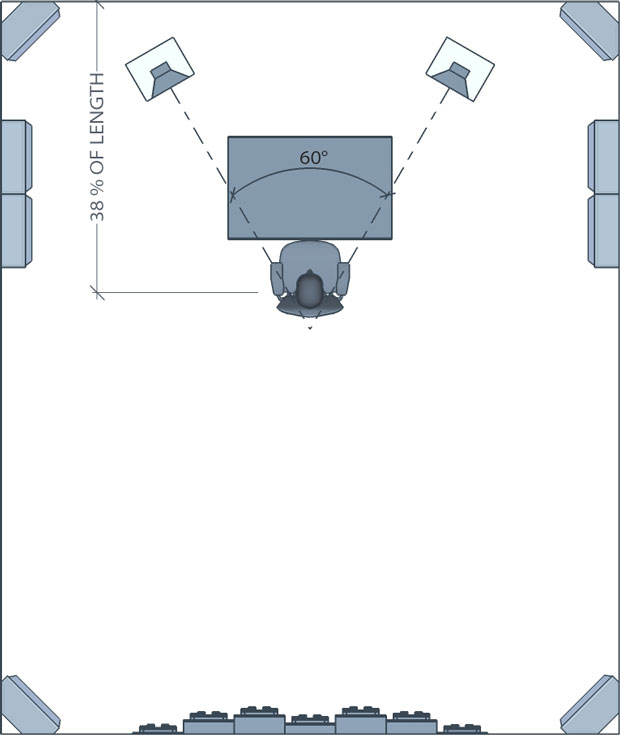


Figure 1. Control Room Diagram

### Single point vs multi-point

There is to note that inside an enclosed environment the energy levels of the frequencies for a specific point in the room will be different from all of those in the other points in the room. A solution that focuses on equalize a single point may not be valid for different positions and angles within the room, so here comes the multi-point approach where the focus relies in finding an average equalization for two or more points.

For example, the figure 2 illustrates a room response on a speech audio signal:

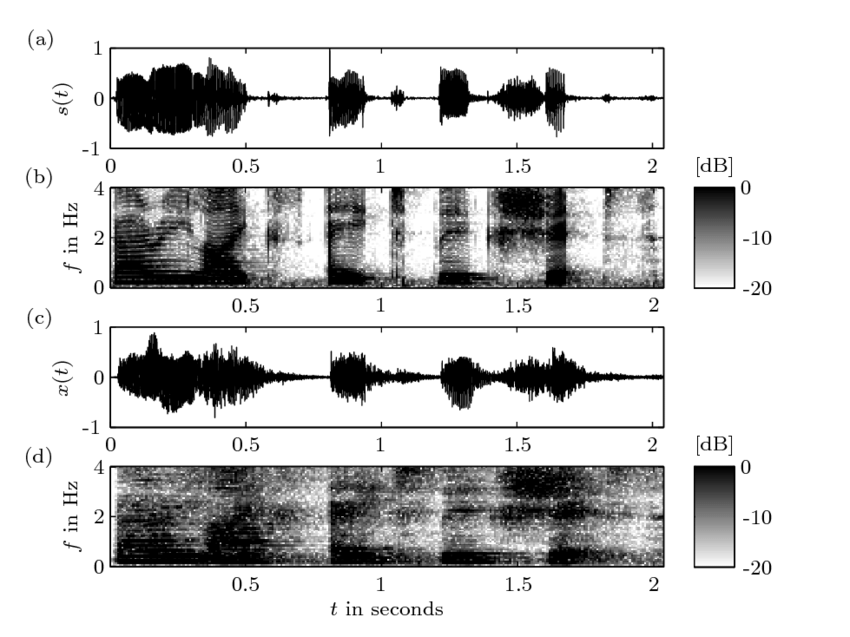


Figure 3. Room Response on a Speech Signal. A) original signal, c) signal after convolution with a room impulse response.

As it can be seen in the spectrograms b) and d), which displays energy content for each frequency, the frequential content changes having less energy in mid and high frequencies (white color) prior to the convolution with the room response which in this case it can be seen it’s adding power on those frequencies (black).

## Mathematical model of the space acoustic effects of audio signal and how it can be treated

The objective of this work is focused on discrete time signals, as the problem is tackled after a phase of recording and thus transforming a real-time signal by sampling it on discrete time instants, dependent on the sampling frequency and the analog-to-digital converter of the audio interface.

### Signal in Discrete Time

A discrete-time signal is represented as and it’s defined only for integer values of n:

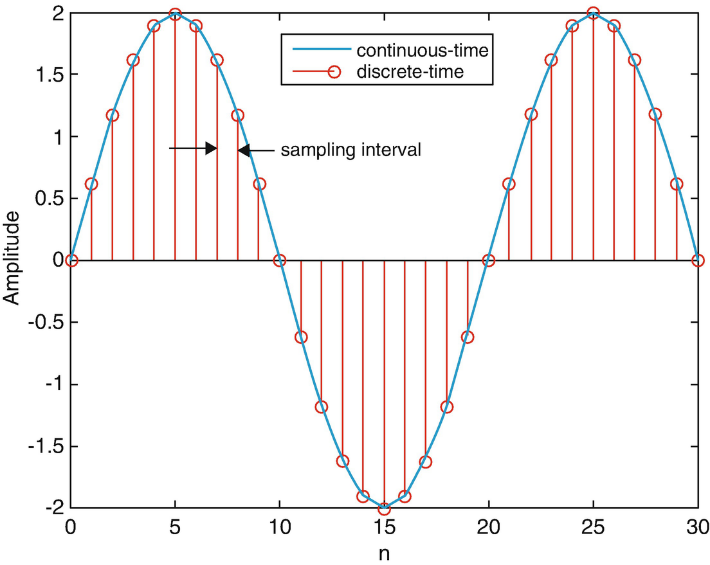


Figure 2. Sampling of a real-time signal

### System’s Impulse Response

A room can be modeled as a system which has its own unique response.

A discrete system can be modeled by its impulse response (**IR**) , and its output can be determined by the convolution of its impulse response and any input signal :

y(n) is how the audio signal sound into the room:

Texto, Chat o mensaje de texto, Pizarra

Descripción generada automáticamente

### Z-Transform and Complex Plane

One effective tool that helps to analyze lineal time-invariant systems is the z-transform, which is the equivalent of Laplace transform for the discrete-time.

Using the Z-transform, a discrete-time signal is converted to the complex frequency domain.

### System’s Transfer Function

The transfer function of a system models, in theory, every possible output for a system given all its possible inputs. The T.F. in discrete time is equivalent to the ratio between the Z-transform of the output and the z-transform of the input.

For the impulse response, its transfer function is given also by the Z transform:

And it’s called the system function or transfer function of the system.

For the convolution, its **Z- transform** results in:

Where , and are the z-transforms of , and ; which means that the z-transform of a system’s output is the multiplication of the z-transform of its input by its system function.

According to this result, the system function can also be calculated as:

The system function is a general case that includes the so-called frequency response of a system or H(w). H(w) corresponds to the values of H(z) for all *z*= ejw, or, in other words, for all z values with a magnitude equal to 1:  |*z*|=1

### Inverse Impulse Response of a System

In the case of the effects of a room over the audio signal propagated into it, such effects could be theoretically eliminated by applying and additional system which transfer function, , performs the inverse effects of the room:

Diagrama

Descripción generada automáticamente

Figure 3. Scheme of theoretical attenuation system of room impulse response.

### Classic Approach

In that way, our problem is pointed into finding that **inverse transfer function** in a mathematical way, and then we will be theoretically able to apply the inverse room response to the signal prior to its reproduction in the room, altering the sound perception once it has been modified and played. This is the **classical approach** to solve this problem.

## Previous works on the theme

This topic has been studied for approximately 40 years at the time of writing this, so approaches to this problem has changed along the time with different names for the same problem: “Room equalization”, “Room correction”, “Room compensation”, “Room inversion”, “Room dereverberation”, “Dereverberation”, “Reverberation reduction”, etc.

Prior classic approaches involve the idea of finding the **transfer function** of a room through obtain its **impulse response** and **invert** it, to find a specific equalizer.

Another methods are: non-parametric least squares, frequency response inversion (no parametric, ARMA, ARIMA models (auto-regressive moving average, parametric), control of temporal decay on low frequencies (parametric) and others, which objective is to equalize the **magnitude** of the RTF.

### Minimum Phase vs Mixed Phase

If the system takes into consideration the phase, it can actuate not only in magnitude but also in the room transfer function (RTF) that has phase excess.

### Single point and multi-point approaches

Following an approach focused on the point or points to equalize, room response can be classified as follows:

|  |  |  |
| --- | --- | --- |
| **ROOM RESPONSE EQUALIZERS** | | |
| **Input/Output relation** | **Focus** | **Description** |
| SISO  (single input – single output) | Single-point | They estimate the equalization filter based on a **unique single point of measuring of the transfer function**. It is valid only for a limited and small zone (which is the size of a fraction of and acoustic wavelength) the measuring point. |
| MISO  (multiple inputs – singleoutput) |
| MIMO  (multiples inputs – multiples outputs) | Multi-point | They use multiple measures of transfer function in different fixed points on a given time to estimate the equalizer, which enlarge effective zone of the equalizer. |

As RTF varies significatively depending on rom’s position and the time, the room can be considered as a weakly non-stationary.

## Room response and its perception

As the acoustical properties of the environment determine the room’s impulse response and the human perception, this involves some knowledge of human perception of sound and psychoacoustics, for a correct analysis of the RIR and the equalizer design.

Impulse response components:

* Direct sound
* Early reflections
* Late reflections

Being more specific, perception of higher frequencies gets affected by the resolution of the auditive human system, which is non-linear and non-uniform, being it almost logarithmic. That’s the reason why psychoacoustic scales are used:

* Mel scale
* Bark scale (critical band rate)
* ERB scale (Equivalent Rectangular Bandwidth)

## Invertibility of the Room Response

The first paper is from 1979 by Neely and Allen, who demonstrated that, if a wall reflectivity is les tan 36%, the Room Transfer Function is of mínimum pase and thus not invertible. If it’s greater (70 to 90%) the response is noto f mínimum pase, but still can be equalized (only the minimum phase part) by factorizing the transfer function of the room into a product of a term of minimum phase and a stable all-pass filter

### Non-linear techniques

Impulse response in a determined point is the sum of the signal and its reflections [], which is algo conditioned by frequency response of the audio reproduction system, which is tried to be mitigiated by applying a linear filter to the signal before being transduced and played by the speakers, inverting the impulse response.

The complexity generated by the number of sources and measuring microphones is such that, although there are linear optimization and inversion algorithms approaches, here comes the **necessity of aboard nonlinear techniques as evolutive algorithms, machine learning and neural networks**.

**Particle Swarm Optimization** has been demonstrated to provide a better solution tan mean least-squares and recursive least-squares. [] A key point is that, on audio, time-invariant is assumed as room’s impulse response is tackled by a static approach.

Algorithms that has been used on contemporary days include:

* ***Time Delay Neural Network:*** to equalize taking as input the same input delayed by a time unit, and as the system’s output the signal recaptured by the microphone.
* ***K-Nearest Neighbour:*** for timbre equalization based on user preferences as training data.
* ***Dilated Residual* *Network****:* to automatize resonance equalization.
* ***Convolutional Neural Network:*** for a point-point equalization, without previous knowledge of filter parameters (gain, cut-off frequency, quality factor, etc.).
* ***Particle Swarm Optimization:*** IIR filter design. [] obtaining filter parameters with a plane impulse response on magnitude and linear phase.
* ***Gravitational Search Algorithms***
* ***Artificial Inmune Algorithm:*** IIR filter design. []
* ***Ant Colony Optimization:***IIR filter design. []
* ***Tabu Search:***IIR filter design. []

### Neural Networks

Neural networks arquitectures that has been previously used to solve the problem []:

* Radial Basis Function (RBF)
* General Regression Neural Networks (GRNN)
* Radial Basis Exact (RBE)
* Back Propagation Neural Network (BPNN)
* Multilayer Perceptron (MLP)

# General objective

To model a room in conjunction with an audio reproduction system and correct acoustical problems on it, to be able to predict the frequency response on any given input signal.

Automatically control a reproduction system on a given room after measuring its IR/RTF to obtain a plain response, mitigating the acoustical problems.

# Specific objectives

1. Use a machine learning/deep learning framework to solve the problem.
2. That the final system can be generalized and valid to more environments and not only to the room where the system was trained and tested on. (Enclosed and controlled spaces where acoustical problems occur) but being possible to have different dimensions and materials.
3. Explore and propose the best machine learning/deep learning based on metrics.

# Bibliography

[1] Invertibility of a room impulse response (1979) S.T. Neely, J. B. Allen

[2] Inverse Filtering of Room Acoustic M

[3] Gerzon, M. (1991). Digital room equalization.

[4] Owsinski, B. (2008). The Mastering Engineer’s Handbook: The Audio Mastering Handbook. Second Edition U.S.A.: Thomson Course Technology.

On the Combination of Systems for Listening-Room Compensation and Acoustic Echo Cancellation in Hands-Free Telecommunication Systems - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Influence-of-a-room-impulse-responses-on-a-speech-signal-a-anechoic-speech-signal\_fig2\_267449238 [accessed 10 Mar, 2022]