TAMPERE UNIVERSITY OF TECHNOLOGY

SHARATH ADAVANNE

Room Surface Estimation Using Reflection Coefficients Measured In-situ.

Master of Science Thesis

Examiners: Dr.Tech. Ari Visa
Dr.Tech. Pasi Pertilä
Examiners and topic approved in
the Department of Signal Processing
meeting on 17th August 2011

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology
SHARATH ADAVANNE: Room Surface Estimation Using Reflection Coefficients Measured In-situ.
Master of Science Thesis, 43 pages, 3 Appendix pages
September 2011
Major: Signal processing
Examiner: Dr.Tech. Ari Visa, Dr.Tech. Pasi Pertilä
Keywords: Room Surface Estimation, Reflection Coefficient, Reflection Method, Sine Sweep, Room Impulse Response

Room surface estimation is the process of estimating and characterizing the surfaces of a room from the measurement provided by an array of microphones. In any normal room, wavefront radiated by the source reaches the microphones after reflecting from the surfaces of the room. On performing the post processing on this recorded signal, depending on the type of source signal used; we obtain the unique signature of the room called the room impulse response (RIR). Historically room impulse responses have been used to calculate the acoustical parameters. These are generally used for objective evaluation of the rooms. Here is an effort to extract more information out of the RIR by understanding the physics of it.

In this thesis the reflection from the surfaces is employed as useful information and methods to use this information are presented. The reflective information can be used in the determination of the reflection coefficient of the surface. The reflection coefficients of four common room materials are obtained from their respective impulse response at oblique incidence. The obtained reflection coefficients are studied for the classification of materials.

The problem of estimating the room surface using the reflection coefficients has been approached in a systematic way, accounting for parasitic reflections and background noise. The room impulse response is obtained using the sine sweep signal transmitted through the speaker attached to one microphone array, and recorded by another microphone array. The recorded signal was used to identify possible surfaces using clustering on the obtained reflection coefficients. The classification so obtained is compared with the ground truth to calculate the performance and practicability.

PREFACE

I would like to express my sincere thanks and appreciation to Dr.Tech. Pasi Pertilä, for giving me the opportunity to work at the Department of Signal Processing and for his attention, guidance, insight, and support during this research and the preparation of this thesis. I would also like to gratefully acknowledge the supervision of my examiner Dr.Tech. Ari Visa.

My final words go to my parents. I want to thank my parents, whose love and guidance is with me in whatever I pursue.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

OPERATORS, SYMBOLS AND NOTATION

$*$	Convolution
$\times$	Multiplication
$\|...\|$	Euclidean distance between two points in space or absolute value
$\sum_a^b$	Sum from a to b
α	Absorption coefficient of a reflective surface
$\delta(.)$	Dirac delta function
ρ_0	Static gas density
λ	Wavelength
ζ	Specific acoustic impedance
c	Speed of sound
dB	Decibel, a logarithmic unit expressing the magnitude of a quantity
F_{lim}	Nyquist frequency, which is equal to $F_S/2$
F_S	Sampling frequency
Hz	Hertz = number of cycles per second, unit of frequency
m	meters
ms	milliseconds
$\mathbf{m}$	Location of microphone
p	Sound pressure
R	Reflection Coefficient
$s(t)$	Signal emitted by the sound source $\mathbf{s}$
$\mathbf{s}$	Location of sound source
S_V	Location of virtual sound source
T_{60}	Time passed in seconds until the magnitude of echoing has decayed below 60dB
v	Particle velcoity
$z(t)$	Signal recorded by the microphone
Z	Wall Impedance

LIST OF ABBREVIATIONS

DOA	Direction of Arrival
IR	Impulse Response
LTI	Linear Time Invariant
MLS	Maximum Length Sequence
RIR	Room Impulse Response
SNR	Signal to Noise Ratio
TF	Transfer Function

1. INTRODUCTION

The Acoustical parameters of a room are uniquely identified by its room impulse response, which gives parameters such as temporal, energetic, spatial and quantitative speech intelligibility. These parameters vary on the size, compositions and the *materials* used in the room. Conversely, it should be possible to determine the type of *materials* found in the room using the room impulse response. Presented in this thesis is a study on how the material properties such as reflection and absorption coefficients can be determined from an impulse response; and how well this information can be deduced from a general room impulse response.

There are three basic methods of measuring Reflection coefficients: (i) reverberation room, (ii) standing wave tube, and (iii) reflection.

The *reverberation room* method [14] is still the reference for many authors, but needs expensive facilites and is affected by unavoidable errors[15], because the assumptions made in the derivation of the reverberation formula are never achieved in practice.

The *wave tube* method has many variants: pure tones [16] or broad-band random noise [17] can be used as input signal; one[18], two[19] or more [20, 21] microphones, fixed or moving, can be used to pick up the sound pressure in the tube; finally, many processing techniques are possible[22, 23].

Almost all these tube methods can be slightly modified to measure the assumption that places an upper limit on the frequency range. They are also extremely sensitive to the mounting of the test materials[24]. Furthermore, only small samples can be tested, and hence effects which depend on the dimensions and edge conditions of the materials will not be observed correctly. Samples with large resonant cavities cannot be fitted to the tube.

Often wave tube results are transformed in reverberation room values; such transfomation requires several assumptions[25].

The reverberation room and the wave tube methods are essentially laboratory methods, not suitable for measurements on the reflecting surfaces in realistic conditions, i.e., *in situ*.

True in situ measurements can be devised by starting from what can be called *reflection methods*[39], because the reflection coefficient is deduced by detecting the acoustic signals impinging on, and reflecting from the test surface.

At first, Ingards and Bolt[26] and Ando[27] used this basic scheme with pure tones in an anechoic room. Then a number of authors investigated pulsed techniques which, in principle, allow the separation of the reflection from the test surface from other unwanted reflections, like those existing in closed rooms: Kinstl[28] used a spark source, Yuzawa[29] used a tone burst of $10ms$, Davies and Mulholland[30] used a recorded impulse, and Cramond and Don[31] used blank shots. Problems of impulse repeatability, non-linearity and difficult signal processing have till now prevented a general acceptance of this basic impulse method.

Hollin and Jones[32] showed that the application of a correlation technique between the signal received by the microphone and the noise fed into the loudspeaker allows the substitution of the pulsed signal with stationary broad-band random noise and a relative background noise immunity. Bolton and Gold[33] applied cepstral analysis to extract the impulse response of the surface being tested, even in enclosed spaces; their technique can improve the method to a great extent, but it requires complex mathematical and computational procedures.

More elaborate techniques exist, like Tamura's decompositions in plane waves[34] of the complex pressure distributions on two parallel planes close to the test surface by using two-dimensional Fourier transforms, or like the comparision suggested by Allard and Champoux[35], of the data given by a two-microphone probe with the values obtained throught an iterative computation of the complex Nobile-Hayek integral for a point source over an impedance plane.

In principle, indirect methods of characterising porous reflecting surfaces and an impedance model of the surface material[36] could also be rearranged to obtain the sound-reflection coefficient, although they imply a considerable work of least-squares fitting and of course rely on the assumption of a more or less general impedance model.

Finally, sound intensity techniques could, in principle, be used to measure the sound-reflection coefficient of a surface, although the usefulness of this method was proved only for frequencies greater than about $500\,Hz$ and after a careful tuning of the experimental set-up to the local acoustic field[37].

The above-mentioned reflection methods are well suited for free-field measurements; most of them work properly also in enclosed spaces, and oblique incidence. Nevertheless, they adopt different input signals and different processing techniques, sometimes not so easy to implement in the field, while little attention has till now been devoted to potential practicalities of the obtained reflection coeffecients. Here is an approach in which the the so obtained reflection coeffecients have been used to identify the type of surface; and classify the same to estimate the different surfaces of a given enclosed room.

1.1 Thesis Outline

This thesis is divided in five chapters. Chapter 2 deals with the theoretical background material required for building this thesis. This defines the sound source & measurement, reflection factor and states the assumptions made henceforth. Chapter 3 presents the selected methods that are utilized to obtain the impulse response from the test material. And further analysis methods on the so obtained impulse response. In the fourth chapter, the results of the proposed methods have been presented and have been compared with existing methods. In addition, the precision of results were calculated and has been evaluated in an elaborative manner. Finally, the implementation process, practical usage with possible pros and cons, and future work is discussed in chapter 5.

2. THEORETICAL BACKGROUND

2.1 Sound

Sound can be defined as the kinetic energy carried by atoms or molecules of the medium, as a travelling wave. Although the kinetic energy is carried through movement and interaction of individual particles, the medium does not move with the wave. The medium of sound propagation itself defines a lot of activities like reflection, refraction and attenuation. This can be due to the type of medium it is. Sound is known to travel faster in solids than in fluids. This is because of the fact that molecules of a solid medium are much closer together than those in a liquid or gas, allowing sound waves to travel more quickly through it. Even within the same medium the speed of propagation is known to vary depending on the temperature of the medium. And hence we come to our first assumption:

Assumption 1. The medium is homogeneous and lossless.

That is during the experimental tests, the arrival time of the wave front depends only on the length of propagation path.

The occurence of a physical event starts a chain of interactions in the particles of the surrounding medium that expands away from the initial event. This event, or a sound source can be about anything: the mechanical impact of two objects, turbulence from interaction with a medium or voices created by human speech organs.

2.2 Sound Source

Sound sources, whether human speech or devices, are complex mechanisms which possess certain attributes. In real world sound sources are usually directional. For example in the human speech, the sound is produced as an action of the pulmonary pressure provided by the lungs which creates sound in the glottis, which further is modified by the vocal tract to produce speech. And thus source of sound cannot be treated as arising from a point like source. Moreover, because of the anatomy of the human head, the pressure of sound waves during speech may be 15dB lower behind the head than in front of the mouth [1]. Directivity of the source can be expressed as a function of angle. A polar plot of various directivity patterns are shown in figures (2.2). For speech sounds the vertical directivity pattern is mostly cardoid to semi-cardoid for frequencies above 1000 Hz and below this frequency the pattern

is from semicardioid to omnidirectional[2, 3]. Sound sources such as loudspeakers also have frequency ranges and directivity patterns that depend on the desing of the source. At this point we have two more assumptions:

Assumption 2. The sound source is a pointlike source, whose location can be represented in three dimensional Cartesian coordinates (x, y, z).

This ensures that the location of the sound source can be expressed as a single point in cartesian coordinate system, rather than a volume, which resembles the real world situation, but would be much difficult for analysis and estimation.

Assumption 3. The sound source is omnidirectional.

This is made because the proposed method does not consider the sound pressure level while estimating the transfer function.

Consider a sound source in a two dimensional geometry and a linear microphone array consisting of N acoustic sensors. The geometry of the array is represented by the sensor postions $(k_0(x_o, y_o), ..., k_{N-1}(x_{N-1}, y_{N-1}))$. We assume that the source located at the position (x_s, y_s) in figure (2.1) generates an acoustic event $s(t)$ which is recorded by the N microphones as signals $s_0(t), .., s_{N-1}(t)$.

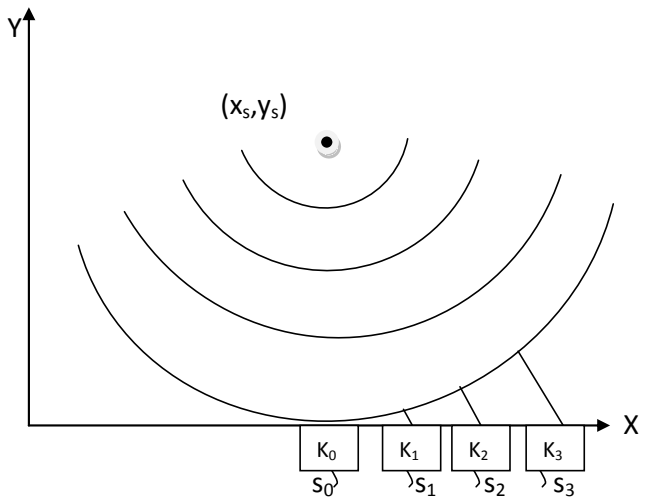


Figure 2.1: Wavefront propagation of an acoustic stimulus generated in position (x_s, y_s). Signals s_0, s_1, s_2, s_3 are acquired through an array of microphones placed in positions k_0, k_1, k_2, k_3.

For the given source signal $s(t)$, propagated in a generic noisy environment. The signal acquired by the acoustic sensor i , can be expressed as follows :

$$s_i(t) = \alpha_i * s(t - \tau_i) + n_i(t) \tag{2.1}$$

where α_i is the attenuation factor due to propagation effects, τ_i is the propagation time and n_i includes all the contaminating noises, which are assumed to be

uncorrelated with $s(t)$ and between microphones. We also indicate with $\delta_{ij}(x, y)$ the relative delay of wavefront arrival between microphones i and j, assuming a source in position (x, y), where :

$$\delta_{ij} = (\tau_j - \tau_i) \tag{2.2}$$

When the microphones array is kept quite far from the sound source, the wave fronts impinging the microphones can be considered planar and hence $\delta_{ij} = 0$(Considering microphone is facing the sound source.), this situation is called as far field. When this is not the case, that is, when the spherical curvature of the wave front can be detected within the microphone array's aperture it is called the near-field. It is an obvious conclusion from this that, in the case of near field the sound source is close enough to the microphone arrays. And hence we define our last assumption:

Assumption 4. The sound source is assumed to be near-field.

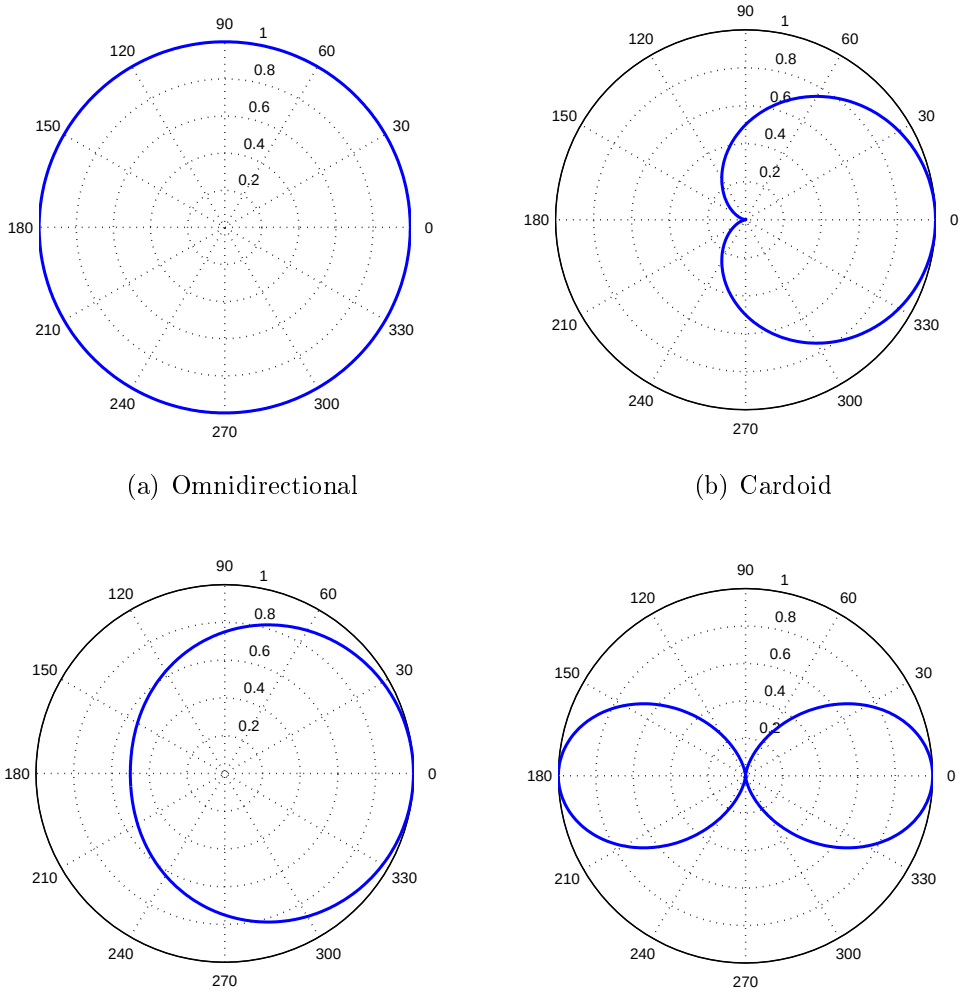
2.3 Sound Measurement

Measurement is the process of connecting a theoretical model to the real world. For a sound wave it is the process of measuring the change in pressure or even the mean velocity of particles. The devices which measure sound are called microphones.

The principle of a basic pressure microphone has remained the same since the very first prototype by Graham Bell in 1876. Changes of the pressure around a diaphragm (a moving membrane), causes a displacement which is converted to a change of electrical current or capacitance [4].

This continuous analogical feed is converted into discontinous digital representation. Such a conversion makes the signal handling easy, since it induces quantization to both time and the amplitude. The time quantization is generally expressed as a sampling frequency F_s. This gives the amount of samples in the data per second($1/s = Hz$). The chosing of sampling frequency also binds us to an upper frequency limit for the feed at $F_{lim} = F_s/2$, which is known as the Nyquist frequency. Frequencies in the feed which are higher than the nyquist freqency are bound to get aliased. And hence it is ideal to filter the feed with a low pass filter of cutoff F_{lim}, before the digitization. Similar to sound sources, even microphones have directivity. This dictates their sensitivity to the direction the sound is approaching. Directivity pattern depends on the design of the microphone and some of them have been shown in figures (2.2).

All microphones have characteristic amplitude and phase response. Amplitude response states the sensitivity for each frequency. Ideally it has to be uniform within the range of interest. Phase response, on the other hand indicates the phase shift for each frequency and ideally there should not be any phase shift.



(a) Omnidirectional

(b) Cardoid

(c) Subcardoid

(d) Polar or Dipole

Figure 2.2: Different polar patterns. A microphone's directional or polar pattern indicates how sensitive it is to sounds arriving at different angles about its central axis. The polar patterns illustrated above represent the locus of points that produce the same signal level output in the microphone if a given sound pressure level is generated from that point. The concentric circles represent the normalized scale of attenuation, and the circumference of the circle represent the direction of sound arrival in degrees.

Microphones are very sensitive devices with addition to the signal of interest, it is also susceptible to measure noise caused by thermal or electromagnetic interference. The power ratio between a signal and noise is called signal-to-noise ratio (SNR)expressed in decibels.

$$SNR = 10 \log_{10} \left(P_{signal} / P_{noise} \right) \tag{2.3}$$

Where P_{signal} is the signal power and P_{noise} is the noise power.

2.4 Reflection factor, absorption coefficient and wall impedance

Consider the propagation of sound in an enclosed room. The sound conducting medium is bounded on all sides of wall, floor and ceiling. These room boundaries are not ideal surfaces, they absorb a portion of the sound energy impinging on them, and reflect the rest. The absorbed energy is either converted into heat, or transmitted to the outside of walls. These numerous reflected componets is responsible for what is known as *the acoustics of a room*.

Before discussing the properties of such involved sound fields, lets consider the fundamental for their occurences: the reflection of a plane sound wave by a single wall or a surface. Any free edge of a reflecting panel will scatter some sound energy in all directions. The same happens when the sound wave hits any obstacle with limited extent, such as a pillar or a listener's head. Henceforth we assume,

Assumption 5. Surface of reflection is unbounded.

In reality all waves originate from a sound source and are therefore spherical waves or superposition of spherical waves. The derivations for the reflection of a spherical wave from a plane wall is highly complicated unless we assume that the wall is rigid. For the present discussion the sound source is assumed to be not too close to the reflecting wall so that the curvature of the wave fronts can be neglected without too much error.

When a plane wave strikes a plane and uniform wall of infinite extent, generally a part of the sound energy will be reflected in the form of a reflected wave originating from the wall, the amplitude and the phase of which is different from the original incident wave. Both these waves interfere with each other and form the *standing waves*, atleast partially.

The change in amplitude and phase which take place during the reflection of a wave can be expressed by a complex reflection factor,

$$R = |R| \exp(i\chi) \tag{2.4}$$

which is a property of the reflecting surface. Its absolute value and the phase angle depends on the frequency and the angle of incidence of sound wave.

Another important quantity is sound intensity, which is a measure of the energy transported in a sound wave. It can be defined as the energy of a sound wave passing perpendicularly through a $1\,m^2$ window, per second. Generally the intensity is a vector parallel to the vector $\mathbf{v}$ of the particle velocity and is given by

$$\mathbf{I} = p\mathbf{v} \tag{2.5}$$

In a plane wave the sound pressure p and the longitudinal component of the particle velocity are related by $p = \rho_0 c v$, and the same holds for a spherical wave at a large distance from the centre. Hence we can express the particle velocity in terms of the sound pressure. Then the intensity is given by

$$\mathbf{I} = \frac{p^2}{\rho_0 c} \tag{2.6}$$

Thus the intensity of a plane wave is proportional to the square of the pressure amplitude. Therefore the intensity of the reflected wave is smaller by a factor of $|R|^2$ than the incident wave and the fraction $1 - |R|^2$ of the incident energy is lost during the reflection. This quantity is called as the *absorption coefficient* of the wall:

$$\alpha = 1 - |R|^2 \tag{2.7}$$

For a wall with zero reflectivity ($R = 0$) the absorption coefficient has its maximum value of 1. Then the wall is said to be totally absorbent or *matched to the sound field*. If $R = 1$ (in phase reflection $x = 0$), the wall is *rigid* or *hard*; in the case of $R = -1$ (phase reversal $x = \pi$), we speak of a *soft* wall. In both cases there is no sound absorption($\alpha = 0$). Soft walls, however are very rarely encountered in room acoustics and only in limited frequency ranges.

Another quantity which is closely related to the physical behaviour of the wall is based on the particle velocity normal to the wall which is generated by a given sound pressure at the surface. It is called the wall impedance,

$$Z = \left(\frac{p}{v_n}\right)_{surface} \tag{2.8}$$

where p denotes the sound pressure and v_n denotes the velocity component normal to the wall. Like, the reflection factor, the wall impedance is generally complex and a function of the angle of sound incidence.

Another common term is the *specific acoustic impedance*, which is the wall impedance divided by the characteristic impedance of the air :

$$\zeta = \frac{Z}{\rho_o c} \tag{2.9}$$

where ρ_o is the static value of the gas density and c is the velocity of sound.

The reciprocal of wall impedance is the *wall admittance* ; the reciprocal of ζ is called *specific acoustic admittance* of the wall.

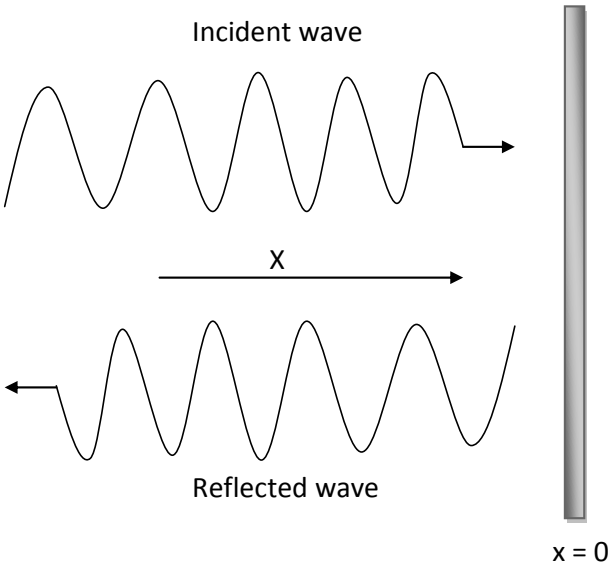


Figure 2.3: Reflection of a normally incident sound wave from a plane surface[3].

2.4.1 Sound Reflection at normal incidence

Consider a plane wall intersecting a sound wave along x-axis perpendicularly. The wall intersects the x-axis at x=0 as showin in figure (2.3). The wave is coming from the left and is sound pressure is given by,

$$p_i(x,t) = \hat{p_o} \exp[i(\omega t - kx)] \tag{2.10}$$

and the particle velocity of the incident wave is given by

$$v_i(x,t) = \frac{\hat{p_o}}{\rho_o c} \exp[i(\omega t - kx)] \tag{2.11}$$

The reflected wave has a smaller amplitude and a different phase, which are described by the reflection factor R of the wall. Furthermore the sign of k is changed as the travelling direction of sound wave is reversed. The sign of the particle velocity is also changed since p/v has opposite signs for positive and negative travelling waves.

The so obtained reflected wave is

$$p_r(x,t) = R\hat{p_o}\exp[i(\omega t + kx)] \tag{2.12}$$

$$v_r(x,t) = -R\frac{\hat{p_o}}{\rho_o c}\exp[i(\omega t + kx)] \tag{2.13}$$

The total pressure and particle velocity at the surface of wall is given by substituting $x = 0$ (point of intersection of wall and sound wave) in the overall equation obtained by adding the above expressions.

$$p(0,t) = \hat{p_o}(1+R)\exp[i\omega t] \tag{2.14}$$

$$v(0,t) = \frac{\hat{p_o}}{\rho_o c}(1-R)\exp[i\omega t] \tag{2.15}$$

Now the wall impedance is obtained by,

$$Z = \rho_o c\frac{1+R}{1-R} \tag{2.16}$$

and from this

$$R = \frac{Z-\rho_o c}{Z+\rho_o c} = \frac{\zeta - 1}{\zeta + 1}. \tag{2.17}$$

A rigid wall ($R = 1$) has impedance $Z = \infty$; for a soft wall ($R = -1$) the impedance will vanish. For a completely absorbent wall the impedance equals the characteristic impedance of the medium.

Inserting 2.17 into the definition of absorption coefficient , we get

$$\alpha = \frac{4Re(\zeta)}{|\zeta|^2 + 2Re(\zeta) + 1} \tag{2.18}$$

The combination of the incident and reflected wave form the standing wave. The pressure amplitude of which is found by adding the equations (2.10)& (2.12), and evaluating the absolute value:

$$|p(x)| = \hat{p_o}[1 + |R|^2 + 2|R|\cos(2kx + \chi)]^{1/2}. \tag{2.19}$$

Similarly for the amplitude of particle velocity we find

$$|v(x)| = \frac{\hat{p_o}}{\rho_o c}[1 + |R|^2 - 2|R|\cos(2kx + \chi)]^{1/2}. \tag{2.20}$$

From equations (2.19) & (2.20) we can deduce that the pressure amplitude and velocity amplitude in the standing wave vary periodically between the maximum

values

$$p_{max} = \hat{p_o}(1 + |R|) \quad and \quad v_{max} = \frac{\hat{p_o}}{\rho c}(1 + |R|)$$

and the minimum values

$$p_{min} = \hat{p_o}(1 - |R|) \quad and \quad v_{min} = \frac{\hat{p_o}}{\rho c}(1 - |R|)$$

in such a way that each of the maximum of the pressure amplitude coincides with

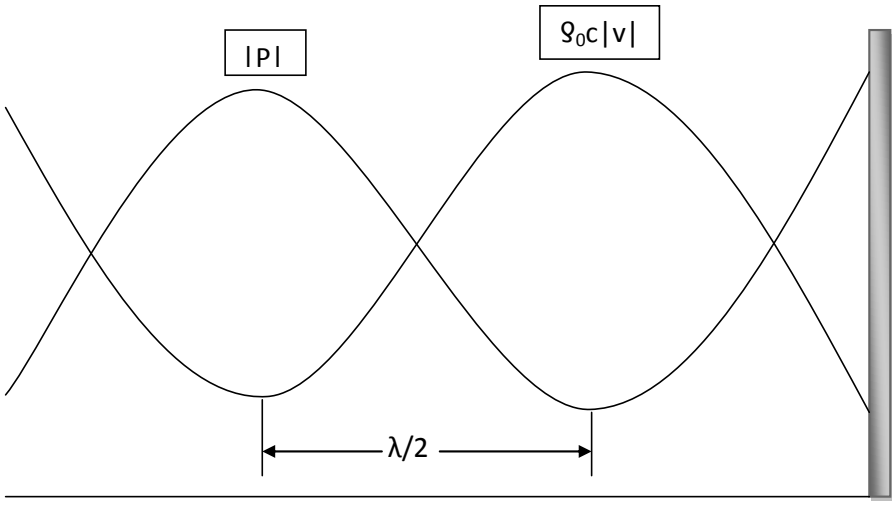


Figure 2.4: Standing sound wave in front of a plane surface with the real reflection factor $R = 0.7$; $|P|$ is the magnitude of sound pressure and $\rho_0 c|v|$ is the magnitude of particle velocity[3].

a minimum of the velocity amplitude and vice versa, as shown in figure (2.4). The distance between two consecutive peaks is $\pi/k = \lambda/2$. On measuring the pressure amplitude as a function of x, we can determine the wavelength. Furthermore, the absolute value and the phase angle of the reflection factor can also be evaluated. This is the basis of a standard method of measuring the absorption coefficient of wall materials in the tube method [10].

2.4.2 Sound Reflection at oblique incidence

In this section we consider a much generalized case, in which the angle of incidence theta may be any value between 0 degrees to 90 degrees. Consider a case in which the wave normal and wall normal lie on the $x - y$ plane. The setup is depicted in figure (2.5).

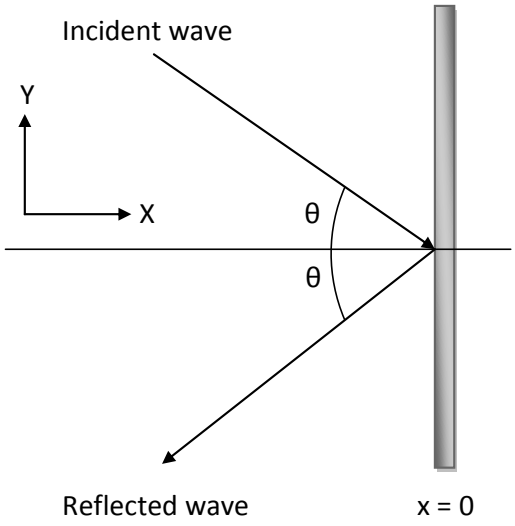


Figure 2.5: Sound reflection at oblique incidence[3].

Now the pressure equation for this system is given by replacing x in equation (2.10) with x', where

$$x' = x\,\cos\theta + y\,\sin\theta \tag{2.21}$$

$$p_i(x,t) = \hat{p_o}\,\exp[-ik(x\,\cos\theta + y\,\sin\theta)]. \tag{2.22}$$

In the above equation and the following equations the factor $exp(iwt)$ has been omitted for the sake of simplicty, since this is common to all pressures and particle velocities. For the calculation of wall impedance the x-component of velocity which is normal to the wall is required. Which is given by,

$$v_x = -\frac{1}{iw\rho_o}\frac{\partial p}{\partial x} \tag{2.23}$$

applied to eqn. (2.22) we get,

$$(v_i)_x = \frac{\hat{p_o}}{\rho c}\,\cos\theta\,\exp[-ik(x\,\cos\theta + y\,\sin\theta)] \tag{2.24}$$

On reflection of this wave from the wall, the sign of x in the exponent of eqns. (2.22) and (2.24) is reversed, since the direction with respect to this axis is altered. Furthermore, the pressure and the velocity are multiplied by the reflection factor R and $-R$, respectively:

$$p_r(x,t) = R\hat{p_o}\,\exp[-ik(-x\,\cos\theta + y\,\sin\theta)] \tag{2.25}$$

$$(v_r)_x = \frac{-R\hat{p_o}}{\rho c}\,\cos\theta\,\exp[-ik(-xcos\theta + ysin\theta)] \tag{2.26}$$

The wall impedance is given by,

$$Z = \frac{p_i + p_r}{v_i + v_r} = \frac{\rho_o c}{\cos\theta}\frac{1+R}{1-R} \tag{2.27}$$

and reflection factor is given by,

$$R = \frac{Z\cos\theta - \rho_o c}{Z\cos\theta + \rho_o c} = \frac{\zeta\cos\theta - 1}{\zeta\cos\theta + 1} \tag{2.28}$$

The pressure amplitude in front of the wall resulting from the addition of equations (2.22) and (2.25) is

$$|p(x)| = \hat{p_o}[1 + |R|^2 + 2|R|\,\cos(2kx\cos\theta + \chi)]^{1/2} \tag{2.29}$$

This again corresponds to the standing wave, whose maxima's are separated by a distance $(\lambda/2)\,\cos\theta$. The common factor $\exp(-iky\,\sin\theta)$ in equations (2.22) and (2.25) tells that the pressure distribution moves parallel to the wall with a velocity

$$c_y = \frac{\omega}{k_y} = \frac{\omega}{k\sin\theta} = \frac{c}{\sin\theta} \tag{2.30}$$

And the absorption coefficient is given by

$$\alpha = \frac{4Re(\zeta)\,\cos\theta}{|\zeta|^2\cos^2\theta + 2Re(\zeta)\cos\theta + 1} \tag{2.31}$$

2.5 Small Room Acoustic

So far we considered the propagation of sound in an unbounded environment. In contrast to this, room acoustics deals with the propagation and behaviour of sound in an enclosed environment; whose boundaries like walls, ceiling and floors are sound-conducting medium. Usually, the sound signals sensed by the microphone are a sum of reflected and delayed secondary singals and a primary signal propagated directly from the sound source. These distortions contain all the information about the distance of the reflective surface and the type of material which has reflected it. Much about this is discussed in coming chapters. In the following subsections, the phenomenon caused by the surroundings and the methods for estimating the room characteristics are considered.

2.5.1 Reverberation

The Reverberation time is defined as the time taken for an attenuation of 60dB of sound energy from the instant the sound source is turned off. This rate of sound decay is related to the volume of the space and the amount of absorption from the reflecting surfaces and the medium.

Rooms with short reverberation times enhances the speech intelligibility and are ideal for conferencing. Conversely, rooms with longer reverberation times is ideal for concerts as the sound persists in the medium for longer time and hence it sustains musical qualities like richness and fullness. The International Standards Organization has maintained different guidelines for acoustic measurements for ordinary rooms [6] and performance spaces [5].

2.5.2 Image Source Method

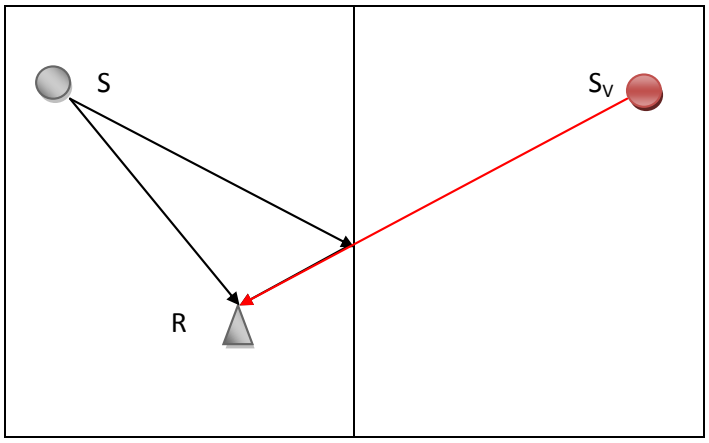


Figure 2.6: Image source method -Single reflection. The reflected path between the source S and reciever R, can also be percieved as sound coming from another virtual source S_V from the other side of the wall(sound propagation shown in red line).

The image source method, coined by Allen and Berkley, is one of the most used room acoustic model. It introduces the concept of virtual sound sources to model the reflections in a simple, shoebox environment [7]. Within the enclosure - the walls, floor and ceiling behave as the acoustic reflectors, and the sound wave arrives to the microphone using many different propagation paths enabled by these reflections. The image source method allows fast and simple calculation of a box like environment even with large number of reflections.

Consider a source S and a reciever R located within a rectangular room as shown in figure (2.6). The line between the source and the reciever is the path taken by the sound wave, which is called the direct sound. As we assumed in section 2.2 the

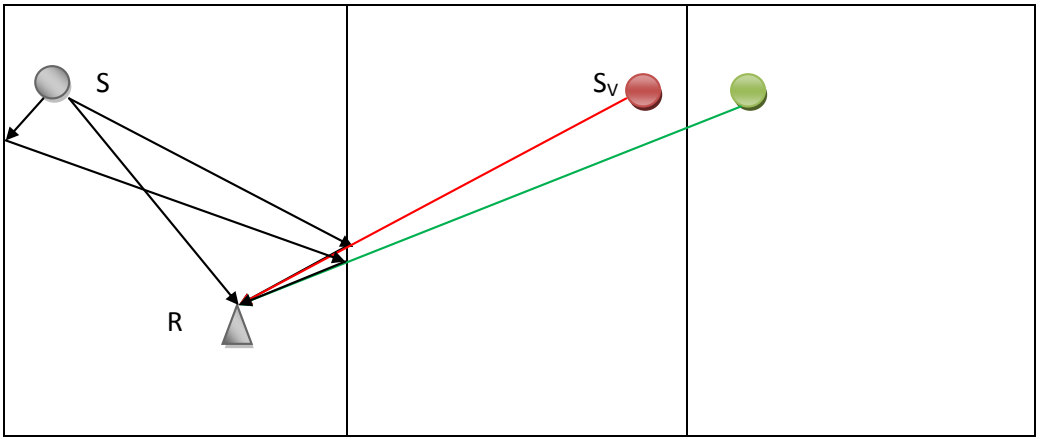


Figure 2.7: Image source method -Double reflection. The double reflected path between the source S and reciever R, can also be percieved as sound coming from a virtual source S_V from beyond two virtual walls, achieved by double flipping the room(sound propagation shown in green line).

sound source is a point-like source emitting spherical sound waves in all directions with equal power. Assuming a point like source, it can be stated that there are sound waves propagating not just directly towards the microphone but also away from the microphone. The sound waves eventually interact with walls and gets reflected off the wall and then impinges upon the microphone. This reflected wave will be referred to as an echo. The microphone perceives this echo as radiating from a point S_V past the wall as shown in the figure (2.6). If we were at the location of the microphone R, we could see the sound source as well as its mirror image. This mirror image will be referred to as a virtual source. It is also the location from which we will perceive the echo to be radiating.

This process can be repeated by making a mirror image of the room's mirror image. Each mirror image of the source represents another virtual source. A diagram of the real sound source along with two virtual sources is shown in the figure (2.7). Furthermore, it can be extended into two dimensions. A map of the virtual sources is illustrated in the figure (2.8).

2.5.3 Room impulse response model

Acoustic spaces are approximately linear-time-invariant(LTI) and generally passive systems. A linear system maps an input to an output using only linear operations. For example, if an input, $x_1(t)$ produces the output $y_1(t)$ and another input, $x_2(t)$ produces the output $y_2(t)$, the input $a_1x_1(t) + a_2x_2(t)$ will always produce $a_1y_1(t) + a_2y_2(t)$. Time invariance means that the system response does not alter with time.

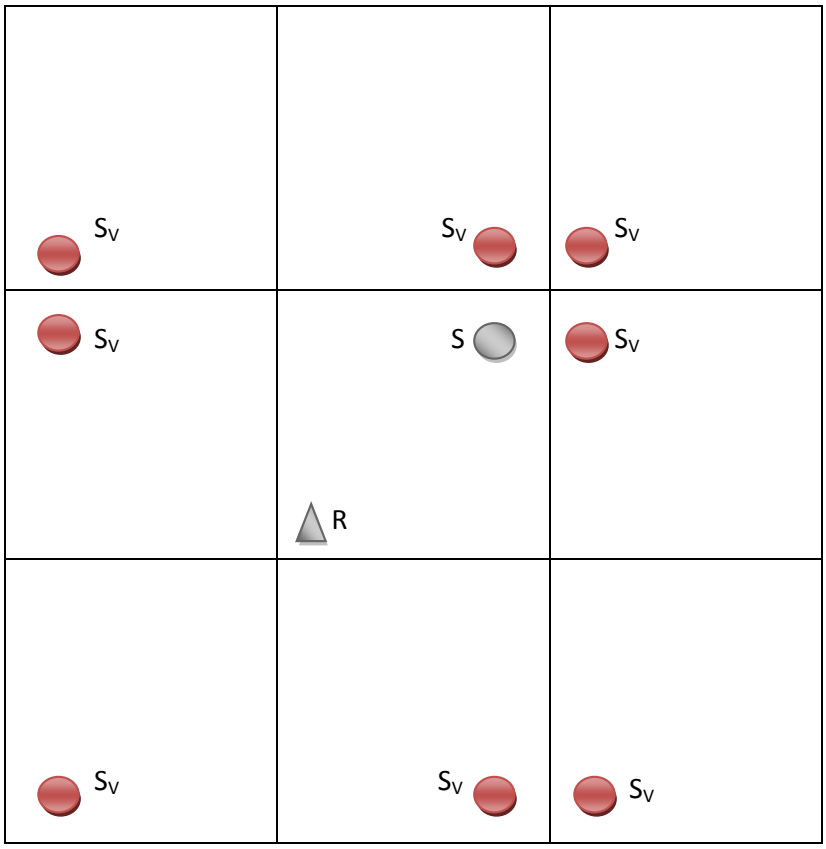


Figure 2.8: Image source method - A map of virtual sources.

As a result, if the input signal is delayed by a certain amount the output signal will be delayed by that same amount. The room response to a sound source is pictured in figure (2.9).

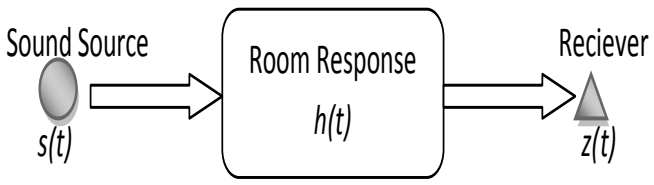


Figure 2.9: Effect of room response on a sound source.

The room response can be quantified theoretically by its response to a Dirac delta function, this is known as the system's impulse response, h(t). A Dirac delta function is a short pulse of energy at time zero which is infinitesimally narrow and its integral of all time is unity. The response of the room to a sound source is the convolution of the impulse response, h(t), with the source signal s(t);

$$z(t) = h(t) * s(t) = \int_{\infty}^{\infty} h(\tau)s(t-\tau)d\tau \tag{2.32}$$

This process has the useful property that when the system is described in the frequency domain, the convolution operation becomes a multiplicative one

$$Z(f) = H(f)S(f). \tag{2.33}$$

Where $S(f), Z(f)$ and $H(f)$ are the Fourier transforms of $s(t), z(t)$ and $h(t)$ respectively. The Fourier transform of $h(t), H(f)$ is known as the Room impulse response(RIR), which can be solved by linear deconvolution

$$H(f) = \frac{Z(f)}{S(f)}. \tag{2.34}$$

The transfer characteristcs of a room are completely described, for a single source and listener orientation, by the room impulse response(RIR). This assumes linearity and time invariance, however the RIR is known to change with conditions within the space such as temperature (but this generally happens slowly over long periods of time) and often time-invariance is assumed. Public address systems can introduce significant levels of distortions but the RIR is a measure of only the passive system response and does not include any active systems that may be in use.

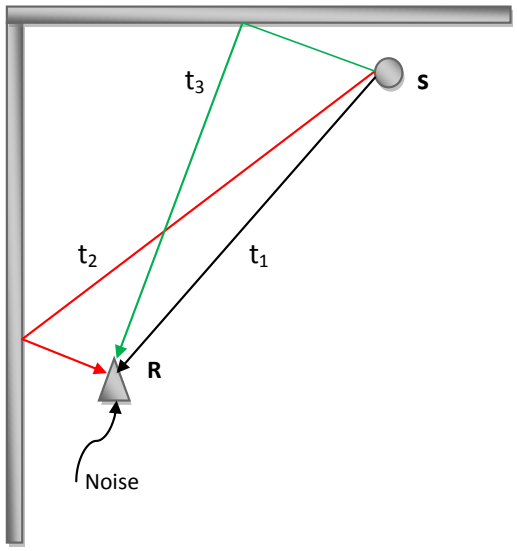
Based on the assumptions of the image source model, the reciever signal is the sum of shifted and scaled versions of the source signal $s(t)$:

$$z(t) = \alpha_1 s(t-t_1) + \alpha_2 s(t-t_2) + ... + \alpha_N s(t-t_N) + n(t) \tag{2.35}$$

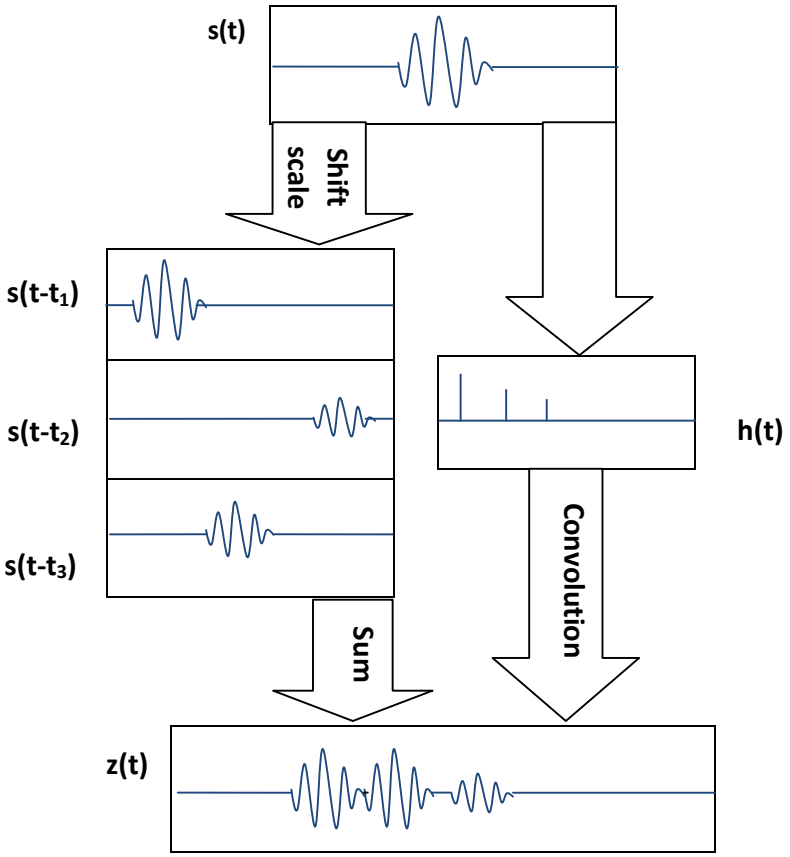
$$= \sum_{i=1}^{N} \alpha_i s(t-t_i) + n(t) \tag{2.36}$$

where N is the number of distinct components, t_i and α_i are the delay and the attenuation associated with each component $i \in [1, ..., N]$. The term $n(t)$ represents the noise and is assumed to have a zero mean and be independent of the signal. The attenuation term α is assumed to account for both inverse-range dependent spreading and absorption due to reflections from surfaces and it is here frequency independent. An example of the propagation paths and the resulting signal components is depicted in figure(2.10).

The impulse response of the channel between the sound source and the receiver is best estimated by using a short impulse-like excitation signal approximating Dirac delta function $\delta(n)$, such as a balloon pop or a clap of hands, and measuring the resulting signal from a distance. However, the reliability of such a measurement is poor because of the nature of the excitation signal and the signal-to-noise ratio (SNR) is low due to the short length and finite amplitude of the excitation signal.



(a) Propagation



(b) Signal model

Figure 2.10: Simplified multipath signal model. (a) The signal measured at sensor position **R** consists of delayed multipath components, each with own delay t_i. The superimposed image is a mockup of the signal as a traveling wave at the moment it reaches the sensor. (b) The measured signal $z(t)$ is defined as sum of shifted and scaled versions of the source signal $s(t)$ [40]. It is equivalently represented as a signal convolution with the impulse response $h(t)$.

Other methods have been developed in order to overcome these problems, such as the so called maximum length sequence(MLS) method [8]. In this method a pseudo-random signal is used as excitation, and the signal is measured with only one microphone. The impulse response estimation is based on auto-correlation function of the MLS, and if the system is considered linear and time-invariant(LTI), it is shown that the cross-correlation between the time reversed excitation and measured signal is the impulse response of the system. Because of the cross-correlation, most of the non-correlated noise will be rejected, thus improving the SNR of the measurement. The problem with this method is that it is sensitive to non-linearities. For example, if the exictation signal is played through a loudspeaker, the harmonic distortions caused by the loudspeaker appear as spurious peaks in the impulse response[13].

The distortions caused by the loudspeaker can be cancelled out, if swept sine technique pioneered by Angelo Farina is used [9]. In this method, a logarithmically rising frequency sweep is used as excitation, and the distortion harmonics are generated ahead of the each frequency appearing in the signal. Farina showed, that after deconvolution as shown in equation (2.34), the impulse responses caused by the harmonic distortions are temporally separated and can be edited out. In the presented thesis, experiments have been carried out using both MLS and swept sine technique. Though the resulting difference between the two input signals does not vary much for our setup. The figures and results discussed in this thesis are of swept sine technique.

2.5.4 Transfer Function Measurement with Sweeps

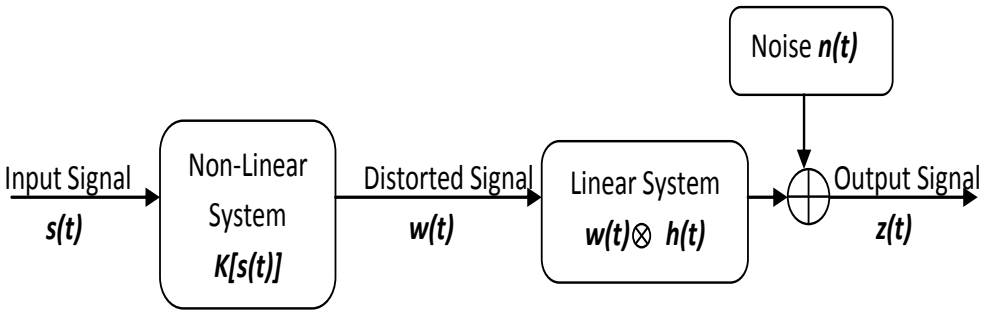


Figure 2.11: Theoretical block diagram of the impulse response measurement system including the loudspeaker (considered as a non-linear element) and the acoustical space (considered as a perfectly linear system).

Compared to using pseudo noise signals, transfer function measurements using sweeps as excitation signal show significantly higher immunity agains distortion and time variance. Capturing binaural room impulse responses for high-quality auralization purposes requires a signal-to-noise of greater than 90dB that is unattainable

with MLS-measurements due to loudspeaker non-linearity, but fairly easy to reach with sweeps due to the possibility of completely rejecting harmonic distortion [11].

The sinesweep technique developed by Farina [12] is based on the following idea: by using an exponential time-growing frequency sweep, it is possible to simultaneously deconvolve the linear impulse response of the system and to selectively separate each impulse response corresponding to the harmonic distortion orders considered. The harmonic distortions appear prior to the linear impulse response. Therefore, the linear impulse response measured is assured exempt from any non-linearity and, at the same time, the measurement of the harmonic distortion at various orders can be performed.

Figure (2.11) illustrates the black box modelization of the measurement process. In this modelization it is assumed that the measurement system is intrinsically not linear but, on the other hand, perfect linearity is considered regarding the acoustical space from which the impulse response is to be derived.

As pointed out by Farina[12], the signal emitted by the loudspeaker is composed of harmonic distortions(considered here without memory) and may be thus represented by the following equation(see figure(2.11)).

$$w(t) = s(t) \otimes k_1(t) + s^2(t) \otimes k_2(t) + s^3(t) \otimes k_3(t) + ... + s^N(t) \otimes k_N(t) \quad (2.37)$$

where $k_i(t)$ represents the i^{th} component of the volterra kernel [12] which takes into account the non-linearities of the measurement system.

In practice, it is relatively difficult to separate the linear part (reverberation part in the impulse response) from the non-linear part (distortions). In the following, we will consider the response of the global system(the output signal from the system represented in figure(2.11)) as being composed of an additive gaussian white noise $n(t)$ and a set of impulse response $h_i(t)$, each of them being convolved by a different power of the input signal

$$z(t) = n(t) + s(t) \otimes h_1(t) + s^2(t) \otimes h_2(t) + s^3(t) \otimes h_3(t) + ... + s^N(t) \otimes h_N(t). \quad (2.38)$$

Where $h_i(t) = k_i(t) \otimes h(t)$.

Equation (2.38) underlines the existence of the non-linearities at the system's output.

In the case of the logarithmic sinesweep technique, the excitation signal is generated on the basis of the gollowing equation:

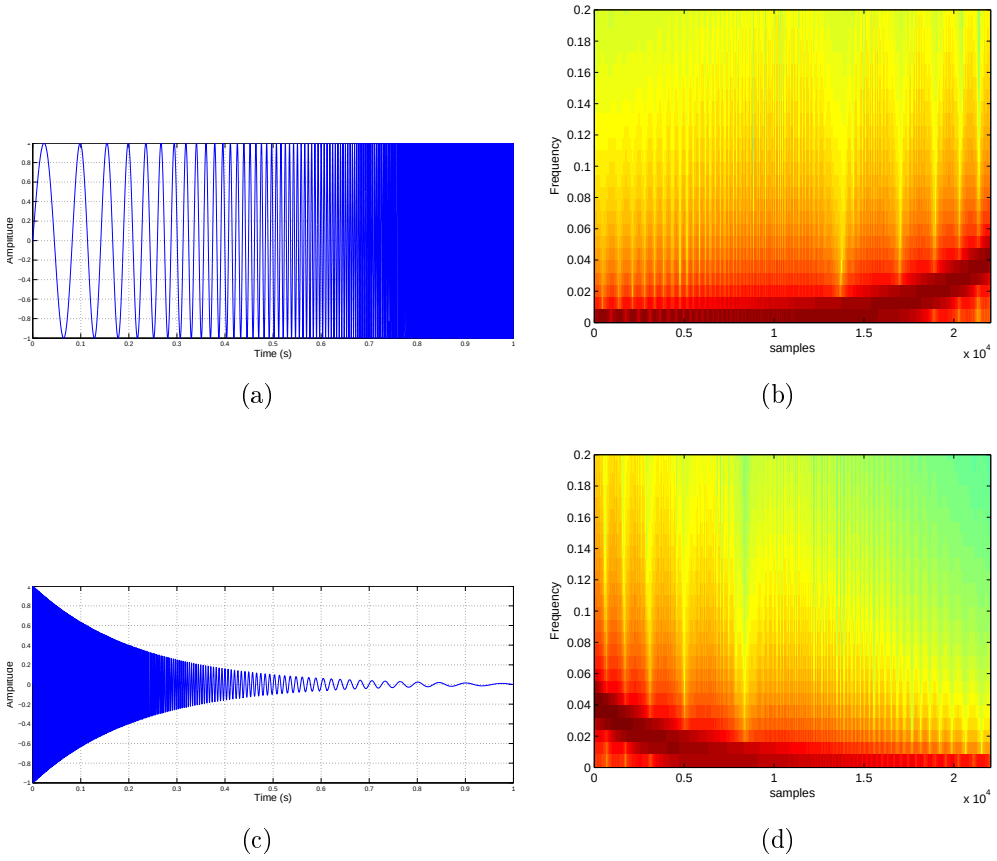


Figure 2.12: (a) Time representation of a Sine Sweep excitation signal ($w_1 = 2\pi 10\, rad/s$ and $w_2 = 2\pi 1000\, rad/s$). (b)Corresponding Magnitude spectrum, (c) Time representation of the inverse filter corresponding to the Sine Sweep signal presented in (a). (d) Corresponding Magnitude spectrum.

$$s(t) = \sin\left[\frac{T\omega_1}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\left(e^{\frac{t}{T}ln\left(\frac{\omega_2}{\omega_1}\right)} - 1\right)\right] \tag{2.39}$$

where ω_1 is the initial radian frequency and ω_2 is the final radian frequency of the sweep of duration T.

Figure (2.12(a)) (2.12(b)) shows the time and spectral representation of a logarithmic sweep with initial and final frequency at 10 Hz and 1000 Hz respectively.

The impulse response deconvolution process is realized by linear convolution of the measured output with the analytical inverse filter preprocessed from the excitation signal. Using linear convolution allows time-aliasing problems to be avoided. In fact, even if the time analysis window has the same length as the emitted sinesweep signal (and is shorter than the impulse response to be measured). The tail of the system response may be lost, but this will not introduce time aliasing.

In practice, a silence of sufficient duration is added at the end of the sinesweep signal in order to revocer the tail of the impulse response.

The deconvolution of the impulse response requires the creation of an inverse filter $f(t)$ able to "transform" the initial Sweep into a delayed Dirac's delta function:

$$s(t) \otimes f(t) = \delta(t - K) \tag{2.40}$$

The deconvolution of the impulse response is then realized by linearly convolving the output of the measured system $z(t)$ with this inverse filter $f(t)$:

$$h(t) = z(t) \otimes f(t) \tag{2.41}$$

The inverse filter $f(t)$ is generated in the following manner:

1. The logarithmic Sweep (which is a causal and stable signal) is temporally reversed and then delayed in order to obtain a causal signal (the reversed signal is pulled back in the poisitive region of the time axis). This time reversal causes a sign inversion in the phase spectrum. As such, the convolution of this reversed version of the excitation signal with the initial sinesweep will lead to a signal characterized by a perfectly linear phase (corresponding to a pure delay) but will introduce a squaring of the magnitude spectrum.

2. The magnitude spectrum of the resulting signal is then divided by the square of the magnitude spectrum of the initial sinesweep signal.

The time and spectral representations of the inverse filter corresponding to the sinesweep in figure (2.12(a)) (2.12(b)) is given in figure (2.12(c)) (2.12(d)).

In order to minimize the influence of the transients introduced by the measurement system and appearing at the begining and the end of the emission of the excitation signal, the ends of the sinesweep signal are exponentially attenuated (exponential growth at the begining and exponential decrease at the end).

3. RESEARCH METHODS AND MATERIAL

3.1 System Overview

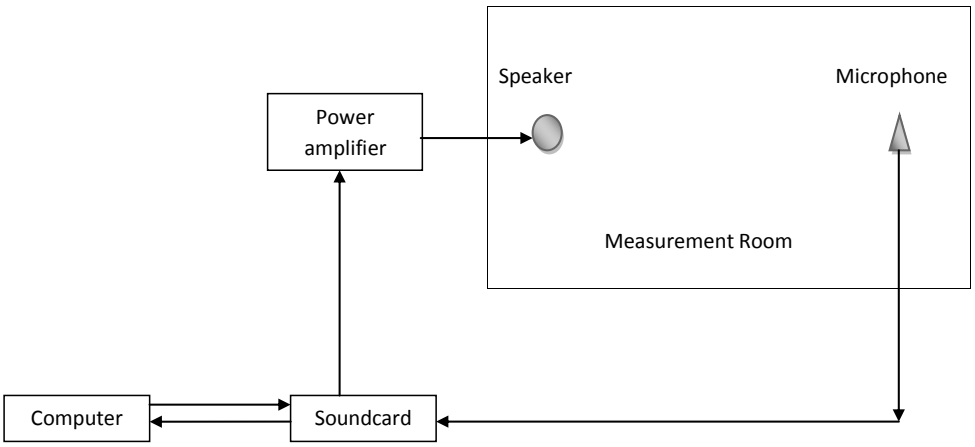


Figure 3.1: Schematic representation of the measurement system.

A complete measurement system has been designed and realized to enable fast, reliable and simple method for deriving impulse response of a given system.

Though several dedicated measurement systems already exist, these are generally expensive and bulky. Therefore, a highly configurable, portable program written in C was developed by the group for the automatic acquisition of the impulse response with common elements such as a microphone, a loudspeaker and a computer was used. This has led to the obtention of a global, cheap and adaptable measurement system(see figure(3.1)) allowing fast and accurate measurement of the impulse response.

The C program controls the generation of the different excitation signals, their emission through the loudspeaker connected via the power amplifier to the full-duplex soundcard, and the simultaneous recording of the signal at the microphone. The time required for one measurement is very short: only a few seconds are necessary to obtain the impulse response of an acoustical space.

The microphone array in combination with its inbuilt speaker was used for all the following experiments. That is, two separate microphone array's were used; one of which acted as a speaker and the other one as the microphone. The implementation

was wantedly carried out using microphone array's which are much inferior when compared to standard superior speakers and microphones to show that the presented implementation is very much practical even with low end devices.

3.2 Generation of sine sweep signal

A common method for measuring the impulse response of an acoustical system is to apply a known input signal and to measure the system's output. The choice concerning the excitation signal and the deconvolution technique enabling the obtention of the impulse response from the measured output is of essential importance:

1. The emitted signal is perfectly reproducible, unlike blank shots or white noise;
2. The signal should have a flat frequency spectrum, like an ideal impulse;
3. The excitation signal and the deconvolution technique have to maximize the signal-to-noise ratio of the deconvolved impulse response;
4. The excitation signal and the deconvolution technique must enable the elimination of non-linear artifacts in the deconvolved impulse response.

And the sine sweep method which has the perfect and complete rejection of the harmonic distortions prior to the linear impulse response, and an excellent signal-to-noise ratio makes it the best impulse response measurement technique in an in-occupied and quiet room.

In order to obtain comparable measurements, the following parameters were used:

1. Sampling frequency : 48000 Hz
2. Initial sine sweep frequency : 1000 Hz
3. Final sine sweep frequency : 16000 Hz
4. Sine sweep duration : 30 seconds
5. Duration of silence after each sweep : 2 seconds

for the estimation of reflection coefficients of individual surfaces. Materials of the size $200 \times 50\,cms$ were used. And hence to avoid the diffraction noise in the final impulse response caused by low frequencies, whose wavelength's are much larger than the material size; the initial sine sweep frequency was chosen to be $1000Hz$ which is way above the theoretical limit given by,

$$f > c/\lambda = 34000/50 = 680 \simeq 700\,Hz$$

The above equation gives us the limit of frequency considering normal sound wave incidence. In our experiments we are dealing with oblique incidence. In case of oblique incidence the surface area of contact for the oblique incident wave is much larger than for a normal incident wave. Thus we have provided enough buffer $((c/1000) - (c/700) = 145cm)$ by considering frequencies beyond 1000 Hz only, there by avoiding the inaccurate results casued by diffraction of low frequencies at oblique incidence. Also initially a second set of measurements were taken in high frequency range $16\,k - 20\,kHz$ which did not provide consistent results. This was accounted to the non-linear properties of the low end microphone array's at higher frequencies. And hence the rest of the experiments were carried out in the stable audible frequency range of $700 - 16\,kHz$. The measured reverberation time for the room was 0.25 seconds, and hence the 2 seconds duration of silence after each sweep was chosen to be a risk free parameter. The length of the sine sweep signal is chosen based on the size of the room, shorter sequences for room with short impulse responses(small reverberation time) and longer sequences for rooms with longer impulse responses(large reverberation time).

3.3 Measurement of reflective coefficient

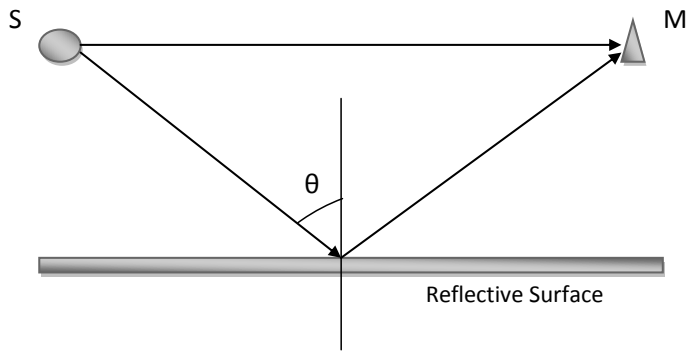


Figure 3.2: Experimental setup for oblique reflective coefficient measurements. S - speaker; M -microphone

The experimental disposition for the measurement of reflective coefficient is represented in figure(3.2). A speaker S and microphone M pair is placed in front of the reflective surface to be tested, at an acute incident angle θ. The incident and the reflected signals are detected by the same microphone. In the audible frequency range, and for reasonably small (no more than a few meters) microphone distances from the loudspeaker and from the test surface, air absorption can be safely neglected. The impulse response so achieved from the setup is shown in figure(3.3(a)).

Where the first peak from the left is called *Direct response* which is a result of direct wave from the speaker to microphone. This can also be used to determine the actual distance between the speaker and microphone. The second peak is called the *Reflected response*, this is the response of interest for us. As it is the response of the wave reflecting from the material of our study.

3.3.1 Reflection Method

To start with a study was carried out to determine the relation between the amplitudes of the direct and reflected impulse response(see figure(3.3(b))). It was observed that the reflected impulse response was much attenuated in comparision to the direct impulse response. Though the method was crude, different surfaces resulted in different level of attenuations. On performing the recording again with a different orientation of speaker-microphone setup (sensor devices facing the surface perpendicularly instead of facing the surface at a particular incident angle) to check the consistency of the method, it was observed that the reflected impulse response was much amplified in comparision with the direct response. This was accounted for the directivity property of the microphone arrays. And hence this method entirely depends on the orientation of the microphone array with respect to surface and the speaker.

A much robust way of determining the properties of a reflective surface is using the *reflection method*[39]. Consider the scenario in figure(3.2), where the signal $s(t)$ from the speaker and the reflected signal from the material is recorded by the microphone M. This can be explained in frequency domain as follows,

$$R_D(\omega) = H_d(\omega) \cdot S(\omega) \tag{3.1}$$

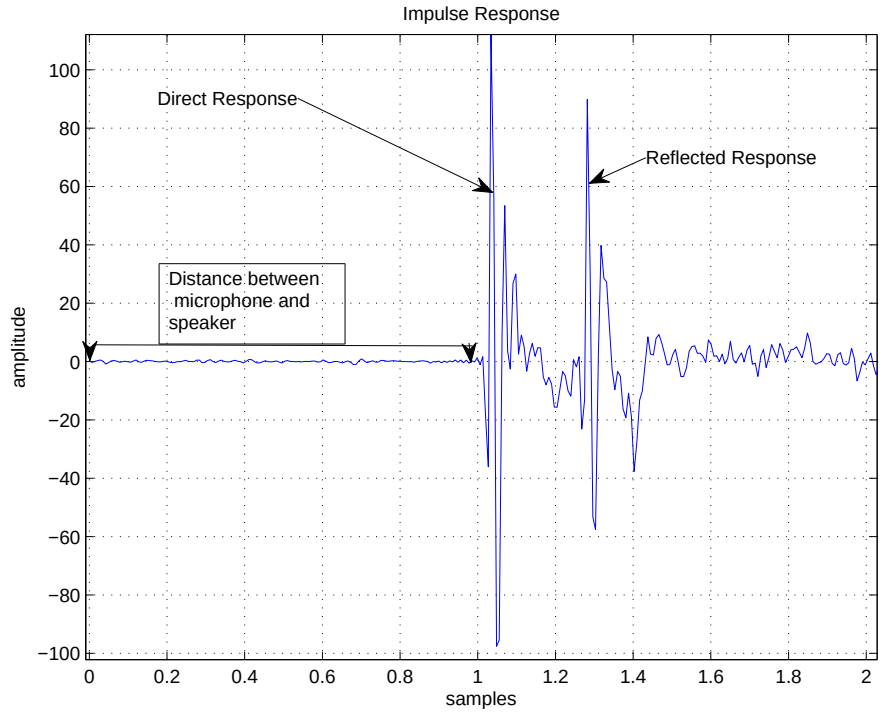
where $R_D(\omega)$ is the direct response, ω is the angular frequency, and has been emitted in future equations for brevity. $S(\omega)$ is the source signal and $H_d(\omega)$ is the transfer function responsible for direct response, which is the product of the propagation loss H_{loss} and the microphone response H_{mic} to the incoming signal.

$$H_d = H_{mic} \cdot H_{loss} \tag{3.2}$$

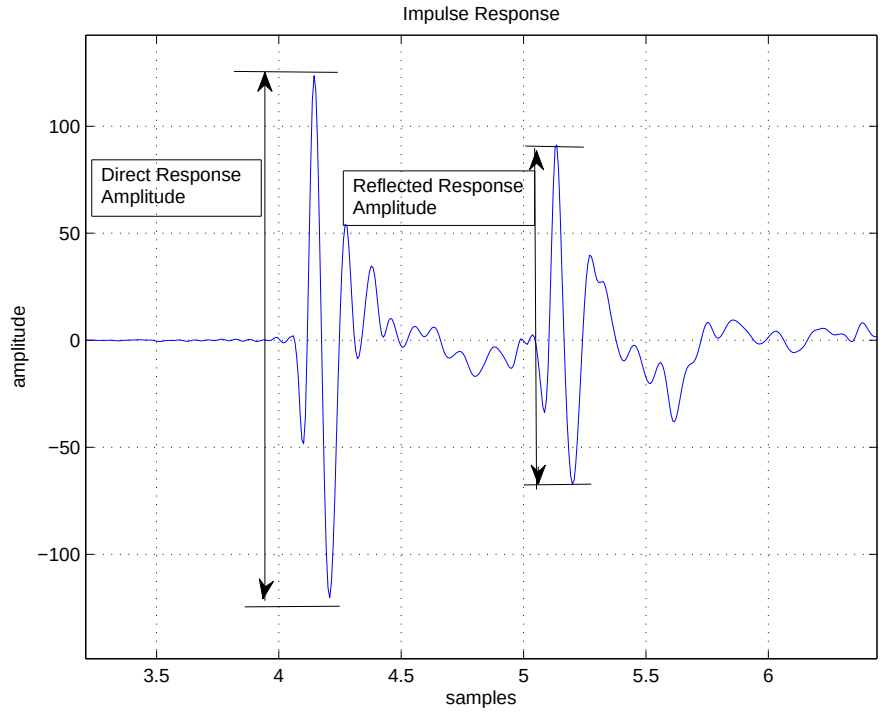
Now the Reflected Response R_R can be explained as,

$$R_R = H_r \cdot S \tag{3.3}$$

where H_r is the transfer function responsible for reflected response, which is the product of the reflective surface material H_{surf}, the propagation loss H_{loss} and the



(a)



(b)

Figure 3.3: (a)A typical Impulse response for the setup shown in figure(3.2), (b) Methodology for amplitude relation of direct and reflected response.

microphone response H_{mic} to the incoming signal.

$$H_r = H_{surf} \cdot H_{mic} \cdot H_{loss} \tag{3.4}$$

now the reflection coefficient of the surface can be obtained by,

$$T(\omega) = \Big(\frac{ReflectedResponse\,(R_R(\omega))}{DirectResponse\,(R_D(\omega))} \Big) = \frac{H_r(\omega)}{H_d(\omega)} = H_{surf}(\omega). \tag{3.5}$$

For a given microphone we know that H_{mic} is constant and the loss in direct and reflected path are considered to be almost comparable. And hence we deduce the transfer function of the reflective surface.

Now from equation(2.5) we can also deduce the absorption coefficient of the material which is given by,

$$\alpha(\omega) \; = \; 1 - |R(\omega)|^2 \; = \; 1 - |T(\omega)|^2. \tag{3.6}$$

In this way the absorption coefficient could be obtained for a large range of frequencies from a single measurement.

Also these transfer function estimations were carried out on highly interpolated data to achieve a considerably good consistency. The applications of the transfer function so obtained are innumerable. We can simulate the material on computer and find out how it responds to different signals.

The operation of isolating a pulse and extracting it from the global impulse response is a windowing operation. The presented thesis has been implemented using a rectangular window. The problem with windows is that often it is not so easy to place their limits, in our approach we have derived the average limit, from a large set of observations. Also it is preferable to use windows of equal length both for the incident pulse and the reflected pulse, as long as the surface behaviour is not too reactive.

3.3.2 Surface Classification and Room Surface Estimation

One of the principle applications using the above determined transfer function is to solve the problem of estimating room surfaces from the acoustic room impulse response, the results of which can be further used for the room geometry estimation. In this lines experiments were carried out in a semi-anechoic room to classify 4 basic materials (see figure (3.5)) in a room environment, plywood, glasswood, ceiling (perforated material, false roof) and floor (plastic material, surface quite similar to that of a polished plywood) using different classification features. The results of which has been really promising and has been discussed in chapter 4.

Furthermore, using the room impulse response of an enclosed room, experiments

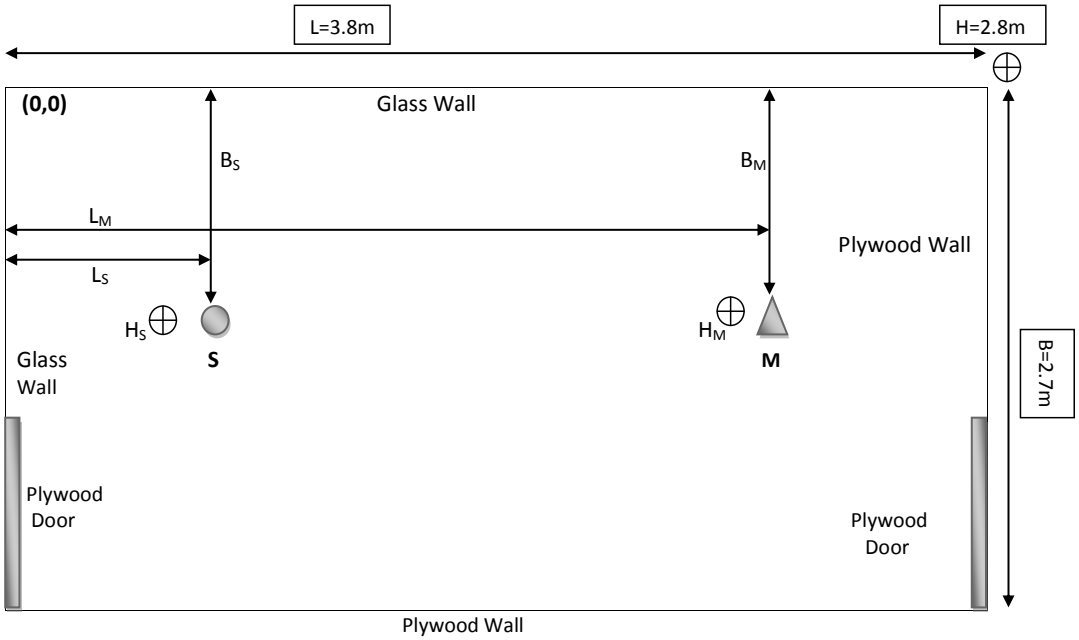
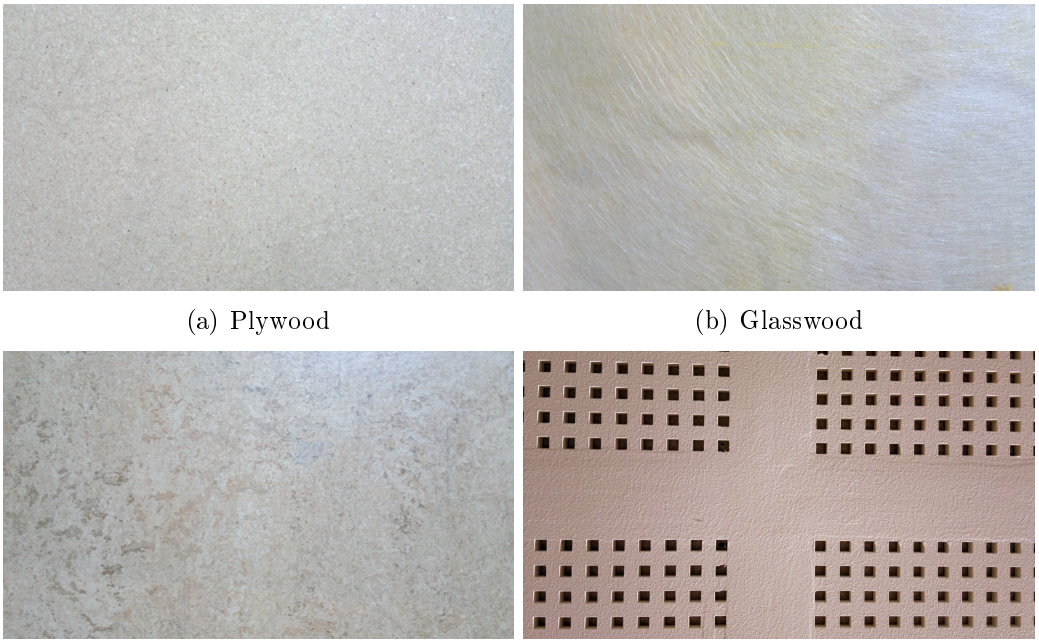


Figure 3.4: Empty study room with array speaker(**S**) and microphone(**M**) positions. The surfaces of which are glass on two sides and plywood wall on two sides. The right top corner of the room is considered as origin for all measurements, the cartesian coordinates of the source is given by (L_S, B_S, H_S) and that of microphone is given by (L_M, B_M, H_M).

were carried out to classify different materials existing in that room. The presented studies were carried out in a small room of dimensions $2.7 \times 3.8 \times 2.8m$. This is because of the fact that the accuracy of reflective coefficient estimation using the reflection method decreases, as the distance of the surface from the speaker increases. The setup of the room is as shown in the figure (3.4) and a picture of the actual experimental setup is shown in figure (3.6). The materials in the room were identified to be- glass, wall, plywood, false-roof(glasswood) and floor.

A typical room impulse response looks like in figure (3.7), which includes direct response, single reflection response, and also responses which are a result of multiple reflections (see fig.(3.8)). Such a transfer function obtained as a result of multiple reflections is the product of two or more transfer functions. The classification of such a transfer functions is a tidious job, and hence precautions have been taken to consider transfer functions of single reflections only for the classification of materials, which is explained in section (3.3.3).

It is also common in the measurement of room impulse response to have a reflected response, which is a result of two sound waves being reflected from two different walls, but reaching the microphone simultaneously (see fig.(3.8)). This occurs when the reflected paths of the two are the same. Precautions have been taken to not consider such reflected response in the study, to keep the problem statement simple and direct.



(a) Plywood (b) Glasswood

(c) Floor (d) Ceiling

Figure 3.5: Basic room materials.

Figure 3.6: Picture of RIR measurement experiment setup

The so classified materials and there positions obtained from the room impulse response were cross checked with the actual scenario. The results of which has been discussed in chapter 4.

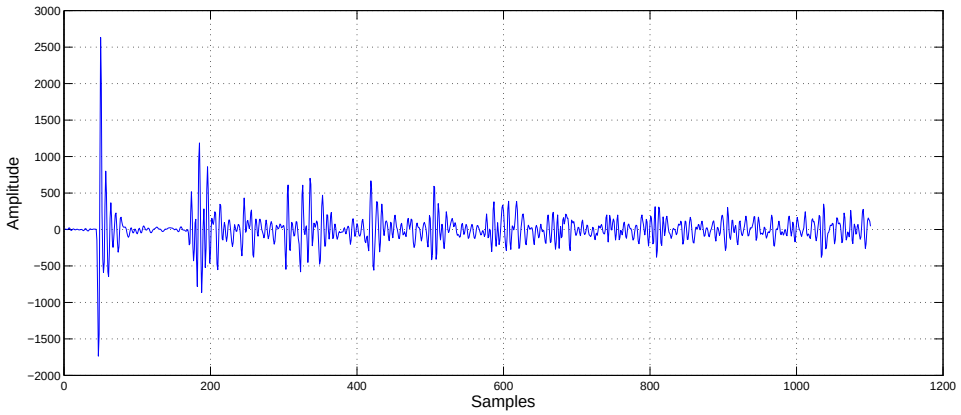


Figure 3.7: Room impulse response of the experimental room at Sampling frequency of 48000Hz.

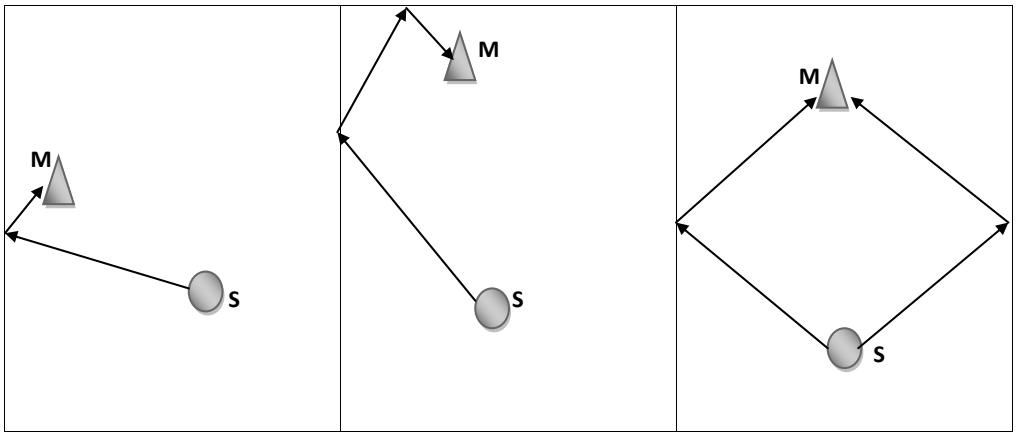


Figure 3.8: Potential errors in transfer function estimation, where S- is the sound source and M is the microphone location. Left image : This is the expected ideal case, where the transfer function is obtained from a single wall reflection. Center image: This is an erroneous situation in which the transfer function is a result of reflection from two different walls. Right image: This erroneous situation arises due to symmetry in the positions of the recording setup in the room. Here the resultant transfer function is of two sound waves being reflected from two different walls, but reaching the microphone simultaneously.

3.3.3 Ground Truth of Room Surfaces

In order to avoid the possible errors and increase the precision in the room surface classification. Erroneous transfer functions obtained from multiple reflection and simultaneous reflections as seen in figure (3.8) has to be removed from the obtained data set. The approach taken to do this is as follows. The room dimensions and the positions of the recording setup are known before hand, this data is used to calculate

the approximate arrival time of all the possible single and double reflections using the image source method explained in section (2.5.2). Further reflections which arrive at close interval of time, less than 1ms has been handpicked manually to avoid the effect of simultaneous reflections. There by we obtain a reduced data set with pure transfer functions.

In the presented thesis the ground truth has been obtained manually. Only to avoid possible errors, since this is possibly the first effort of classification of surfaces using room impulse response. But there have been established researches on self localization of speaker and microphone setup [41] which can be used for automaically obtaining the ground truth of room surfaces in coordination with the image source method.

3.4 Classifier

In the presented thesis the classifier used is the K-means clustering [42]. This partitions data into k mutually exclusive clusters. K-means algorithm treats each observation in the data as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible.

K-means works iteratively such that it minimizes the sum of distances from each object to its cluster centroid, over all clusters. This way it moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible.

4. RESULTS AND DISCUSSION

4.1 Material Classification

The outlined reflection technique was tested in a semi-anechoic room of the university laboratory, with no external devices(computers) in the room. The reverberation time of the room was measured to be , $T_{60} = 0.25s$.

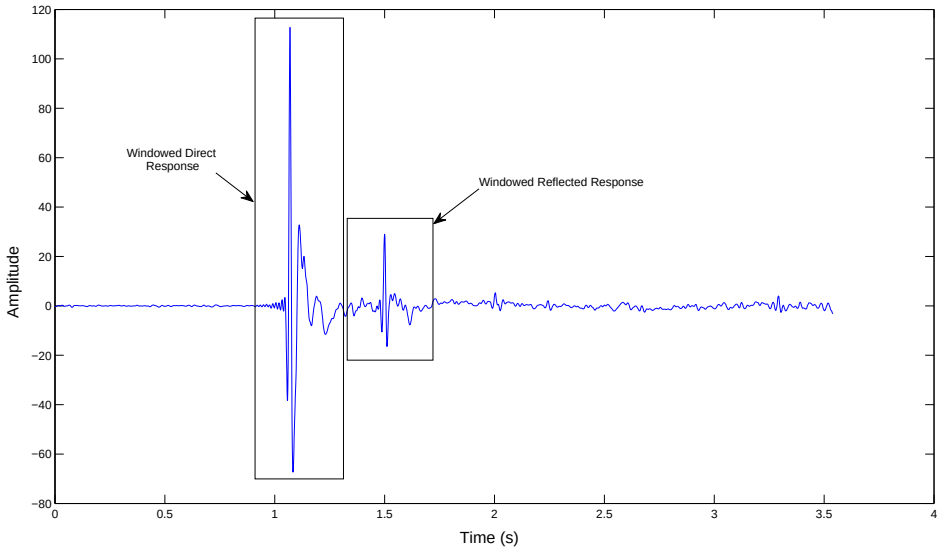
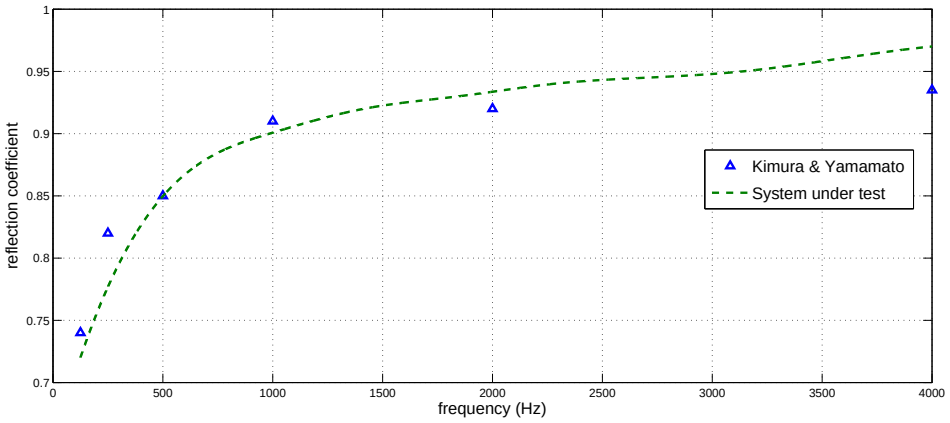


Figure 4.1: Windowing of the impulse response obtained from reflective surface.

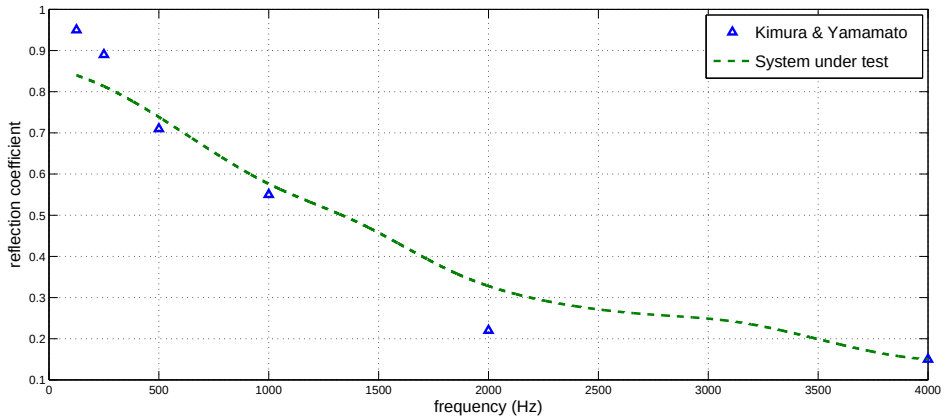
The measurement setup was as shown in figure (3.2). Initially the experiments were conducted on two different materials-plywood(thickness $\simeq 25mm$) and glass fibre(thickness $\simeq 25mm$) to check the consistency of the reflection method. A typical impulse response is shown in figure (4.1), where the framed portions are the direct pulse and the pulse reflected from the surface.

The incident and reflected pulse were selected by windowing, in order to apply equation (3.5). The resulting reflection coefficients in oblique sound incidence (approx. 45°) are compared with the values derived from stretched pulse technique done by Kimura and Yamamoto [38] (shown in figure (4.2)). The agreement between the reflection method and the stretched pulse technique is fairly good. With the presented method it is also possible to determine the phase response, group and

phase delay, bode plots, more about this is presented in appendix A.



(a) Plywood



(b) Glass Fibre

Figure 4.2: Reflection coefficient's of (a)plywood and (b)glass fibre panel at oblique incidence. ($\triangle$) comparision values from stretched pulse technique, ($- - -$) Average of reflection coefficient's recorded in one microphone array. Response only upto 4kHz has been displayed here due to unavailability of comparision data beyond this frequency.

The ideal window lengths in this case were about $1.04\,ms$; other computations with different window lengths did not show substantial deviations, until the energy of the reflected pulse was considered correctly. The selection of impulses was done using an algorithm which automatically recognizes impulses in steps. It first lists the energy of the entire impulse response in a fixed window size of $1\,ms$. In the windows with high energy, the algorithm searches for the maxima in an extended window size of $2\,ms$ and then selects the impulse of the length $1.04\,ms$ around the maxima.

Based on the these early results, a study was carried out to classify the four basic materials in a room environment - plywood, glasswood, ceiling (or false roof, with

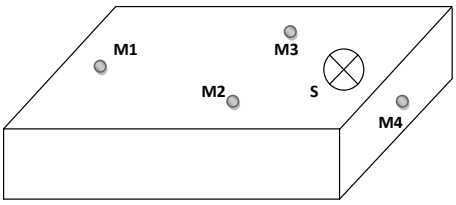


Figure 4.3: Figure showing the apparent positions of microphones and speaker in a microphone array. Note the position of **M4** it is on a different side of the microphone array, and hence the recordings of which may not be identical to the recordings by **M1**, **M2**, **M3**.

perforated material) and floor (plastic material, surface similar to that of a polished plywood). Recordings were taken in a similar setup (see figure (3.2)) and the magnitude respose so obtained was used for classification. The magnitude response of these materials, as recorded in four different microphones of a microphone array, in 3 different heights ($0.385\,m$, $0.583\,m$, $0.8\,m$) has been displayed in figure (4.4). In the image the DC offset (zero frequency) of the respective magnitude responses have been removed to get a better perspective of the different responses. We see that the overall deviation of magnitude response of a particular material is not out of bounds. There are few cases of outshoots, which can be accounted to the position of the 4th microphone in the microphone array(see figure (4.3)), which is at a different position and hence recieves less sound pressure then the rest. Another reason would be the windowing operation. It is not easy to place the limits of the window, where the windowed function is about zero, in order to avoid aliasing effects (these arise from the convolution of the window with the framed function in the frequency domain).

To classify the four materials based on the magnitude response, we've to select the best feature. It was observed that the response of surfaces to mid-frequencies (upto 10kHz) was different from the higher frequencies (beyond 10kHz). Classification was tested using features like cepstral coefficients, moments, gradient and energy of the signal in different intervals. Though these features could classify the given data with a good separation between classified groups, the most basic feature like a mean operator could also achieve the same level of classification. Thus the two features used for the classification of the above four material was mean of first 10kHz and mean of frequencies beyond 10kHz. The resulting classification using the K-means classifier is shown in figure (4.5). It is to be noted that for the classification process the actual magnitude responses have been considered, that is the responses including their DC offset, which gives the level of attenuation. We see that the separation of floor material and plywood is not distinct and unique. This can be accounted to

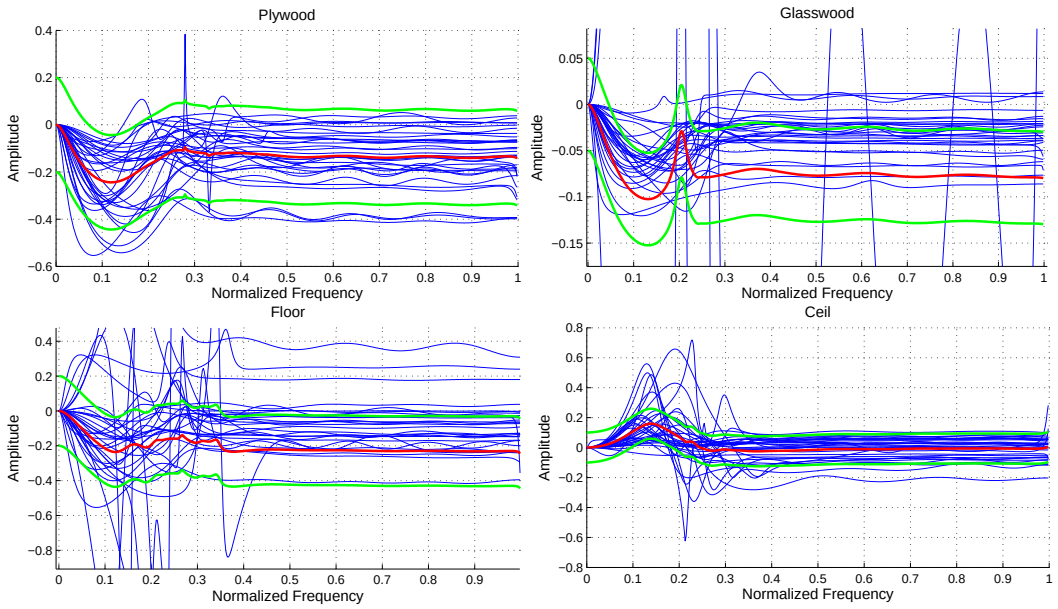


Figure 4.4: The magnitude response without the DC offset for the four basic room materials - plywood, glassfibre, floor and ceiling. The y-axis represents the linear attenuation scale, x-axis represents normalized frequency. Observe the slope of first 100 samples of each material. This itself identifies these materials. The red line represents the mean and the green line represents the deviation of data.

the similarity in the surface texture of plywood and floor material. We can observe the same in figure (4.4) most of the magnitude response of floor material mimics the plywood magnitude response, and hence selection of any feature wouldn't make a difference. This can be one of the possible setbacks of this method. That is materials with similar surface texture will offer similar responses.

4.2 Room Surface Estimation

Now that we could successfully classify different materials based on the reflection coefficients; a study was carried out to estimate the room surfaces using the reflection coefficients derived from room impulse response. The first step of this study was to select an ideal room environment for the proposed case. We chose a study room of our college whose dimensions are $2.7 \times 3.8 \times 2.8m$. Studies were carried out after emptying the room of all the chairs and tables. The floor and roof material of the study room were of the same material as our laboratory classification experiment. The walls of the study room were composed of glass on two sides and plywood on the other two sides (see figure (3.4)).

Recordings were taken with two microphone arrays, one of which was used as a speaker at a time and at positions $(L_S\,,\ B_S\,,\ H_S)$ and $(L_M\,,\ B_M\,,\ H_M)$ as listed in table (4.1). The room impulse response was obtained using the sine sweep. As discussed in chapter 3; during the calculation and classification of reflection coefficients

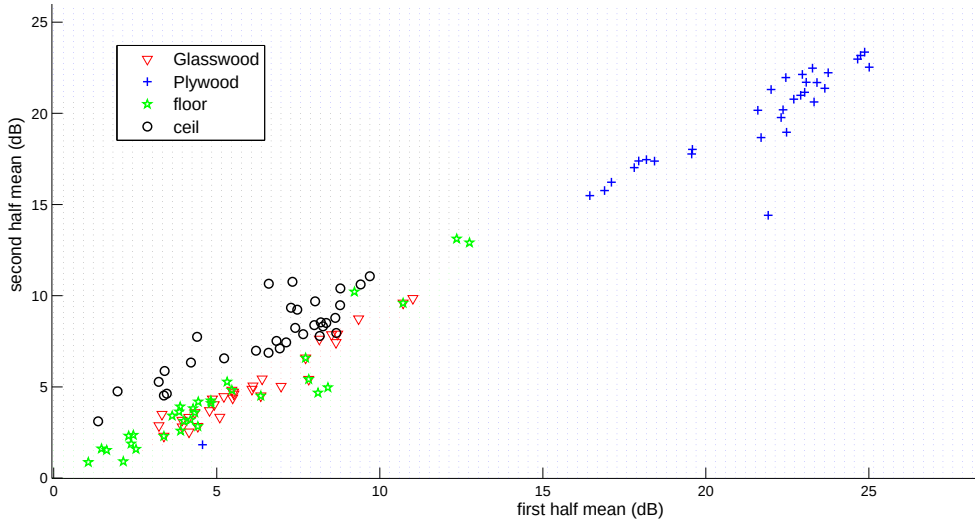


Figure 4.5: Classification of common room materials based on two magnitude features of the derived transfer functions. The type of finishing/coating over the material accounts a lot in the presented method of reflection coefficient estimation. Since the floor material had a finishing identical to the plywood, we see that the seperation between plywood and floor is not distinct.

precautions were taken to not consider reflected response as a result of multiple reflections, and reflected response which can be a result of two different walls reaching the microphone simultaneously. Recordings were totally taken in four different combinatorial positions of speaker($\mathbf{S}$) and microphone($\mathbf{M}$). Where in each recording consisted of two sets, first set in which one of the microphone array acts as the speaker and the other acts as the microphone, and the second set in which their work is exchanged with the position remaining a constant. Thus we achieve enough data for our studies.

Table 4.1: Different positions of Speaker($\mathbf{S}$) and Microphone($\mathbf{M}$) in the study room.

	$(L_S\,,\,B_S\,,\,H_S)$	$(L_M\,,\,B_M\,,\,H_M)$
Position 1	(0.979, 1.368, 1.295)	(3.304, 1.288, 1.306)
Position 2	(0.979, 1.368, 1.736)	(3.304, 1.288, 1.696)
Position 3	(1.047, 1.960, 1.736)	(2.684, 0.943, 1.696)
Position 4	(1.233, 0.462, 1.736)	(2.337, 2.398, 1.696)

The achieved classification of materials using the room impulse response, at different speaker-microphone positions has been shown in figures (4.7),(4.8),(4.9)& (4.10). The two features used for classification were gradient or first 50 samples and gradient of consecutive 100 samples. The bottom image of these figures show the respective material in real world (the ground truth), and top image shows the material as classified using the reflection coefficient. We see that most of the classified surfaces are

correctly classified in real world as well. There has been certain outliers, which are a result of similarity of surfaces- in this case the floor material and the door material have the same kind of outer finishing, and also the angle of incidence- though the difference in angle of incidence, changes only the attenuation margin by a couple of decibel's and not the entire response[38]; there are chances that this might induce some error in the classification. Table (4.2) gives a numerical proof of the precision of the presented method. The precision value is calculated as follows

$$precision = \frac{total\,hits - misses}{total\,hits}. \tag{4.1}$$

Where for each position of the speaker-microphone setup, *total hits* is the total number of events detected and *misses* is the total number of false events in the detection.

Table 4.2: Numerical presentation for the results of the wall classification using the formula (4.1)

	total hits	*misses*	*precision*
Position 1	20	3	85.00%
Position 2	19	3	84.21%
Position 3	22	4	81.82%
Position 4	20	3	85.00%

Considering the complexity of the problem, and the simplicity of the solution, the achieved precision is commendable. This result with additional information from multiple microphones such as Direction of Arrival [43] can be further used to localize walls/reflecting surfaces. The outline of the proposed plan for room surface, and further geometry estimation in presence of the direction of arrival data from microphone array's is shown in figure (4.6).

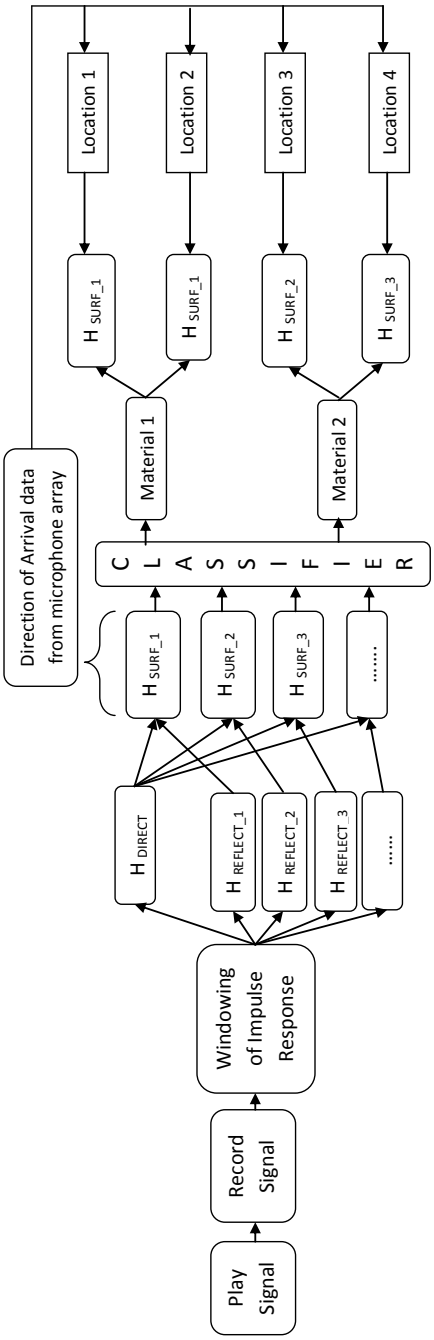


Figure 4.6: Outline of Room Surface Estimation.

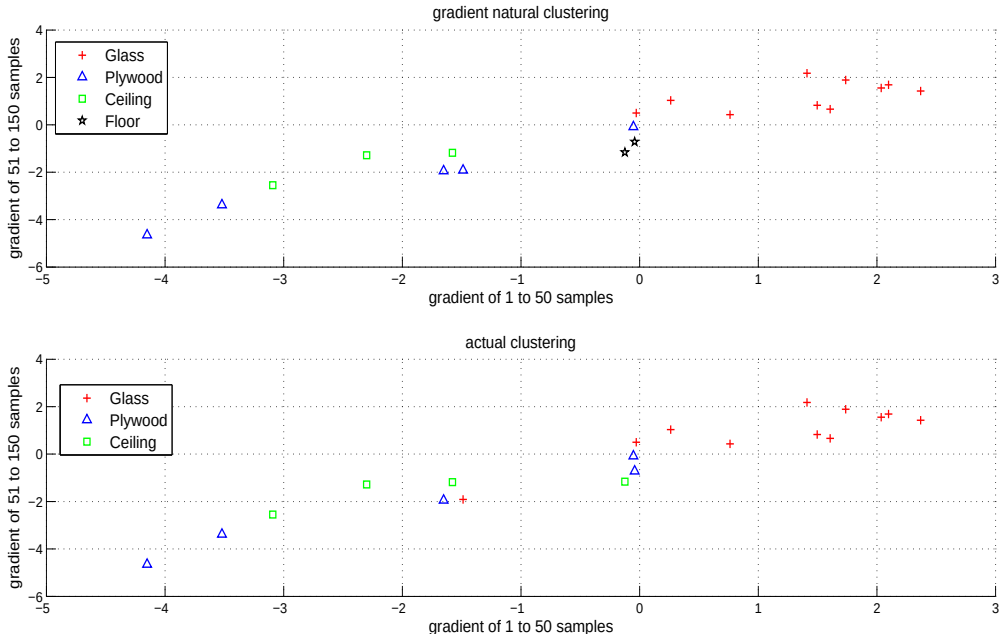


Figure 4.7: Classification of study room wall materials for position 1 listed in table(4.1) based on selective features of the derived transfer functions. Top image : Classification of walls, Bottom image: truth value.

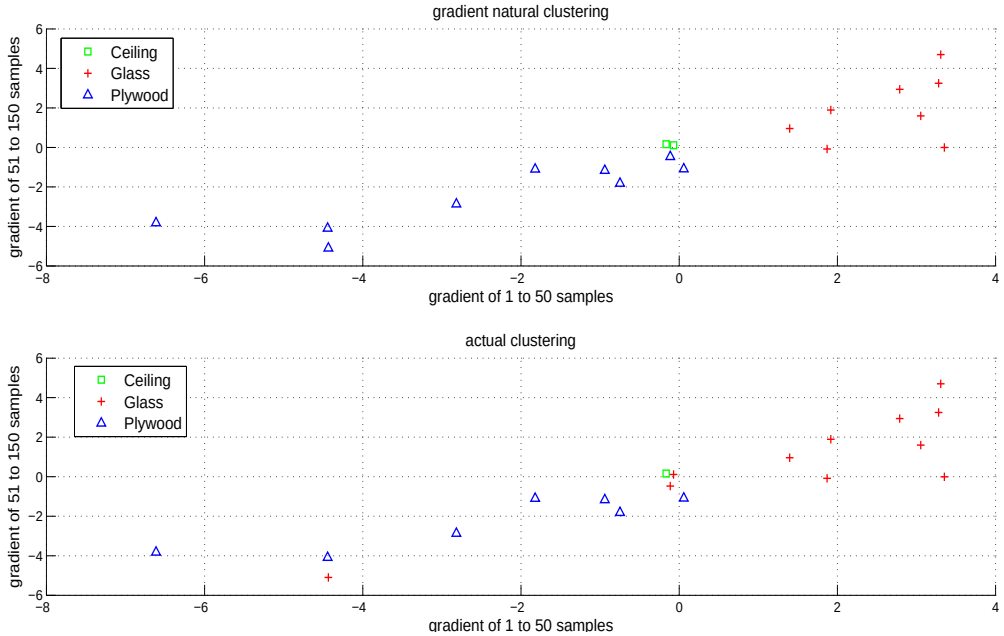


Figure 4.8: Classification of study room wall materials for position 2 listed in table(4.1) based on selective features of the derived transfer functions. Top image : Classification of walls, Bottom image: truth value.

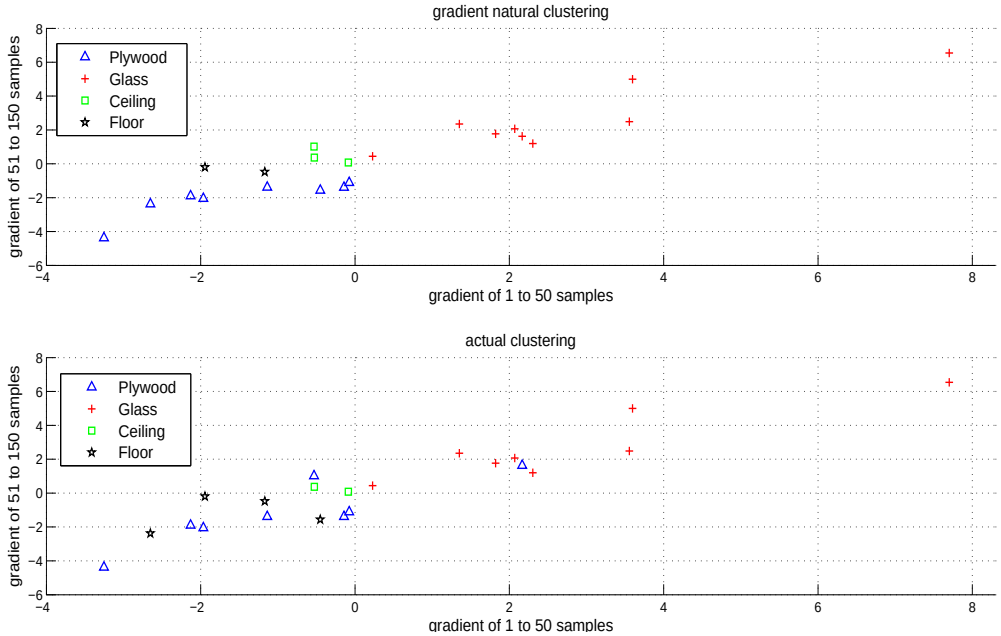


Figure 4.9: Classification of study room wall materials for position 3 listed in table(4.1) based on selective features of the derived transfer functions. Top image : Classification of walls, Bottom image: truth value.

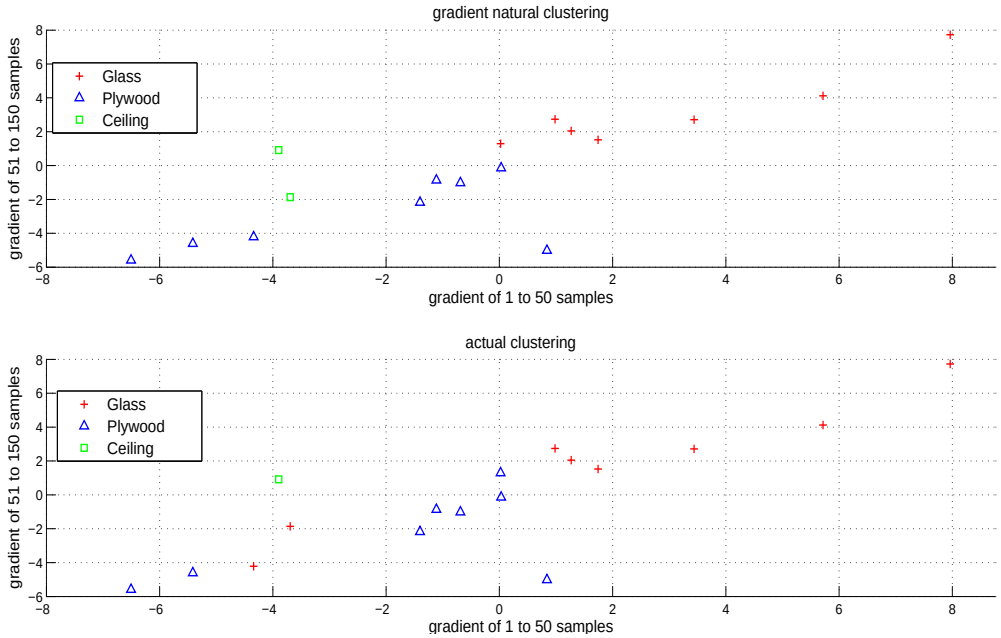


Figure 4.10: Classification of study room wall materials for position 4 listed in table(4.1) based on selective features of the derived transfer functions. Top image : Classification of walls, Bottom image: truth value.

5. CONCLUSION

The reflection method of measuring the reflection coefficient *in situ* has been described in general terms, taking into account unwanted reflections and background noise. It was shown that the reflection method can, in principle, give good measurements, choosing the right input signal and taking advantages from modern digital-processing techniques.

The proposed use of sine sweep as test signal can greatly improve the overall immunity against non linear distortions and background noise. Thus, quick in-situ measurements become effective even in reverberant and noisy conditions.

Some errors typical of digital-processing techniques such as possible diffraction of low frequencies and window length for windowing impulses were identified and commented upon.

Experiments conducted with a low-cost, portable instrumentation showed oblique-incidence results which are in good agreement with those obtained by using the stretched pulse technique.

The reflection coefficient of basic room materials were succesfully classified using simple operators. Which shows that the results achieved are unique and easily identifiable. Though some errors occured, these were accounted for the similarity between the surface of the reflecting material. The handling of which can be considered as a future work.

The proposed reflection method was further used to identify the wall materials in a test room using the room impulse response. The performance of which was commendable with an average precision of over 80%. And the possible sources of error has been accounted for two main reasons, similarity of reflecting surfaces and the angle of incidence.

The results of the material classification in a room using room impulse response have been promising, with additional information like direction of arrival, which can be obtained using the microphone array data; two walls with identical surfaces can be separated. Further, studies can be carried out to identify the room geometry using the proposed method and the direction of arrival data as shown in the figure(4.6).

REFERENCES

[1] H.K.Dunn and D.W.Farnsworth, "Exploration of pressure field around the human head during speech," *The Journal of the Acoustical Society of America*, vol.10, no. 1, pp. 83,1939.

[2] W.Chu and A.Warnock, "Detailed directivity of sound fields around human talkers," *IRCCNRC,* Canada, Tech. Rep. IRC-RR-104,2002.

[3] H.Kuttruff, "Room Acoustics," *Taylor & Francis,* 3rd edition, 1991.

[4] J.Eargle. "The Microphone Book," *Focal Press,* 2nd edition, 2005.

[5] ISO Standard 3382-1. "Acoustics - Measurements of room acoustic parameters - Part 1: Performance spaces," *International Standards Organization*, 2000.

[6] ISO Standard 3382-2. "Acoustics - Measurements of room acoustic parameters - Part 2: Reverberation time in ordinary rooms," *International Standards Organization*, 2000.

[7] J.Allen and D.Berkley. "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, **65**:943-950, 1979.

[8] J.Vanderkooy, "Aspects of MLS measuring systems," *The Journal of Audio Engineering Society,* vol. 42, no. 4, pp. 219-231,1994.

[9] A.Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *in 108th AES Convention,* 2000.

[10] E.C.Wente and E.H.Bedell, "The measurement of acoustic impedance and the absorption coefficient of porous materials," *The Bell System Technical Journal*, January, 1928.

[11] S.Muller and P.Massarani, "Transfer function measurement with sweeps," *Journal of Audio Engineering Society,* **49**(6):443-471,2001.

[12] A.Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique (preprint 5093)," *Presented at the* 108^{th} *AES Convention, Paris, France,* February 19-22 2000.

[13] G.B.Stan, J.J.Embrechts and D.Archambeau, "Comparision of different impulse response measurement techniques," *Journal of Audio Engineering Society, Volume 50 Issue 4 pp.249-262,* April 2002.

[14] ISO 354, "Acoustics - Measurement of Sound Absorption in a Reverberation Room." *ISO, Geneva, Switzerland*, 1985.

[15] C.W.Kosten, "International comparision measurement in the reverberation room," *Acustica*, **10** (1960) 400-11.

[16] L.L.Beranek, "Precision measurement of acoustic impedance," *Journal of Acoustical Society of America,* **12**(1940) 3-13

[17] ASTM E1050, "Standard test method for impedance and absorption of acoustical materials using a tube, two microphones and a digital frequency analysis system," *ASTM,*Philadelphia, USA, 1990.

[18] F.J.Fahy, "Rapid method for the measurement of sample acoustic impedance in a standing wave tube," *Journal of Sound Vibration,* **97**(1)(1984) 168-170.

[19] A.F.Seybert and D.F.Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," *Journal of Acoustical Society of America,* **61**(5)(1977) 1362-70.

[20] J.Y.Chung and D.A.Blaser, "Transfer function method of measuring in-duct acoustic properties **I** Theory," *Journal of Acoustical Society of America,* **68**(3)(1980) 907-13.

[21] J.Y.Chung and D.A.Blaser, "Transfer function method of measuring in-duct acoustic properties **II** Experiment," *Journal of Acoustical Society of America,* **68**(3)(1980) 914-21.

[22] A.Cops and W.Lauriks, "Characterization of sound absorbing materials from two microphone experiments," *Proceedings INTER-NOISE 90, Acoustical Society of Sweden, Gothenberg, Sweden,* 1990, pp. 117-22.

[23] W.T.Chu, "Impedance tube measurements - A comparitive study of current practices," *Noise Control Engineering Journal,* **37**(1)(1991) 37-44.

[24] A.Cummings, "Impedance tube measurements on porous media: the effects of air-gaps around the sample," *Journal of Sound and Vibration,* **151**(1)(1991)63-75.

[25] A.London, "The determination of reverberant sound absorption coefficients from acoustic impedance measurements," *Journal of Acoustical Society of America,* **22**(2)(1950) 263-9.

[26] U.Ingard and R.H.Bold, "A free field method of measuring the absorption coefficients of acoustic materials," *Journal of Acoustical Society of America,* **23**(1951) 509-16.

[27] Y.Ando, "The interference pattern method of measuring the complex reflection coefficient of acoustic materials at oblique incidence," *Proceedings of 6th ICA,* Paper E33, 1968.

[28] Z.Kintsl, "Investigation of the sound absorption of wall sections by pulse technique," *Soviet Physics-Acoustics,* **21**(1975) 30-2.

[29] M.Yuzawa, "A method of obtaining the oblique incident sound absorption coefficient through an on-the-spot measurement," *Applied Acoustics,* **8**(1975) 27-41.

[30] J.C.Davies and K.A.Mulholland, "An impulse method of measuring normal impedance at oblique incidence," *Journal of Sound Vibration,* **67**(1)(1979) 135-49.

[31] A.J.Cramond and C.G.Don, "Reflection of impulses as a method of determining acoustic impedance," *Journal of Acoustical Society of America,* **75**(2)(1984) 382-9.

[32] K.A.Hollin and M.H.Jones, "The measurement of sound absorption coefficient in situ by a correlation technique," *Acustica,* **37** (1977) 103-10.

[33] J.S.Boltion and E.Gold, "The application of cepstral techniques to the measurement of transfer functions and acoustical reflection coefficeints," *Journal of Sound Vibration,* **93**(2)(1984) 382-9.

[34] M.Tamura, "Spatial fourier transform method of measuring reflection coefficients at oblique incidence. 1. Theory and numerical examples," *Journal of Acoustical Society of America,* **88**(5)(1990) 2259-64.

[35] J.F.Allard and Y.Champoux, "In situ two-microphone technique for the measurement of the acoustic surface impedance of materials," *Noise Control Engineering Journal,* **32**(1)(1989) 15-23.

[36] H.M.Hess, K.Attenborough, N.W.Heap, "Ground characterization by short-range propagation measurements," *Journal of Acoustical Society of America,* **87**(5)(1990) 1975-86.

[37] M.S.Atwal and M.J.Crocker, "Measurement of the absorption coefficient of acoustical materials using the sound intensity method," *CETIM, Senlis, France,* 1985, pp 485-90.

[38] K.Kimura and K.Yamamoto, "A method for measuring oblique incidence absorption coefficient of absorptive panels by stretched pulse technique," *Applied Acoustics,* **62**(2001) 617-632.

[39] M.Garai, "Measurement of the sound-absorption coefficient in situ: The reflection method using periodic pseudo-random sequences of maximum length," *Applied Acoustics,* **39**(1993) 119-139.

[40] Teemu Korhonen, "Acoustic Source Localization Utilizing Reflective Surfaces," *Doctoral Thesis,* TUT, 2010.

[41] P.Pertila, M.Mieskolainen and M.S.Hamalainen, "Closed-form self localization of asynchronous microphone arrays," *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), Edinburgh,* May 30-June 1, 2011.

[42] Richard O. Duda, Peter E. Hart and David G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, Inc., 2001.

[43] Yli-Hietanen, K.Kalliojarvi and J. Astola, "Robust time-delay based angle of arrival estimation," *Proc. 1996 IEEE Nordic Signal Processing Symposium, Espoo, Finland,* September 1996, pp.219-222.

A. APPENDIX

Presented in this appendix are the additional results obtained from the reflection coefficient estimation using the reflection method. The figures here are obtained from the setup of a plywood panel ($thickness \simeq 25mm$, panel size 50×200) placed at a height of $.447m$ from the speaker and microphone array plane. The speaker and microphone array are placed at an oblique angle of $45°$ as shown in setup (3.2).

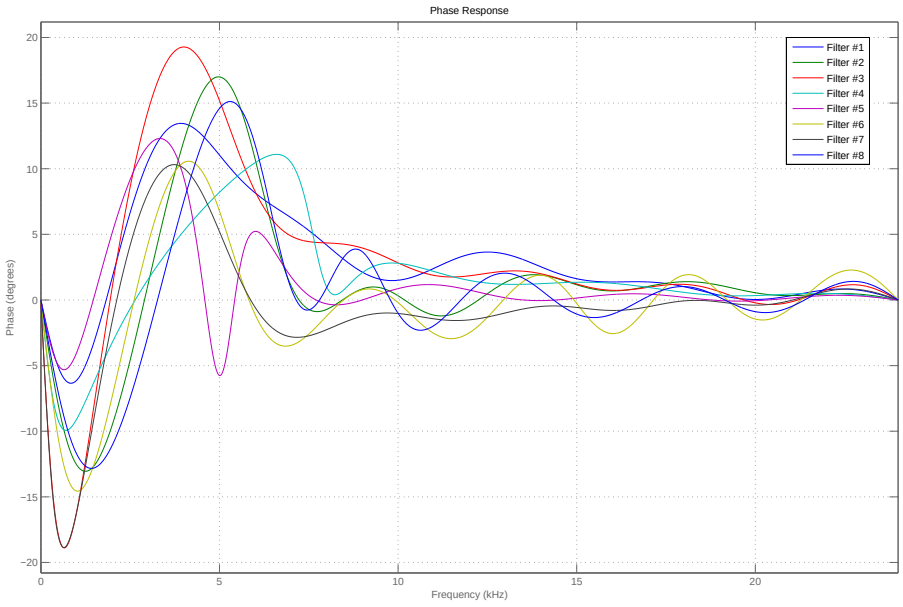


Figure A.1: Phase plot of transfer function obtained through different channels of a microphone array for plywood.

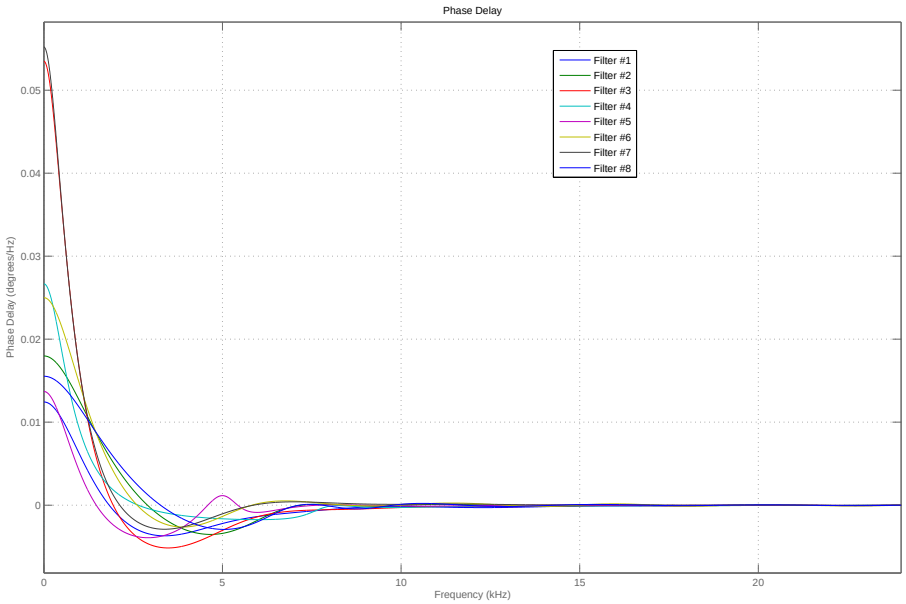


Figure A.2: Phase delay plot of transfer function obtained through different channels of a microphone array for plywood.

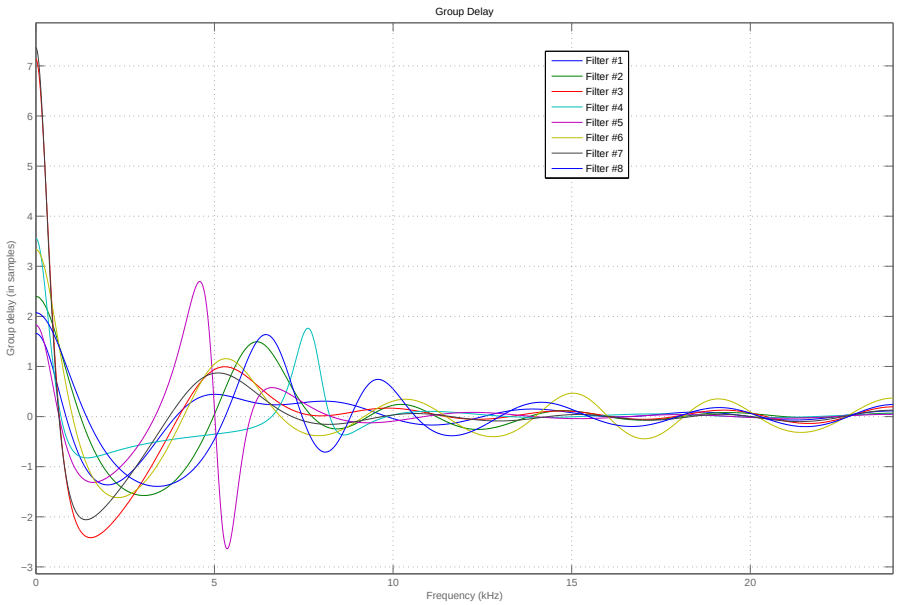


Figure A.3: Group Delay plot of transfer function obtained through different channels of a microphone array for plywood.

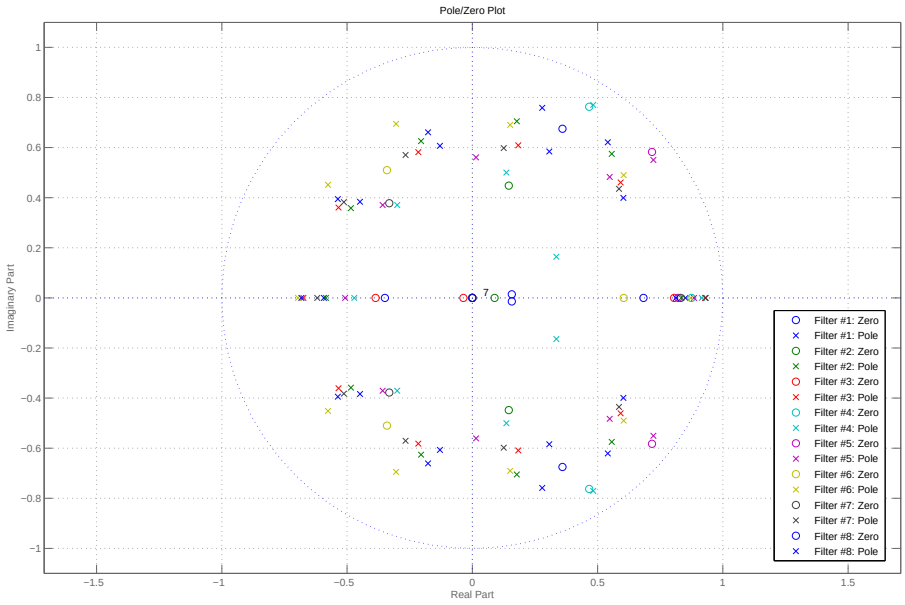


Figure A.4: Z-plot of transfer function obtained through different channels of a microphone array for plywood.