Predicting noise attenuation level in the earplugs using Gaussian Process Regression

Stephy Annie Curie

ALPINE HEARING PROTECTION

TUDelft

Predicting noise attenuation level in the earplugs using Gaussian Process Regression

by

Stephy Annie Curie Rakesh Arya

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Monday October 10, 2022 at 09:30 AM.

Student number:	5265886	
Project duration:	November 16, 2021 – October 10, 2022	
Thesis committee:	Dr. R. C. Hendriks,	TU Delft, Chair
	Dr. J. Martinez,	TU Delft, Daily supervisor
	Dr. C. J. M. Verhoeven,	TU Delft
	Arthur van Keeken,	ALPINE, Company supervisor

Cover: Women - Photo by Moustafa Khamis on behance (Modified). Laptop - Alex Varanese on behance (Modified)

Preface

This thesis is made as a completion of the master education in Electrical Engineering (Signals and Systems) and concludes my journey at Delft University of Technology. Though this journey was a bumpy road, I am really grateful for the exposure and opportunities provided by the University.

This thesis has been done in collaboration with ALPINE, Hearing Protection Company. I would like to thank Dr. J. Martinez and Arthur van Keeken (CEO of ALPINE) for giving me this tremendous opportunity. With having a very little knowledge on hearing protectors during the initial phases, this project intrigued me on investigating the vast potentials that hearing protectors can offer.

Firstly, I would like to express my gratitude to Dr. J. Martinez, my daily supervisor for constantly supporting me throughout this process. Your valuable suggestions and guidance helped me to refine my thesis to a greater extent. Also, I would like to show my deep appreciation to Dr. R. C. Hendriks who helped me finalize my project. I want to especially thank Arthur van Keeken and Fanny Hofstra, my company supervisors for offering the freedom to work on new ideas, providing deep insights into the study and all the resources that were needed for the project.

This endeavor would not have been possible without the help of my family members (Gnanavelu, Manonmani, Rakesh, Nirmala, and Patrick). Thank you all for encouraging me in pursuing my dreams. I would be remiss in not mentioning Lavinia Paoletti, you were there for me with a word of encouragement through my difficult times. Grazie mille for making my experience in the Netherlands so beautiful. Finally thank you Godwin, your support and encouragement is worth more than I can express on paper.

Nandri Vanakkam!

Stephy Annie Curie
Schiedam, September 2022

Summary

This thesis is done in collaboration with the Hearing protection company, ALPINE. The earplug development followed in ALPINE is a trial and error method. Due to which a large number of earplug samples get wasted, and the entire process is time-consuming and expensive. So, the objective of this thesis is to ease the error earplug development process by implementing a prediction model. The model must be able to predict the noise attenuation provided by an earplug based on the material and design specifications.

Firstly, the factors that influence the sound attenuation in an earplug are researched so that those factors can be included in the dataset. The dataset is separated into two: Venturi and Mesh datasets due to the diversity in the earplug design configurations. Both datasets are multidimensional due to several factors that influence the earplug noise attenuation. Eventually, the venturi dataset is built with 1160 data points and the mesh dataset with 1003 data points.

The noise attenuation data of all the earplugs in ALPINE are analysed to estimate the contribution of each earplug component. The component contribution analysis helped in understanding which component in an earplug actually works and provides sound attenuation.

Finally, to implement the most suitable prediction model, different regression models are analysed. With the available data and based on the analysis, Gaussian Process Regression (GPR) model is selected. With RBF kernel function and hyperparameter optimization, the GPR model is trained and tested on both datasets. A Mean Absolute Error of 3.5 and 3.26 is observed in venturi and mesh datasets respectively.

The main finding during the course of this thesis is that noise attenuation provided by an earplug is highly subjective. Testing the same earplug on two different target populations with different ages, gender, and ethnicity resulted in entirely different noise attenuation levels (A difference of 10 dB is noted). Sound perception is crucial in earplug design. Even though the most ideal sound attenuation prediction model for earplugs can be developed, in the end, everything relies on individual sound perception. Due to individual sound perception, prediction of noise attenuation in earplugs has higher level of uncertainty. However, if we have accurate age, gender, and ethnicity information, the earplugs can be modelled and designed for each group. In such cases, the sound attenuation prediction model can be developed with higher certainty.

The work done in this thesis should be considered the first step in developing a model for predicting noise attenuation levels in earplugs. For model validation, the earplugs have to be manufactured from scratch with new design specifications and it will take much longer than this thesis period. So model validation is suggested as an extension to this thesis.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
REAT	Real Ear-Attenuation at Threshold
IL	Insertion Loss
SPL	Sound Pressure Level
AVP	Assumed Value Protection
NRR	Noise Reduction Rating
SNR	Single Number Rating
HSE	Health and Safety Executive
RBF	Radial Basis Function
GPR	Gaussian Process Regression
ARD	Automatic Relevance Determination
L-BFGS-B	Limited-memory - Broyden - Fletcher - Goldfarb - Shanno with bound constraints
TNC	Truncated Newton Method
SCG	Scaled Conjuagte Gradients
MAE	Mean Absolute Error
SE-ARD	Squared Exponential - Automatic Relevance Determination

List of Figures

List of Tables

1
Introduction

World Health Organization estimated around 430 million people with hearing loss [1]. This hearing loss is mainly due to noise pollution [2]. The commonly used solution to protect people from over noise exposure is by using hearing protection devices such as earplugs and earmuffs [3]. Hearing protectors reduce the risk of hearing loss and the noise exposure level. The earplugs are designed for several applications such as Sleeping, Concerts, Motor Sports, etc and so each earplug is designed to provide a certain level of protection based on the application. ALPINE is one of the hearing protection companies that manufacture earplugs for the aforementioned applications [4]. However, the earplugs in ALPINE are developed by trial and error methods which have their own disadvantages and the main idea of this thesis is to overcome the disadvantages faced by the company due to the trial and error method of earplug development.

The following sections explains the current way of earplug development in ALPINE, the problems associated with it and the solutions to overcome it in order to understand the main objective of this thesis.

1.1 Current way of earplug development in ALPINE

The current way of earplug development is shown in Figure 1.1. The research and development team in ALPINE, first conceptualise the earplug design along with some prior knowledge. After analysing and designing the earplug for a certain application, the earplug samples are prototyped. Then these earplugs are tested on subjects and on ear simulators to assess the attenuation level. If the earplug doesn't achieve the required attenuation level, then these earplugs are discarded and redesigned until the required noise attenuation level is achieved. Hence, the earplug development in ALPINE is entirely a trial and error process. The disadvantages of this process are explained in the next section.

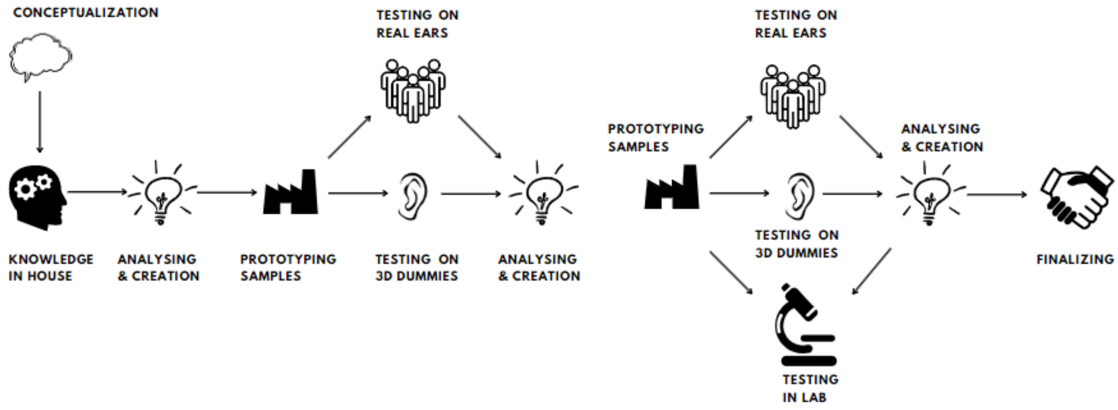


Figure 1.1: During the earplug designing phase, the earplugs are conceptualised, analysed and prototyped. Then these earplug are tested on subjects and on ear simulators to assess the attenuation level to check if the earplug provides the required attenuation level. If not, these earplugs are discarded and a new earplug is redesigned until the required noise attenuation level is achieved. Hence, the earplug development and testing process iterative and is a trial and error method [4]

1.2 Downsides of trial and error testing method

As mentioned in Section 1.1, the drawbacks of earplug development and testing methods carried out in ALPINE are as follows: For testing the samples on subjects, the earplugs are tested on 16 or 20 subjects. Since this process is iterative, it becomes expensive to do the testing on subjects and ear simulators for every iteration. The earplug design goes through several iterations of testing phase until it provides the required attenuation level. Hence, **large number of samples get wasted and it is expensive**. Another major downfall is that the entire process is **time-consuming**.

1.3 A Prediction model over trial and error testing method

To overcome the challenges mentioned in Section 1.2, a noise attenuation prediction model can be implemented. This model should be able to predict the noise attenuation level of the earplug given the material specifications and earplug designs. Based on the predicted noise attenuation level from the model, the product developers can fine-tune the earplug design before prototyping the samples and sending it to the testing laboratories. In this way the the number of iterations taken for earplug development process is drastically reduced and can easily overcome the downsides mentioned in Section 1.2.

1.4 Objective

The objective of this thesis is to

- Study the factors that influence the noise attenuation level in an earplug.
- Analyse the noise attenuation data to find how much each of the earplug component contributes to the total attenuation
- Analyse different prediction models
- Implement the most suitable prediction model that can predict the noise attenuation level of the earplugs given the parameters that has an influence on noise attenuation such as material specification and earplug design

1.5 Model Validation

It has to be noted that the newly developed model proposed in this thesis will be implemented and tested with the available data from the company. However, to validate the developed model, the earplugs have to be manufactured from scratch with new design specifications. The time involved in validating the model will take longer than the time allocated to carry out this thesis. Hence, the work done in this thesis should be considered as a first step in developing a model for predicting noise attenuation level in earplugs. The validation of the model is proposed as an extension of this thesis.

1.6 Thesis structure

This report is organised as follows; Chapter 2 first gives the preliminary background information required for this research such as testing method involved in an earplug testing, factors that influence the attenuation level in an earplug. Then various literature based on ear anatomy and physiology of hearing, material properties used for sound dampening are reviewed. Finally different prediction models are analyzed and compared. Chapter 3 explains data analysis along with how the earplug components' contribution to attenuation are calculated, why and how the Gaussian process regression model is used for predicting the attenuation of earplugs. Chapter 4 explains the analysis of data and the associated findings are presented. Finally the results of the predicted model are shown and analysed. Chapter 5, summarises the thesis and the future research prospects are given.

2

Preliminary Review

This chapter focuses on reviewing various literature on the factors that influence the noise attenuation level in earplugs (other than earplug components) and various regression models. However, to understand various terminologies associated with the earplugs, this chapter first introduces the required preliminary information to the reader. The basic information such as the earplug testing methods, construction of an earplug and its effect on providing sound attenuation are explained. Then the review starts by exploring the ear anatomy, physiology of hearing followed by the material properties that influences the sound dampening and the earplug performance in real world.

As mentioned in Section 1.5, since we cannot validate the developed prediction model, various prediction models are analysed in this chapter. By considering the advantages and disadvantages of those prediction models, the most suitable model for this application is selected.

The motivation of this chapter is summarised as follows:

- To understand which of the factors influence the noise attenuation level in earplugs
- To analyse different prediction models

2.1 Earplug Testing methods

Since this thesis revolves around earplug development and testing methods, it is important to understand how those earplugs are tested. The amount of noise reduction provided by an earplug is referred as its attenuation [5]. Hence, the earplug manufacturers provide the noise attenuation value in decibels (dB) for different frequencies in hertz (Hz). The EN352-2:2020 standards mentions the construction, design, performance and the minimum attenuation requirements for earplugs [6]. As per these standards, these earplugs are tested on 16 subjects for the frequencies 63, 125, 250, 500, 1000, 2000, 4000, 8000 Hz.

The **Real Ear-Attenuation at Threshold (REAT)** is the subjective earplugs testing method and sometimes referred as the gold-standard since it is used in many parts of the world [7]. The REAT measures the attenuation by determining the subject's binaural threshold of hearing without an earplug and then remeasure the subject's binaural threshold of hearing with an earplug [5]. Then the difference between the two hearing level is taken. This difference gives the attenuation level of the earplug (for the aforementioned frequencies) which is called the Insertion Loss (IL). Insertion Loss (IL) is the difference between sound pressure levels (SPL) that is measured at reference before and after a noise reduction treatment (in our case is the Earplugs) [5]. The subjective test method setup is shown in the Figure 2.1. The subject is seated at the centre of four loud speakers. The subject is given a button to press when he/ she first hears the sound that is played at a particular frequency and it is his/ her threshold of hearing. This testing is carried out one by one for all 16 subjects.

Figure 2.1: This is picture is taken from the PZT testing laboratory and certification body in Germany. This setup shows the subjective testing method where the subject is seated at the centre of four loud speakers. The subject is given a button to press when he/ she first hears the sound that is played at a particular frequency and it is his/ her threshold of hearing. The hearing threshold is measured with and without the earplugs. This testing is carried out one by one for all 16 subjects.

Once the earplugs are tested for 16 subjects, the standard deviation of the attenuation is subtracted from the mean, which gives the Assumed Value Protection (AVP) level for each frequency of that particular earplug.

Another method of testing is the objective method called as an Acoustic/ Hearing protector test fixture (Artificial ear). This method incorporates a microphone with a diaphragm, and measures the insertion loss by measuring the sound pressure level with and without the earplug. The earplug attenuation data provided by the acoustic test fixtures does not match with human earplug attenuation data. This might be because the objective methods do not directly account for all of the sound paths to the occluded ear, however bone conduction is either accounted through post measurement adjustments or ignored completely [5]. The objective tests are mostly done by the company during the product development phase to check if the earplug designed reached the required attenuation level. However, it has to be taken into account that most of the attenuation data in some frequencies are greater than the subjective test data. The Acoustic Test Fixture is shown in the Figure 2.2 and the small circular section is where the earplugs are fitted and tested.

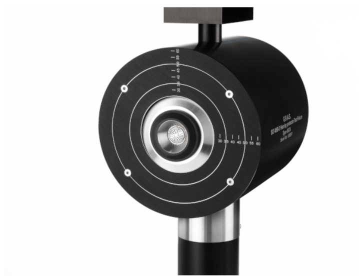


Figure 2.2: This figure shows the Acoustic Test Fixture with a microphone that measures the insertion loss by measuring the sound pressure level with and without the earplug. The earplug is inserted into the small circular opening in the middle [8].

2.2 Factors that influence the noise attenuation in earplugs

In Section 2.1, the methods of earplug testing are explained in brief. Now to implement a prediction model, first it is important to understand how the earplugs are designed to filter the unwanted noise. The sound that hits the earplugs can be absorbed, reflected or transmitted and it depends on the type of medium the sound travels through. Hence the construction and composition of earplug materials play a major role in sound attenuation. Earplugs consists of two main components: *Thermoplug* and *Filters* as seen in the Figure 2.3. The noise attenuation in the earplugs is greatly influenced by the thermoplug and filter. Based on the thermoplug material and the type of the filter used, the noise attenuation can be increased or decreased.

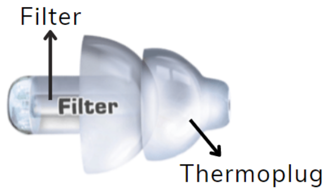


Figure 2.3: This figure shows the components in an earplug produced by the company ALPINE [4], where the outer shell like part is the thermoplug and the filter is inserted inside the thermoplug

2.2.1 Thermoplug Material

As mentioned in Section 2.2, the sound is either absorbed or reflected and it depends on the material used. The thermoplug is made up of rubber-like, soft material that is easily fitted into the ear canal of an individual. At first, it might seem to be just a material to hold the filter in place and fit the earplug into an ear. But thermoplug provides some sound dampening, and it can be confirmed by the noise attenuation test results of just the thermoplug without any filter inside (just air) which can be seen from the Figure 2.4. The x-axis is the Frequency in Hertz (Hz) and y-axis is the Noise attenuation in Decibels (dB). From the Figure 2.4, it can be seen that some sound dampening is provided by the thermoplug especially between 1000 - 4000 Hz. Between 125-500 Hz, attenuation is in negative decibels , this is because the equation for calculating the sound pressure level is relative to the threshold of hearing. The equation for calculating sound pressure level is given in Equation 2.1

$$\mathrm{SPL} = 20\log\frac{P}{P_o} \tag{2.1}$$

where P is the sound pressure of interest and P_o is the reference sound pressure of 0.00002 pascals = 0 dB the threshold of hearing [9]. Since sound pressure of interest is relative to reference sound pressure, if reference pressure P_o is larger than P, we end up in negative values. It means the sound pressure we measured is lesser than the threshold of hearing.

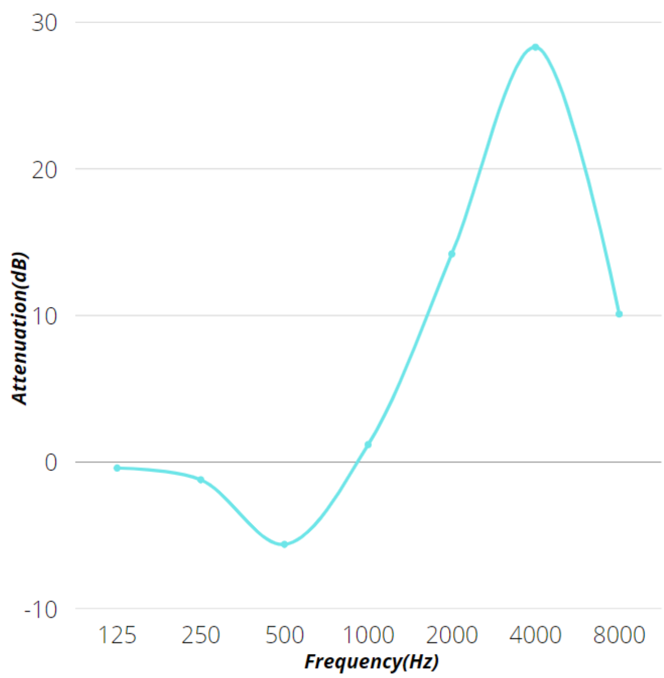


Figure 2.4: This graph shows the sound attenuation provided by the thermoplug component produced by the company ALPINE [4]. The thermoplug is tested without any filter inserted inside it (just an open area filled with air). The x-axis denotes the frequency in Hz and y-axis denotes the attenuation level in dB. It is observed that the thermoplug itself provides some sound dampening between 1000 - 4000 Hz

The sound dampening provided by just the thermoplug is because of the sound transmission and absorption properties of the material used. The thermoplug material used in ALPINE is a blend of polymers (both rubber and plastic). This material is widely used as damping materials because of the viscoelastic polymer phases (elastic behaviour, ability to recover after deformation) isolated from stiff polymer matrix (stiff segment, ability to resist the deformation) that can attenuate energy and dissipate it to heat [10].

2.2.2 Filter types

In Section 2.2.1, it was mentioned that the thermoplug offers sound dampening to certain extent due to its material properties. Also, the construction of filter inside the thermoplug contributes to sound dampening. One type of filter can be thought of as a cylindrical tube with a small opening (hole) of a certain diameter. So as this diameters increases, more sound waves are transmitted through the filter which results in less sound dampening. On the other hand, as the diameter of the opening decreases, less sound waves are transmitted and more sound dampening is achieved. This filter type, from now on will be mentioned as Venturi filter.

The second type of filter comprises mesh (with many small openings). As the size of openings is large, more sound waves are transmitted, therefore it results in less attenuation and vice versa. This filter type will be mentioned as a Mesh filter in the rest of the report.

2.2.3 Other factors

The noise attenuation level of the earplug is basically a number given by a laboratory tested on a group of subjects under ideal conditions. The noise attenuation is a frequency dependant, subjective random process. So, no two noise attenuation measurements for the same earplug tested on two different person will be the same. It is because, the ear shape and the hearing level differs from person to person. So the way each individual perceive sound might be different. Also, from the sound attenuation report provided by the testing laboratory, it can be seen that the sound pressure level measured in decibels differs per frequency. Hence, this has to be taken into account while predicting the noise attenuation level of the earplugs.

Based on analysing the sound attenuation data and studying the construction of the earplug from ALPINE, the factors that has an influence on noise attenuation level in earplug are shown in Figure 2.5. The noise attenuation provided by an earplug has a higher level dependency on subjects and frequency. Further, per subject and per frequency the noise attenuation depends on the earplugs components such as thermoplug and filter.

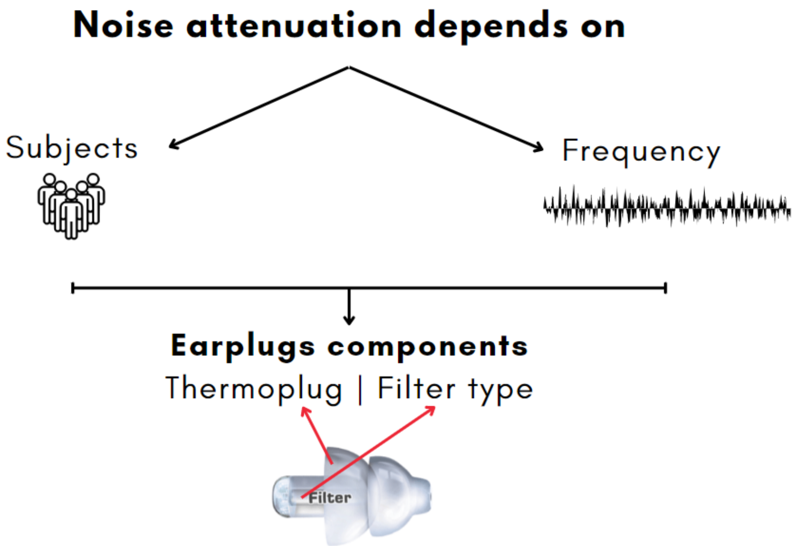


Figure 2.5: The noise attenuation provided by an earplug has higher level dependency on subjects and frequency. Further, per subject and per frequency, the noise attenuation depends on the earplugs components such as thermoplug and filter.

Hence the effects of the earplug components are mainly analysed to understand how much each of the components contribute to the total attenuation based on prior data available from the company. This analysis is further explained in Chapter 3

2.3 Ear Anatomy

In Section 2.2, the factors that provide noise attenuation in earplugs are explained. Apart from that, it is also important to understand how hearing works in humans which is useful in designing an earplug. If an earplug is designed in a such a way that is neither easily inserted into the ear canal nor sufficiently protective, it might not provide the level of noise attenuation proposed by the product manufacturer [11]. Earplug insertion and comfort are dependent on the external auditory canal shape, which may be represented in a variety of shapes and sizes in humans [11]. Hence, it is important to explore the ear anatomy and how it influences the earplug.

The parts of the ear are external, middle and inner as shown in Figure 2.6. Firstly, the external ear comprises of the pinna, ear canal and ear drum. The middle ear has three small bones called malleus, incus and stapes and these bones are collectively called as the ossicles. The ear drum connects to the malleus, and the stapes connects to the inner ear. The inner ear is mainly focused on hearing and balance. Cochlea is filled with special fluids and is responsible for the hearing. It has sensory hair cells which are connected to the central hearing system by auditory nerve [12].

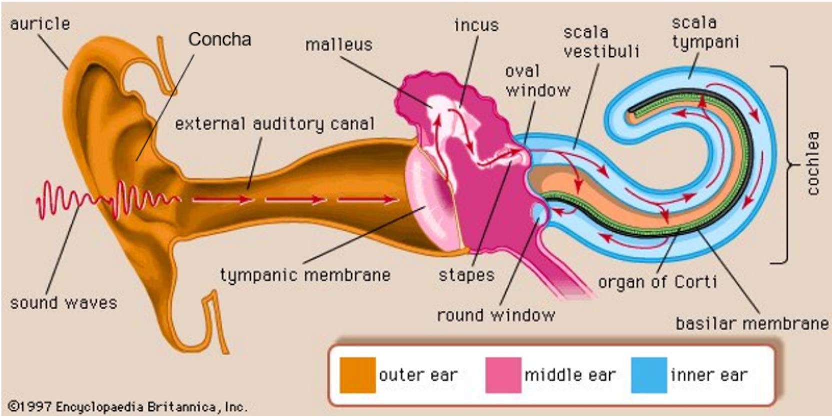


Figure 2.6: Ear Anatomy depicting the sound propagation from outer ear to inner ear [13]. The sound waves are collected by the pinna in the outer ear. These waves travels through the ear canal, hits the eardrum and causes the eardrum to vibrate. This vibration of eardrum moves the ossicles in the middle ear thereby transferring the sound vibration into the inner ear [12].

Now that we have seen the anatomy of the ear, it is important to understand the physiology of hearing and the auditory function because it helps in designing an earplug focusing on attenuation in a particular frequency range.

2.3.1 Physiology of hearing

The sound waves are collected by pinna. Pinna focuses the sound energy into the eardrum. These sound waves travel through the ear canal and hits the eardrum and causes the eardrum to vibrate. This vibration of eardrum moves the ossicles thereby transferring the sound vibration into the cochlea. As the stapes vibrates, the fluids in cochlea moves in a wave-like manner thereby stimulating the hair cells [12]. The hair cells are arranged in such a way that it responds to different sounds based on the frequency. High frequency sounds stimulates the lower part of the cochlea and low frequency sounds stimulates the upper part of the cochlea. Figure 2.7 shows how the hair cells in cochlea detects different frequency ranges. Based on the stimulation of the hair cells in the cochlea with respect to the frequency of the sounds, the hair cells sends nerve impulses through the auditory nerve that reaches the central hearing system called the auditory cortex. Auditory cortex is where the nerve impulses are converted into meaningful sound [12].

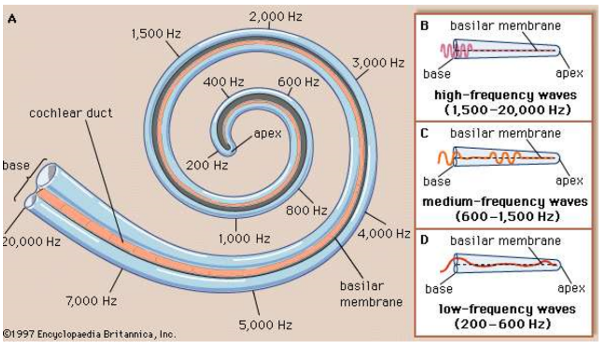


Figure 2.7: Cochlea depicting how the hair cells in cochlea detects different frequency ranges. High frequency sounds stimulates the lower part of the cochlea and low frequency sounds stimulates the upper part of the cochlea [13].

There are some important auditory functions which might be useful in designing an earplug for a particular application and it is stated as follows:

- The ear canal amplifies the incoming sound waves around 3000 Hz (Common human speech range) [14]

- The outer ear amplifies the incoming sound in the frequency range of 1500-7000 Hz about 10-15 dB [15]
- The outer and middle ear together amplifies the incoming sound about 30dB [16]
- The ear is most sensitive to the frequency around 3000-4000 Hz because of the aforementioned amplifying mechanism and so an intense stimulus is produced at these frequencies and the hair cells which responds to these frequencies are at high risk from damage [16]. It has to be noted that hearing loss caused by the noise exposure first occurs at 3000-4000 Hz and this is because of this amplifying mechanism [16]

Based on the aforementioned amplifying mechanisms of ear, it is seen that ear is most sensitive around 3000-4000Hz and the damage occurs first that frequency range. So, **it is an important point to be noted while designing an earplug, where the focus must be on choosing an earplug material and designing an filter in such a way that an earplug provides more attenuation around 3000-4000 Hz**.

2.3.2 Ear Canal length

The ear canal is curved and it is about 2.5 cm (approximately 25mm) long [17]. Figure 2.8 shows the ear canal as a straight tube [18]. The ear canal has two bends, first bend around 8mm and second bend around 15mm. The narrowing of ear canal is between the first and second bend [19]. When the earplug is fitted deeply enough, the narrowing between the first and second bends facilitate an airtight seal [20]. Based on the experimentation results given in [21], it was observed that attenuation increased with canal segment length at a rate of approximately 1.5 dB/mm up to a length of 14 mm. Also, it was also observed that the second bend is around 15mm, and by inserting an earplug up until 15mm the attenuation increased [21]. **So from the aforementioned research, it is inferred that designing an earplug with longer canal segment which reaches the second bend of the ear canal, increases the attenuation**.

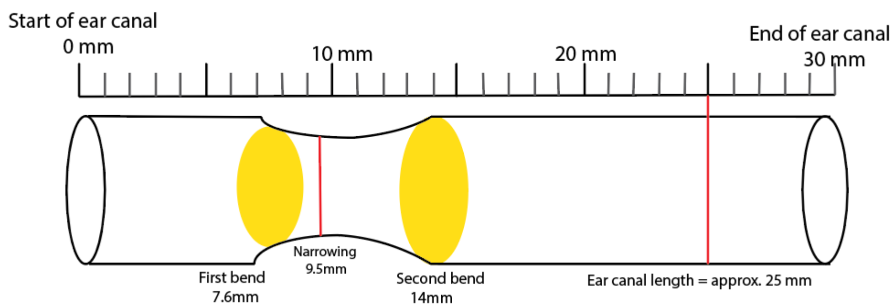


Figure 2.8: Ear canal elongated as a straight tube, showing the first and second bend of the ear (yellow shaded region) and the airtight seal provided by the narrowing of the tube between first and second bend [18].

The aforementioned factors such as ear canal length, earplug size and earplug length involves a diverse and highly varying values since these parameters vary from person to person. There are not many data available from the company regarding the earplugs that has different ear canal length, earplug size and earplug length. Hence, it is assumed that these factors does not affect the predictions. So, these factors are not included as input features in the dataset. However, these factors were given as earplug design suggestions to the company, ALPINE.

2.4 Material Properties

In Section 2.3, the Ear anatomy is explored in detail to understand the working of a human ear and to make earplug design considerations based on that. Additionally, material selection plays an important role not only in providing comfort to the user but also in noise attenuation. To the best of my knowledge, there are not many studies conducted on the material selection of earplug which provides noise reduction. However, with the focus on reducing noise pollution, there are various materials available for soundproofing and sound dampening in buildings and acoustic rooms. These materials work mainly on the sound absorption technology which transforms the vibration energy of the sound into heat. Hence, these studies can be used to understand the material properties of earplugs.

ALPINE focuses on using materials that don't affect the comfort of the user i.e. it should be easily fitted inside the ear, doesn't cause any discomfort while using it and so soft rubber-like materials are used. Thermoplug is the part that goes into the ear canal hence, the thermoplug material should be made of soft rubber-like material. Then, depending upon the application the filter varies. For example, in the case of developing an earplug for sleeping, then the entire earplug should not have any hard part since it might cause discomfort to the user because of the sleeping posture (the ear is squeezed while sleeping). In such a case, the filter can be made of a soft material. In other applications, such as developing an earplug for concerts, the user might be mostly in an upright posture and in such a case the filter material does not need to be soft. So, the filters is made up of hard materials such as plastic. Now we can see that the thermoplug is made of soft materials and the filter is made of either soft or hard materials depending on the application. So, to understand how these materials contribute to sound dampening, it is important to understand the mechanisms of sound propagation in materials.

2.4.1 Mechanism of Sound propagation

As the sound energy hits a surface (Incident energy), part of its energy is reflected (Reflected energy) and the non-reflected part is absorbed (Absorbed energy) by the material and the rest is transmitted (Transmitted energy). Hence, the incident energy is given as the sum of reflected energy, absorbed energy and transmitted energy. If more energy is absorbed then less energy is transmitted and vice versa [22]. Absorbent materials are mostly elastic, and these materials are soft or fibrous [22]. If the sound wave hits a soft material sound energy is absorbed due to the deformation of the material and if the material is porous, the sound wave is absorbed due to the vibration of the air inside the pores thereby losing the energy by friction against the edges [22]. In ALPINE, elastomers are used for thermoplug and amorphous polymers are used for the filter. In [23], it has been mentioned that typically the elastomers and amorphous polymers show higher sound absorption properties . So, the inference is that the materials used for thermoplug and filters in ALPINE provide noise attenuation by sound absorption. In the case of hard materials used in filters, general idea is that hard materials reflect sound [24]. But, a hard material like plastic which is below 10 millimetres in thickness cannot reflect sound [25]. The filter material used by ALPINE falls under amorphous type polymers and based on the paper [23], these types of hard polymers have sound absorption properties. So it is concluded that the hard polymers used for filter manufactured in ALPINE provide noise attenuation by sound absorption.

It is now clear that the two types of materials used for earplugs in ALPINE are hard polymers and soft, flexible elastomers. These two materials provide sound attenuation mostly by sound absorption.

The study on material properties is done to understand how the material contributes to sound attenuation and it will be useful information for the company to design further products based on this study.

2.5 Performance of earplugs in Real world

In Section 2.4, an explanation to sound attenuation due to material properties is given. We design an earplug and test it in laboratories expecting the earplug to provide a certain level of attenuation. However, laboratory attenuation measurements and real-world field measurements are not the same [26]. Since the Noise Reduction Rating (NRR) (in our case it is a Single Number Rating) is based on laboratory testing, it does not take into account the loss of protection that occurs when hearing protectors are not fit properly or when they are not worn for the entire time that the wearer is exposed to noise [27]. For most wearers, the Single Number Rating identified on the current product label significantly overestimates the protection of the hearing protector in the workplace [27]. The results of the experiments done in [28], confirm the overestimation of the values indicated by the manufacturers. From 3 to 5 dB at high frequencies, the discrepancy reaches up to 8 to 10 dB at medium and low frequencies. Health and Safety Executive (HSE) recommends derating the attenuation of hearing protection by 4dB when estimating attenuation provided under real-world conditions [29]. Hence based on the aforementioned study on the performance of earplugs on real-world, a suggestion to ALPINE is that **while designing an earplug, the company must take into account the real-world issues and design an earplug that might be 4dB higher than the attenuation required for a particular application. This ensures that the user can wear an earplug without having to worry about real-world discrepancies.**

2.6 Analysis of Prediction models

The goal of this thesis is to predict the noise attenuation provided by the earplugs based on design specifications. So, a prediction/ regression model is required. The selection of a particular regression technique depends on the relationship between input and output variables, the output variable type, the number of input variables and so on. As mentioned in Section 1.5, it is not possible to validate the developed prediction model in this thesis. So, to use a prediction model for our application with a certain degree of trustability, various prediction models are reviewed in this section. By analysing the advantages and disadvantages of those prediction models, the most suitable model for this application is selected.

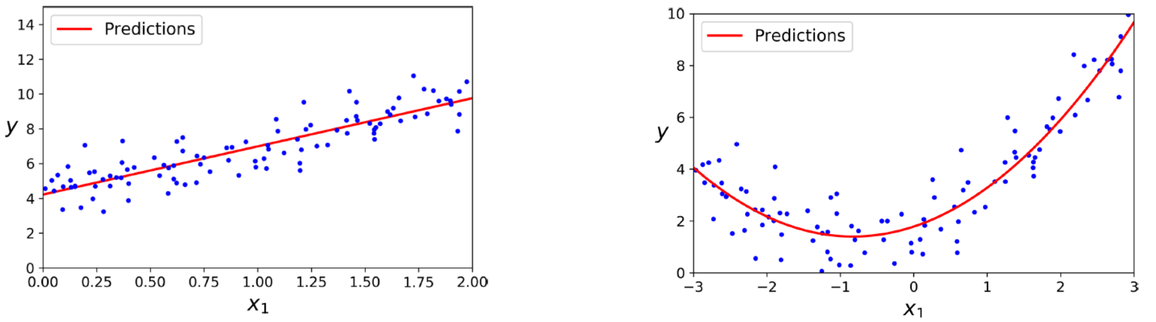
2.6.1 Linear Regression

The most commonly used regression model is the Linear Regression, wherein the relationship between input and output variables is linear as shown in Figure 2.9a. The linear model makes predictions by calculating a weighted sum of input features plus a constant [30] and the equation is given in Equation 2.2

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_1 + ... + \theta_n x_n \tag{2.2}$$

where, $\hat{y}$ is the predicted value, n is the number of input features, x_i is the i^{th} input feature value, θ_0 is the constant term and θ_j is the j^{th} model parameter (input feature weights).

In linear regression, the input feature weights are set in such a way that the model best fits the data, so that we have accurate predictions. Though linear regression is a simple technique that is implemented easily, this method can easily underfit, meaning linear regression assumes a linear relationship between the input and output variables. In such a case, **if the relationship between the input and output variables is not linear or if the dataset is complex then the model fails to capture the relationship and results in low accuracy prediction** [31]. Also, in some cases **the inputs variables are not independent (multicollinearity), but linear regression assumes that the input data is independent and so the correlation between the input variables must be removed before using linear regression technique** [31].



(a) Blue dots shows the observations and red line shows the predicted line using Linear Regression

(b) Blue dots shows the observations and red curve shows the predicted curve using Polynomial Regression

Figure 2.9: Plots showing how the data points are fitted with Linear and Polynomial Regression [30]

2.6.2 Polynomial Regression

In case of a non-linear relationship between the input and output variables, the linear regression cannot capture the data relationship in such cases polynomial regression can be used as shown in Figure 2.9b. Polynomial regression is a way of using a linear model to fit non-linear data. This method adds powers to each input feature as shown in Equation 2.3

$$\hat{y} = \theta_0 + \theta_1 x + \theta_2 x^2 \tag{2.3}$$

where, $\hat{y}$ is the predicted value, x is the input feature value, θ_0 is the constant term and θ_1, θ_2 is the model parameter (input feature weights).

In this equation, the weights are still linear however the curve that fits the data is quadratic in nature. We can further increase the degree from 2 to how much ever we want so that the curve best fits the data. But high degree polynomial regression sometimes overfits the data [30], which means the model captures the noise/ outliers in the data. To avoid this we can increase the number of training data. **In the case of multivariate polynomial regression i.e. multiple input features, the problem is again multicollinearity** [32] as in linear regression. Also, **if the number of input features in the dataset is high, then multivariate polynomial regression gets quite complex**.

2.6.3 Ridge Regression

Overfitting is a serious issue and it is reduced by regularization i.e. to constrain the degrees of freedom a model has, In polynomial regression, to regularize it, we reduce the number of polynomial degrees and in linear regression we constrain the weights of the model [30]. Ridge regression is a regularised version of Linear regression meaning a regularization term is added to the cost function, by which the model not only fits the data but also constrains the weights to be as small as possible [30]. The ridge regression cost function is given by Equation 2.4

$$J(\theta) = \mathrm{MSE}(\theta) + \frac{\alpha}{2} \sum_{i=1}^{n} \theta_i^2 \tag{2.4}$$

where MSE is the mean square error between the true target values and predicted values and it is a function of θ, and θ is the model parameter, α is the hyperparameter/ the penalty term which controls how much we want to regularize the model. When data has multicollinearity, the α solves it [33]. By changing α values, the penatly term is controlled. If α is large, penalty increases and the magnitude of weights decreases and ends up close to zero [30], [34]. As α increases the bias increases and variances decreases.

If we have a vector of input feature weights, the the regularization term is given by Equation 2.5 [30]

$$J(\theta) = \mathrm{MSE}(\theta) + \frac{\alpha}{2} \|x\|_2 \tag{2.5}$$

where $||x||_2$ is the l_2 norm of the weight vector.

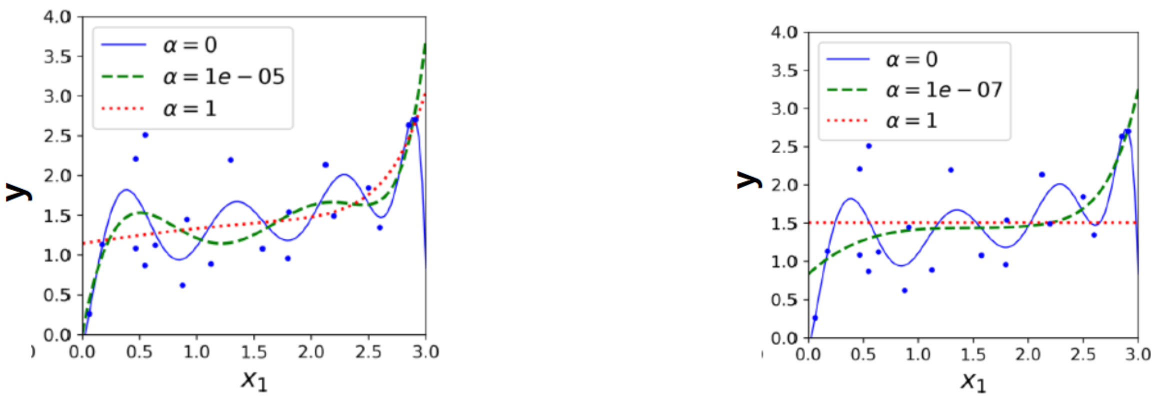
2.6.4 LASSO Regression

LASSO regression is same as ridge regression but the regularization term uses l_1 norm of the weight vector [30] and it given by Equation 2.6

$$J(\theta) = \text{MSE}(\theta) + \alpha \sum_{i=1}^{n} |\theta_i| \tag{2.6}$$

where MSE is the mean square error between the true target values and predicted values and it is a function of θ, and θ is the model parameter/ weight vector, α is the hyperparameter/ the penalty term which controls how much we want to regularise the model.

LASSO regression, eliminates the weights of less important input features in the dataset [30]. The selection of input features is arbitrary in nature and usually, prediction performance is worse than ridge regression [35]. Figure 2.10a shows polynomial regression with ridge regularization, where, as α increases, the predictions become more flat and reasonable, with reduced variance and increased bias. Figure 2.10b shows polynomial regression with LASSO regularization with smaller α values. From Figure 2.10, it can be seen that at $\alpha = 1e-05$, the prediction accuracy is better in ridge regularization when compared to LASSO.



(a) Polynomial regression with ridge regularization, where, as α increases, the predictions become more flat and reasonable

(b) Polynomial regression with lasso regularization, where, as α increases, the predictions become completely flat

Figure 2.10: Plots showing how the data points are fitted with Ridge and Lasso Regression [30]

2.6.5 Decision trees Regressor

Decision trees are capable of performing regression on complex datasets [30]. It is a tree-structured regressor with a Root node, Interior nodes and finally Leaf nodes. The root node represents the entire sample and it gets split up into interior nodes which represent the input features of the dataset. The branches are the decision rules and finally, the leaf nodes represent the output. The leaf node is reached by answering yes/ no questions from the root node. If the leaf node cannot be split up anymore, then that particular leaf node is taken as the final prediction. The final predicted value is the average of the value of the output in that the particular leaf node and this goes through multiple iterations [36]. For example, Figure 2.11 shows the dataset and related decision tree regressor where the Root node i.e. Outlook has three branches of interior nodes namely Sunny, Overcast and Rainy. Each of the interior nodes is further split up into the interior nodes based on the dataset, The leaf node represents the output/ target variable [37]. Decision trees are easy to interpret and such models are called white-box models

[30]. It can handle both categorical (names) and numerical variables. Also, decision tree is not affected by missing values. However, **when the input variables are uncorrelated, this regressor does not work well** [35]. Also, this algorithm can be **inadequate for predicting continuous values**. The main issue with the decision tree algorithm is that it is **very sensitive to small variations in the dataset** [30]

Outlook	Temp	Humidity	Windy	Hours Played
	Predictors			Target
Rainy	Hot	High	False	26
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	62
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

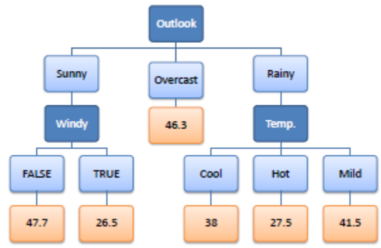


Figure 2.11: On the left, an example dataset with predictors and target is shown. On the right, related decision tree regressor with root node, interior nodes and the leaf nodes (predicted values) is shown [37]

2.6.6 Random Forests

Decision trees are the fundamental concept of random forests and random forests are an ensemble of decision trees and each tree is different as shown in Figure 2.12. Each tree makes its own prediction and the final prediction is the average of all prediction output values from all the trees. Decision tree searches for best input feature while node splitting whereas random forest splits then node randomly which results in great tree diversity and has high bias and low variance [30]. Random forests are considered to be a **black box approach and it becomes hard to explain why those predictions were made** [30]. Random forests can **overfit the data and might not work with small dataset**.

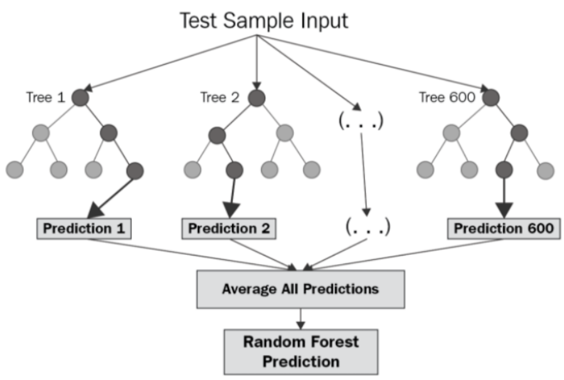


Figure 2.12: An example of Random Forest Regressor is shown, which is an ensemble of decision trees where each tree makes its own prediction and the final prediction is the average of all prediction output values from all the trees [38]

2.6.7 Bayesian Inference for Gaussian Process Regression

The goal of any regression model is to best fit a line/ curve through the given data and use it for prediction purposes. In such cases, for example in linear regression $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, the goal is to estimate the weights i.e. θ that best fits the data. Here, θ is the unknown parameter, and so we have to estimate it.

Based on the mathematical treatments of the unknown parameters in the model, the machine learning models are categorized into deterministic/ Frequentist method and Random/ Bayesian method [39]. In frequentist approaches the unknown parameters that have to be determined are considered to be nonrandom variables and the goal is to estimate a single value for them [39]. In the Bayesian approach, the unknown parameters are random variables and the goal is to estimate the probability distributions of them [39]. Usually, the Bayesian approach allows us to incorporate prior knowledge that we have on our data into the model, unlike the frequentist approach which works on hypothesis and not considering prior beliefs.

Bayesian statistics works based on Bayes' Theorem and it calculates conditional probability (Probability of one event given another event). Assume $\mathbf{X}, \theta$ are statistically dependant random vectors, then Bayes' theorem is given by the Equation 2.7

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} \tag{2.7}$$

where $p(\theta)$ is the prior distribution which is the prior knowledge we have on θ, $p(\mathbf{X}|\theta)$ is the likelihood function and $p(\mathbf{X})$ is called the evidence i.e. the probability of the input data $\mathbf{X}$ also called as the marginal probability which means the probability of particular event irrespective of the other random variables' outcome. Finally, $p(\theta|\mathbf{X})$ is the posterior distribution

In short the Bayes' theorem is given as

$$P_{po} = \frac{L * P_{pr}}{E}$$

where, P_{po} is the Posterior function, L is the likelihood function, P_{pr} is the prior function and E is the Evidence or the Marginal probability.

So Bayesian workflow is given as follows:

- **Prior Distribution**: Capturing the prior knowledge available about a given parameter [40].
- **Determine Likelihood Function**: This is determined using the parameters' information in the observed data [40].
- **Posterior Distribution**: Combines the prior distribution and likelihood function using Bayes' theorem. Posterior distribution updates one's knowledge on a particular problem using the prior knowledge [40].

Reasons to choose Bayesian Inference over Frequentist approach:

- As mentioned earlier, if we have some knowledge about our data, we can encode this information directly into the prior distribution.
- In case, if we have a small dataset, the frequentist approach can incorrectly fit the data. However, in Bayesian inference, prior distribution over the parameters helps to prevent incorrect fit and this prior distribution acts as a regularization term [41].
- Bayesian inference captures the uncertainty in our data. With the 95% confidence intervals, we can say that we are 95% sure that the unknown parameter that we estimated lies between some data values [41].

Gaussian Process Regression (GPR)

A Gaussian process is a stochastic process and it is characterised by the mean and covariance function [42]. The parameter definitions and equations in this section follow the textbook [42].
If we have two or more random variables that follows Gaussian distribution, then it is said to be Multivariate Gaussian distribution. If a vector $\mathbf{Y} \in \mathbb{R}^d$ has a multivariate Gaussian distribution, then it is given as

$$\mathbf{Y} \sim N_d(\mu, \Sigma)$$

where, $\mu \in \mathbb{R}^d$ is the mean vector and $\Sigma \in \mathbb{R}^{d \times d}$ is the covariance function.

The probability density function for multivariate Gaussian distribution is given as follows:

$$p(\mathbf{x}|\mu,\Sigma) = |\Sigma|^{-\frac{1}{2}}(2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$$

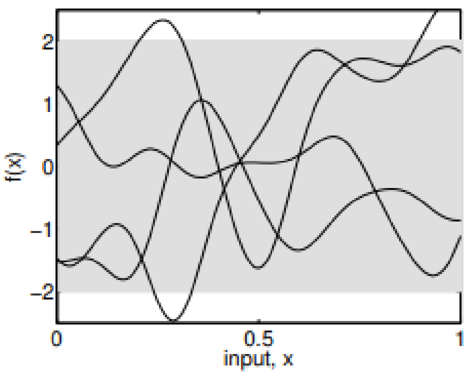
where d is the number of dimensions, $\mathbf{x}$ is the vector of input variables, μ is the mean vector and Σ is the $d \times d$ covariance matrix.

A multivariate Gaussian regression function is given as follows:

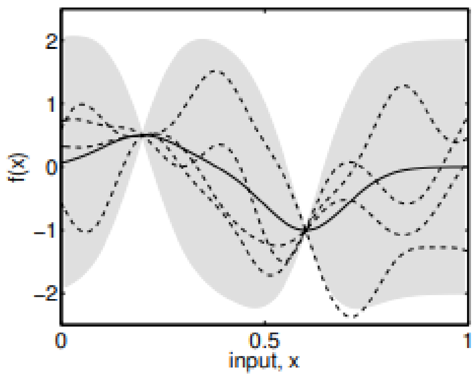
$$p(\mathbf{f}|\mathbf{X}) = N(\mathbf{f}|\mu,\mathbf{K})$$

where $\mathbf{X} = [x_1, x_2, ..., x_n]$ is the vector of input features/ data points, $\mathbf{f} = [f(x_1), f(x_2), ..., f(x_n)]$ is the latent, $\mu = [m(x_1), m(x_2), ..., m(x_n)]$ is the mean function and $\mathbf{K}_{ij} = k(x_i, x_j)$ is the covariance/ kernel function.

Figure 2.13, shows the GPR, where Figure 2.13a shows sample functions drawn at random from prior distributions specified by Gaussian process [42]. Let us say we have no knowledge of our data, so we assume the mean to be zero. The shaded region in Figure 2.13a, denotes the prior covariance we have assumed and in this example, the prior covariance does not depend on x [42]. Let's say we observe two data points $(x_1, y_1), (x_2, y_2)$. From Figure 2.13b, we see that the function passes through the two data points exactly and solid line is the mean values of the prior functions. The uncertainty is reduced near the data points and the shaded region depicts the uncertainty is observing the points other than the given data points [42].



(a) Sample functions drawn at random from prior distributions specified by Gaussian process with zero mean and shaded region being the covariance function



(b) The function passes through two new data points. Uncertainty reduces near the two data points. Uncertainty (shaded region) is large where the data points are not observed. Solid line is the mean values of the prior functions

Figure 2.13: Gaussian Process Regression with prior and posterior distributions [42]

Covariance/ Kernel functions

Covariance function is a similarity measure between two different target values and it determines the nature of the Gaussian process [42]. The parameter definitions and equations in this section follow the textbook [42]. The covariance function is given by

$$k(x,x') = \text{cov}\left(y(x), y(x')\right)$$

where cov is the covariance function, $y(x)$ is the target value of at x and $y(x')$ is the target value at x' The properties of the covariance function are:

- k must be symmetrical and positive semi-definite function
- k is assumed to be stationary i.e. covariance function depends on the distance between the inputs i.e.

$$\text{cov}(y(x), y(x')) = k(x - x')$$

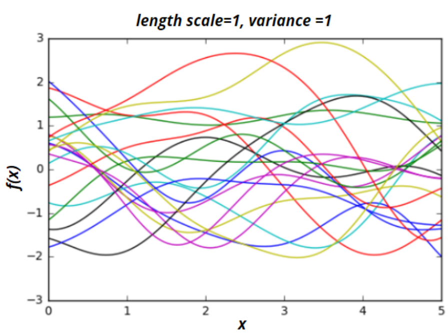
- Sometimes, covariance function is assumed to be isotropic, which means covariance only depends on the magnitude of the distance and not on the direction.
- If input data has more than two features, the multidimensional covariance function is a product of one dimensional covariance function for each of the input feature.

Types of kernel functions are

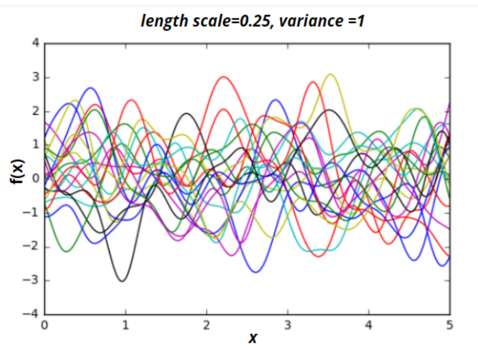
- **RBF/ Squared exponential kernel function**: It is a squared distance between two input values. It is most commonly used kernel function and given by:

$$k(x, x') = \exp\left(-\frac{1}{2\sigma^2}\frac{(x - x')^2}{l^2}\right) \tag{2.8}$$

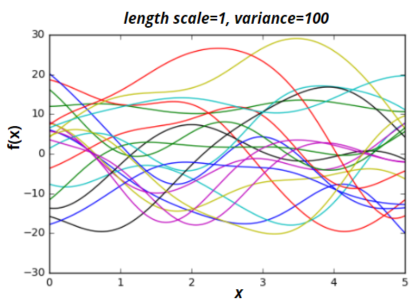
  where l is the length scale parameter and it determines the length in which the function should vary. If $l = 1$, the curves vary at the distance of 1. As the length scale decreases, the kernel function becomes more wiggly as shown in Figure 2.14a and Figure 2.14b. σ^2 is the variance which scales the y-axis as variance increases as shown in Figure 2.14a and Figure 2.14c. RBF kernels are infinitely differentiable, if we know the value at one location and all the derivatives then we will know the function value everywhere [42].



(a) Random functions drawn from Gaussian process with RBF kernel with lengthscale =1, variance=1 [42]



(b) Random functions drawn from Gaussian process with RBF kernel with lengthscale =0.25, variance=1 [42]



(c) Random functions drawn from Gaussian process with RBF kernel with lengthscale =1, variance=100 [42]

Figure 2.14: Random functions drawn from Gaussian process with RBF kernel with with varying lengthscale and variance

- **Matern kernel function**: Often, it is not required to use the models that are so smooth like RBF kernel functions. So in such cases, Matern kernel function is used and it is given by:

$$
\begin{aligned}
k(x,x') &= \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|x-x'|}{l})^{\nu}K_{\nu}(\frac{\sqrt{2\nu}|x-x'|}{l}\right) \\
k(x,x') &\sim ((1+|x-x'|)exp\left(-\frac{|x-x'|}{l^2})\right)
\end{aligned}
\tag{2.9}
$$

where K_ν is a modified Bessel function, l is the length scale and ν is an hyper parameter that determines the smoothness of the function and says how many times the kernel function can be mean square differentiated. Making $\nu = \infty$ in Equation 2.10 will then make it a RBF kernel. The equation for Matern 3/2 is given in Equation 2.10 and can be mean square differentiated only once. The equation for Matern 5/2 is given in Equation 2.11 and can be mean square differentiated twice.

$$
k_{\nu=3/2}(x,x') = \left(1+\frac{\sqrt{3}|x-x'|}{l})\exp(-\frac{\sqrt{3}|x-x'|}{l}\right) \tag{2.10}
$$

$$
k_{\nu=5/2}(x,x') = \left(1+\frac{\sqrt{5}|x-x'|}{l}+\frac{5r^2}{3l^2})\exp(-\frac{\sqrt{5}|x-x'|}{l}\right) \tag{2.11}
$$

The kernel becomes one when two inputs are the same and become zero if the distance between two inputs is big. The prior distribution has to be specified in GPR. This prior is specified by the covariance function.
Choosing the hyperparameters is the crucial step in choosing a kernel function. In RBF and Matern kernel functions, the hyperparameters are the process variance and the lengthscale. Hence, to choose the best hyperparameter that best fits the data, hyperparameter optimisation is done. The optimal hyperparameters are chosen by log marginal likelihood [42] by maximising the log of the likelihood function. Once the hyperparameters are found, the predictions are done. The covariance function for our application will be selected based on the nature of our dataset and its performance on the model.

To summarise, GPR is used in many machine learning applications because of its flexibility and uncertainty measures over predictions [43]. If we have a small dataset with lots of uncertainty in the measurements and we have some prior knowledge on our data then GPR can be a viable option to use as a model for making predictions.

The following section will compare and analyse the advantages and disadvantages of reviewed prediction models. Based on the analysis, the most suitable model will be selected for the noise attenuation prediction in earplugs.

2.7 Advantages and Disadvantages of reviewed prediction models

Table 2.1: Advantages and Disadvantages of reviewed prediction models

Prediction Models	Advantages	Disadvantages
Linear Regression	Simple Technique	Prone to underfitting. If the input-output relationship is not linear and complex then it results in low accuracy predictions. It assumes that the input data is independent
Polynomial Regression	Fits non-linear data	Prone to overfitting. Sensitive to outliers and it assumes that the input data is independent
Decision Tree	Handles both categorical and numerical values. Not affected by missing values.	Does not work well if the input variables are uncorrelated [35]. Insufficient for continuous value prediction. Sensitive to small variations in the dataset [30]
Random Forest	Works well with non-linear data. Handles both categorical and numerical variables. Not sensitive to outliers	Black box approach. Prone to overfitting and works well only with large dataset. Might be biased to categorical variables [44]
Gaussian Process Regression	Non-linear regression and a non-parametric model. Gives uncertainty measures over predictions. Works well on small dataset. Incorporates prior knowledge. Different kernels can be specified [45]	Kernel tuning [46].

2.8 Main Inferences

The main inferences based on the review of the existing literature related to this thesis are summarised by the following points

- Ear is most sensitive at 3000-4000 Hz and the damage occurs first at this frequency range. Hence, an earplug can be designed in such a way that it provides maximum protection around 3000-4000 Hz [16]
- Studies show that an earplug with a longer ear canal segment that can reach up until the second bend of the ear provides high attenuation [21]. Hence, an earplug can be designed with an ear canal segment of length around 14/ 15mm. However, comfort should be taken into account while doing so, since earplugs with longer ear canal segments might be uncomfortable for the user to wear
- Based on the review of various prediction models, GPR might work well for the data available. Since, the input-output relation is non-linear, the gathered dataset is small and it has high subjective uncertainty in the noise attenuation measurement, GPR might be the optimal choice. Detailed explanation on the dataset and GPR model is given in Chapter 3

2.9 Research Questions

1. What is the contribution of each earplug component to the total attenuation?

2. Due to high subjective uncertainty in the noise attenuation measurements in earplugs, is it possible to predict the noise attenuation of earplugs?
3. How the noise attenuation level in the earplugs can be predicted?

2.10 Goals and Objectives

1. Analysing the noise attenuation data of different products tested on both subjects and artificial ears provided by the company
2. Estimating the contribution of each earplug component i.e. Thermoplug contribution and Filter contribution to the total attenuation in percentage
3. Building a dataset with noise attenuation per frequency for different products and its related material specifications
4. Running the dataset through the Gaussian process prediction model to check if it produces viable prediction results
5. Testing the dataset
6. Analysing the prediction results

The next chapter, Chapter 3 presents the data analysis, explains the calculations involved in estimating the contribution of the earplug components to the total attenuation and results of estimated earplug components' contributions. Finally, the dataset with the input features that are selected based on the factors that influence the noise attenuation is shown along with an explanation to GPR.

3 Methodology

This chapter explains how the dataset is built and the selected input features based on the factors that influence the noise attenuation. The first research question 1, revolves around finding the contribution of each earplug component. So, the noise attenuation data of various products in ALPINE is analysed. Then, based on the data analysis, the calculations involved in estimating how much each of the earplug components contributes to the total attenuation are shown along with the estimated contribution result plots (in %). The second research question 2 and third research question 3, are about the prediction model. Based on the analysis of the prediction models 2.8 in Chapter 2, it is inferred that the GPR model is suitable for our application. So, in this chapter, a brief explanation of GPR based on Bayesian inference is given. Finally, the steps involved in the regression task along with the framework used to perform the regression are explained.

The motivation of this chapter is given as follows:

- To build the input dataset based on the factors that influence the noise attenuation
- To analyse the noise attenuation data of different earplugs in ALPINE and estimate the contribution of the earplug components individually
- To implement the GPR model and make predictions

3.1 Building the Dataset

The first step in developing a prediction model is to build the dataset. The test results usually have the noise attenuation value (in dB) per frequency (Hz) for 16 subjects. The average of the noise attenuation value per frequency is taken and the standard deviation is subtracted from the mean value to calculate the final Assumed Protection Value (APV). Instead of taking this APV value in the dataset, the attenuation values of all 16 subjects per frequency are accounted in the dataset to capture the subjective variations. So, for each product and each frequency, the noise attenuation values of 16 subjects are present. An example of the noise attenuation values provided by the testing laboratory is given in Table 3.1. Actual noise attenuation values are not mentioned in the table due to confidentiality purposes.

Table 3.1: Sound attenuation report provided by the laboratory after testing the earplugs on 16 subjects per frequency. The values in the table are interpreted as follows: **For Subject 1, at 63 Hz the noise attenuation of the earplug is 20 dB**. StD refers to the standard deviation and APV is the Assumed Protection Value. Actual noise attenuation values are not mentioned in the table due to confidentiality purposes

Subjects \ Frequency(Hz)	**63**	**125**	**250**	**500**	**1000**	**2000**	**4000**	**8000**
1	20	26	27	26	25	30	25	28
2	22	28	27	27	22	31	35	18
3	24	29	19	25	24	29	32	20
4	25	25	20	24	26	32	26	30
5	19	27	26	26	27	28	34	20
6	19	28	21	27	31	33	36	31
7	23	30	25	30	23	27	25	25
8	18	31	23	22	30	30	26	29
9	22	24	24	19	29	31	27	19
10	18	25	18	31	27	28	30	20
11	21	26	19	21	28	29	31	31
12	22	27	26	21	25	33	28	17
13	26	31	27	23	25	30	29	19
14	25	28	28	25	24	28	35	25
15	20	29	25	26	19	18	27	27
16	19	24	18	27	20	31	28	25
Mean (dB)	21.4	27.4	23.3	25	25.3	29.3	29.6	24
StD (dB)	2.6	2.3	3.6	3.2	3.4	3.5	3.8	4.9
APV (dB)	**18.8**	**25.1**	**19.7**	**21.8**	**21.9**	**25.8**	**25.8**	**19**

The thermoplug and filter material specifications are collected for each product. As mentioned in Section 2.2.2, there are two types of filter; Venturi and Mesh. The dataset is separated for these two types of filters since there are lots of variations in the product design, and it is hard to combine these two filter types in one dataset. Hence, the prediction models are separately applied for each of these two datasets.

3.1.1 Venturi type earplugs' Dataset

In the Venturi type dataset, there are two filter materials; Soft and Hard. As mentioned in Section 2.2.2, the venturi filter diameter plays an important role in noise attenuation and so it is considered as an input feature in the dataset. In the venturi earplug type, only the filter material varies and for all the products same thermoplug material is used. The thermoplug material is not taken as an input feature in the dataset since it is the same for all the products and do not contribute much. However, the canal in which the filter is inserted inside the thermoplug differs in diameter. Also, from Section 2.2.3, it is observed that sound pressure level differs per frequency, so it is taken as an input feature as well. To sum up, the following input features are considered in the venturi dataset:

- Frequency
- Filter material
- Filter diameter
- Thermoplug canal diameter

The final dataset with input features for a few data points of venturi type earplugs is shown in Figure 3.1 and the dataset has a total of 1160 data points. Each frequency has attenuation values of 16/ 20 subjects so it is possible for the model to capture the subjective uncertainty. The Figure 3.1 shows an example of how the venturi dataset looks. The exact values of each input feature are not given due to confidentiality reasons. The target variable i.e the output is the attenuation values of all the products.

Product	Frequency(Hz)	Filter Canal diameter(mm)	Filter Material	Filter Type	Filter Hardness(Shore A-scale and R-scale)	Venturi diameter(mm)
Product 1	63	k	Material 1	Venturi	y A	x
Product 1	125	k	Material 1	Venturi	y A	x
Product 1	250	k	Material 1	Venturi	y A	x
Product 1	500	k	Material 1	Venturi	y A	x
Product 1	1000	k	Material 1	Venturi	y A	x
Product 1	4000	k	Material 1	Venturi	y A	x
Product 1	8000	k+1	Material 1	Venturi	y A	x
Product 2	63	k+1	Material 2	Venturi	z R	x+0.10
Product 2	125	k+1	Material 2	Venturi	z R	x+0.10
Product 2	250	k+1	Material 2	Venturi	z R	x+0.10
Product 2	500	k+1	Material 2	Venturi	z R	x+0.10

Figure 3.1: Each frequency has attenuation values of 16/ 20 subjects so it is possible for the model to capture the subjective uncertainty. The input features are Frequency (Hz), Filter material, Filter diameter (mm), and Thermoplug canal diameter (mm)

3.1.2 Mesh type earplugs' Dataset

As mentioned in Section 2.1, the tests are done either on ear simulators (Objective tests) or subjects (Subjective tests). In the Mesh dataset, we have both of these two noise measurement data values. The attenuation values from the subjective test and objective tests are not always the same and there is a difference of a maximum 10 dB. This difference in values varies per frequency and the product. So, it is hard to generalise and come up with a certain value per frequency. Hence, to account for this difference, the test method is given as an input feature in the dataset. This test method takes two values: Subjective and Objective.

Since sound pressure level differs per frequency as seen in Section 2.2.3, it is taken as an input feature. The thermoplug material and the canal in which the filter is inserted are the same. Since it does not vary, it is not taken as an input feature in the mesh filter type dataset. In mesh filter type, the entire earplug is a composition of 3 layers including the mesh filter. However, due to the product design confidentiality of the company, the other layers cannot be explained in this report. However, those layers are named Layer 1, Layer 2 - The Mesh filter and Layer 3. Since the mesh filter contributes to noise attenuation as seen in Section 2.2.2, the mesh opening percentage and material are taken as input features. Additionally, the specification of Layers 1 and 3 are accounted in the dataset as well since they vary for different products and contribute to noise attenuation which is seen in Section 3.2.2.

To summarise, the following input features are considered in the mesh dataset:

- Test method
- Frequency
- Layer 1 type
- Layer 2 - Mesh material

- Layer 2 - Mesh opening percentage
- Layer 3 material
- Layer 3 thickness

The final dataset with input features for a few data points of mesh earplugs is shown in Figure 3.2. This dataset has total data points of 1003. Each frequency has attenuation values of 16/ 20 subjects so it is possible for the model to capture the subjective uncertainty. The Figure 3.2 shows an example of how the mesh dataset looks. The exact values of each input feature are not given due to confidentiality reasons. The target variable i.e the output is the attenuation values of all the products.

Product	Test method	Frequency(Hz)	Layer 1 type	Mesh Material	Mesh Opening (%)	Layer 3 Material	Layer 3 thickness (mm)
Product 1	Objective	125	A100	P20	Mesh	LD20	x
Product 1	Objective	250	A100	P20	Mesh	LD20	x
Product 1	Objective	500	A100	P20	Mesh	LD20	x
Product 1	Objective	1000	A100	P20	Mesh	LD20	x
Product 1	Objective	4000	A100	P20	Mesh	LD20	x
Product 1	Objective	8000	A150	P20	Mesh	LD20	x
Product 2	Subjective	125	A150	H2	Mesh	LD50	x+10
Product 2	Subjective	500	A150	H2	Mesh	LD50	x+10

Figure 3.2: Each frequency has attenuation values of 16/ 20 subjects so it is possible for the model to capture the subjective uncertainty. The input features are Test method, Frequency (Hz), Layer 1 type, Layer 2 - Mesh material, Layer 2 - Mesh opening (in %), Layer 3 material, and Layer 3 thickness

3.2 Earplug components' contribution to total attenuation

To answer the Research Question 1, it is important to know how much each of the components in the earplug contributed to the total attenuation separately. So, with the available noise attenuation data from objective and subjective tests, the contribution of each component is calculated. The following sections explain the data analysis and the calculations involved in estimating the component's contribution in both venturi and mesh earplugs.

3.2.1 Venturi type Earplug components' contribution to total attenuation

Venturi earplug type has two components; Thermoplug and the filter. ALPINE has conducted some tests on just the thermoplug to know how the thermoplug attenuates the incoming sound by itself as mentioned in Section 2.2.1. Hence for this test, just the thermoplug without the filter is taken and tested in the ear simulator. The sound attenuation values for just the thermoplug are reported in decibels per frequency as shown in Table 3.2. The values in the table are for representation purposes only.

Table 3.2: ALPINE tested the thermoplug without any filter in an ear simulator (Objective test). The table below shows the sound attenuation provided by just the thermoplug in dB for various frequencies.

Frequency (Hz)	Attenuation (dB)
125	0.1
250	0
500	-1.3
1000	1
2000	13.1
4000	26
8000	5.6

As shown in, Table 3.2, the attenuation value is reported to be 0 dB and it also has negative values. It is because, from the sound pressure level equation Equation 2.1 it can be seen that the sound pressure of interest is relative to reference sound pressure and so if the sound pressure of interest is larger than the reference sound pressure, we will end up in negative values. If the sound pressure of interest is equal to the reference sound pressure, then we end up at 0 dB.

From the company we have thermoplug attenuation data from objective tests, however, there are no subjective tests conducted on just the thermoplug. Also, there are no objective and subjective tests conducted on just the filter. But we have subjective tests conducted on the entire earplug (Thermoplug with the filter). So, by comparing the objective test results of the thermoplug and the subjective test results of the earplug, the thermoplug and filter contribution are estimated. Here the comparison is between objective and subjective test methods. Because, in the objective test we use a digital machine to find the noise attenuation and in this case, the measured attenuation data does not vary much. However, in subjective tests, attenuation data is highly random. Also, the objective and subjective test results are not always the same as mentioned in Section 2.1. From the data, it is observed that mostly the objective test results are higher than the subjective test results. But due to inadequate data in hand, we assume that both the objective and subjective data can be compared and used for component contribution estimation.

Goal: To find the thermoplug and filter contribution(in %) of all the earplugs in ALPINE

Data available:

- Thermoplug attenuation data from objective tests
- Earplug (Thermoplug with filter) attenuation data from subjective test

Assumption: Objective and Subjective test data can be compared and used for component contribution estimation

The attenuation values are on a decibel scale. Direct mathematical operations such as addition, and subtraction cannot be performed on the decibel values. So, the decibel values have to be first converted into linear scale i.e. the decibel values must be converted into sound pressure values in pascals. From the sound pressure level equation, the sound pressure of interest in pascals is found by keeping the reference pressure as 20μ pascals(threshold of hearing). The equations are given in Equation 3.1.

$$\begin{aligned} \text{SPL} &= 20\log\frac{P}{P_{ref}} \\ P &= (10^{\frac{SPL}{20}}) * P_{ref} \end{aligned} \tag{3.1}$$

where SPL is the Sound Pressure Level in decibels, P is the sound pressure of interest in pascals, P_{ref} is the reference pressure in pascals.

Let's see the components contribution calculation with an example. The table in the Figure 3.3, shows the attenuation value of just thermoplug A in decibels (second column) and the sound pressure of interest in pascals (third column) calculated using the Equation 3.1 at 125 Hz and the attenuation value of an earplug i.e. Product 1 in decibels (second column) and the sound pressure of interest in pascals(third column) calculated using the Equation 3.1 at 125 Hz.

Thermoplug attenuation value

Frequency (Hz)	Attenuation(dB)	Sound pressure (Pascals)
125	0	20μPa

(a) Conversion of sound attenuation value in decibels to pascals of Thermoplug A. The table shows the attenuation value of thermoplug A in decibels (second column) and the sound pressure of interest in pascals (third column) calculated using the Equation 3.1 at 125 Hz

Product 1 attenuation value

Frequency (Hz)	Attenuation(dB)	Sound pressure (Pascals)
125	20	200μPa

(b) Conversion of sound attenuation value in decibels to pascals of Product 1. The table shows the attenuation value of an earplug i.e. Product 1 in decibels (second column) and the sound pressure of interest in pascals (third column) calculated using the Equation 3.1 at 125 Hz

Figure 3.3: Conversion of sound attenuation values in decibel values to pascals (Linear scale)

By using the data given in the Figure 3.3, the contribution of thermoplug A to the total attenuation of product 1 is calculated using the following equation,

$$\begin{aligned} T_A &= \left(\frac{P_A}{P_{P1}}\right) * c \\ T_A &= \left(\frac{20\mu}{200\mu}\right) * c \\ T_A &= 10\% \end{aligned} \tag{3.2}$$

where T_A is the contribution of thermoplug A, P_A is the Sound pressure in thermoplug A, P_{P1} is the Sound pressure in Product 1 and $c = 100$.

The thermoplug contribution is then normalised between 0-100. We have taken the total contribution to be 100%. The venturi earplug only consists of two components, so the thermoplug A contribution is subtracted from the total contribution to find the filter contribution.

$$\begin{aligned} F_v &= T_c - T_A \\ F_v &= 100 - 10 \\ F_v &= 90\% \end{aligned} \tag{3.3}$$

where F_v is the venturi filter contribution in %, T_c is the total contribution of a product which is always 100% and T_A is the contribution of thermoplug A.

From the results, it is inferred that thermoplug A contributes only 10% to the total attenuation and the filter contributes 90% to the attenuation of product 1 at 125Hz.

In Venturi filter type of earplugs, ALPINE has 6 products. Let's name those products as Product 1, Product 2, Product 4, Product 5, and Product 6 (for confidentiality purposes). The Same thermoplug i.e. Thermoplug A is used for Products 1 to 5, but only the filter diameter varies. Using the thermoplug

A attenuation data, the thermoplug A contribution (in %) is estimated for all the products separately for all the frequencies. Likewise, the filter contribution is also found.

Product 6 has a different thermoplug, let's say thermoplug B. Product 6 has three filter variations of different diameters. There are no tests conducted separately on the filter for this product. However, ALPINE has conducted some tests to find how much thermoplug B attenuates the incoming sound. So, the calculation for finding the thermoplug B and filter contribution is similar to Equation 3.2 and Equation 3.3. Using the Thermoplug B attenuation value and Product 6 attenuation value, the contributions are estimated as follows:

$$T_B = \frac{P_B}{P_{P6}} * c \tag{3.4}$$

where T_B is the contribution of thermoplug B, P_B is the Sound pressure in thermoplug B, P_{P6} is the Sound pressure in Product 6 and $c = 100$.

$$F_v = T_c - T_B \tag{3.5}$$

where F_v is the venturi filter contribution in %, T_c is the total contribution of a product which is always 100% and T_B is the contribution of thermoplug B.

All these calculations are done per frequency, because the sound pressure varies per frequency. The contributions of each components for Products 1 to 4 (Soft range filters) is shown in Figure 3.4 and Figure 3.5. The x-axis is the frequency in Hz and y-axis is the estimated contribution of earplug components in %. The green curve is the thermoplug contribution and the blue curve is the filter contribution.

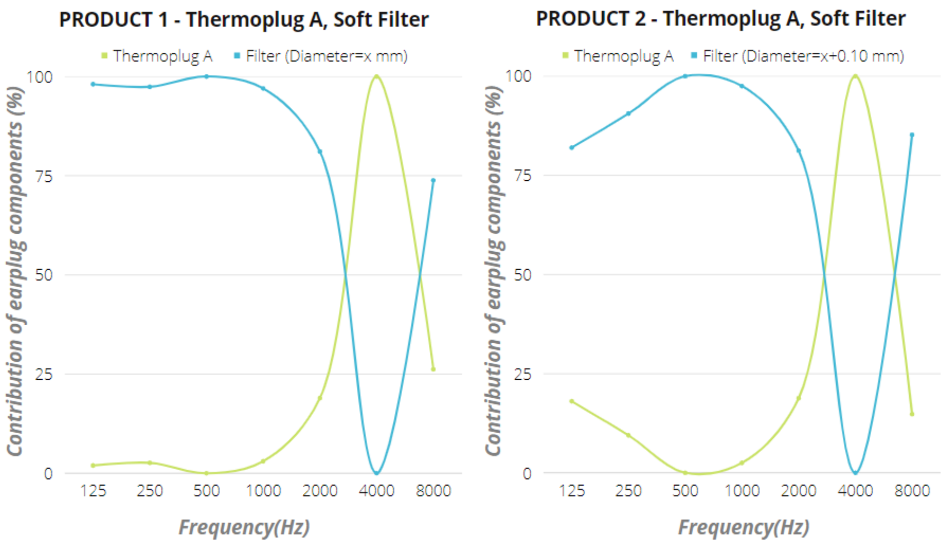


Figure 3.4: Contributions of thermoplug A and soft filter of the Products 1 and 2. The filter diameter of Product 1 and 2 varies by 0.10 mm. The x-axis is the frequency in Hz and y-axis is the estimated contribution of earplug components in %. The green curve is the thermoplug contribution and the blue curve is the filter contribution. At 4000 Hz most of the sound attenuation is from the thermoplug. Between 125 - 500 Hz and around 8000 Hz most of the contribution to sound attenuation is provided by the filter.

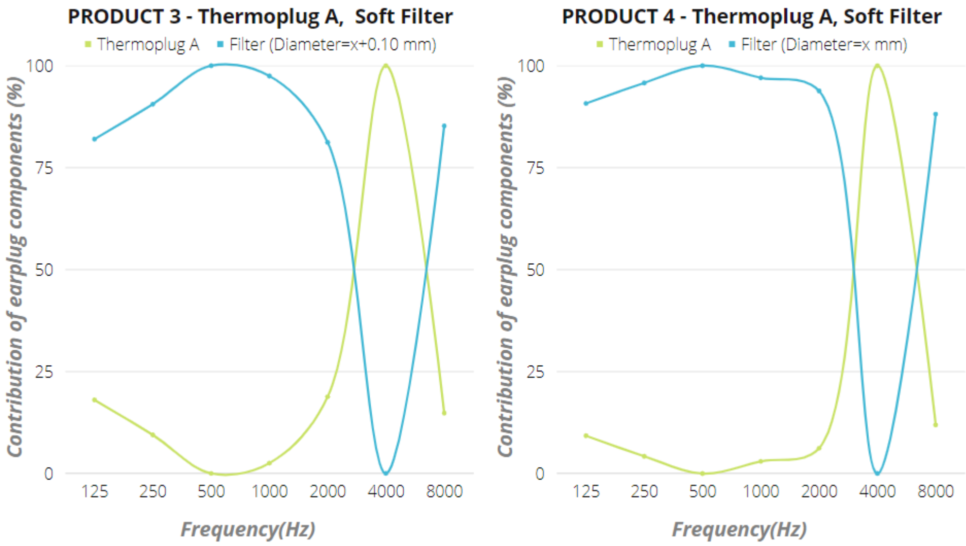


Figure 3.5: Contributions of thermoplug A and soft filter of the Products 3 and 4. The x-axis is the frequency in Hz and y-axis is the estimated contribution of earplug components in %. The green curve is the thermoplug contribution and the blue curve is the filter contribution. The filter diameter in these two products varies by 0.10 mm. Like Product 1 and 2, at 4000 Hz most of the sound attenuation is from the thermoplug. Between 125 - 500 Hz and around 8000 Hz most of the contribution to sound attenuation is provided by the filter.

The contributions of each components for Products 5 and 6 (Hard range filters) can be seen in Figure 3.6.

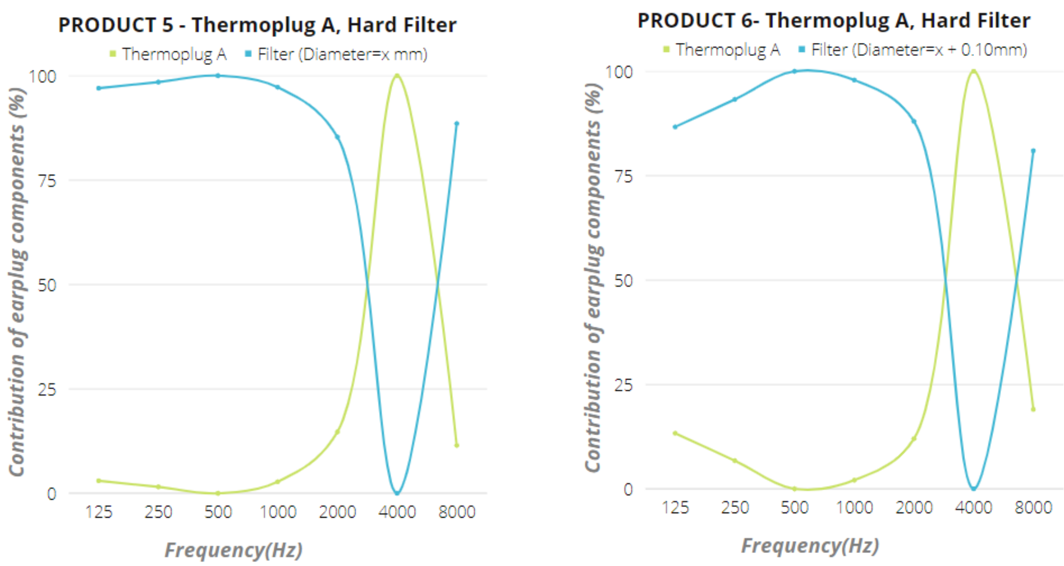


Figure 3.6: Contributions of thermoplug A and hard filter of the Products 5 and 6. The x-axis is the frequency in Hz and y-axis is the estimated contribution of earplug components in %. The green curve is the thermoplug contribution and the blue curve is the filter contribution. The filter diameter in these two products varies by 0.10 mm. Like the soft range filters, most of the sound attenuation from the thermoplug is at 4000 Hz. Most of the contribution to sound attenuation provided by the filter is between 125 - 500 Hz and 8000 Hz

The contributions of each components for Products 7(three variations) can be seen in Figure 3.7.

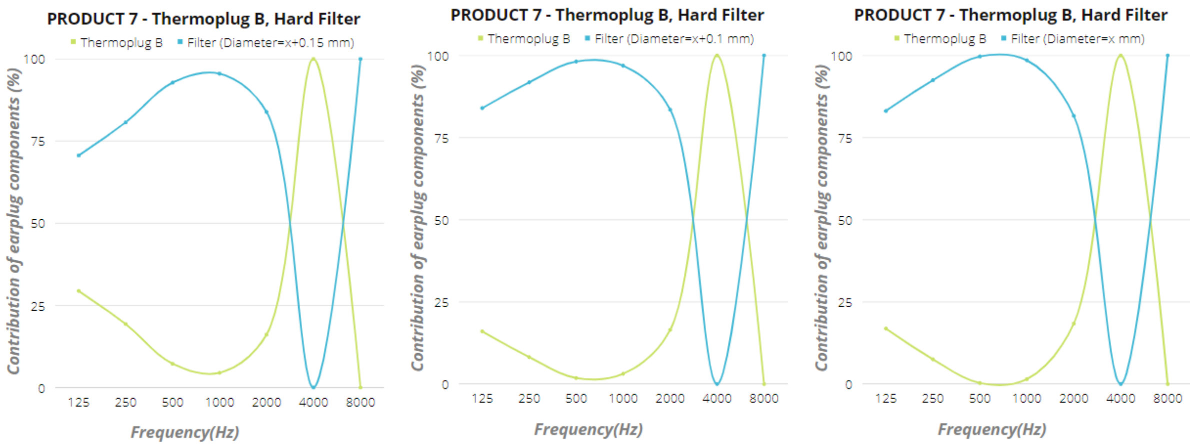


Figure 3.7: Contributions of thermoplug B and hard filter of the Product 7. The x-axis is the frequency in Hz and y-axis is the estimated contribution of earplug components in %. The green curve is the thermoplug contribution and the blue curve is the filter contribution. The filter diameter in these three products are x mm, x+0.1 mm and x+0.15 mm. Like the previous products, most of the sound attenuation from the thermoplug is at 4000 Hz. Most of the contribution to sound attenuation provided by the filter is between 125 - 500 Hz and 8000 Hz

From the graphs, it is clear that at 4000 Hz most of the sound attenuation is from the thermoplug. Around 125-500 Hz and 8000 Hz, the minimal contribution is provided by the thermoplug. This behaviour is supported by comparing the actual attenuation values of the thermoplug and product 1. From Figure 3.8, it is seen that the attenuation provided by just the thermoplug A and product 1 (thermoplug A with filter of x mm diameter). Most of the attenuation at 4000 Hz is from the thermoplug (Blue curve).

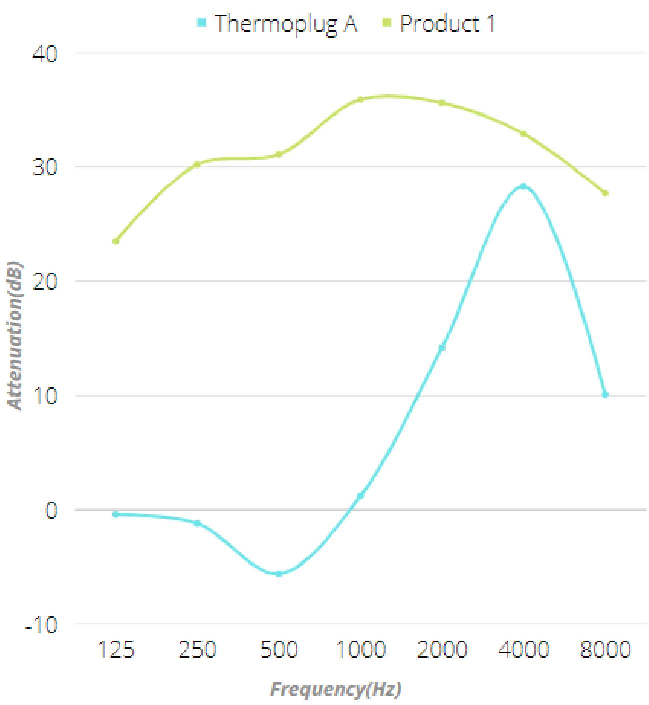


Figure 3.8: Sound attenuation provided by just the Thermoplug A (without filter) vs Sound attenuation provided by Product 1 (Thermoplug A with filter of x mm diameter). he x-axis is the frequency in Hz and y-axis is the sound attenuation in dB. The blue curve is the sound attenuation provided by the thermoplug A and the green curve is the sound attenuation provided by Product 1. Most of the attenuation in Product 1 is from the thermoplug at 4000 Hz

The estimation of components' contribution is carried out to understand how much each of the components in an earplug contributes to the total attenuation. Hence, it will be useful for the company to understand the functioning of each component per frequency and modify the design if needed.

3.2.2 Mesh type Earplug components' contribution to total attenuation

In this section, the component contribution is calculated for mesh earplugs. Unlike Venturi earplugs, in mesh type of earplugs, the entire filter part is a composition of three layers. In ALPINE, the mesh earplugs are designed to achieve flat attenuation i.e. providing the same attenuation at all frequencies. Hence, multiple layers are added to achieve that goal. So here the aim is to find the contribution of all those layers namely, Thermoplug, Layer 1, mesh filter layer and Layer 3 to the total attenuation provided by the products. Thermoplug B is used for all the products in this mesh type. The other layers have quite a lot of variations. The number of variations in each layer is as follows:

- Layer 1 has two variations, namely A100 and A150
- Mesh filter layer has three variations, namely Mesh H2, Mesh P3, and Mesh P20
- Layer 3 has seven variations, namely LD20, LD50, LD80, FO, LA6, LA20, and LA250

Hence, mesh type earplugs' composition is **Thermoplug B + Layer 1 + Mesh layer + Layer 2**. There are several products based on the different variations of each layers. Firstly, it is important to calculate the contribution of each of these layer's variations. Based on the data analysis it is found that following objective test attenuation data are available from ALPINE and the components of each product is given in Table 3.3.

Table 3.3: Components list of each product in mesh type of earplugs

Product	**Components**
-	Thermoplug B
Product 1	Thermoplug B + Mesh P20
Product 2	Thermoplug B + A100 + Mesh P20
Product 3	Thermoplug B + A100 + Mesh P20 + LD20
Product 4	Thermoplug B + A100 + Mesh P20 + LD50
Product 5	Thermoplug B + A100 + Mesh P20 + LD80
Product 6	Thermoplug B + A100 + Mesh P20 + FO
Product 7	Thermoplug B + A100 + Mesh P20 + LA6
Product 8	Thermoplug B + A100 + Mesh P20 + LA20
Product 9	Thermoplug B + A100 + Mesh P20 + LA250
Product 10	Thermoplug B + A100 + Mesh P3
Product 11	Thermoplug B + A100 + Mesh H2
Product 12	Thermoplug B + A150 + Mesh P20 + LA6

There are no tests conducted separately on each of these layers and each layer is always in combination with another layer. So, before calculating the contribution of each component in percentage, it is first required to calculate the effects of each layer i.e. How much each of these layers is attenuating. Before doing any mathematical operations all the attenuation data in decibels must be converted into linear scale i.e. Pascals (Pa).

The thermoplug B effects are already available from the objective tests. So there is no requirement to calculate it separately. Firstly, the effects are A100 (Layer 1) are calculated. Product 2 is a combination of Thermoplug B, A100 and Mesh P20 and Product 1 is a combination of Thermoplug B and Mesh P20. So, by subtracting the sound pressure value of product 2 from product 1 we get the A100 sound pressure value i.e. A100 effects. The calculations are as follows:

$$\begin{aligned} A_{100} &= P_{P2} - P_{P1} \\ A_{100} &= (T_B + A_{100} + P_{20}) - (T_B + P_{20}) \end{aligned} \tag{3.6}$$

where A_{100} is the sound pressure of A100 layer, P_{P2} and P_{P1} is the sound pressure of Product 2 and Product 1. T_B, A_{100}, and P_{20} is the sound pressure of thermoplug B, A100 layer, and Mesh P20 layer respectively.

The effects of Mesh filter layer P20 is calculated as follows:

$$\begin{aligned} P_{20} &= P_{P1} - T_B \\ P_{20} &= (T_B + P_{20}) - (T_B) \end{aligned} \tag{3.7}$$

where P_{P1} is the sound pressure of Product 1. T_B, and P_{20} is the sound pressure of thermoplug B, and Mesh P20 layer respectively.

Likewise, based on the product compositions, the effects of each layer is found. The calculations for finding the effects of different layer 3 are as follows:

$$\begin{aligned} LD_{20} &= P_{P3} - P_{P2} \\ LD_{50} &= P_{P4} - P_{P2} \\ LD_{80} &= P_{P5} - P_{P2} \\ FO &= P_{P6} - P_{P2} \\ LA_{6} &= P_{P7} - P_{P2} \\ LA_{20} &= P_{P8} - P_{P2} \\ LA_{250} &= P_{P9} - P_{P2} \end{aligned} \tag{3.8}$$

where LD_{20}, LD_{50}, LD_{80}, FO, LA_6, LA_{20} and LA_{250} is the sound pressure of LD20, LD50, LD80, FO, LA6, LA20 and LA250 Layer 3 variations respectively. P_{P2}, P_{P3}, P_{P4}, P_{P5}, P_{P6}, P_{P7}, P_{P8} and P_{P9} is the sound pressure of Products 2, 3, 4, 5, 6, 7, 8 and 9 respectively.
To find different Mesh filter layer effects, the sound pressure of A_{100} is calculated from Equation 3.6 and the following calculation is done

$$\begin{aligned} P_3 &= (P_{P10}) - (T_B + A_{100}) \\ H_2 &= (P_{P11}) - (T_B + A_{100}) \end{aligned}$$

where P_3, H_2, P_{P10} and P_{P11} is the sound pressure of Mesh layer P3, Mesh layer H2, Product 10 and Product 11 respectively.

To find A150 effects, sound pressure of Mesh P20 layer calculated from Equation 3.7 and sound pressure of LA6 layer calculated from Equation 3.8 are used and the calculations are as follows:

$$A_{150} = P_{P12} - (T_B + P_{20} + LA_6)$$

where A_{150} is the sound pressure of A150 layer, P_{P12} is the sound pressure of Product 12.

However, by doing the aforementioned calculations, there is a problem of ending up in negative sound pressure values. For example, let's take the sound pressure value of Product 1 = 0.0193 mPa and Product 2 = 0.0189 mPa at 125 Hz, so as per Equation 3.6,

$$\begin{aligned} A_{100} &= 0.0189 - 0.0193 \\ A_{100} &= -0.0004\ mPa \end{aligned} \tag{3.9}$$

Since the value of Product 2 is lesser than Product 1, it results in a negative value as shown in Figure 3.9. However, the negative sound pressure value does not make any sense. Such a small value does not contribute to hearing as well. Hence, this value is approximated to 0. So, while finding the effects of each layer in the earplug, if it results in a negative value, then it will be approximated to zero as shown in Figure 3.10 (Values in the figure are for representation purposes only).

Frequency(Hz)	Thermoplug B effects(mPa)	A100 effects(mPa)	Mesh P20 effects(mPa)	L20 effects(mPa)
125	0.02	-0.0004	-0.0007	0.3666

Figure 3.9: Since the sound pressure value of Product 2 is lesser than Product 1, it results in Negative sound pressure values (highlighted box in grey)

Frequency(Hz)	Thermoplug B effects(mPa)	A100 effects(mPa)	Mesh P20 effects(mPa)	L20 effects(mPa)
125	0.02	-0.0004	-0.0007	0.3666

↓ **Approximation**

Frequency(Hz)	Thermoplug B effects(mPa)	A100 effects(mPa)	Mesh P20 effects(mPa)	L20 effects(mPa)
125	0.02	0	0	0.3666

Figure 3.10: Negative sound pressure value does not exist and contribute to hearing. Hence, this value is approximated to 0 (highlighted box in grey)

Based on the aforementioned calculations, the effects of each layer are found. Using these effects, the contributions of each component in the mesh earplugs are calculated. The contribution of each component is calculated for each product per frequency. For example, the contribution of the components for product 3 is calculated. Firstly, the contribution of thermoplug B is calculated in the same way as Equation 3.2 and the equation is given as follows:

$$TB_3 = \frac{P_{TB}}{P_{P3}} * c \tag{3.10}$$

where TB_3 is the contribution of thermoplug B (in %) in Product 3, P_{TB} is the sound pressure of Thermoplug B, P_{P3} is the sound pressure of Product 3 and $c = 100$.

To find the contribution of Layer 1, Mesh and Layer 2 in product 3, A100 effects, Mesh P20 effects and LD20 effects are used in the calculation and they are calculated as follows:

$$A100_3 = \frac{A_{100}}{P_{P3}} * c \tag{3.11}$$

where $A100_3$ is the contribution of A100 layer (in %) in Product 3, A_{100} is the sound pressure of A100 layer, P_{P3} is the sound pressure of Product 3 and $c = 100$.

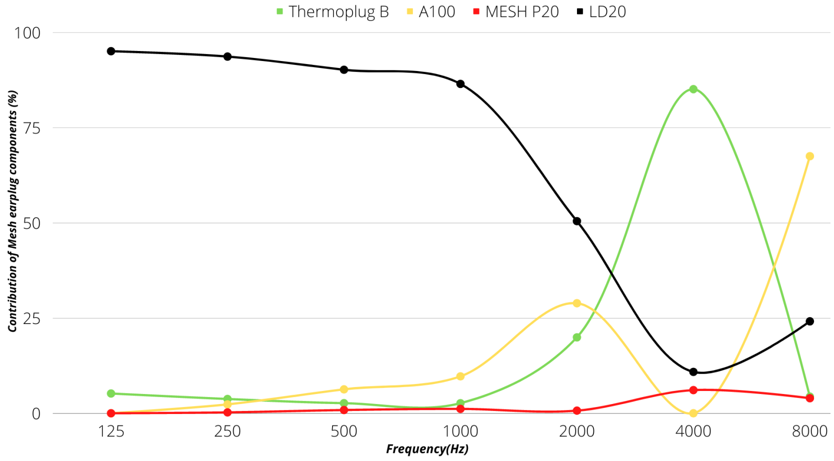
$$P20_3 = \frac{P_{20}}{P_{P3}} * c \tag{3.12}$$

where $P20_3$ is the contribution of Mesh P20 layer (in %) in Product 3, P_{20} is the sound pressure of Mesh P20 layer, P_{P3} is the sound pressure of Product 3 and $c = 100$.

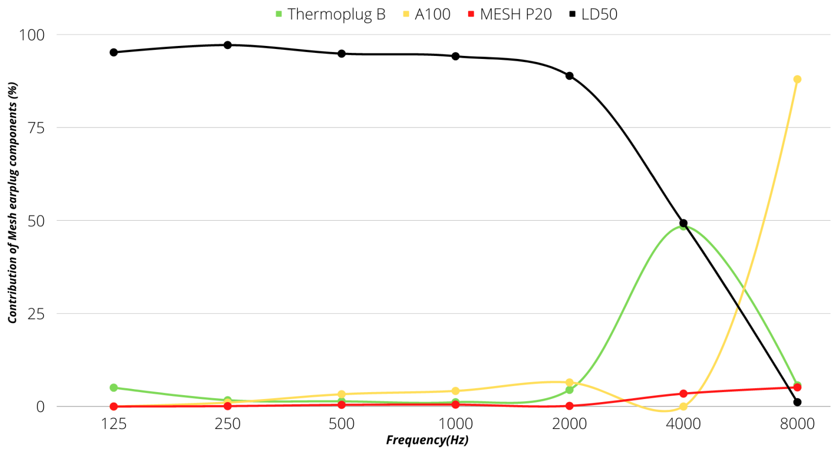
$$LD20_3 = \frac{LD_{20}}{P_{P3}} * c \tag{3.13}$$

where $LD20_3$ is the contribution of LD20 layer (in %) in Product 3, LD_{20} is the sound pressure of LD20 layer, P_{P3} is the sound pressure of Product 3 and $c = 100$. Likewise, the contribution of components is estimated for all other mesh type of earplugs.

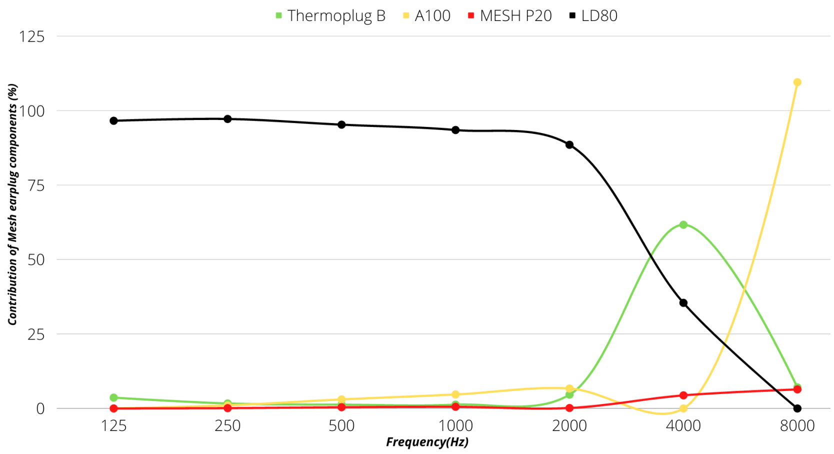
The contributions of components in product 3, 4, and 5 are shown in Figure 3.11, the products have different 3rd layer and all other layers, i.e. Thermoplug, Layer 1 and Layer 2 are the same. From, the graphs, it is seen that at low frequencies, Layer 3 contributes the most to the total attenuation of the earplug and at 4000 Hz, mostly thermoplug contributes and finally around 8000 Hz, Layer 1 contributes the most. Also, the mesh filter layer's contribution to the total attenuation is very minimal when compared to all other components in products 3, 4 and 5.



(a) Contribution of Thermoplug, Layer 1 - A100, Layer 2 - Mesh filter P20 and Layer 3 - LD20 in Product 3



(b) Contribution of Thermoplug, Layer 1 - A100, Layer 2 - Mesh filter P20 and Layer 3 - LD50 in Product 4



(c) Contribution of Thermoplug, Layer 1 - A100, Layer 2 - Mesh filter P20 and Layer 3 - LD80 in Product 5

Figure 3.11: Plot shows the contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve), Layer 2 - Mesh filter P20 (Red curve) and Layer 3 - LD20 (Black curve) in Product 3, 4 and 5. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. At low frequencies, Layer 3 contributes the most to the total attenuation of the earplug and at 4000 Hz, mostly thermoplug contributes and finally around 8000 Hz, Layer 1 contributes the most. The mesh filter's contribution to the total attenuation is very minimal when compared to all other components.

From Figure 3.12, it can be seen that most of the contribution to total attenuation is from FO layer when compared to all other components.

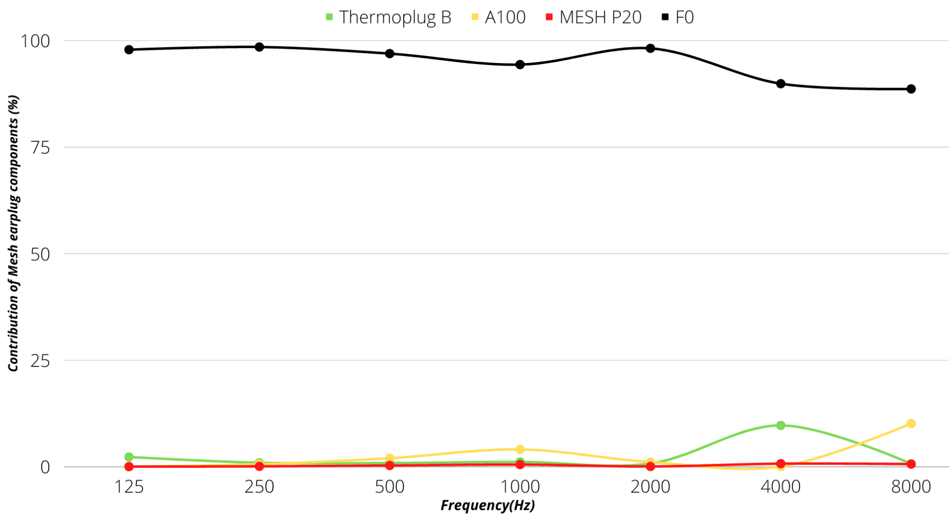
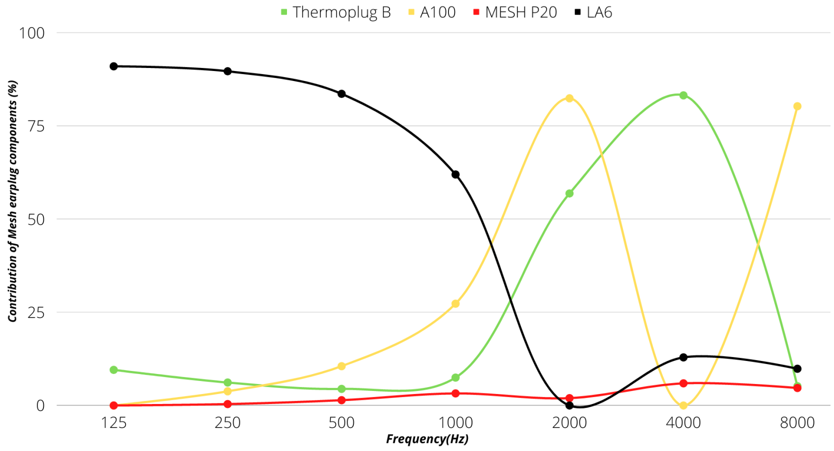
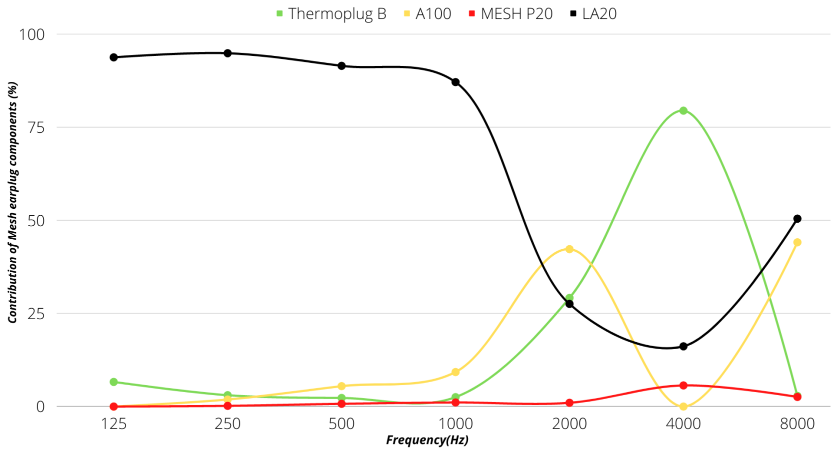


Figure 3.12: Contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve) , Layer 2 - Mesh filter P20 (Red curve) and Layer 3 - FO (Black curve) in Product 6. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. Most of the contribution to total attenuation IN Product 6 is from FO layer when compared to all other components

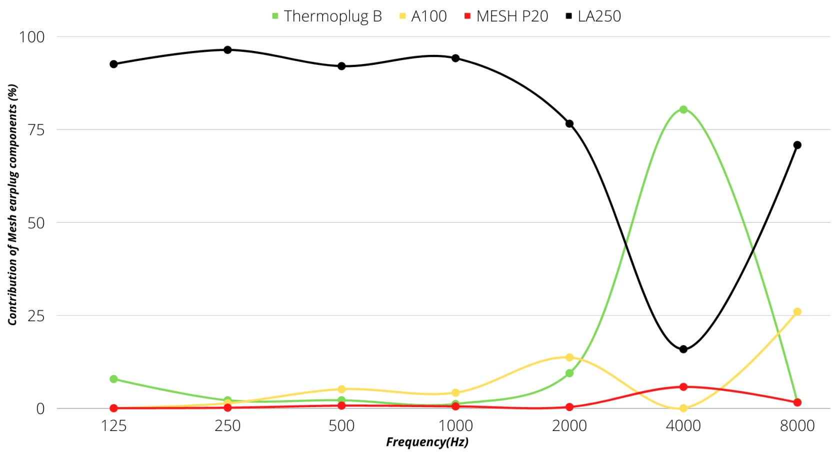
The Figure 3.13 shows the components contribution in product 7,8 and 9 with variations in layer 3. Here, in low frequencies layer 3 contributes the most and at 4000 Hz, thermoplug contributes the most. However, in Figure 3.13a, at 2000 Hz and 8000 Hz, Layer 1 - A100 contributes the most. In Figure 3.13b, at 8000 Hz, contribution of layer 3 - LA20 is better than Layer 1 - A100. In Figure 3.13c, at 8000 Hz, Layer 3 - LA250 contributes the most. Again in these products, the contribution of mesh layer is very minimal when compared to the other components.



(a) Contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve), Layer 2 - Mesh filter P20 (Red curve) and Layer 3 - LA6 (Black curve) in Product 7. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. In low frequencies layer 3 - LA6 contributes the most and at 4000 Hz, thermoplug contributes the most. However, at 2000 Hz and 8000 Hz, Layer 1 - A100 contributes the most



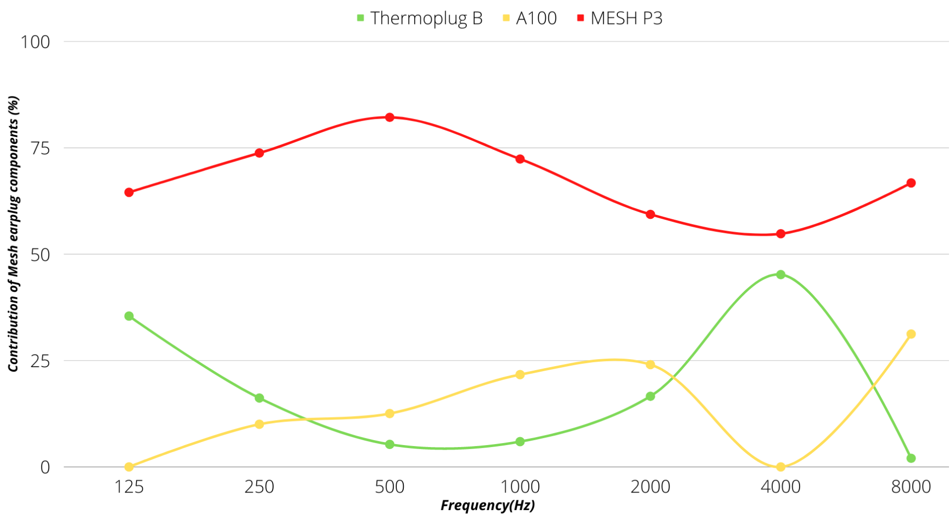
(b) Contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve), Layer 2 - Mesh filter P20 (Red curve) and Layer 3 - LA20 (Black curve) in Product 8. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. Like Product 7, in low frequencies layer 3 - LA20 contributes the most and at 4000 Hz, thermoplug contributes the most. However, at 8000 Hz, contribution of layer 3 - LA20 is better than Layer 1 - A100.



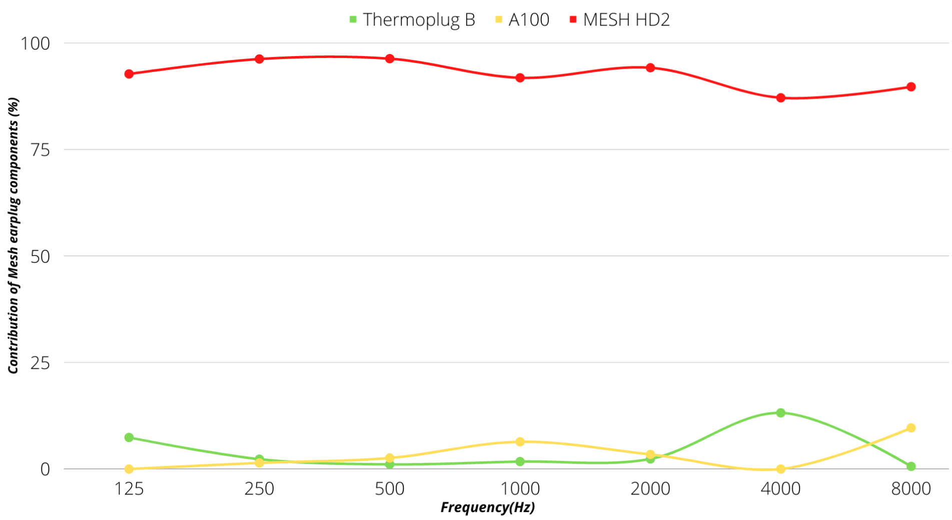
(c) Contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve), Layer 2 - Mesh filter P20 (Red curve) and Layer 3 - LA250 (Black curve) in Product 9. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. Like Product 7 and 8, in low frequencies layer 3 - LA250 contributes the most and at 4000 Hz, thermoplug contributes the most. At 8000 Hz, Layer 3 - LA250 contributes the most.

Figure 3.13: Contribution of mesh earplug components in Product 7,8 and 9 with Layer 3 variations

The Figure 3.14 shows the contribution of the components in products 10 and 11 with variations in the Mesh filter layer. These products do not have Layer 3 and without layer 3 most of the contribution to total attenuation is from the mesh layer. However, the attenuation provided by a product with layer 3 is always larger than that of a product without layer 3. Also, the goal of the mesh type earplugs is to achieve a flat attenuation and so with the current mesh filter configuration, it is not possible to achieve a flat attenuation. Hence, it is a good option to add the 3rd layer to the current product line.



(a) Contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve), and Layer 2 - Mesh filter P3 (Red curve) in Product 10. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. Most of the contribution to total attenuation is from the mesh filter



(b) Contribution of Thermoplug (Green curve), Layer 1 - A100 (Yellow curve), and Layer 2 - Mesh filter HD2 (Red curve) in Product 10. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. Most of the contribution to total attenuation is from the mesh filter

Figure 3.14: Contribution of mesh earplug components in Product 10 and 11 with mesh layer variation and without layer 3. These products does not have Layer 3 and it can be seen from the figure that without layer 3 most of the contribution to total attenuation is from the mesh layer. However, the attenuation provided by a product with layer 3 will be always larger than that of a product without layer 3

The Figure 3.15 shows the components contribution in product 12 with layer 1 being A150. It is clear that at low frequencies, the contribution of layer 3 is maximal, around 4000 Hz, thermoplug contributes the most and around 8000 Hz contribution of layer 1 - A150 is more than the other components.

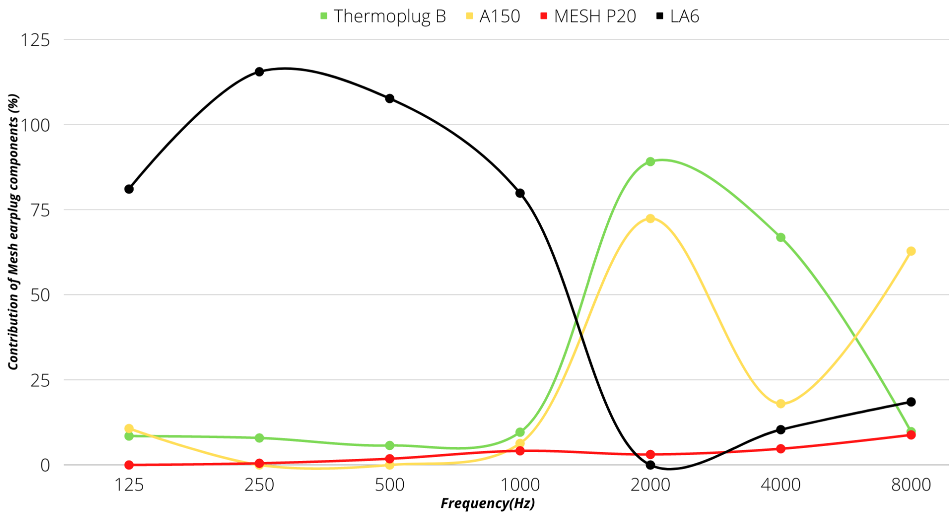


Figure 3.15: Contribution of Thermoplug (Green curve), Layer 1 -A150 (Yellow curve), Layer 2 - Mesh filter P20 (Red curve) and Layer 3 -LA6 (Black curve) in Product 12. The x-axis is the frequency in Hz and y-axis is the contribution of mesh earplug components in %. It is clear that at low frequencies, the contribution of layer 3 - LA6 is maximum. Around 4000 Hz, thermoplug contributes the most. Around 8000 Hz contribution of layer 1 - A150 is more than the other components

The contribution of all the products is estimated to understand how the component in each earplug design performs and upgrade the design if some component performs poorly. Thus, the estimation of earplug components contribution done so far, answers the Research Question 1. It has to be noted that in mesh type earplugs, the mesh filter is the main component. But from the analysis done in answering the Research Question 1, it is found that the mesh filter's contribution to the total attenuation is very minimal in mesh type of earplugs. Hence the company must consider some design variations in the mesh filter to improve the sound attenuation provided by these types of earplugs.

3.3 Gaussian process regression (GPR)

As mentioned in Section 2.6.7, the GPR works based on Bayesian inference and in this section the steps involved in predictions are given along with the associated equations. All the parameter definitions and equations follow the textbook [42] and the tutorial [43].

The multivariate Gaussian regression function is modelled as

$$p(\mathbf{f}|\mathbf{X}) \sim N(\mathbf{f}|\mu, \mathbf{K}) \tag{3.14}$$

where $\mathbf{X} = [x_1, x_2, ...x_n]$, $\mathbf{f} = [f(x_1), f(x_2), ...f(x_n)]$, $\mu = [m(x_1), m(x_2), ...m(x_n)]$ and $\mathbf{K}$ is the covariance of the function outputs at two points $f(x_i)$ and $f(x_j)$. $\mathbf{X}$ is the observed data points, m is the mean function. With no data observation, the mean function is made zero, given that the data is normalised to a zero mean.

Now, if we have to make predictions on new test points, $\mathbf{X}_*$, then we have to derive the predictive distribution function, $\mathbf{f}(\mathbf{X}_*)$. Let's say the new test points are characterised by the Gaussian distribution and given by $N(m(\mathbf{X}_*), K(\mathbf{X}_*))$, where $m(\mathbf{X}_*)$ is the mean of the new test points and $K(\mathbf{X}_*)$ is the covariance function of of the new test points. The joint distribution of $\mathbf{f}$ and $\mathbf{f}_*$ is given by

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim N\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) & K(\mathbf{X},\mathbf{X}_*) \\ K(\mathbf{X},\mathbf{X}_*)^T & K(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix} \right) \tag{3.15}$$

where, the mean $m(\mathbf{X}), m(\mathbf{X}_*) = 0$ and let $K(\mathbf{X},\mathbf{X}) = \mathbf{K}$, $K(\mathbf{X},\mathbf{X}_*) = K(\mathbf{X},\mathbf{X}_*)^T = \mathbf{K}_*$, $K(\mathbf{X}_*,\mathbf{X}_*) = \mathbf{K}_{**}$. Now, we have the latent function $\mathbf{f}$, observed input training data $\mathbf{X}$ and the new test points $\mathbf{X}_*$. The

Equation 3.15 is the joint probability distribution equation $p(\mathbf{f},\mathbf{f}_*|\mathbf{X},\mathbf{X}_*)$. For regression, we need the conditional distribution [42] i.e. find the probability distribution of $\mathbf{f}_*$ given $\mathbf{f},\mathbf{X},\mathbf{X}_*$ and it is given by

$$\mathbf{f}_*|\mathbf{f},\mathbf{X},\mathbf{X}_* \sim N(\mathbf{K}_*^T\mathbf{K}\mathbf{f},\mathbf{K}_{**}-\mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{K}_*)$$

where $\mathbf{K}_*^T\mathbf{K}\mathbf{f}$ is the predicted mean and $\mathbf{K}_{**}-\mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{K}_*$ is the predicted covariance.

In reality, the true function values are not available hence, we take noisy realisation of the true function [43]. So, it is given as $y = f(x)+\varepsilon$, where ε is the additive Gaussian noise. The covariance function then becomes $\mathrm{cov}(y) = \mathbf{K}+\sigma_n^2 I$. Hence, the joint distribution of the observed values i.e. (X, y) and the function value at new test points is given as

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{f}_*\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\mathbf{K}+\sigma_n^2 I & \mathbf{K}_*\\ \mathbf{K}_*^T & \mathbf{K}_{**}\end{bmatrix}\right) \tag{3.16}$$

Now, by calculating the conditional distribution of Equation 3.16, the predictive equation is given as follows:

$$\bar{\mathbf{f}}_*|\mathbf{X},\mathbf{y},\mathbf{X}_* \sim N(\bar{\mathbf{f}}_*,\mathrm{cov}(\mathbf{f}_*))$$

where $\bar{\mathbf{f}}_* = \mathbf{K}_*^T[\mathbf{K}+\sigma_n^2 I]\mathbf{y}$ is the predicted mean and $\mathrm{cov}(\mathbf{f}_*) = \mathbf{K}_{**}-\mathbf{K}_*^T[\mathbf{K}+\sigma_n^2 I]^{-1}\mathbf{K}_*$ is the predicted covariance.

3.3.1 Hyperparameters optimization

As mentioned in Section 2.6.7, tuning the hyperparameters to be at the optimal value is crucial and how those hyperparameters affected the kernel function. The parameter definitions and equation given in this section follow [42] and [43]. The optimal hyperparameters are estimated by log marginal likelihood

$$\Theta^* = \arg\max_{\Theta} \log p(\mathbf{y}|\mathbf{X},\Theta) \tag{3.17}$$

where Θ has the optimal parameters. Equation 3.17 maximises the log of the likelihood function. It marginalise the likelihood function over f values and the equation is given in Equation 3.3.1. Instead of maximising the log marginal likelihood it is also possible to minimise the negative log marginal with respect to the needed parameters i.e. the process variance, the noise variance and the lengthscale as given Equation 3.3.1.

$$\log p(\mathbf{y}|\mathbf{X}) = -\log(N(\mathbf{y}|0,\mathbf{K}_y))$$
$$\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2}\mathbf{y}^T\mathbf{K}_y^{-1}\mathbf{y}-\frac{1}{2}\log|\mathbf{K}_y|-\frac{N}{2}\log(2\pi)$$

After hyperparameter optimisation, prediction equation is given as follows:

$$\bar{\mathbf{f}}_*|\mathbf{X},\mathbf{y},\mathbf{X}_*,\Theta \sim N(\bar{\mathbf{f}}_*,\mathrm{cov}(\mathbf{f}_*))$$

where $\bar{\mathbf{f}}_*|\mathbf{X},\mathbf{y},\mathbf{X}_*$ is the predicted latent function given the input data points, target values, new test points and the optimal hyperparameters Θ.

An optimization algorithm is used to find the optimal hyperparameters. When the input space is more than 1 dimension, the kernels are modelled by multiplying the kernels over different dimensions, where each dimension has a different lengthscale value, this is called Automatic Relevance Determination (ARD) [47]. In the case of a categorical variable, first, it has to be encoded. The variables are one hot encoded and represented as binary values. Any multi-dimensional kernel can be used while using ARD. For example, the RBF kernel is used for a multi-dimensional input space, then it is given as follows:

$$\prod_{d=1}^{D}\mathrm{RBF}(x_d,x_d') = \prod_{d=1}^{D}\sigma_d^2\exp\left(-\frac{(x_d-x_d')^2}{2{l_d}^2}\right)$$
$$\prod_{d=1}^{D}\mathrm{RBF}(x_d,x_d') = \sigma_f^2\exp\left(-\sum_{d=1}^{D}\frac{(x_d-x_d')^2}{2{l_d}^2}\right)$$

where x_d and x'_d are two different input data points at dimension d, σ_d^2 is the process variance at dimension d, l_d is the lengthscale at dimension d and σ_f^2 is the overall process variance. The required design choice here to make is the selection of the base kernel of each dimension of the input and the hyperparameters are estimated using the log marginal likelihood [47].

3.4 Gaussian Process framework

The GPy, a Gaussian Process frame from the Sheffield machine learning group [48] has been used in our application to carry out the regression task. The concept of Gaussian Process is shown in the Figure 3.16

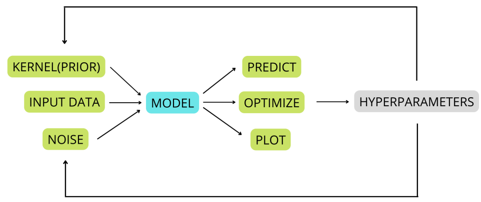


Figure 3.16: Gaussian Process Concept in GPy framework [48]. A kernel function, input data and, noise are given to the model. The kernel and noise are controlled by hyperparameters and the optimal hyperparameter values are estimated via optimization algorithm. After training the model, it can be used to make predictions [48]

The kernel function (prior), input data and the noise are given to the model. The optimal hyperparameters of the kernel and noise are estimated via optimization algorithms in GPy. GPy supports ARD and uses Limited-memory - Broyden - Fletcher Goldfarb Shanno algorithm with bound constraints (L-BFGS-B) optimization method to find the optimal hyperparameters. Other supported optimization methods are Truncated Newton Method (TNC) and Scaled Conjugate Gradients (SCG). The results from the optimizer will be compared and the best one will be chosen. Once the model has been trained, predictions are made on new test points.

3.4.1 Regression on Venturi dataset

In the venturi data, the attenuation (in dB) has to be predicted based on the input features frequency (Hz), the filter specifications such as material, hardness and diameter (mm). The goal is to model attenuation as a function of frequency and filter specifications. Firstly, let's see the plots for each of the input features vs the Attenuation in dB.

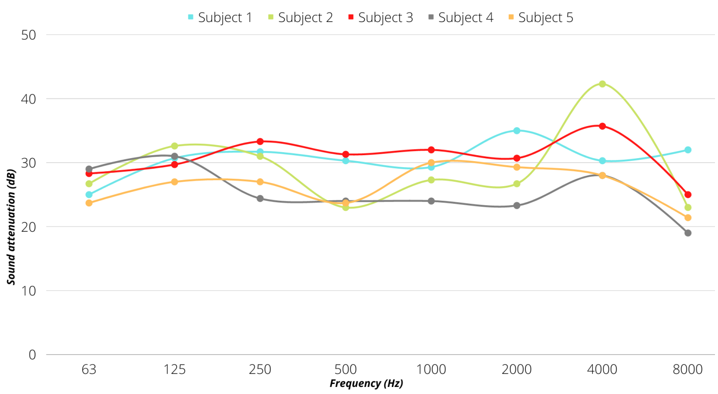
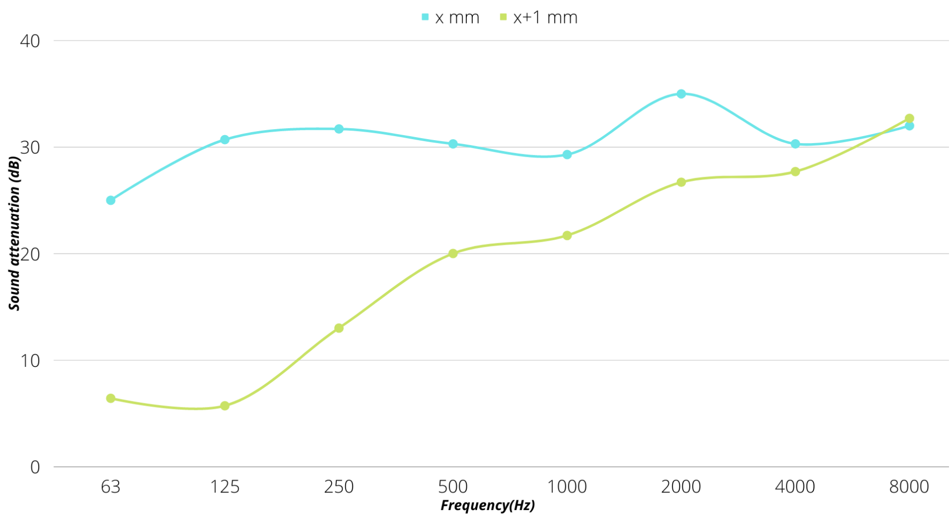
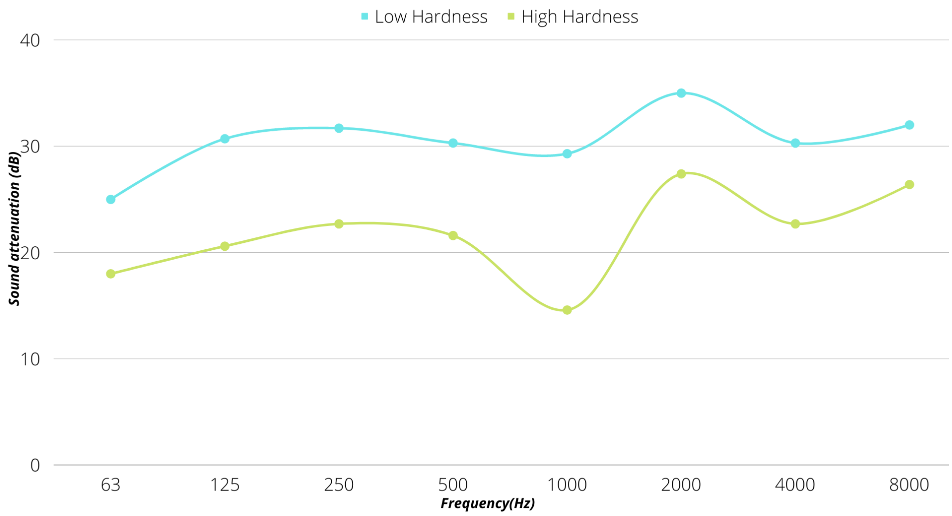


Figure 3.17: Frequency (Hz) vs Attenuation (dB) plot for first five subjects of Product 1 in soft filter Range. The same product, tested on different subjects results in different attenuation values for each frequency

The Figure 3.17 shows the plot of Frequency (Hz) vs Attenuation(dB) for different subjects for the same product and it is clear that the noise attenuation is highly varying per subject.



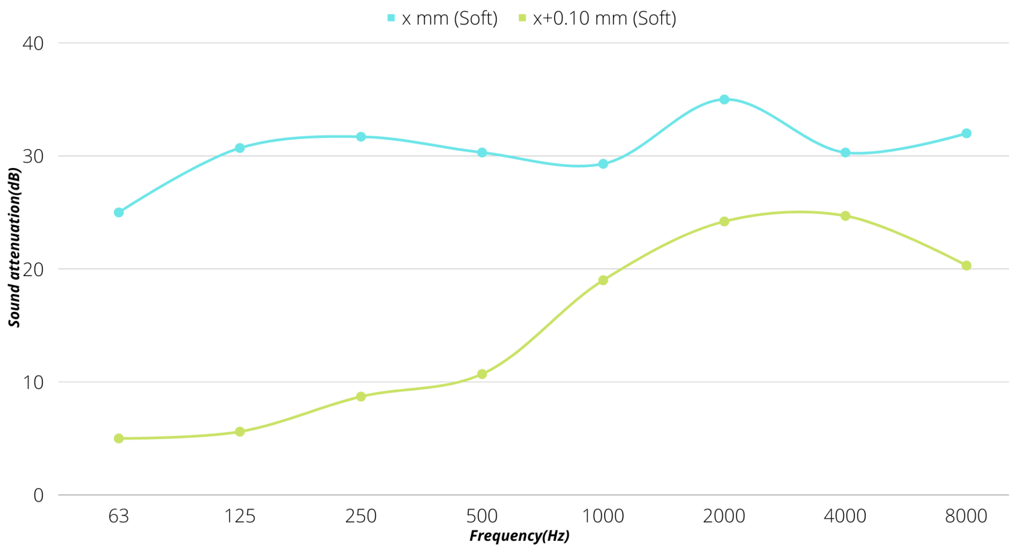
(a) Plot of Frequency (Hz) vs Sound Attenuation (dB) for different thermoplug canal length x mm (Blue curve), x+1 mm (Green curve) of two different subjects. The plot shows that the attenuation decreases with an increase in the thermoplug canal diameter



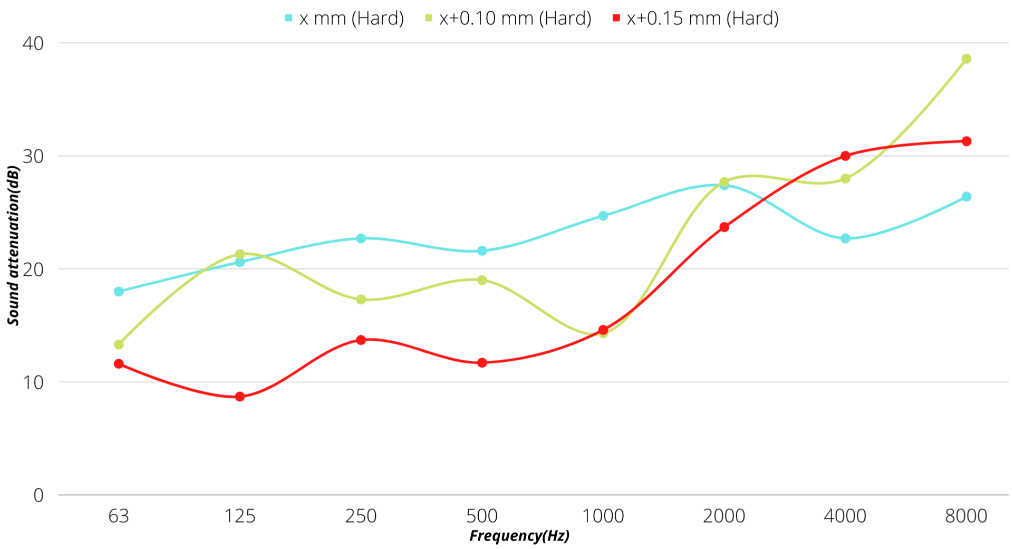
(b) Plot of Frequency(Hz) vs Sound Attenuation(dB) for low (Blue curve) and high (Green curve) filter Hardness of two different subjects. From the plot it is clear that the material with low hardness has better attenuation than a material with high hardness

Figure 3.18: Plot of Frequency (Hz) vs Sound Attenuation (dB) with varying thermoplug canal diameter and filter hardness tested on two different subjects

Figure 3.18a shows the plot of Frequency (Hz) vs Attenuation (dB) for different thermoplug canal diameter (x, x+1 mm) tested on two different subjects. It is clear that attenuation decreases with an increase in thermoplug canal diameter. Figure 3.18b shows the plot of Frequency (Hz) vs Attenuation (dB) for different filter hardness and material tested on two different subjects, and it is clear that the material with low hardness has better attenuation than a material with high hardness.



(a) Plot of Frequency (Hz) vs Attenuation (dB) for different filter diameters x mm (Blue curve) and x+0.10 mm (Green curve) of two different subjects in soft range filters

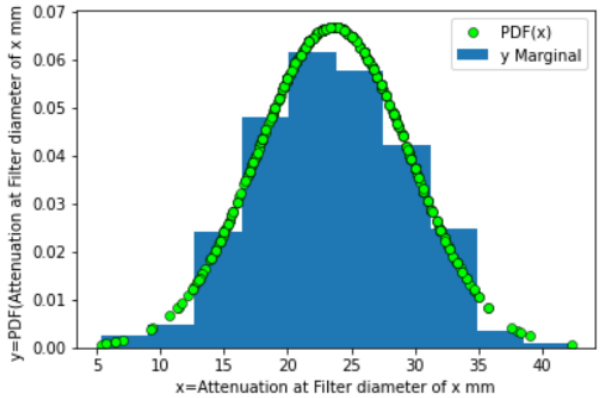


(b) Plot of Frequency (Hz) vs Attenuation (dB) for different filter diameters x mm (Blue curve), x+0.10 mm (Green curve) and x+0.15 mm (Red curve) of two different subjects in hard range filters

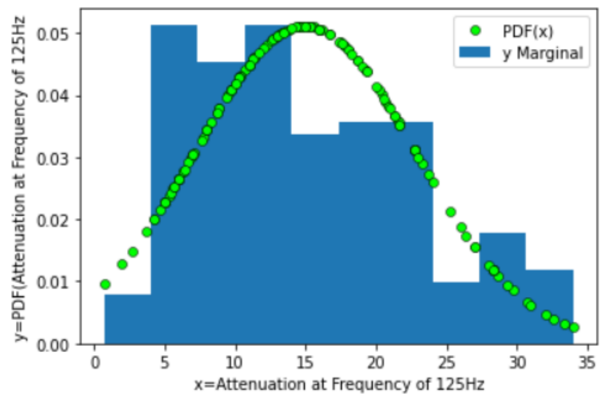
Figure 3.19: Plot of Frequency (Hz) vs Attenuation (dB) with varying filter diameter in soft and hard range filters. The observation is, as the filter diameter increases the attenuation decreases and vice versa

Figure 3.19a and Figure 3.19b shows the plot of Frequency (Hz) vs Attenuation (dB) for different filter diameter (x, x+0.1, x+0.15 mm) of both soft and hard range filters tested on three different subjects and the observation here is that the as the filter diameter increases the attenuation decreases and vice versa.

Now that the input-output relationship is analysed, the next step is to check if the marginal of the target variable has a Gaussian distribution. Because we have assumed that the likelihood is Gaussian as well. If it turns out to be non-Gaussian likelihood then the posterior calculation becomes intractable [49]. Hence it is important to check if the observations are Gaussian. In Figure 3.20a, the histogram plot for Filter diameter of x mm is seen and the green point represents the probability density function of attenuation at Filter diameter = x mm. From the plot, it is clear that the marginals are Gaussian distributed. The Figure 3.20b shows the histogram plot for the frequency at 125 Hz and the green point represents the probability density function of attenuation at frequency = 125 Hz. From the figure, it is observed that the marginals are not a good match for Gaussian distribution. However, the observations are assumed to be Gaussian distributed since the probability density function more or less looks like Gaussian.



(a) The histogram plot for filter diameter of x mm and the green point represents the probability density function of attenuation at Filter diameter = x mm. It is clear that the histogram (marginals) are Gaussian distributed and so the likelihood must be Gaussian as well



(b) The histogram plot for frequency at 125 Hz and the green point represents the probability density function of attenuation at frequency = 125 Hz. From the figure it can be observed that the marginals are not a good match for Gaussian distribution. However, the observations are assumed to be Gaussian distributed since the histogram more or less looks like Gaussian

Figure 3.20: Assumption is that the likelihood is Gaussian distributed. If it turns out to be non-Gaussian likelihood then the posterior calculation becomes intractable [49]. Hence it is important to check if the observations are Gaussian and so the observations are checked to see if their marginals are Gaussian distributed. Two observations are plotted here for reference. The histogram of sound attenuation at filter diameter = x mm and frequency = 125 Hz are plotted.

Encoding

In Figure 3.1, we have two categorical variables, Filter Hardness and Filter Material. Since the GPR model does not work with categorical variables, before giving it to the model, these variables must be encoded. The encoding method used here is one hot encoding where the categorical variables are converted to binary values. Figure 3.21 shows one hot encoding method, where it creates new columns of binary values, where the value of 1 represents the presence of a variable and the value of 0 represents the corresponding variable's absence. So, it creates additional features in the input dataset. If we have 2 columns of 4 categorical variables, after encoding we will end up with 4 columns.

Filter Material	Filter Hardness
Material 1	Hardness 1
Material 2	Hardness 2

Encoding →

Filter Material 1	Filter Material 2	Filter Hardness 1	Filter Hardness 2
1	0	1	0
0	1	0	1

Figure 3.21: One Hot Encoding of filter material and hardness. The categorical variables are converted to binary values where the value of 1 represents the presence of a variable and the value of 0 represents the corresponding variable's absence. So, it creates additional features in the input dataset. In this figure we have 2 columns (Filter Material and Filter Hardness) of 4 categorical variables (Material 1, Material 2, Hardness 1 and Hardness 2), after encoding we will end up in 4 columns: Filter Material 1, Filter Material 2, Filter Hardness 1, Filter Hardness 2

Kernel Function

Before giving the dataset to the prediction model, an important step is to select the kernel function based on the input data trend. For example, the data trend in the venturi dataset is that, as the frequency increases the attenuation increases with frequency and as the filter diameter decreases the attenuation increases. To model this trend and because of the nature of the dataset both RBF and Matern kernel functions are used. The MAE scores of the model after using these kernels are checked and based on the better MAE scores the kernel is selected. The model is trained three times first with RBF kernel, then with Matern 3/2 kernel and finally with Matern 5/2 kernel. After the training period, the Mean Absolute Error (MAE) between the actual attenuation values and the predicted attenuation values is checked. From Figure 3.22, based on the mean MAE, it is clear that the MAE is relatively low when the RBF kernel is used. It might be because the RBF captured the trend of input data well. So, the RBF kernel is selected.

Trials	MAE_RBF	MAE_Matern32	MAE_Matern2
1	3.25	3.26	3.21
2	3.59	3.27	3.57
3	3.3	4	3.6
Mean_MAE	3.28	3.51	3.46

Figure 3.22: MAE scores between the actual attenuation values and the predicted attenuation values for RBF and Matern kernels. The model is trained first with RBF kernel, then with Matern 3/2 kernel and finally with Matern 5/2 kernel. After the training period, the MAE values are checked. The MAE is relatively low when the RBF kernel is used. It might be because the RBF captured the trend of input data well. So, the RBF kernel is selected.

Since we have an input dimension of 7 (after encoding), i.e. the number of features is 7, hence a multi-dimensional kernel, Squared Exponential - Automatic Relevance Determination (SE-ARD) is used. The kernel hyperparameters are different for each input feature as seen in Section 3.3.1. Before optimising the kernel, initial hyperparameter values are as follows:

- Length scale for input dimensions of 7 = [1, 1, 1, 1, 1, 1, 1]
- Process variance = 1

- Gaussian noise Variance = 1

To select the optimizer and for estimating the optimal hyperparameters of the kernel function for each input feature, again different optimizers are tried. The run time and the number of evaluations for each optimizer are checked. From Figure 3.23, it can be seen that the run time and number of evaluations of the L-BFGS-B optimizer are much lesser when compared to the other optimizers and the process variance and Gaussian noise variance turns out to be almost the same. The length scale for the last 4 dimensions is the same and relatively higher than the other three dimensions because those are the binary values after encoding since there is not much variation between the binary values, the lengthscale values are relatively higher. Dimension 3 has a length scale of 0 because, this input feature has a filter diameter of very low values, for example, 0.1, 0.11 (in mm). So, to capture these variations the length scale is low.

Optimizer	Run time	Number of evaluations	Lengthscale for input dimension of 7	Process variance	Gaussian Noise variance
L-BFGS-B	11s	36	[1, 1, 0, 11.7, 11.7, 11.7, 11.7]	0.883	0.351
TNC	49s	147	[1, 1, 0, 8.9, 8.9, 8.9, 8.9]	0.84	0.357
SCG	2m 35s	394	[1, 1.4, 0, 13.7, 13.7, 13.7, 13.7]	0.795	0.344

Figure 3.23: Comparison of different optimizers available in the GPy framework. The figure shows the run time and number of evaluations of L-BFGS-B optimizer are much lesser than other optimizers. Hence, L-BFGS-B optimizer is used for the hyperparameter optimization.

Likelihood Function

Likelihood function, tells us how well the model has predicted the target given the latent function and input data. Marginal likelihood marginalise the latent function from the model to find the optimal values for the kernel and noise hyperparameters. Since, the marginal turns out to be Gaussian, during the model training, the likelihood is mentioned as Gaussian Likelihood in the GPy framework. Since, the likelihood is Gaussian, the inference of the latent function is tractable i.e. to predict the posterior latent function distribution. So, in GPy framework the inference is also mentioned as Exact Gaussian inference. Finally, the posterior latent function is returned which then is used to make predictions. An overview of the entire process of the training and testing the model is given in Figure 3.24.

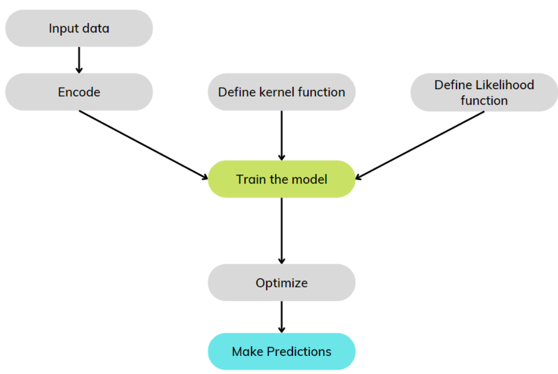


Figure 3.24: Steps involved in training the model and making predictions using noise attenuation data in earplugs

Testing and Validation

Venturi dataset has a total of 1160 data points. The data points are split for training (80% of the total data points), and testing (20% of the total data points). Once the model is trained, new data points in the test data are tested and the attenuation values are predicted. To evaluate a regression model, the Mean Absolute Error (MAE) metric is used [50] and the equation is given as follows:

$$\text{MAE} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} |y_a - y_p|$$

where MAE is the Mean Absolute Error, n is the number of data points, y_a is the actual target values and y_p is the predicted target values. The MAE calculates the absolute difference between the predicted values and the actual values. The evaluation metric used for our model is MAE. The results of the model is given in Chapter 4

3.4.2 Regression on Mesh dataset

In the mesh dataset, the goal is to model the attenuation as a function of frequency (Hz), Layer 1, 2 and 3 specifications. Since the attenuation data provided by the company has both subjective and objective test results, both are included in the dataset. The steps involved for regression in mesh data are the same as mentioned in Figure 3.24. The dataset consists of categorical variables (Test method, Mesh Material, and Layer 3 Material) so it must be encoded first. As mentioned in Section 3.4.1, one hot encoding is used and after encoding the input dimension becomes 13. The Figure 3.25 shows the plot of Attenuation against frequency for different subjects of product 3. Again from this plot, it is clear that attenuation is highly varying with subjects.

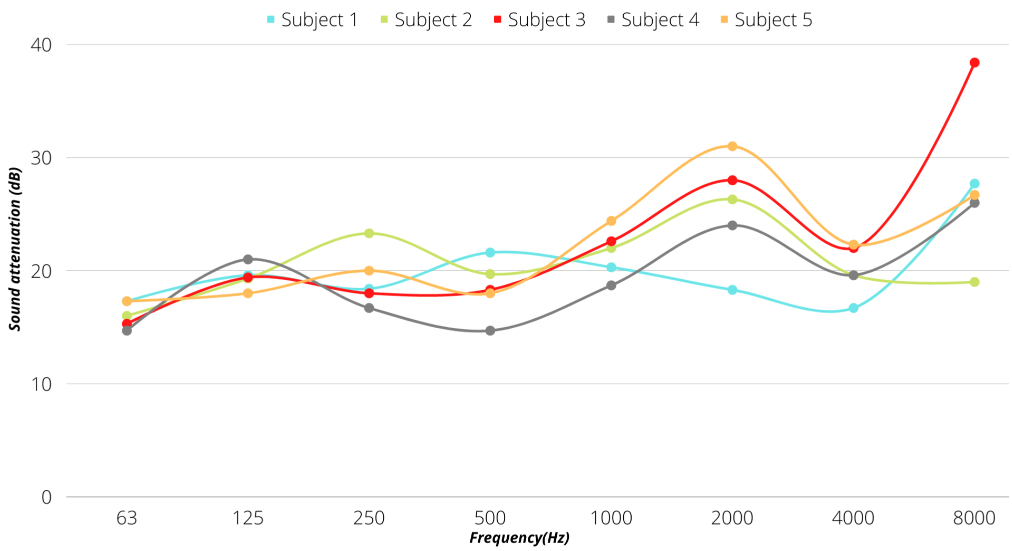


Figure 3.25: Frequency (Hz) vs Attenuation (dB) plot for first five subjects of Product 3 in in Mesh earplugs. The same product tested on different subjects results in different attenuation values for each frequency

From the Figure 3.26, it is clear that the y marginal at frequency= 250 Hz is not a great match for Gaussian distribution. However, the histogram, more or less resembles the Gaussian distribution. So, the observation are assumed to be Gaussian distributed. Since the marginals are Gaussian, the likelihood function and the Inference over latent functions are specified with Gaussian distribution. Once the kernel and likelihood function, are defined the model is trained and tested. RBF kernel and L-BFGS-B optimizer are used. The test results are given in Chapter 4

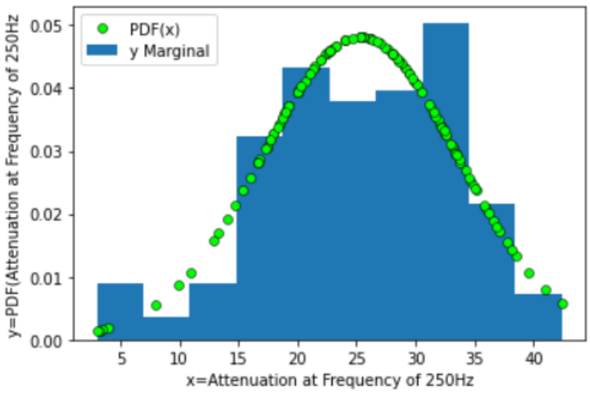


Figure 3.26: Histogram of attenuation at frequency=250 Hz and the green point represents the probability density function of attenuation at frequency=250 Hz. The histogram of attenuation at frequency= 250 Hz more or less resembles the Gaussian distribution. So, the marginals are assumed to be Gaussian distributed

In Chapter 4, the main findings of this research are presented along with the prediction results of the Gaussian Process model.

4

Results

4.1 Findings

Noise attenuation is highly subjective

On analysing the earplug's noise attenuation data, it is found that there exist two products on market. The two products have the same thermoplug and filter configurations but only the name of the product differs. The interesting finding is that the noise attenuation provided by these two products is entirely different. A difference of 10 dB is noted. To identify the root cause behind this difference, the sound attenuation report provided by the laboratory on these two products is checked. From the report, it is observed that the two earplugs are tested on two different target populations (each target population consists of 16 subjects). The two different target populations are varying with age, ear canal size and gender. This means that these two products are not tested with the same set of people and on the same day (testing the same earplug on the same person at two different periods will result in different attenuation values) leading to a huge difference in the noise attenuation level.

To support the aforementioned findings, the paper on Anthropometric analysis of 3D ear scans of Koreans and Caucasians for ear product design [51] is studied. In [51] the following points were mentioned:

- The ear measurements for Koreans were larger than Caucasians in most of the ear dimensions
- The average ear length and breadth in males were longer and wider than in female
- The height of pinna increased with age while the width of concha decreased with age

It is clear from the paper [51] that we have ethnicity, gender and age variations in the ear measurements. So, we can infer that these variations can affect the noise attenuation measurements because the earplug testing method is merely a measurement of an individual's threshold of hearing. So, if the same product is tested on different subjects with different ethnicity, gender and age the noise attenuation will be different.

The main finding here is that sound perception is more important than we think and crucial to earplug design. Even though we can model the physical nature of an earplug system or develop the most ideal sound attenuation prediction model for earplugs it all boils down to the individual sound perception. Due to individual sound perception, prediction of noise attenuation in earplugs will have higher level of uncertainty. However, if we have accurate age, gender, and ethnicity information, the earplugs can be modelled and designed for each group. In such cases, the sound attenuation prediction model can be developed with higher certainty.

4.2 Prediction results of Venturi Dataset

The model is trained on 928 data points and tested on 232 data points. Figure 4.1, shows the actual attenuation values versus the Predicted attenuation value. The **MAE** is observed to be **3.5**.

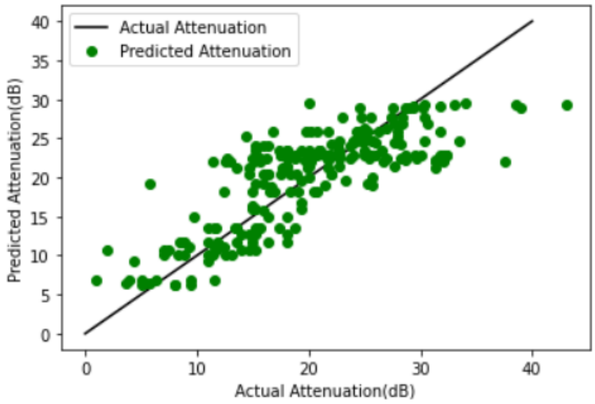


Figure 4.1: Actual Attenuation values (Black line) vs Predicted Attenuation values (Green dots) of venturi dataset with MAE of 3.5

As mentioned in the Section 4.1, products with the same thermoplug and filter configurations ended up in different attenuation values. As this information is fed to the model, the model tried to fit a curve and predict the attenuation value in between the attenuation values of the two products which are evident from the Figure 4.2 and Figure 4.3. Products 1 and 4 (Soft range filters) have the same thermoplug and filter configurations, but different attenuation values. So from Figure 4.2, it is noted that the predicted values (black curve) are in between the actual attenuation values of Product 1 and 4 (Soft range filters). Similar results are observed in Products 2 and 3 (Soft range filters) as shown in Figure 4.3.

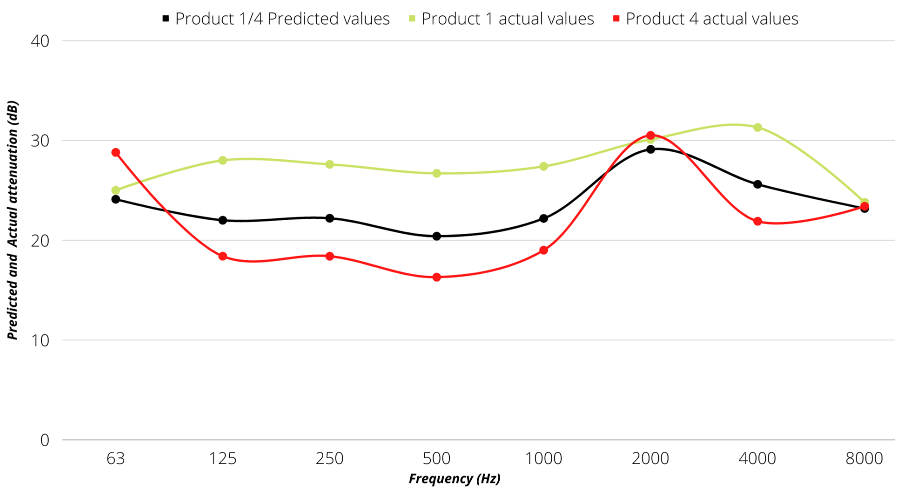


Figure 4.2: Comparison of predicted (Black curve) and actual attenuation values of Soft Range Filter products - Product 1 (Green curve) and 4 (Red curve). The x-axis is the frequency in Hz and y-axis is the predicted (Red curve) and actual (Green curve) sound attenuation in dB. Products 1 and 4 have the same thermoplug and filter configurations, but different attenuation values. The model tried to fit a curve and predict the attenuation value in between the attenuation values of the two products. It is observed that the predicted values are in between the actual attenuation values of Products 1 and 4

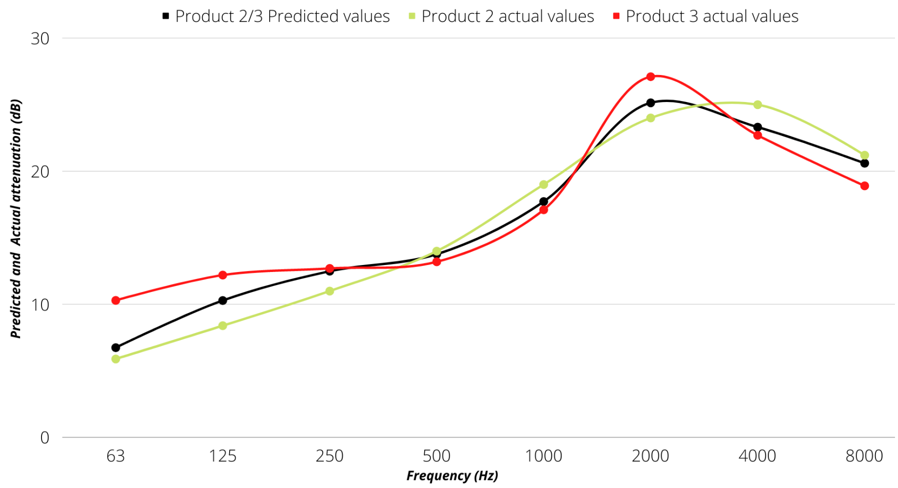


Figure 4.3: Comparison of predicted (Black curve) and actual attenuation values of Soft Range Filter products - Product 2 (Green curve) and 3 (Red curve). The x-axis is the frequency in Hz and y-axis is the predicted (Red curve) and actual (Green curve) sound attenuation in dB. Products 2 and 3 have the same thermoplug and filter configurations, but different attenuation values. The model tried to fit a curve and predict the attenuation value in between the attenuation values of these two products. The predicted values are in between the actual attenuation values of Products 2 and 3

However, this huge error is not observed in hard range filters because no products in this range has same thermoplug and filter configurations. So the error is minimal as shown in Figure 4.4

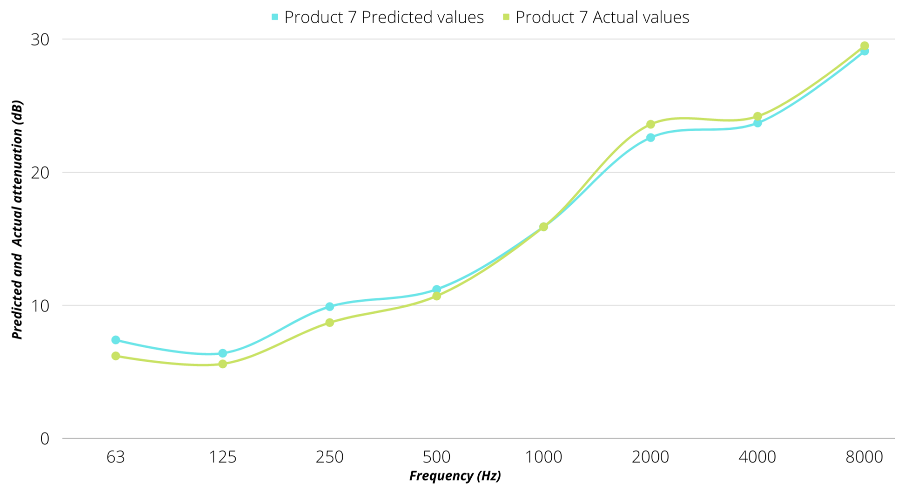


Figure 4.4: Comparison of predicted (Blue curve) and actual attenuation values (Green curve) in Product 7 (Hard Range Filter). The x-axis is the frequency in Hz and y-axis is the predicted (Blue curve) and actual (Green curve) sound attenuation in dB. The predicted values are almost the same as the actual values

Let's test the model by first predicting the attenuation values for an increased filter diameter of x+0.05 mm. As the filter diameter increases, the attenuation must decrease and vice versa. So for filter diameter = x+0.05 mm the attenuation values must be less than the attenuation values of filter diameter = x mm. By testing this on the trained model, it is noted that as the filter diameter increases the attenuation decreases as shown in Figure 4.5

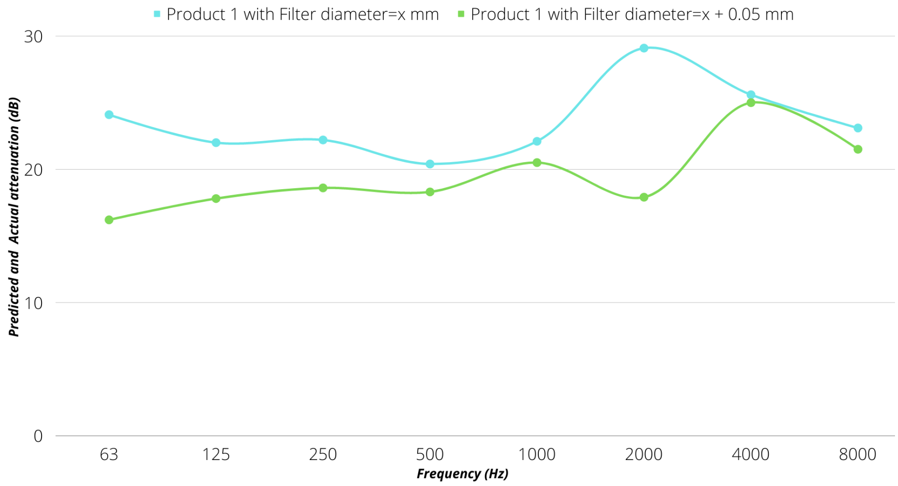


Figure 4.5: Comparison of predicted attenuation values for different filter diameters of x mm (Blue curve) and x+0.05 mm (Green curve). The x-axis is the frequency in Hz and y-axis is the predicted (Green curve) and actual (Blue curve) sound attenuation in dB. As the filter diameter increases the predicted attenuation decreases

4.3 Prediction results of Mesh Dataset

The model is trained on 802 data points and tested on 201 data points. Figure 4.6, shows the actual attenuation values versus the Predicted attenuation value and Figure 4.7 shows the plot of Actual Attenuation values vs Predicted Attenuation values of Product 3 in mesh dataset. The **MAE** is observed to be **3.26**.

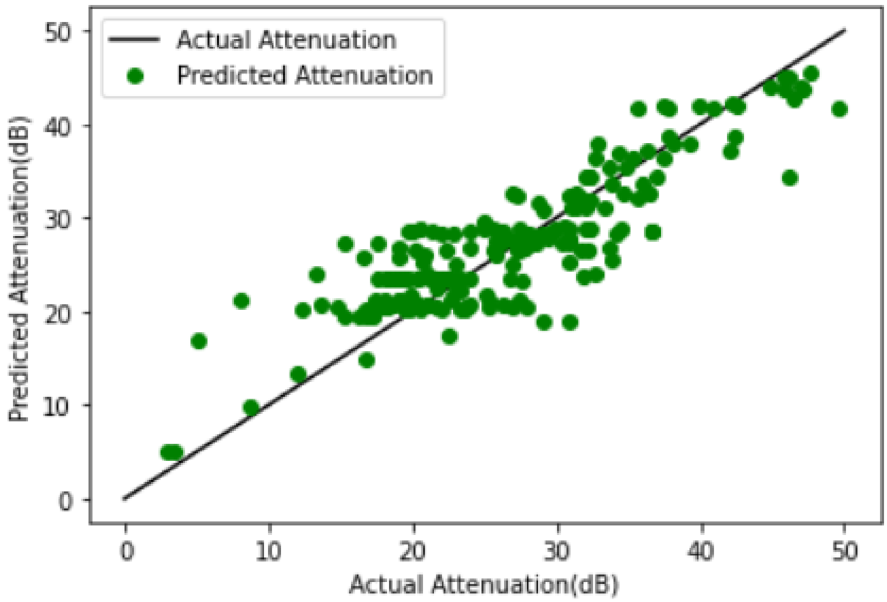


Figure 4.6: Actual Attenuation values (Black line) vs Predicted Attenuation values (Green dots) of mesh dataset with MAE of 3.26

Figure 4.7: Actual Attenuation values (Green curve) vs Predicted Attenuation values (Blue curve) of Product 3 in mesh dataset. The x-axis is the frequency in Hz and y-axis is the predicted (Blue curve) and actual (Green curve) sound attenuation in dB. The predicted values are almost the same as the actual values.

To test the model, the mesh filter's (Layer 2) specification is changed and the values are predicted. As per theory, the attenuation of Mesh P3 must be higher than that of P20 because the percentage of opening in P3 is lesser than P20. It has to be noted that P3 is not used as a mesh layer in any of the products of ALPINE. It is included for testing the model. From the prediction results of the model, it is observed that the attenuation of mesh P3 is higher than P20 as shown in Figure 4.8

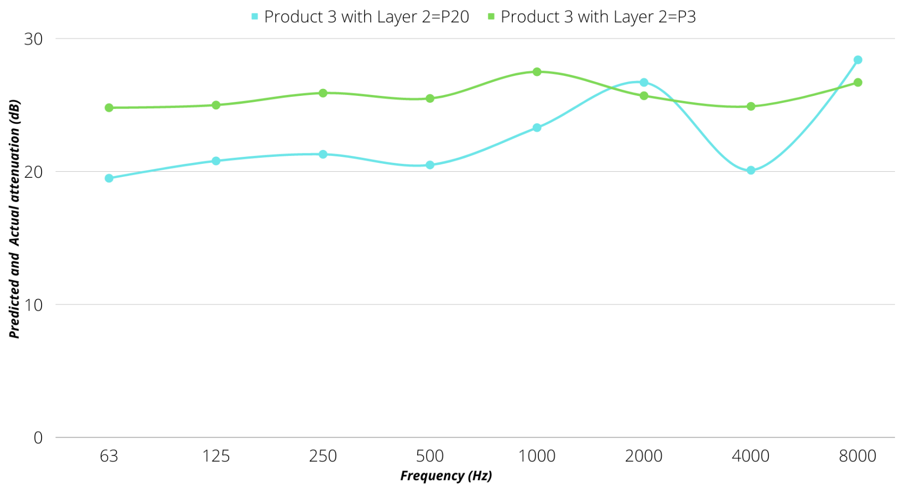


Figure 4.8: Comparison of attenuation values for different Mesh filter layer of P20 (Blue curve) and P3 (Green curve). The x-axis is the frequency in Hz and y-axis is the predicted (Green curve) and actual (Blue curve) sound attenuation in dB. As per theory, the attenuation of Mesh P3 must be higher than that of P20 because the percentage of opening in P3 is lesser than P20. The predicted attenuation values by the developed model resulted in higher attenuation for P3 than P20 aligning with the theory of mesh filter

4.4 Observations

The MAE remains between 3 to 4 for both venturi and mesh datasets. Due to high subjective variations in the attenuation data, it might not be possible to predict the attenuation with higher certainty. This might always result in some errors in the predictions. The available data is quite low with not many variations in the filter and thermoplug specifications. However, the accuracy of the predictions can be increased by adding more data points with varying input features. For example, by adding new products with different filter specifications such as different venturi diameters, materials and so on. In this way, the model will be able to learn better and make good predictions.

Thus, this chapter answers the Research Questions 2 and 3, by predicting the attenuation values using GPR.

5 Conclusion

5.1 Goal

This thesis is mainly focused to find a solution to overcome the difficulties faced by the trial and error method of earplug development in ALPINE. So, this research is mostly oriented towards the company ALPINE. However, the findings mentioned in Section 4.1 can be generalised to the entire earplug manufacturing industry.

The objective of this thesis is to predict the noise attenuation level in earplugs. Hence a regression model is developed to predict the noise attenuation level of the earplug before sending it to the testing laboratories. In such a way, the materials and money used for prototypes can be reduced along with less time consumption. The first research question mainly focused on estimating the individual components' contribution to the total attenuation of an earplug. By analysing the attenuation data available from the company, the contributions of each component in an earplug were estimated. These estimations provided an insight into which component contributed the most and the least to attenuation thereby answering the research question 1. From Section 4.1, it is clear that noise attenuation results are highly subjective. The second and third research questions are as follows: With this subjective uncertainty, whether the noise attenuation in earplugs can be predicted and if yes, how? Based on Section 4.2 results, it is noted that with the GPR model, we can to predict the noise attenuation level in earplugs with some level of accuracy and thereby answering the research questions 2 and 3.

5.2 Important Findings

During the course of research and data analysis, there are some important findings. The main findings of this thesis are summarised as follows:

- The attenuation provided by earplugs is highly subjective. These variations are due to the different hearing capabilities of the subjects which are greatly influenced by Gender, Ethnicity, Age and the associated difference in the ear shape and dimensions. Hence, it might not always be possible to predict the attenuation of the earplugs with higher certainty
- Sound perception is crucial to an earplug design. Regardless of developing the most ideal sound attenuation prediction model for earplugs, it all boils down to individual sound perception. Due to individual sound perception, noise attenuation in earplugs will not be predictable to a higher certainty. However, if we have accurate age, gender, and ethnicity information, the earplugs can be modelled for each group. Developing earplugs for specific groups can be an intensive process but noise attenuation in earplugs will be predictable to higher certainty.

Other findings which are steered towards the products in ALPINE are as follows:

- The contribution of the filter in Venturi type of earplugs around 4000 Hz is very minimal. Most of the contribution is from the thermoplug around 4000 Hz
- In mesh type earplugs, the mesh filter is the main component. But from the analysis, it is found that the mesh filter's contribution to the total attenuation is very minimal when compared to the other components.

The main goal of this thesis is to predict the noise attenuation level in the earplugs using Gaussian Process Regression (GPR). The assumption behind GPR is that the prior, likelihood and posterior are Gaussian distributed. Based on this assumption the dataset is analysed and using GPy framework [48], the GPR model is trained and tested on different datasets. The model predicts the attenuation value of earplugs with mean absolute error fluctuating between 3 to 4. By using the data provided by ALPINE, a constant error will be always present in the prediction results. This is due to the presence of some products in ALPINE with the same design configurations but different attenuation levels. In the perspective of the prediction model, the same input features will have different target variables (for two different products) as mentioned in Figure 4.2 and Figure 4.3 which results in the error. To test the model, the filter specifications are changed with the values which the model has not seen before. With the new filter specifications, the predicted attenuation values are reasonable and aligned with the earplug working theory as shown in Figure 4.5 and Figure 4.8. It might be possible to increase the accuracy of prediction results with the addition of new data points with more information on the thermoplug and filter specifications.

The time involved in the validation for the newly developed model is greater than this thesis period as mentioned in Section 1.5. Hence, the model validation is proposed as an extension of this thesis.

5.3 Recommendations

This thesis might be steered towards researching how sound is perceived by different individuals based on gender, age, and ethnicity differences. These factors greatly influence the ear canal length and shape which is highly diverse and changes from person to person. Doing this research will help develop the earplugs considering the diverse ear dimensions and sound perception of the people. Such information will be useful in incorporating more knowledge in developing the prediction model further. Also, the developed model should be validated to ensure that it is producing reliable results that correspond with reality.

References

[1] World Health Organization. *Deafness and hearing loss*. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (visited on 08/01/2022).

[2] Mahendra. Taneja. "Noise-induced hearing loss ". In: *Indian Journal of Otology* 20.4 (2014), pp. 151–154. DOI: 10.4103/0971-7749.146928. eprint: https://www.indianjotol.org/article.asp?issn=0971-7749;year=2014;volume=20;issue=4;spage=151;epage=154;aulast=Taneja;t=6. URL: https://www.indianjotol.org/article.asp?issn=0971-7749;year=2014;volume=20;issue=4;spage=151;epage=154;aulast=Taneja;t=6.

[3] National Institute on Deafness and Other Communication Disorders. *Hearing Protectors*. 2021. URL: https://www.nidcd.nih.gov/health/hearing-protectors (visited on 08/01/2022).

[4] ALPINE. *ALPINE Hearing Protection*. 2021. URL: https://www.alpine.nl/ (visited on 08/01/2022).

[5] E. H. Berger. "Methods of measuring the attenuation of hearing protection devices". In: *The Journal of the Acoustical Society of America* 79.6 (1986), pp. 1655–1687. DOI: 10.1121/1.393228. eprint: https://doi.org/10.1121/1.393228. URL: https://doi.org/10.1121/1.393228.

[6] WISEWORKSAFE.

[7] Jérémie Voix and Frédéric Laville. "The objective measurement of individual earplug field performance." In: *The Journal of the Acoustical Society of America* 125 6 (2009), pp. 3722–32.

[8] GRAS. *Gras 45ca Headphone/Hearing-Protector Test Fixture.* URL: https://www.grasacoustics.com/products/test-fixtures/product/734-45ca (visited on 07/29/2022).

[9] WTW (Bill) Cory. "Fan noise". In: *Fans and Ventilation, Elsevier Science,* 1.14 (2005). ISBN 9780080446264, pp. 215–237. URL: https://doi.org/10.1016/B978-008044626-4/50016-4.

[10] Wei Fang et al. "Enhanced sound insulation and mechanical properties based on inorganic fillers/thermoplastic elastomer composites". In: *Journal of Thermoplastic Composite Materials* 32.7 (2019), pp. 936–950. DOI: 10.1177/0892705718766382. eprint: https://doi.org/10.1177/0892705718766382. URL: https://doi.org/10.1177/0892705718766382.

[11] Clara Marchini and Antony Joseph. "A Method Paper for the External Auditory Canal Anthropometrics Study". In: *AuD Capstone Projects - Communication Sciences and Disorders. 32.* (2021).

[12] Hearing Link Services. *How the ear works*. URL: https://www.hearinglink.org/your-hearing/about-hearing/how-the-ear-works/ (visited on 07/30/2022).

[13] Britannica. *The physiology of hearing*. URL: https://www.britannica.com/science/ear/The-physiology-of-hearing (visited on 07/30/2022).

[14] Bruce M. Carlson. "Chapter 7 - Special Senses-Vision and Hearing". In: *The Human Body*. Ed. by Bruce M. Carlson. Academic Press, 2019, pp. 177–207. ISBN: 978-0-12-804254-0. DOI: https://doi.org/10.1016/B978-0-12-804254-0.00007-7. URL: https://www.sciencedirect.com/science/article/pii/B9780128042540000077.

[15] Bagus Tris Atmaja. “The Physiology, Mechanism, and Nonlinearities of Hearing”. In: (Jan. 2019). DOI: 10.13140/RG.2.2.24895.97444.

[16] Peter W. Alberti. “2 THE ANATOMY AND PHYSIOLOGY OF THE EAR AND HEARING”. In: 2006.

[17] Ears and Hearing UK. *Ear Anatomy*. 2022. URL: https://earsandhearinguk.com/ear/about-the-ear/the-ear-anatomy_of_the_ear/ (visited on 07/30/2022).

[18] Wayne Staab. *Ear Impression Otoblock Myth Part 2*. 2016. URL: https://hearinghealthmatters.org/waynesworld/2016/otoblock-myth-part-2/ (visited on 07/30/2022).

[19] Farmer B. L. Alvord L. S. “Anatomy and orientation of the human external ear.” In: *Journal of the American Academy of Audiology* 8(6) (June 1997), pp. 383390.

[20] Jennifer Tufts, Siyuan Chen, and Lynne Marshall. “Attenuation as a function of the canal length of custom-molded earplugs: A pilot study”. In: *The Journal of the Acoustical Society of America* 133 (June 2013), EL446–51. DOI: 10.1121/1.4802896.

[21] Chambers R. Kattamis N. Davis B. Norris J. and J. Bieszczad. “Effects of custom earplug design parameters on achieved attenuation”. In: *Poster presentation at the annual meeting of the National Hearing Conservation Association, New Orleans, LA* (2011).

[22] Alice Elizabeth GonzÃąlez. “How Do Acoustic Materials Work?” In: *Acoustics of Materials*. Ed. by Zine El Abiddine Fellah and Erick Ogam. Rijeka: IntechOpen, 2019. Chap. 1. DOI: 10.5772/intechopen.82380. URL: https://doi.org/10.5772/intechopen.82380.

[23] Anshuman Shrivastava. “3 - Plastic Properties and Testing”. In: *Introduction to Plastics Engineering*. Ed. by Anshuman Shrivastava. Plastics Design Library. William Andrew Publishing, 2018, pp. 49–110. ISBN: 978-0-323-39500-7. DOI: https://doi.org/10.1016/B978-0-323-39500-7.00003-4. URL: https://www.sciencedirect.com/science/article/pii/B9780323395007000034.

[24] Abstracta. *Acoustic guide*. 2022. URL: https://abstracta.se/acoustic-guide/#:~:text=Generally%20speaking%2C%20porous%20or%20soft,sound%20and%20thus%20create%20echoes. (visited on 08/01/2022).

[25] Mazharul. *Does Plastic Absorb or Reflect Sound? Ist it Soundprrof! Why?* 2022. URL: https://mrsoundnoise.com/does-plastic-absorb-or-reflect-sound/ (visited on 08/01/2022).

[26] ELLIOTT H. BERGER. “Hearing Protector Performance: How They Work - and - What Goes Wrong in the Real World”. In: *Earlog, Fifth in a comprehensive series of technical monographs covering topics related to hearing and hearing protection.* (1980).

[27] 3M Education. *How to Use the Noise Reduction Rating (NRR)*. 2000. URL: https://multimedia.3m.com/mws/media/918670/3m-hearing-protection-how-to-use-the-noise-reduction-rating-nrr.pdf (visited on 02/15/2022).

[28] Alain Kusy and Jacques Chatillon. “Real-world attenuation of custom-moulded earplugs: Results from industrial in situ F-MIRE measurements”. In: *Applied Acoustics* 73 (2012), pp. 639–647.

[29] MIOA Liz Brueck BSc. “Real world use and performance of hearing protection”. In: *Research Report, Health and Safety Laboratory* (2009).

[30] Aurlien Gron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. O’Reilly Media, Inc., 2019. ISBN: 1491962291.

[31] Naman Singh. *Advantages and Disadvantages of Linear Regression*. 2022. URL: `https://iq.opengenus.org/advantages-and-disadvantages-of-linear-regression/` (visited on 08/01/2022).

[32] Priyanka Sinha. "Multivariate Polynomial Regression in Data Mining: Methodology, Problems and Solutions". In: *International Journal of Scientific and Engineering Research* 4 (Dec. 2013).

[33] Prashant Sharma. *Different Types of Regression Models*. 2022. URL: `https://www.analyticsvidhya.com/blog/2022/01/different-types-of-regression-models/` (visited on 08/01/2022).

[34] Great Learning Team. *What is Ridge Regression?* 2020. URL: `https://www.mygreatlearning.com/blog/what-is-ridge-regression/` (visited on 08/01/2022).

[35] Gokul Elumalai. *Pros and cons of common Machine Learning algorithms*. 2019. URL: `https://medium.com/@gokul.elumalai/pros-and-cons-of-common-machine-learning-algorithms-45e05423264f` (visited on 08/01/2022).

[36] Gurucharan M K. *Machine Learning Basics: Decision Tree Regression*. 2020. URL: `https://towardsdatascience.com/machine-learning-basics-decision-tree-regression-1d73ea003fda` (visited on 08/01/2022).

[37] Saedsayad. *Decision Tree - Regression*. 2022. URL: `https://www.saedsayad.com/decision_tree_reg.htm#:~:text=Decision%20tree%20builds%20regression%20or,decision%20nodes%20and%20leaf%20nodes.` (visited on 08/01/2022).

[38] Chaya Bakshi. *Random Forest Regression*. 2020. URL: `https://levelup.gitconnected.com/random-forest-regression-209c0f354c84` (visited on 08/01/2022).

[39] Sergios Theodoridis. *Machine Learning: A Bayesian and Optimization Perspective*. 1st. USA: Academic Press, Inc., 2015. ISBN: 0128015225.

[40] Depaoli S. King R. et al. van de Schoot R. "Bayesian statistics and modelling. Nat Rev Methods Primers 1". In: *Poster presentation at the annual meeting of the National Hearing Conservation Association, New Orleans, LA* 1 (2021). URL: `https://doi-org.tudelft.idm.oclc.org/10.1038/s43586-020-00001-2`.

[41] Max Reynolds. *When to use Bayesian*. 2017. URL: `https://towardsdatascience.com/when-to-use-bayesian-8723c818b742#:~:text=As%20alluded%20to%20earlier%2C%20Bayesian,a%20prediction%20given%20new%20data.` (visited on 08/01/2022).

[42] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006, pp. I–XVIII, 1–248. ISBN: 026218253X.

[43] Jie Wang. *An Intuitive Tutorial to Gaussian Processes Regression*. 2020. DOI: `10.48550/ARXIV.2009.10862`. URL: `https://arxiv.org/abs/2009.10862`.

[44] Towards AI. *Why Choose Random Forest and Not Decision Trees*. 2020. URL: `https://towardsai.net/p/machine-learning/why-choose-random-forest-and-not-decision-trees` (visited on 08/01/2022).

[45] Scikit Learn. *Gaussian Process*. 2007-2022. URL: `https://scikit-learn.org/stable/modules/gaussian_process.html` (visited on 08/01/2022).

[46] Ivan De Boi. *Gaussian Processes Literature Study 2019*.

[47] D. Duvenaud. "Automatic model construction with Gaussian processes (Doctoral thesis)". In: (2014).

[48] GPy Authors. *GPy - A Gaussian Process (GP) framework in Python*. URL: `https://gpy.readthedocs.io/en/deploy/index.html` (visited on 08/01/2022).

[49] Alan Saul. *Lecture notes Non-Gaussian likelihoods for Gaussian Processes*.

[50] "Mean Absolute Error". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 652–652. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_525. URL: https://doi.org/10.1007/978-0-387-30164-8_525.

[51] Wonsup Lee et al. "Anthropometric Analysis of 3D Ear Scans of Koreans and Caucasians for Ear Product Design". In: *Ergonomics* 61 (Sept. 2018), pp. 1480–1495. DOI: 10.1080/00140139.2018.1493150.