Universidade de Évora – Escola de Ciências e Tecnologia

Mestrado em Engenharia Mecatrónica

Dissertação

Automatic System for Enviromental Noise Characterization

EL-HAOUAS ABDELMOULA

Orientador(es) | Mouhaydine Tlemcani
Fernando Manuel Janeiro
Maria Teresa Batista

Évora 2024

Universidade de Évora – Escola de Ciências e Tecnologia

Mestrado em Engenharia Mecatrónica

Dissertação

Automatic System for Enviromental Noise Characterization

EL-HAOUAS ABDELMOULA

Orientador(es) | Mouhaydine Tlemcani

Fernando Manuel Janeiro

Maria Teresa Batista

Évora 2024

A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências e Tecnologia:

Presidente	|	João Manuel Figueiredo (Universidade de Évora)
Vogais	|	Mouhaydine Tlemcani (Universidade de Évora) (Orientador) Oumaima Mesbahi (Universidade de Évora) (Arguente)

Évora 2024

Acknowledgement

The success and final outcome of this work required a lot of guidance and assistance from many people, and I am extremely privileged to have got this all along with the completion of my work. All that I have done is only due to such supervision and assistance and I would not forget to thank them.

I respect and thank Pr. Mouhaydine Tlemçani, Pr. Maria Teresa Batista and Pr. Fernando Manuel Tim Tim Janeiro, for providing me an opportunity to be in University of Evora and giving us all support and guidance, which made me complete the work duly. I am extremely thankful to them for providing such a nice support and guidance.

I would not forget to remember Pr. João Manuel Gouveia Figueiredo for their encouragement and moreover for their timely support and guidance.

One more time, I heartily thank my internal supervisor Pr. Mouhaydine Tlemçani, for his guidance, suggestions, and shared knowledge during this project work.

Finally, I would like to convey my sincere gratitude to my family and friends for giving me their support and guidance and for inspiring me to continue my education.

Abstract

Automatic System for Environmental Noise Characterization

This thesis delves into a comprehensive investigation of environmental noise impact at Quinta do Carmo, Évora, Portugal. Leveraging cutting-edge techniques, the study entails the production of noise maps, source identification, and rigorous sound level characterization using the PS-3227 sound level meter at ten strategic locations during operational hours. The essence of this research lies in the in-depth analysis of sound levels, encompassing both measurement and mapping methodologies, crucial for a nuanced assessment of noise pollution in landscape areas. This affords valuable insights into the specific activities contributing to heightened sound levels. Moreover, our approach employs a sophisticated supervised learning algorithm to discern nuances among five distinct colored noise sound signals. This algorithm meticulously preprocesses known noise signals, extracting salient features from both time and frequency domains for precise classification into predefined classes. Stringent validation processes ensure the stability and accuracy of predictions. In the testing phase, the classifier is applied to previously unexplored datasets, amplifying the scope and robustness of our findings. Beyond algorithmic intricacies, the study ventures into the exploration of potential correlations between real landscape data and identified noise color patterns, adding a layer of complexity and real-world relevance to our research endeavor.

Keywords: sound; noise pollution; sound perception; environmental noise; noise mapping; colored noise; machine learning; supervised learning; classification; feature extraction

Resumo

Sistema automático de caraterização do ruído ambiente

A presente tese investiga o impacto do ruído ambiental na Quinta do Carmo, Évora, Portugal. Recorrendo a técnicas de ponta, o estudo envolve a produção de mapas de ruído, a identificação de fontes e a caraterização rigorosa dos níveis sonoros, utilizando o sonómetro PS-3227, em dez locais estratégicos durante o horário de funcionamento. A essência desta investigação reside na análise aprofundada dos níveis sonoros, englobando metodologias de medição e de mapeamento, cruciais para uma avaliação matizada da poluição sonora em áreas de paisagem. Isto permite obter informações valiosas sobre as actividades específicas que contribuem para o aumento dos níveis sonoros. Além disso, a nossa abordagem emprega um algoritmo sofisticado de aprendizagem supervisionada para discernir nuances entre cinco sinais sonoros de ruído colorido distintos. Este algoritmo pré-processa meticulosamente sinais de ruído conhecidos, extraindo características salientes dos domínios do tempo e da frequência para uma classificação precisa em classes predefinidas. Processos de validação rigorosos garantem a estabilidade e a exatidão das previsões. Na fase de teste, o classificador é aplicado a conjuntos de dados previamente inexplorados, ampliando o âmbito e a robustez das nossas descobertas. Para além das complexidades algorítmicas, o estudo aventura-se na exploração de potenciais correlações entre dados paisagísticos reais e padrões de cores de ruído identificados, acrescentando uma camada de complexidade e relevância do mundo real ao nosso esforço de investigação.

Palavras-chave: som; poluição sonora; perceção sonora; ruído ambiental; mapeamento de ruído; ruído colorido; aprendizagem automática; aprendizagem supervisionada; classificação; extração de características

List of contents

List of figures

List of tables

Introduction:

In an era where urbanization and industrialization continue to transform the world's landscapes, the issue of environmental noise pollution has risen to the forefront as a significant challenge. According to the world health organization (WHO), The consequences of excessive noise on human health, well-being, and overall quality of life are profound and multifaceted. One critical aspect of this auditory landscape is the pervasive issue of environmental noise pollution, which resonates with significance in an ever-evolving world. The first facet of this study revolves around the Characterization of Environmental Noise Pollution through Noise Measurement and Mapping, while the second delves into the identification of Colored Noise Signals using Supervised Learning Algorithms.

The research, set in landscape of Quinta do Carmo, Évora, Portugal, unfolds as an analysis of noise pollution through the creation of noise maps, the identification of noise sources, and the measurement of sound levels at ten strategic points. This study, conducted during working hours, meticulously employs sound level meters to collect data. The results unearth a compelling narrative ten sampling points examined, the noise levels did not exceed the maximum permissible limit of 55 dB(A), with the peak measurement reaching a staggering 66.3 dB(A).

In the second part of our study, we ventured into the realm of Colored Noise Signal Identification, employing advanced supervised learning algorithms. This phase commenced with the generation of colored noise signals using the versatile MATLAB platform. Subsequent feature extraction was meticulously conducted to capture the distinguishing characteristics of each noise color. The heart of this segment lay in the application of linear and quadratic discriminant analysis, sophisticated techniques that enabled us to identify and classify distinct noise colors effectively. Post-learning phase, we undertook rigorous validation procedures to ascertain the classifier's accuracy. This investigation aimed to discern patterns and correlations within the collected sound data. Our subsequent analysis has uncovered intriguing insights into the prevalence and distribution of various noise colors at specific data points. The following discussion delves into the outcomes of our LDA and QDA analyses, shedding light on the dominant noise types and unique acoustic characteristics within our study area.

This multifaceted approach aims not only to contribute to the current understanding of environmental noise pollution but also to offer an innovative and practical system for its automatic characterization. The fusion of real-world measurements with state-of-the-art machine learning techniques stands to provide a holistic and in-depth perspective on this pervasive issue. Our research holds the promise of enhancing our grasp of environmental noise pollution and, by extension, paving the way for more effective strategies to manage and mitigate this problem in urban settings. Ultimately, our Automatic System for Environmental Noise Characterization represents a significant step towards creating healthier and more harmonious urban environments in the face of mounting noise pollution challenges.

Chapter 1: State of Art

1. The physique of sound

Sound production depends on the exchange of energy among air molecules. Sound energy is generated when a vibrating object, in contact with a gaseous medium (such as air), causes air molecules to move with the surface. This results in a net velocity being imparted to the molecules, which then bunch up as the surface pushes them together in an outward direction. The increased proximity of the molecules leads to an increase in air pressure, which is the cumulative force of the molecules upon each other and any object they encounter. Two forms of energy contribute to the propagation of sound: the net velocity of the molecules (kinetic energy) and the pressure they exert (potential energy). Energy can be exchanged between these two forms, however, it cannot be created or destroyed - only dissipated through the expenditure of energy as heat or through the application of a force over a distance to move a mass [1]. The total sound intensity at a distance from a sound source is the product of the two forms of energy measured at that point:

$$I = pu = \frac{P}{A} \qquad (1)$$

Where:

- I represents the sound intensity (W/m^2).
- p is pressure (Pa).
- u is particle velocity (m/s).
- P is power (W).
- A is the area (m^2) through which the sound energy flows.

The velocity of particles is quantified as the rate of displacement of the mass of moving air, rather than the velocity of individual molecules [2]. **The intensity of sound** is expressed as the power of the sound wave, expressed per unit time, passing through a given area.

As the particles approach one another, the force of interaction between them increases. If the vibrating surface is oscillating, alternating between approaching and receding from the compressed air region, then the force pushing the molecules together decreases. Consequently, as

the surface retreats from the compressed air region, a region of lower pressure is created which results in the particles moving backward to fill this space [3]. Upon completion of this process, their net movement is zero; they return to their initial positions and no displacement occurs unlike in wind where particles move consistently in one direction.

The wavelength (λ) of a sound wave is determined by the period of oscillation and the velocity of propagation and is calculated as the distance traveled between successive peaks of compression or rarefaction. This is a result of sound energy traveling in successive waves of compression and rarefaction.

$$\lambda = cT = \frac{c}{F} \qquad (2)$$

At a fixed c constant velocity of propagation (m/s), an increase in frequency f in ($Hz = \frac{1}{S}$) leads to a decrease in wavelength λ in (m). Conversely, an increase in propagation velocity leads to an increased wavelength due to the wave traveling farther during one cycle of oscillation. Low-frequency sound waves have long wavelengths, up to 17 m at 20 Hz; conversely, high-frequency sound waves have shorter wavelengths, such as 1.7 cm at 20 kHz. This difference is due to how sound waves interact with objects of similar dimensions; the behavior of sound waves varies with different frequencies in the same acoustical environment.

Oscillations of air molecules can cause energy to be transferred to other objects when they come into contact, resulting in the objects beginning to move. This is how sound is detected: the pressure variations in the air cause a microphone or eardrum to vibrate in harmony with the oscillations of the air. Depending on its origin, this vibration may then be interpreted as music, speech, or noise.

2. Categorization of sound

Sound could be categorized commonly in landscape by three divisions biophonic, geophonic, and anthrophonic sound. Biophonic sound is the collective sound produced by all living organisms in a particular biome. Geophonic is the study of the acoustic relationships between living organisms and natural refers to the sounds of the environment, such as wind, running water, and rain. Anthrophonic soundscapes are composed of human-generated sounds, such as traffic and construction noise. They can be evaluated using a reliable approach to assess their

prevalence and acceptability in the urban environment [4]. Anthrophonic sounds are also a crucial factor in protected area conservation, with mountain landscapes often featuring favorite and prevalent sounds. This categorization could be helpful because they are often discernible aurally (by ear).

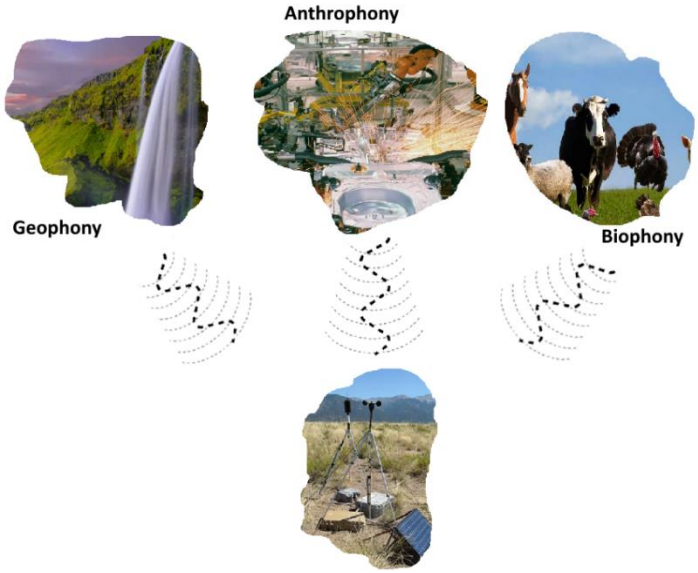


Figure 1: sound transmission models for (a) single and (b) multiple sources of sound

3. Sound synthesis

3.1. algorithms

Sound synthesis uses mathematical functions and algorithms to produce sounds without a direct physical interpretation. They allow the creation of interesting efficient sounds that cannot be achieved physically. Practically, they require low memory and offer a dynamic control of the generated spectrum using very few parameters. However, they are not so easy to understand and manipulate. Most of the techniques for sound synthesis have been developed in the 1970s and 1980s such as the frequency modulation (FM) synthesis and wave-shaping synthesis [5].

The frequency modulation (**FM**) **synthesis** as a means of producing audio sounds, was developed in the 1970s by John Chowning and was the basis of some of the early prototype digital synthesizers like the well-known Yamaha DX7. The principle fundamental of FM is to use a modulator signal to modify the frequency of the carrier audio signal that is played, see Figure 2. Thus, by taking the easiest example when the two oscillators generate sines waves, we obtain the resulting formula:

$$y(n) = a_c \, sin(2\pi(f_c + d \, sin(2\pi f_m n))n) \qquad (3)$$

Where:

- n is the audio sample.
- y is the output of the synthesizer.
- f_c, f_m being the respective frequencies of oscillation of the carrier and the modulator.
- a_c the amplitude of the carrier.
- d the modulation index.

This method permits the generation of signals with a rich-harmonic content from a quite simple waveform and few parameters.

Wave-shaping synthesis, or non-linear distortion. It consists of passing the original signal through a selected non-linear transfer function to distort the waveform and modify its harmonic content to create a richer signal for instance. Practically, it could be easy to get a signal containing only odd or even harmonics with this technique by passing unit-amplitude sinusoid through respectively an odd ($f(-x) = f(x)$) or an even ($f(-x) = -f(x)$) transfer function. Commonly used functions are polynomials (e.g. Chebyshev polynomials) or piecewise functions (e.g. hard clipping) [6].

Most of the abstract techniques for sound synthesis have been developed in the 1970s and 1980s and according to were considered obsolete at some point, as more expressive and efficient methods emerged. However, they remain famous techniques that take a key place in the sound synthesis history.

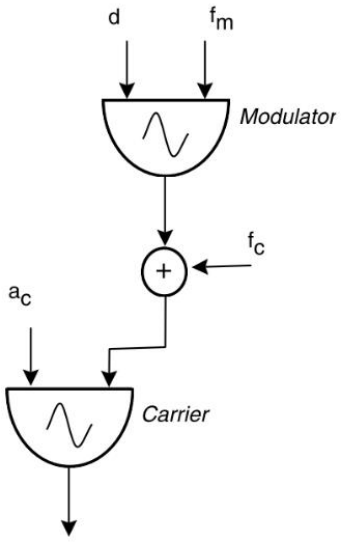


Figure 2: a FM operator typically consists of a sine wave.

3.2. Processed recordings

Processed recording techniques are time-based approaches that manipulate recordings of pre-existing sounds to reproduce or create new timbres. Originally, in the 1950s, sound recordings were stored on magnetic tapes, which could be manipulated by playing them at different speeds, reversing their direction, or editing by cutting, pasting, or looping sections of the tape. With the advent of computers, digital audio samples became the new standard for sound storage, and various time-based techniques for processing them emerged. These techniques allowed for a greater degree of precision and control over the manipulation of sound recordings, resulting in new timbres and sonic textures.

Granular synthesis is a sound synthesis technique that operates on the microsound level. it consists of putting end-to-end brief sounds called grains with a duration near the threshold of human auditory perception, typically from one to a few hundred milliseconds. Each grain contains a waveform shaped by an amplitude envelope as shown in figure 3.

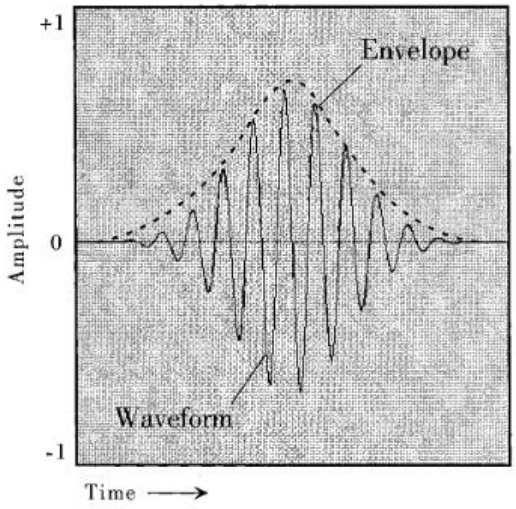
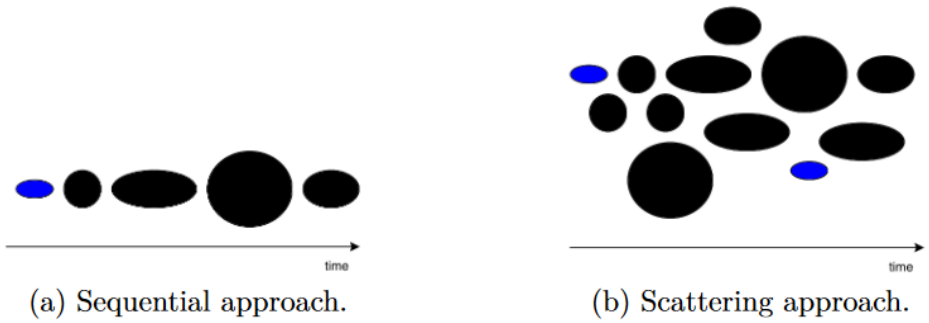


Figure 3: Portrait of a grain in the time domain. The duration of the grain is typically.

These grains can be either extracted from longer recorded sounds or artificially generated. Grains can be aggregated using three different approaches: grains are generated one after the other without overlap (Figure 4, a); in the scattering approach multiple grains can be generated at the same time synchronously or not, forming a "sound cloud" (Figure 4, b); and the granular sampling approach, where a grain is created by selecting a small portion of a sound sample and then applying an envelope to it. To produce a complete sound, either one grain can be replicated many times, or several grains can be aggregated from different portions of the original signal. It is important to choose an appropriate envelope to avoid abrupt changes between consecutive grains when they are combined.



(a) Sequential approach. (b) Scattering approach.

Figure 4:Representations of different granular synthesis approaches

3.3. Concatenative synthesis

Concatenative utilizes a huge database of the source sound, segmented into small *units* along with their corresponding audio *descriptor*. Then, the goal of this technique is to generate *target* sound by selecting the best matching units from the database based on their similarity to the target sound's audio descriptors [7]. To achieve this, a search algorithm is utilized to find the closest matching units, which are then concatenated using short cross-fades. In some cases, additional modifications such as pitch shifting may be applied to further refine the final output. The representative diagram of the concatenative synthesis in Figure 5.

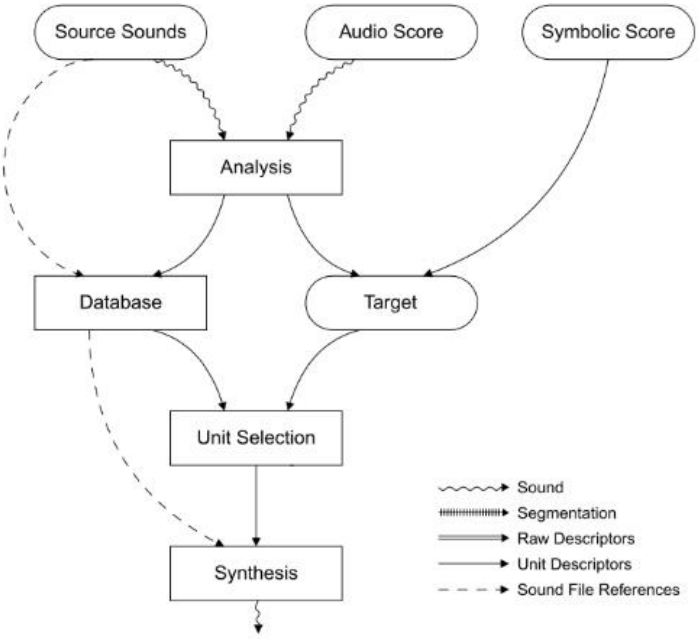


Figure 5: Data flow model of a concatenative synthesis system.

Wavetable synthesis is an example of temporal-based synthesis, which encompasses several synthesis techniques that use signal waveforms stored in a computer memory known as a wavetable or lookup table. These waveforms can either be purely synthetic, such as single-cycle waves that have precisely one cycle/period of the signal, or they can be recorded excerpts. There are various techniques available for generating sounds, such as periodic sound creation through repetitive waveform reading, generating new timbres through cross-fading between two wavetables, and combining multiple wavetables by applying an envelope to each and then

summing them. However, a significant drawback of these techniques is the requirement of a considerable amount of memory for storing the sample database, which may be large. Nevertheless, these techniques are usually computationally efficient due to the pre-existing storage of the necessary information.

3.4. Spectral modeling

Spectral modeling approaches in audio synthesis aim to characterize sound properties by focusing on their spectral content, resulting in models that closely align with human sound perception [8]. These techniques provide parameters that are more representative of psychoacoustic aspects rather than the underlying acoustic mechanisms.

Among the earliest and most prominent methods in spectral modeling synthesis is additive synthesis, historically employed in Telharmonium, the first synthesizer. Additive synthesis, inspired by Fourier analysis, posits that any periodic waveform can be represented as a sum of sinusoids with distinct amplitude envelopes and frequencies, for example in Figure 6. Generation entails the summation of multiple sinusoidal oscillators with varying frequencies ($f_k(n)$ and amplitude envelopes ($a_k(n)$. However, a drawback of this modular and potent approach lies in the significant number of parameters required. Synthesizing signals with intricate harmonic content or noisy characteristics often necessitate a high number of oscillators, leading to complex manual control and computationally demanding processes.

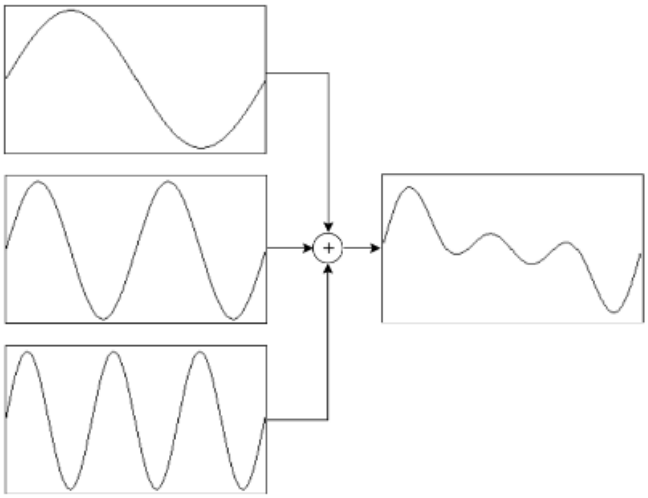


Figure 6: Representation of additive synthesis with 3 sinusoidal oscillators (M=3)

Subtractive synthesis is another well-known spectral modeling technique widely utilized in hardware and software synthesizers since the 1960s. It stands in contrast to additive synthesis, adopting a source-filter synthesis approach. Instead of adding simple and harmonically limited waveforms to create a complex signal, subtractive synthesis begins with a rich excitation signal, such as pulses or noise, and passes it through a filter to selectively remove undesirable harmonics. This method offers fewer controls compared to additive synthesis, primarily focusing on the design and configuration of the filter(s). However, the resulting sounds generated by subtractive synthesis often possess an inherently "artificial" quality.

3.5. Physical modeling

The general principle of this method is to mathematically model the laws of physics responsible for simulating sound-producing processes by means of a set of equations such as laws of acoustics and mechanics. Sound is synthesized by characterizing the behavior of different elements of the instrument or sound source. These elements can include mass, springs, strings, resonators, and more, depending on the specific instrument being modeled. Physical modeling methods aim to simulate the physical mechanisms of synthesis, which may not be directly connected to perception but are based on the laws of physics [9]. Despite this distinction, these methods offer users a realistic instrument-like experience.

There are various categories of physical modeling exist. The first category and oldest one is the numerical solution of differential wave equations. While this method can be applied to any vibrating object, it has primarily been used for modeling string instruments. Another approach is modal synthesis an instrument represented as a combination of modal acoustic structures and their interactions. These structures can include components such as membranes, air columns, metal plates, strings, or bridges. By assembling these elements and incorporating physically unattainable interactions, novel and intriguing timbres can be created. A last example of a physical modeling technique that can be given is waveguide synthesis is a popular technique for real-time sound generation in physics-based commercial instruments due to its efficiency. The wave propagation equation is solved within a waveguide model consisting of a bidirectional delay line and filters to simulate acoustic phenomena. See Figure 7.

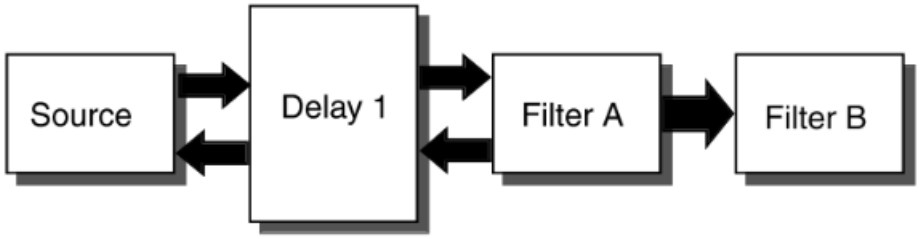


Figure 7:Generic waveguide filter instrument.

However, physical modeling synthesis provides scientists with physically meaningful controls and the possibility to synthesize sound with high quality, the resultant models are highly specialized and strongly influenced by the intended sound characteristics they were designed to produce, such as those of plucked strings or reed instruments. The efficacy of each physical modeling technique is dependent on its particular application in sound synthesis.

4. Microphones

In the process of sound recording, microphones are critical transducers for capturing non-electronically synthesized sounds. A bewildering range of microphones is available, each with its unique suitability for specific applications. Effective microphone selection and placement requires a deep understanding of microphone functionality to achieve optimal results. While experience is key, knowledge of the properties of different types of microphones can help us to expedite this process.

Microphones can be classified based on their physical method of transduction (figure 8) and spatial sensitivity pattern. The interrelationship between these behaviors is complex and requires a comprehensive examination of the various physical transducer types. The two main types of transducers used for recording are dynamic and capacitors. Dynamic microphones utilize a conductor that moves within a magnetic field in response to the forces applied by incident sound waves, while capacitor microphones use one plate of a capacitor that moves relative to the second fixed plate. Both types convert air vibrations into corresponding electrical voltages. While there are variations on these basic types, new types are also being developed. A thorough examination of dynamic and capacitor microphones can serve as a starting point for exploring the topic [3].

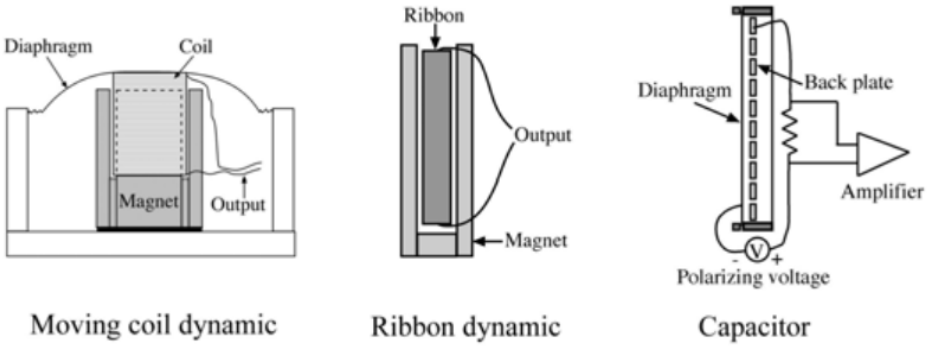


Figure 8:Microphone transducer types.

4.1. Dynamic microphones

Dynamic microphones are a type of passive microphone that operates on the principle of electromagnetism, which is the process of inducing a current in a wire by moving a conductor within a magnetic field. In the case of moving-coil dynamic microphones, a fine-diameter wire coil is placed in a magnetic field and attached to a diaphragm that reacts to changes in air pressure. The coil generates a small voltage as it moves in the magnetic field. This voltage is then transmitted, typically through a transformer, to an external amplifier with low input impedance and high gain. The output voltage is directly proportional to the velocity at which the coil moves through the magnetic field, resulting in the microphone's inherent sensitivity to air particle velocity as shown in equation 4.

$$e(t) = Blu(t) \qquad (4)$$

Where:

- $e(t)$ the s instantaneous output voltage.
- B the magnetic field strength.
- l is the length of the conductor.
- $u(t)$ is the instantaneous velocity of the conductor.

Although conceptually uncomplicated, the implementation of dynamic microphones is a complex process, as the mass of the wire coil used in the microphone element is not negligible. Therefore, great care must be taken in its construction to ensure that the element can move with ease, enabling the detection of small air pressure variations at all audible frequencies. Moreover, acoustical considerations must be considered to produce consistent output levels for sounds

emanating from various directions and frequencies [10]. As a result, elaborate acoustic labyrinths are often integrated into the microphone housing to regulate its frequency response and directional sensitivity.

A second category of dynamic microphones is the **ribbon microphone**, which features a thin, corrugated metal ribbon suspended within a magnetic field, and capable of moving in response to air movement from both the front and rear of the ribbon. The ribbon microphone exhibits lower conversion efficiencies and produces lower output voltages for a given sound level, but possesses a lower mass compared to a moving coil element and can provide superior high-frequency transduction. Since both the front and back of the ribbon are exposed to moving air, the ribbon microphone is sensitive to sounds emanating from the front and rear but not from the sides, as forces from the side of the ribbon are ineffective in inducing motion and producing an output. This design yields a unique bidirectional or figure-eight polar sensitivity pattern, although its characteristic pattern may be altered using acoustical techniques to control the access of the air to the ribbon. Ribbon microphones were pioneered by RCA in the early 20th century and were widely employed in many renowned recordings and radio announcing during that period. With the advent of new ribbon materials and stronger rare-earth magnets, the ribbon microphone has gained renewed popularity due to its enhanced durability and higher output levels than earlier designs.

4.2. Capacitor microphones

A capacitor microphone comprises a thin conductive membrane, supported by a ring, positioned near a stationary plate, separated by an air dielectric. Because v = q/C, by keeping the voltage (v) constant while modifying the capacitance (C), a corresponding alteration in charge (q) is induced, which is equivalent to the current. The current then passes through a high-value resistor, resulting in a voltage drop that is subsequently buffered and amplified to produce the microphone's output [11]. The voltage created by an omnidirectional capacitor microphone as a function of pressure is approximated by Equation 5:

$$e = \frac{E_0 a^2 P}{8hT_0} \qquad (5)$$

Where:

- e is open-circuit output voltage.
- E_0 is the polarizing voltage.
- a is the diaphragm radius (m).
- P is the pressure (Pa).
- h is the distance (m) from the diaphragm to the backplate.
- T_0 is the diaphragm tension (N/m).

Unlike the dynamic microphone, the capacitor microphone requires an external power source in the form of a battery or phantom power to operate. **Phantom power** is a method of applying external power through the **XLR** cable by raising the two signal lines to a common **DC** voltage via the preamplifier. Electret capacitor elements, which use a permanently charged plate, are also available. The construction of the capacitor element involves both art and science, with materials, design, and manufacturing techniques all affecting the quality of the microphone. Tensioning of the diaphragm is particularly critical and requires careful assembly, increasing the cost of high-quality capacitor microphones. The acoustical properties of the microphone can be altered to allow for dynamic and capacitor microphones to behave as either pressure or pressure-gradient microphones, with the use of mechanical damping and acoustic pathways. This also enables the inherent resonances of the sensing elements to be tailored for a flatter response. The variety of microphones available displays the effects of these design approaches and allows for the selection of a microphone suitable for different applications and budgets.

4.3. Microphone sensitivity

In addition to the type of microphone transducer assembly, sensitivity refers to how the performance performs to convert acoustic pressure to output voltage. Generally, there is a trade-off between sensitivity and frequency response. Small microphones tend to have lower sensitivity; however, they operate at both high and low frequencies, so large microphones possess high sensitivity but are effective mainly at lower frequencies [12]. The definition of sensitivity (S) is the following:

$$S = \frac{elctrical\ input}{acoustic\ input} \qquad (6)$$

The voltage sensitivity of a microphone is measured in (mV) per Pascal (Pa) at 1 kHz.

4.4. Microphone directional sensitivity

Microphone directional sensitivity, also known as polar pattern, refers to the microphone's sensitivity to sound waves arriving from different directions. Different microphone designs exhibit different polar patterns, which can affect the microphone's ability to pick up sounds from specific directions or reject sounds from others [13].

The most common directional sensitivity patterns are:

1. Omnidirectional: The microphone picks up sounds equally from all directions.
2. Cardioid: The microphone is most sensitive to sounds coming from directly in front of the microphone, and least sensitive to sounds coming from behind the microphone.
3. Hyper-cardioid: such as the cardioid pattern, but with a narrower sensitivity range in the front and increased sensitivity to sounds coming from the sides.
4. Bidirectional: The microphone is equally sensitive to sounds coming from the front and back while rejecting sounds coming from the sides.

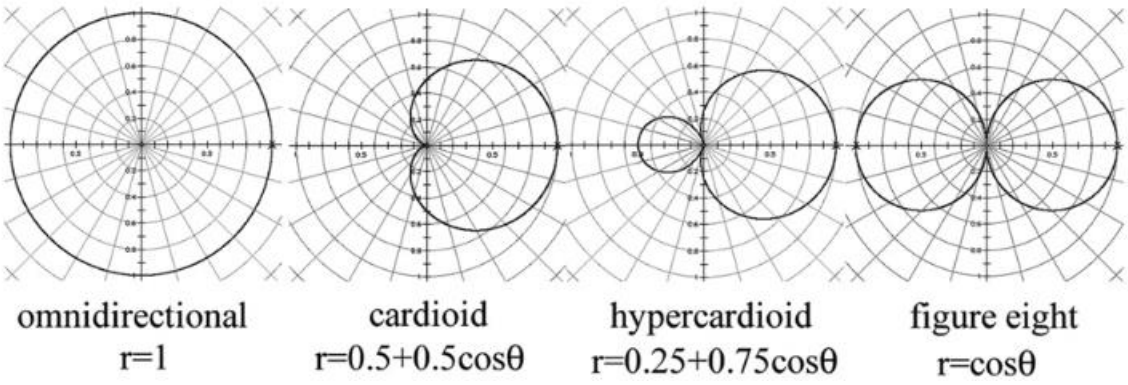


Figure 9:Polar patterns and their polar equations. These are linear plots. Polar patterns are also sometimes presented as log plots.

The polar patterns shown in Figure 9 give the various spatial sensitivity patterns and the mathematical equations that describe them. Often called polar patterns, the radius **r** is a measure of the microphone's sensitivity at the angle **Θ** relative to the front of the microphone, which is defined as the positive **x-axis**. The overall spatial sensitivity is determined by the sum of

contributions from an omnidirectional (pressure) term and a bidirectional cosine (pressure-gradient) term. By varying the ratio of these terms, we can create any desired polar pattern. In fact, variable-pattern microphones achieve this by combining two capsules, although not always the two mentioned: a back-to-back cardioid pair is often used, as it is easier to construct.

Choosing the right directional sensitivity pattern is important for achieving optimal sound quality in each application. For example, a cardioid microphone is often used for recording vocals or individual instruments, while an omnidirectional microphone is better suited for recording ambient sounds or groups of instruments. Hyper-cardioid microphones are often used in live sound reinforcement applications to provide better isolation between different sound sources on stage.

5. Digital

The term 'digital' can be applied to any technology in which sound is created and manipulated using discrete or quantized values, as opposed to continuous values. This typically involves the use of computers and advanced electronics. However, it should be noted that this emphasis on technology is more often a marketing strategy than a result of using digital methods for producing sound. An example of this would be physical modeling synthesizers, where much of the complexity associated with digital processing is deliberately hidden from the user; thus, these synthesizers are perceived by performers as flexible and responsive instruments. Looking towards the future, we may see digital instruments that do not give away their synthetic method of sound production based on external appearance; however, a power supply cable might provide some form of clue.

In the world of digital electronics, real-world values are represented as binary numbers: voltages or currents that can take on only two possible states, either "on" or "off", or "one" and "zero". By combining these two-valued signals, any numerical value can be encoded into digital form. A familiar example of a digital circuit is the light switch; when no dimmer is present, the light will be either on or off.

Gates are simple electronic circuits that take one or more of these digital inputs and generate an output based on their electrical function. For instance, the output may only activate if both inputs are equal, or it may inverse the input state. The rules determining how these interactions occur are known as Boolean algebra.

Registers are circuits that have the capability to store a binary value. Sets of registers can be used to hold whole numbers and are known as memory. To accurately represent real-world values which may change over time, a large amount of memory is typically required. This is especially true for audio signals where high precision and frequent measuring of the value is necessary; synthesizers and samplers often house many memory chips to store audio signals properly.

memory is the most essential element of a computing system it's divided into two forms. Primary memory is permanent storage called read-only memory (**ROM**) Stores crucial information essential to operate the system. Digital information (data) written to individual cells using binary stored in ROM by using a process known as fusing, which involves the application of short bursts of high current to break physical links within the ROM. Second memory is temporary storage called random access memory (**RAM**), allowing for direct access to data as it is required. This stands in contrast to serial memory, such as tape, which requires winding through the entire medium to locate desired data. Certain variants of Read-Only Memory (ROM) can be erased and rewritten; this is accomplished by storing data as electrical charges on capacitors rather than a permanent break in a wire. These reprogrammable ROMs are commonly referred to as Erasable Programmable ROMs (EPROMs), Flash EPROMs, or simply Flash Memory or Flash Drives.

Microprocessors are stand-alone, general-purpose computers that are designed to execute a wide variety of logical operations quickly and efficiently. This is accomplished by utilizing memory stores and registers to store values, an arithmetic section to perform logical and mathematical functions on the stored values, and a control unit to manage the movement and processing of data, usually in the form of a program consisting of instructions.

A digital signal processor (DSP) is a specialized microprocessor chip that has been optimized for the purpose of signal manipulation. They are predominantly used in audio applications (headphones, smart speakers), but can also be employed to process video and other types of signals. The architecture of DSPs is streamlined to provide high performance and efficiency when performing mathematical operations like "add", "subtract", "multiply" and "divide" on these signals [14].

5.1. Digital numbers

Digital techniques are used with sound, there is a way to receive a sound and values as numbers. Digital systems work with binary digits, or bits, to store and manipulate numbers in a basic way. Bits tend to be organized into groups of eight, for various historical and mathematical reasons. An individual bit only has one of two values: on or off, practically given by the values 1 and 0, respectively. Eight bits can represent 256 values, from 0 to 255, or %0000 0000 to %1111 1111 in binary notation, '%' mean binary number, and the binary digits(bits) are bunched into blocks of four for easy reading. A block of 4 is known as a nibble(!), and the collections of 8 bits are called bytes. sixteen bits or two bytes make up 'word' can be used to represent numbers 0 to 65,535, and more bits can give a larger range of numbers. A word is called the most significant and the least significant bytes, which are normally abbreviated to MSB and LSB, respectively. the numbers are integers – only whole numbers can be represented with this method.

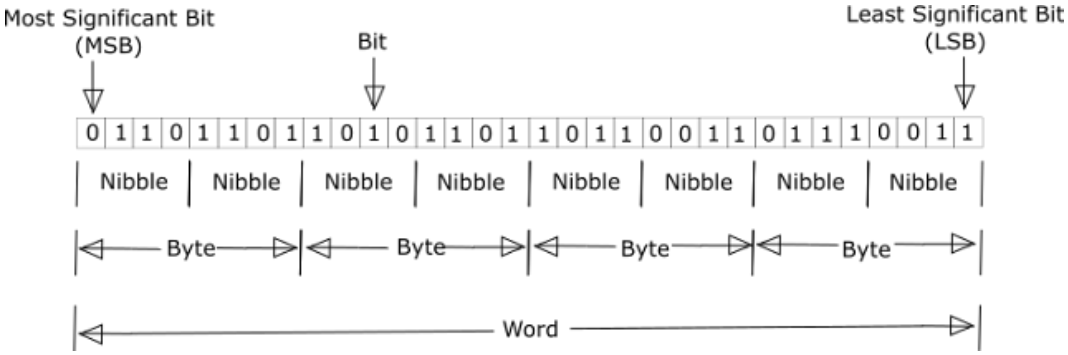


Figure 10:Data binary type transformation representation

Larger numbers, at most decimal numbers, involve a different method of representing them. Floating point numbers split the number into two parts: first a decimal number part 0 to 9.9 and the other a multiplier or exponent part, which is a power of ten. The value 2314 would be stored as $2.314 * 10^3$ in floating point representation. In contrast binary numbers, a power of two is replaced by instead power of ten, However, the concept of representing a numerical value as a decimal number and a multiplier remains invariable. Floating point numbers can be processed using either a microprocessor or specialized chips known as Digital Signal Processors (DSPs). DSPs are designed to carry out complex mathematical operations on numbers; some are designed specifically for integers, while others are designed for floating-pointing numbers. These specialized chips are typically used for signal processing operations such as filtering, equalization, and producing effects such as echo, reverb, phasing, and flanging.

5.2. Computers and time

Our perception of the physical world is one that exhibits continuous behavior, where each moment in time has a unique value. However, this appearance of continuity can be deceiving, as it is often a consequence of the scale at which we observe certain phenomena. For instance, the propagation of sound through a gas seems to be a continuous process to us, but it is the sum of many individual collisions between gas molecules, each imparting a small force. This summation of many discrete events allows us to describe the observed behavior of physical systems using continuous functions.

Computers operate as discrete systems, with each program instruction and measurement requiring a finite amount of time to execute. However, our perception of time enables us to treat computer processes as continuous systems if their speed is significantly faster than our sensory systems. Moreover, mathematical proofs show that we can determine continuous values even if our measurements are not continuous if we make them rapidly enough relative to the rate of change of the observed quantity. Sampling theory, which rests on a few fundamental equations, provides means of converting between continuous and discrete representations of our measured quantities, As shown in figure 11.

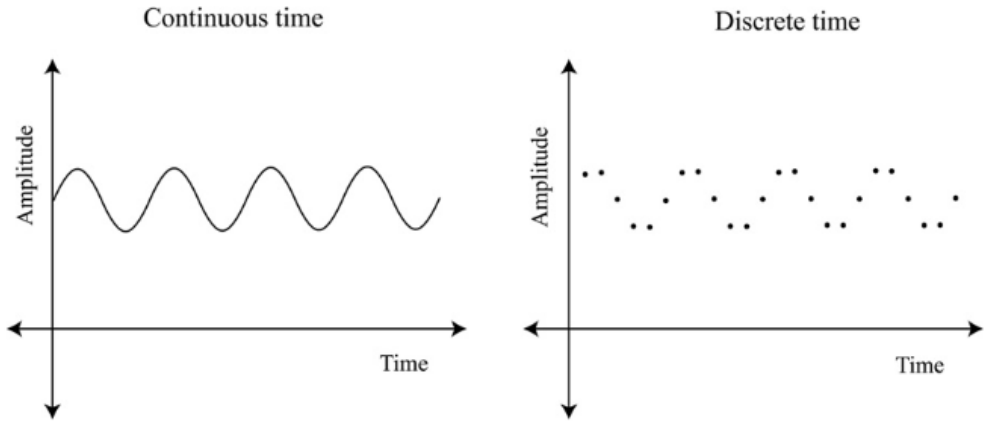


Figure 11:Discretization of continuous signal

6. Sound sensor

Sound sensor usually used for detecting noise in surrounding environment. Arduino can collect the output signal from the sensor and run it simultaneously. Sound sensor can be used to create multiple interactive works such as "clap and ringer" to find lost keys or create a remote control if a buzzer is added. This sensor works by analyzing the sound. Specification tool as follows[3]:

a. Voltage source: 3.3V to 5V.

b. Function: Detects the intensity of the sound quickly.

c. Interface: Analog.

d. Size: 22 x 32 mm

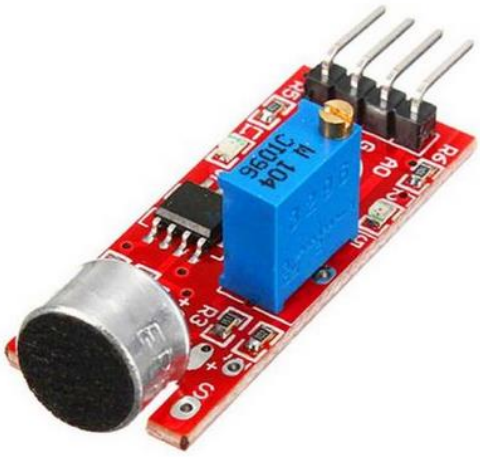


Figure 12: Analog sound sensor type KY-037

Chapter 2: Fourier Transform and Methodology

1. Sampling

Mainly, the signals in real life, such as our voices are called “analog” signals. The sampling process is a conversion from an analog to a digital representation at a fixed time interval T corresponding to a sample rate of $f_s = 1/T$. An audio signal is a continuous sequence of values, in contrast, a digital is a series of numbers. The numbers represent the value (amplitude, or magnitude) of the signal at distinct points in time and these are called samples. As is often the case, there are problems in implementing the ideal theoretical approach using electronic devices. Examining the theory will push us to understand why in part of practice such conversions may not be executed without the same alteration of the signals.

1.1. Nyquist-Shannon sampling theorem.

The Nyquist-Shannon sampling theorem states that if a function $x(t)$ contains no frequencies (at or) above B $hertz$, then it can be completely determined from its ordinates at a series of points spaced less than $1/2B$ seconds apart. Clearly, a sufficient sample rate must be greater than $2B$. This principle is often referred to as the Nyquist theorem, with the Nyquist frequency denoting the limit of $fs/2$. However, the practical implementation of the sampling theorem proved to be challenging once sampling hardware became available [15]. The theorem assumes ideal execution of the sampling period timing and accurate measurement of the signal voltage, which introduces complexities in practice.

Analog audio signals can be described as voltage variations as a function of time. when utilizing analog electronics that generate continuous signals. However, once we sample the data, we can convert it into voltage variations with respect to frequency. This conversion process resembles the way our auditory system transmits sound information to the central nervous system. Although, in theory, analog filters could perform this task, their implementation would prove highly impractical. The time domain corresponds to the representation of voltage variations over time, while the frequency domain pertains to the representation of voltage

variations with respect to frequency. Figure 13 provides a graphical comparison between these domains.

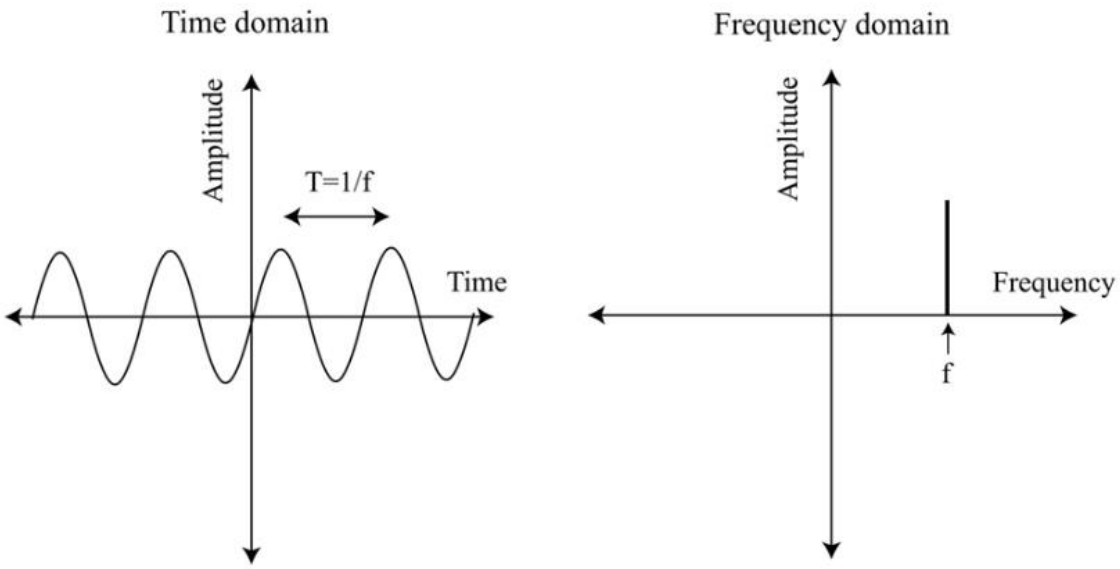


Figure 13:The same sinusoidal signal in time domain and frequency domain plots.

1.2. Fourier transform.

The conversion between the time domain and the frequency domain is achieved through the application of the Fourier transform. This mathematical tool, along with its counterpart, the inverse Fourier transform, enables the representation of data in either domain, thereby facilitating various operations essential for digital signal processing. The Fourier transform's profound utility becomes evident in its applications, such as spectrograms, which dynamically display frequency against amplitude, allowing real-time visualization of signal variations. By isolating, organizing, and manipulating individual spectral components. Moreover, the Fourier transform enables the identification and removal of background noise, as well as the unveiling of hidden sounds that were previously masked by significantly louder ones. These processes heavily rely on the Fourier transform's ability to isolate the distinct spectral components of a signal, which can subsequently be edited and recombined as needed.

The Fourier transform equation is expressed as a sum of a sum of sinusoids. The transform is based on the concept of Fourier series (equation 1), an infinite sum of cosines that can be utilized to represent any periodic signal:

$$\varphi(y) = a_0 \cos\left(\frac{\pi y}{2}\right) + a_1 \cos\left(3\frac{\pi y}{2}\right) + a_2 \cos\left(5\frac{\pi y}{2}\right) + \dots \qquad (1)$$

The a_n coefficient of each term represents the amplitude of a specific frequency component; and the cosine term determines the frequency of that component. However, due to its infinite nature, the series cannot yield an exact value unless we have an infinite number of terms, which is practically unattainable. To effectively utilize the Fourier series, we need to develop a technique that allows us to obtain an approximate solution using a finite number of terms.

1.2.1. Discrete Fourier Transform

The Discrete Fourier Transform (DFT) is simply the Fourier transform with summations instead of integrals and the most powerful tool used to transform a discrete-time signal, such as digital audio, into a representation of its constituent frequencies. The DFT is given by the following equation:

$$X_k = \sum_{n=0}^{N-1} x_n . e^{-\frac{i2\pi kn}{N}} \qquad (2)$$

$$X_k = \sum_{n=0}^{N-1} x_n [\cos\left(\frac{2\pi kn}{N}\right) - i.\sin\left(\frac{2\pi kn}{N}\right)] \qquad (3)$$

Where:

x_n is the discrete-time signal, N is the number of samples in the signal, k is the frequency index, andX_k is the DFT of the signal.

Euler's formula :

$$e^{ix} = \cos(x) + i.\sin(x) \qquad (4)$$

This is more commonly written:

$$X_k = \sum_{n=0}^{N-1} W_N^{kn} x_n \quad (5)$$

Where:

$$W_N = e^{-\frac{i2\pi}{N}} \quad (6)$$

And W_N^k for $k = 0 \ldots N - 1$ are called also the N_{th} roots of unity. They're vertices of a regular polygon inscribed in the unit circle of the complex plane, with one vertex at (1,0). Below are roots of unity for N = 2, N = 4, and N = 8, graphed in the complex plane Figure 14.

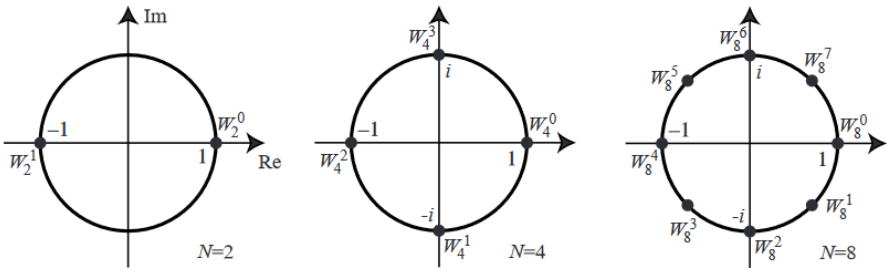


Figure 14: complex plan.

The power of roots of unity with N are periodic, this is because the Nth roots of unity represent points on the complex unit circle every $2\pi/N$ radians apart, multiplying bW_N corresponds to a clockwise rotation by this angle. Multiplying byW_N^Nresults in a complete rotation of 2π radians, which is no rotation at all. Generally, $W_N^k = W_N^{k+jN}$ for any integer j. Hence, when raising W_N to a power, the exponent can be reduced modulo N.

The output of the DFT is a complex number that contains the magnitude and phase information of the signal in the sequency domain. The magnitude represents the strength of the frequency component, and the phase represents the timing of the frequency component. we can

invert to obtain signal original signal back; this process is called Inverse Discrete Fourier Transform (IDFT) and the equation is:

$$x_n = \frac{1}{N}\sum_{n=0}^{N-1} X_k . e^{\frac{i2\pi kn}{N}} \quad (7)$$

$$x_n = \frac{1}{N}\sum_{n=0}^{N-1} W_N^{-kn} X_k \quad (8)$$

The formula remains the same but with a reversal of roles between x and X , as well as between k and n. Additionally, the exponent of W is negated, and there is a normalization factor of 1/N in front.

Two-point DFT (N=2)

$W_2 = e^{-i\pi} = -1$, then

$$X_k = \sum_{n=0}^{N=2-1}(-1)^{kn} x_n = (-1)^{k.0} x_0 + (-1)^{k.1} x_1 = x_0 + (-1)^k x_1$$

So

$$X_0 = x_0 + x_1$$

$$X_1 = x_0 - x_1$$

Four-point DFT (N=4)

$W_4 = e^{-i2\pi}/4 = -i$, then

$$X_k = \sum_{n=0}^{N=4-1}(-i)^{kn} x_n = x_0 + (-i)^{k.1} x_1 + (-i)^{k.2} x_2 + (-i)^{k.3} x_3 \quad (9)$$

$$= x_0 + (-i)^k x_1 + (-1)^k x_2 + (i)^k x_3$$

So

$$X_0 = x_0 + x_1 + x_2 + x_3$$

$$X_1 = x_0 - ix_1 - x_2 + ix_3$$

$$X_2 = x_0 - x_1 + x_2 - x_3$$

$$X_3 = x_0 + ix_1 - x_2 - ix_3$$

To compute X fast, we can pre-compute common subexpressions:

$$X_0 = (x_0 + x_2) + (x_1 + x_3)$$

$$X_1 = (x_0 - x_2) + -i(x_1 - x_3)$$

$$X_2 = (x_0 + x_2) - (x_1 + x_3)$$

$$X_3 = (x_0 - x_2) + i(x_1 - x_3)$$

This significantly reduces the number of additions required. It's important to note that each addition and multiplication mentioned here refers to complex operations, rather than real operations. If we use the diagram below to represent complex multiplication and addition:

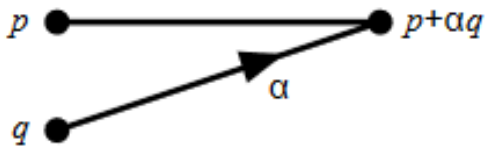


Figure 15: diagram the 2 points.

So, we can diagram the 4-point DFT like so:

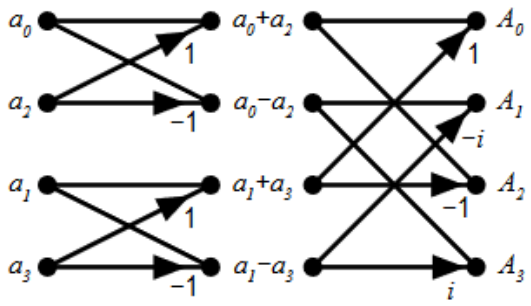


Figure 16: diagram the 4 points.

1.2.2. Fast Fourier Transform

If you were a scientific computer programmer in 1965 and were presented with the challenge of evaluating the discrete Fourier transform (DFT) for many samples, you would soon realize that the task would be extremely difficult to accomplish with the technology available at the time. The DFT is a highly computational algorithm that requires a significant amount of processing power to execute. Unfortunately, the computers of 1965 were not capable of handling the large amount of data that would be required for even a small number of samples. Specifically, the computational effort required for the DFT increases with the square of the number of samples N. So, for a thousand sample sequences, it would require millions of complexes multiplies, but the computers of 1965 were only able to perform a few hundred complex multiplies per second. This mismatch between the computational requirements and the available technology would have made it a formidable challenge to implement and run the DFT. The need for a fast method to perform Fourier analysis was a pressing concern. The goal was to find an efficient algorithm that could quickly analyze large sets of data to extract important frequency information. Until has come in 1965, J. W. Cooley and J. W. Tukey, two mathematicians, introduced the fast Fourier transform (FFT) algorithm in a published paper [16].

1.2.2.1. FFT Algorithm

The FFT algorithm comes especially to fast computing DFT, by taking the 2-point DFT and 4-point DFT concepts and generalizing them to 8-point, 16-point, ….,2^r-point, we obtain the FFT algorithm.

The simplest and most common form of the Cooley–Tukey algorithm is **radix-2** decimation-in-time (**DIT**) FFT. Radix-2 DIT splits a DFT of size N into two interleaved DFTs (hence the name "radix-2") of size *N*/2 with each recursive stage [17].

The discrete Fourier transform (DFT) is described by the formula:

$$X_k = \sum_{n=0}^{N-1} x_n . e^{-\frac{i2\pi kn}{N}} \qquad (10)$$

Radix-2 DIT first computes DFTs

-even-indexed inputs: $x_{2m} = (x_0, x_2, \ldots\ldots, x_{N-2})$

-odd-indexed inputs: $x_{2m+1} = (x_1, x_3, \ldots\ldots, x_{N-1})$

Then combines two results to produce the DFT of the complete sequence. The whole idea can then be performed recursively to minimize the overall runtime to $O(NlogN)$. This simplified form considers N to be a power of two; since the number of sample points N could chose freely by the application (e.g., by changing the sample rate or window, zero-padding, etc.), generally, not an important restriction.

The radix-2 DIT algorithm rearranges the DFT of the functionx_n into two parts: a sum over the even-numbered indices $n = 2m$ and a sum over the odd-numbered indices $n = 2m + 1$:

$$X_k = \sum_{m=0}^{N/2-1} x_{2m} . e^{-\frac{i2\pi k(2m)}{N}} + \sum_{m=0}^{N/2-1} x_{2m+1} . e^{-\frac{i2\pi k(2m+1)}{N}} \qquad (11)$$

We can simplify the second sum by factoring out a common multiplier of$e^{-\frac{i2\pi k}{N}}$, as shown in the equation below. It is then evident that two sums are DFT of the even-indexed part x_{2m} and the DFT of the odd-indexed part x_{2m+1} of the function x_n. By notation DFT of the even-indexed part x_{2m} will be E_k and the DFT of odd-indexed part x_{2m+1} by O_k and we obtain:

$$X_k = E_k + e^{-\frac{i2\pi k}{N}} O_k \qquad \text{for } k = 0, \ldots\ldots, \frac{N}{2} - 1.$$

Please be informed that the equalities hold for $k = 0, \dots, N-1$ but the core is that E_k and O_k are calculated in this way for $k = 0, \dots\dots, \frac{N}{2} - 1$ only. Using the periodicity of the complex exponential, $X_{k+\frac{N}{2}}$ is also obtained from E_k and O_k:

$$X_{k+\frac{N}{2}} = \sum_{m=0}^{N/2-1} x_{2m} e^{-\frac{i2\pi m(k+\frac{N}{2})}{N/2}} + e^{-\frac{i2\pi(k+\frac{N}{2})}{N/2}} \sum_{m=0}^{N/2-1} x_{2m+1} e^{-\frac{i2\pi m(k+\frac{N}{2})}{N/2}} \quad (12)$$

$$= \sum_{m=0}^{N/2-1} x_{2m} e^{-\frac{i2\pi m}{N/2}} e^{-2\pi m i} + e^{-\frac{i2\pi k}{N}} e^{-\pi i} \sum_{m=0}^{N/2-1} x_{2m+1} e^{-\frac{i2\pi m}{N/2}} e^{-2\pi m i}$$

$$= \sum_{m=0}^{N/2-1} x_{2m} e^{-\frac{i2\pi m}{N/2}} - e^{-\frac{i2\pi k}{N}} \sum_{m=0}^{N/2-1} x_{2m+1} e^{-\frac{i2\pi m}{N/2}}$$

$$= E_k - e^{-\frac{i2\pi k}{N}} O_k$$

We could write X_k and $X_{k+\frac{N}{2}}$ such as:

$$X_k = E_k + e^{-\frac{i2\pi k}{N}} O_k$$

$$X_{k+\frac{N}{2}} = E_k - e^{-\frac{i2\pi k}{N}} O_k$$

The core idea, expressing the DFT of length *N* recursively in terms of two DFTs of size $N/2$, is the core of the $radix - 2$ DIT fast Fourier transform. The algorithm gains its speed by re-using the results of intermediate computations to compute multiple DFT outputs. As shown in figure 16.

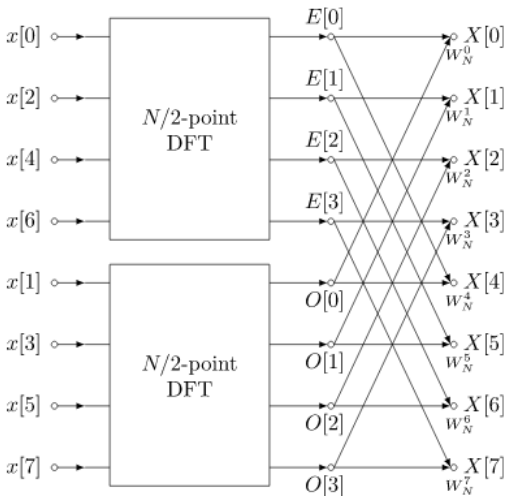


Figure 17:Data flow diagram for N=8: a decimation-in-time radix-2 FFT

To obtain the results, the algorithm combines the outcomes of these smaller DFTs using a simple operation involving $even\ (E)$ and $odd\ (O)$ indexed elements and complex exponentials. This operation is often referred to as a "butterfly" calculation.

1.2.2.2. Performance analysis

To compute the DFT of an N-point sequence using the definition,

$$X_k = \sum_{n=0}^{N-1} x_n . e^{-\frac{i2\pi kn}{N}} \qquad (11)$$

It will be required N^2complex multiplies and adds, that means works out to $4N^2$ real multiplies and $4N^2$real adds.

The basic idea of the computational step of the FFT algorithm is a butterfly. Each butterfly performs two complex number calculations: $p + \alpha q$ and $p - \alpha q$, involving one complex multiplication $(\alpha . q)$ and two complex additions. This results in 4 real multiplications and 6 real additions per butterfly.

There are $N = 2$ butterflies per stage, and log2 N stages, that means about $4\frac{N}{2}\log_2 N = 2N \log_2 N$ real multiplies and $3N \log_2 N$ real adds for an N-point FFT.

Table 1: Cost Comparison

N	$r = \log_2 N$	**BRUTE FORCE** $4N^2$	**FFT** $2N \log_2 N$	speedup
2	1	16	4	4
4	2	64	16	4
8	3	256	48	5
1,024	10	4,194,304	20,480	205
65,536	16	$1.7 \cdot 10^{10}$	$2.1 \cdot 10^{6}$	~10^{4}

The FFT algorithm is a LOT faster for big N. Figure 17

In the case of FFT algorithms for N not a power of two. The algorithms are generally fastest when N has many factors, however.

2. Quantization

Quantization, in mathematics and digital audio processing, is the operation of mapping input values from a large set (often a continuous set) to output values in a (countable) smaller set, often with a finite number of elements. Rounding and truncation are typical examples of quantization processes. Quantization is involved to some degree in nearly all digital signal processing, as the process of representing a signal in digital form ordinarily involves rounding. Quantization also forms the core of essentially all lossy compression algorithms.

At regular intervals (sampling period T), Quantization converts the continuous voltage into a series of discrete binary numbers, representing the measured values. The digital representation of the analog signal utilizes digital words, where the number of bits per word determines the discernible number of discrete voltage values in the resulting data. The greater the number of bits per word, the finer the resolution at which we can differentiate voltage levels. The minimum distinguishable change corresponds to the analog level associated with toggling the least significant bit. Given a fixed maximum input signal amplitude, the ability to resolve this minimum difference governs the overall dynamic range that can be provided.

In analog audio systems, the lower limits of signal amplitude are typically determined by the residual noise present in the system. This noise is primarily caused by thermal effects in circuit components and the inherent noise characteristics of active devices used. Thermal noise generates a consistent power density across the frequency spectrum, resulting in what is commonly known as white noise. However, this noise doesn't serve as an absolute limit since we can still perceive signals that are smaller than the noise until they are completely masked by it.in digital audio, the constraints on low-amplitude signals stem from the limitations in accurately measuring these small signals due to the finite number of bits used for signal encoding. As the input signal level decreases, the measurement error relative to the signal amplitude increases. When dealing with large signals, there are ample bits available to represent the signal's amplitude accurately. However, as the signal level diminishes, it eventually approaches the voltage equivalent of the least significant bit (LSB) in the digital conversion. The discrepancy between the signal voltage and the LSB voltage equivalent represents the error in the conversion process. When the signal falls below the LSB equivalent voltage, the output may even become zero. The error signal in digital audio is correlated with the original signal, which is unlike the constant wideband noise observed in analog electronic systems. Consequently, as the signal level decreases, it may intermittently drop below the LSB threshold and then return, causing the signal to intermittently appear and disappear. As a result, the signal-to-error ratio diminishes as the signal level decreases, making it more challenging to distinguish the desired signal from the associated errors.

To address the issue of signal-correlated errors, a technique involves introducing a small quantity of wideband noise referred to as dither into the signal. This dither, characterized by its random properties, effectively reduces the correlation between the quantization error and the signal itself. Consequently, the result is a more natural-sounding noise floor. While the notion of intentionally adding noise may appear counterintuitive, there exists an optimal amount of dither that effectively decorates the signal from the error, without introducing perceptible artifacts. This optimal dither level typically ranges from 1/3 to 1 least significant bit (LSB) equivalent, depending on the specific dither employed. The choice of either type can also influence the characteristics of the resulting dithered signals, albeit the perceptual differences tend to be minimal when listening to signals at full amplitude. It is crucial to incorporate dither whenever

the word length is reduced in digital audio computations, as well as during the initial sampling stage.

Approximately, the dynamic range of a digital system is equal to 2 to the power of the number of bits, n, in the word, or 20 log 2^n. accurately, the dynamic range of a digital system using n bits per word is:

$$SNR(dB) = 6.02n + 1.76 \qquad (13)$$

Since, the energy in the error generated by the converter is a statistical probability function, the error should be calculated by integrating the product of the error and its corresponding probability. We could use the $20\,log2^n$ approximation to see forcefully how word length affects the dynamic range of digital audio systems in Table1.

In theory, increasing the number of bits used to encode sample words leads to higher confidence in measurement accuracy and improved sound quality. However, in practice, the physical limitations of the conversion process prevent real converters from achieving the full accuracy predicted by theory, as they cannot attain theoretical perfection. Analog-to-digital converters (ADCs) that claim 24-bit quantization typically exhibit linear conversion within a range of only 20 to 22 bits. Nevertheless, employing more bits can still enhance fidelity- as long as the electronic circuitry meets the necessary accuracy requirements.

Table 2: bits and SNR

n Bits	Dynamic Range (dB)
2	12
4	24
6	36
8	48
10	60
12	72
14	84
16	96
18	108
20	120
24	144

3. Timing

The regularity of sampling is a critical factor in accurately measuring analog signals. It is essential to assume that each sample is taken at a precisely consistent interval. Even small deviations in the timing of the sampling process can lead to errors in measurement. This is because analog signals undergo changes over time, and samples taken at incorrect times will yield different values compared to samples taken at the correct time. To ensure accurate sampling, it is necessary to have a stable timing reference and circuitry that can perform the conversion within the allocated time between samples. Variations in the timing of samples are referred to as **jitter**, which is a potential source of error in the conversion process.

The sampling theorem states that a continuous function, such as an analog audio signal, can be accurately represented by a discrete time sequence if the sample rate is at least twice the

highest frequency in the original signal. However, this theory assumes an infinite discrete time sequence, which is not achievable in real-world converters. Nevertheless, if the signal spectrum is limited to less than half the sample rate, we can reconstruct the original signal satisfactorily.

In the realm of motion pictures, when sampling occurs at an insufficient rate is obvious: the spokes of a rotating wheel, for instance, appear to move in a backward direction if the wheel completes more than one full rotation within the span of $1/24th$ of a second, which is the duration between frames. Instead of the backward rotation, frequencies above half the sampling rate become "folded over," resulting in audible frequencies that fall below half the sampling frequency. The dots in Figure 18 mark the sample times, we observe that the high-frequency sine wave completes multiple cycles between samples. Consequently, the resulting sine wave manifests as a lower frequency than both the original signal and the sample rate. Aliasing frequencies arise as sums and differences involving the sample rate and audio frequencies present above half the sample rate.

The only the difference alias falls within the audible range of frequencies.

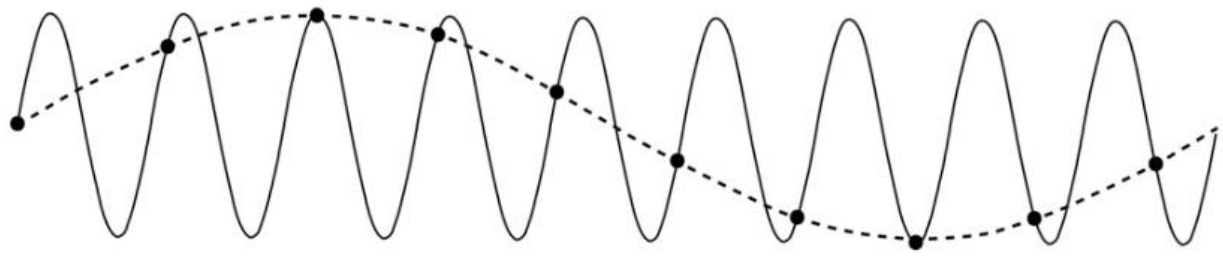


Figure 18: A simple example of aliasing. The alias frequency created is the difference between the signal frequency and the sample rate.

To effectively prevent aliasing, analog low-pass filters are essential at the input of the analog-to-digital converter (A/D). These filters are responsible for removing frequencies that surpass half the sample rate. For 44.1 kHz sample rates, this means that frequencies above 22.05 kHz must be significantly attenuated. The implementation of analog filters with sharp corner frequencies poses considerable challenges, resulting in complex converters. Each filter stage provides an attenuation of 6 dB per octave, necessitating multiple stages to effectively suppress high-frequency content above half the sample rate while preserving frequencies below the

critical threshold. However, the application of these intricate analog filters can introduce audible degradation to the audio signal due to the introduction of phase and amplitude irregularities at frequencies well below the filter cutoff.

After acquiring the sampled data, it can undergo digital filtering, replacing the need for analog anti-alias filters through a process called oversampling. By setting the sample rate at a multiple of the desired rate, excess data can be digitally filtered and decimated to achieve a lower sample rate. Decimation involves averaging the samples to obtain a reduced effective sample rate. This approach permits more relaxed low pass filtering on the input, as raising the input sample rate Fs enables less steep analog filters to still effectively limit signal frequencies below half the sample rate. Further digital filtering is applied during the decimation process, allowing for an increased sample rate without generating additional data.

Theoretically, the sample rate only needs to be higher than twice the highest frequency in the signal. However, determining the necessary sample rate in practice is a matter of discussion. The widely used 44.1 kHz sample rate for CDs and other digital devices was chosen due to storage limitations in the past. Nowadays, we have the capability to handle higher data rates and utilize higher sample rates and longer word lengths. Nevertheless, the optimal sample rate remains uncertain. It involves striking a balance between high sample rates and the amount of data that needs to be stored, which is a trade-off left to the discretion of digital recording system users.

4. Sound Visualization

Acoustic soundscapes contain a wealth of information about the environment. By carefully tuning into the soundscape, one can often infer their location, the time of day, and any activities in the vicinity. For example, one might be able to discern the presence of flowing water, a forest, or a grassland, as well as whether it is day or night. Even when not paying direct attention to the sounds, the brain can still process and interpret the soundscape, providing meaningful insight into the environment. However, due to the sheer volume of recordings collected by soundscape recording systems, it is not feasible to manually review them all, acoustic scientists use visual representations to allow us to visualize the temporal and/or spectral characteristics of sound [18].

4.1. Oscillograms

An oscillogram as shown in figure 19 is a visual representation of a sound wave over time, generated by measuring air pressure fluctuations and plotting them as a function of time, read an oscillogram from left to right. On an oscillogram, the x-axis corresponds to time, typically measured in seconds or milliseconds, and the y-axis corresponds to the amplitude or intensity of the sound wave, usually measured in decibels (dB). This graphical depiction captures changes in the pressure of the sound wave over time, with each peak and trough representing a distinct point in the waveform.

Oscillograms are a fundamental tool in the study of sound waves, allowing for the analysis of frequency, amplitude, duration, and other characteristics. They are used to investigate a variety of acoustic phenomena, including speech, music, animal vocalizations, and environmental sounds. By examining the patterns and properties of sound waves in oscillograms, researchers can gain a better understanding of how acoustic signals behave in different contexts and leverage this knowledge for applications such as noise reduction and speech recognition.

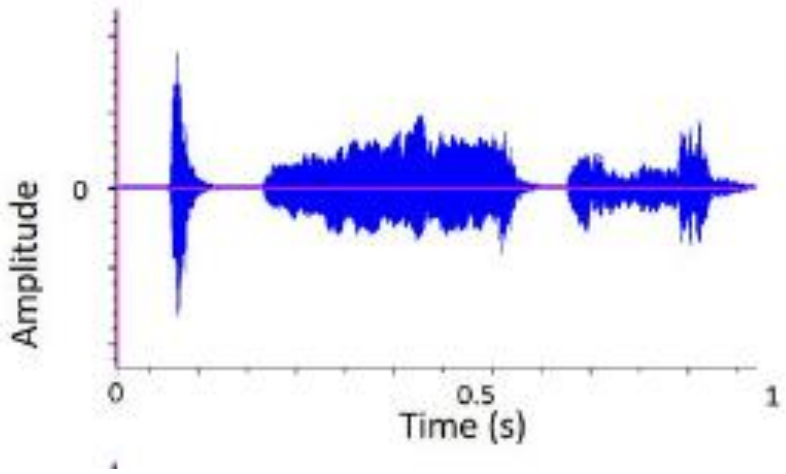


Figure 19:Oscillogram

4.2. Power spectrum

Power spectrum figure 20 is a representation 2D showing how a sound metric such as sound pressure level y-axis varies with frequency on the x-axis. By looking at the power spectrum of a recording signal, the pressure level is usually measured in (dB) and the frequency is measured in

vibrations per second (or hertz, abbreviation Hz) or thousands of vibrations per second (kilohertz, abbreviation kHz).

The power spectrum of a sound signal can be used to infer the amount of sound energy or power present at different frequencies. This can be visualized by plotting the average power in each frequency band against the mid-value of the frequency band. To analyze a wide frequency range, it is often divided into octaves, which are equally sized bands where the frequency doubles with each octave. The Fourier transformation is an effective method for producing power spectrums, providing a set of equal frequency bands that allow the researcher to assess the energy (power) contained within each of them. Power spectrums are a useful tool for summarizing the frequency composition over a given time, rather than examining how it changes over time.

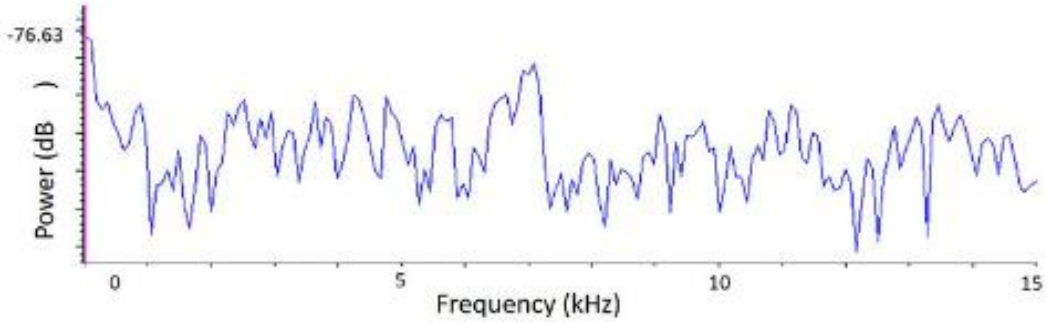


Figure 20:Power spectrum

4.3. Spectrogram

A spectrogram figure 21 is a visual way of representing the signal of sound recording in 3D: frequency on the y-axis, time on the x-axis, and amplitude represented by color intensity, commonly used to display frequencies of sound waves produced by humans, machinery, animals, whales, jets, etc., as recorded by microphones. Warmer or darker colors typically reflect higher amplitudes, and dark blues correspond to low amplitudes.

Spectrograms are used to visualize the temporal pattern of sound, with frequency on the y-axis and time on the x-axis. Frequency is divided into frequency bands, also referred to as bins, each typically having a bandwidth of 20 Hz, with the maximum frequency represented being 7,000 Hz. This results in 350 frequency bands being shown on the spectrogram. While

spectrograms are visually appealing, they only display relative variations in amplitude and do not provide sound pressure levels, making it difficult to compare recordings.

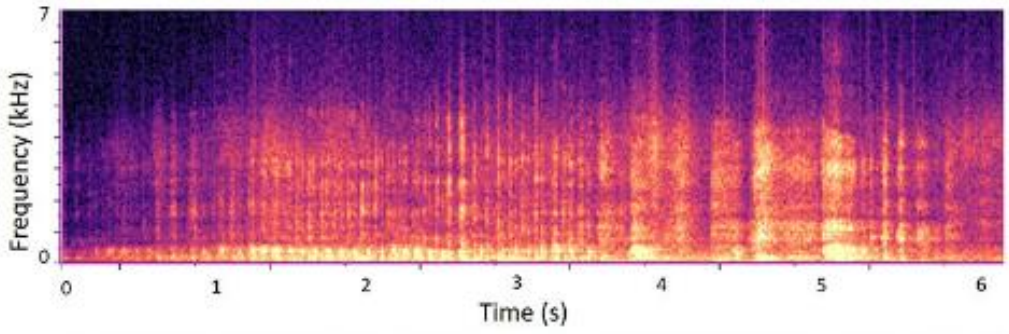


Figure 21:Spectrogram

Digital sound recordings can be used to produce oscillograms, power spectra, and spectrograms for visual inspection. Oscillograms are generated directly from the digital sound recordings, whereas post-recording processing is required to generate power spectra and spectrograms. Furthermore, these acoustic representations can be used to extract acoustic measurements which can be used to calculate acoustic indices.

5. ADC

In a conventional Analog-to-Digital Converter (ADC), the audio waveform undergoes periodic evaluation, synchronized with the sample clock rate, typically at 44.1 kHz, corresponding to intervals of 22 microseconds. This involves utilizing an analog memory component known as a sample-and-hold circuit. The primary function of this circuit is to capture and retain the instantaneous voltage value of the audio signal, ensuring fidelity during the conversion process. It is imperative that the sample-and-hold process occurs swiftly to prevent distortion due to delays or value fluctuations during conversion.

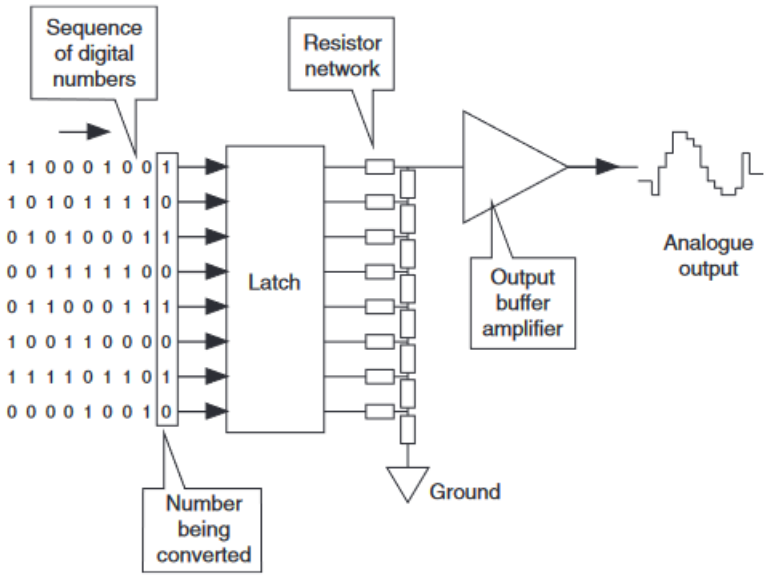


Figure 22: Analog-to-Digital Converter (ADC).

Upon capturing the audio value, a comparison is drawn between the stored value and a signal generated through the cooperation of a counter and a Digital-to-Analog Converter (DAC). This operation shares similarities with the waveform synthesis seen in wavetable synthesizers generating ascending sawtooth patterns. The counter incrementally tallies from zero, with its numerical sequence being translated into an ascending voltage output by the DAC. Simultaneously, a comparator assesses the sampled value from the sample-and-hold circuit against the DAC-generated output. The point of convergence signifies equality, prompting the comparator to signal the alignment. Consequently, the counter's output transfers to the ADC's output and is held in a latch. This yielded output from the ADC constitutes a numerical representation of the sampled value. Following this step, the counter resets, permitting the initiation of ADC processing for the subsequent sample, thereby sustaining a continuous cycle (Figure 22).

While operational nuances among various ADCs may exist, the fundamental principle remains consistent. The procedure encompasses audio signal sampling, conversion of sampled values into numerical equivalents, and the presentation of these numerical values at the ADC's output port. This iterative process harmonizes with the sample clock rate. Diverse ADCs may adopt distinct methodologies for conversion, and the resulting numerical output can manifest in

either serial, involving a single bit stream, or parallel, characterized by multiple streams carrying complete samples.

6. Experimental setup Digitalization of sound signal

6.1. Materials

6.1.1. Pasco model Ps-3227 device and its capabilities in capturing sound signal.

The wireless sound sensor contains two sensors in one package: a sound wave sensor and a sound level sensor. The active sensor can be utilized using PASCO software. Note that it is only possible to take measurements with one sensor at a time. The sound wave sensor is selected by default.

6.1.2. Sound wave sensor.

The sound wave sensor measures relative changes in sound pressure as a function of time. Since the sensor does not measure absolute changes, the measurement is unitless. Use a scope display to analyze the frequency, wavelength, and relative amplitude of sound waves. Use the FFT display to analyze the harmonics of sound waves.

6.1.3. Sound Level Sensor.

The sound level sensor can measure sound level in (dB) using two different scales the dBA or dBC. The dB scale is designed to replicate human perception of sound by filtering out low frequencies. It is the scale used by OSHA for regulatory measurements. On the other hand, The dBC scale measures sound levels in a wide range of frequencies within and outside the frequency range of human hearing. The dBA scale is recommended for most classroom applications. you can utilize various visual representations such as graphs, digital displays, meters, or tables. These displays will provide clear and accessible information about the sound levels being measured Figure 23.

6.1.4. Components

Figure 23: PS-3227 sensor device

1 Sensor ID

Use this ID when connecting the sensor to the software.

2 Bluetooth Status LED

- Red Blink is ready to be paired with software.
- Green blink Paired with software.
- Yellow blink Remotely logging data.

3 Micro USB port

- For charging the battery when connected to a USB charger.
- For transmitting data when connected to the USB port of the computer.

4 Power button

Press and hold for one second to turn on or off.

5 Battery status LED

- The red blink Battery needs to be recharged soon.
- The green solid Battery is fully charged.
- Yellow solid Battery is charging.

6 Threaded hole

For attaching the sensor mounting rod. Also accepts any ¼-20 screw.

7 Microphone

Location of the sound sensing element.

6.1.5. Specifications

Table 3: Specifications details

Microphone Frequency Range	100 – 15,000 Hz
Max Sampling Rate (Sound Wave)	100 kHz
Max Sampling Rate (Sound Level)	20 Hz
Sound Level Range	50 – 110 dB
Sound Level Accuracy	± 2 dB

The Pasco Model PS-3227 sensor was chosen as the microphone for the experiment due to its suitability for the desired sound measurements and specific characteristics that make it well-suited for the experimental setup. With its wide frequency range, the sensor is capable of capturing sound signals across a broad spectrum, including frequencies within and outside the range of human hearing. Its high accuracy and sensitivity ensure precise measurements, even capturing subtle changes in sound levels. The sensor's versatility allows for its use in different environments, both indoor and outdoor, making it adaptable to various experimental setups. Additionally, the Pasco brand instills confidence in the sensor's performance, providing accurate and consistent results for analysis and further study.

6.2. Method

6.2.1. **Representative Location:** The study was carried out around **QUINTA DO CARMO**. this area covers 2333 m^2. Ten points on the area were selected as indicated in Figure 24, to ensure adequate coverage and representation of the entire location.

Figure 24: Map pinpoint the location of 10 noise measurement.

The locations of measurements most of it is forest included the internal streets with main roads, such as the main gate, and some of the branch roads. Type of buildings included: (1) Farming stores (2) services buildings (such as library and restaurants).

6.2.2. **Distance and Height:** For each part of the area, a microphone was positioned within the designated region. The placement of the microphone is important to capture the sound accurately.

The sound level meter was placed at a height of 1.5 meters above the ground level and at a minimum distance of 3 meters from each point.

6.2.3. **Microphone Orientation**: The microphone was oriented appropriately based on its directional characteristics and the nature of the sound source.

The sound level meter Omnidirectional microphone.

6.2.4. **Calibration:** the microphone was calibrated using a sound level calibrator to verify its sensitivity and ensure accurate measurements. This calibration process accounted for any inherent microphone characteristics or variations.

The device was calibrated on 41.1 dB (A) before taking noise measurements and was set to record noise samples at 4-sec intervals during the 5-minutes exposure time.

6.2.5. Measurement time

Measurement time in each one of the 10 measurement points inside the area was of 3 min. To verify whether this time was sufficient for the proposed evaluation of the noise inside the area. and the measurement period for all point were between 10:30 am and 4:00 pm

The data was collected and analyzed by using PASCO Capstone and ArcMap 10.3.1 software to create a noise map [19].

6.3. PASCO Capstone

Capstone is a powerful software solution designed to handle large data sets, high-speed sampling, and customized preferences to fit the needs of your lab. Its user-friendly interface is accessible for beginners. In figure 25, the entry screen of the software is displayed.

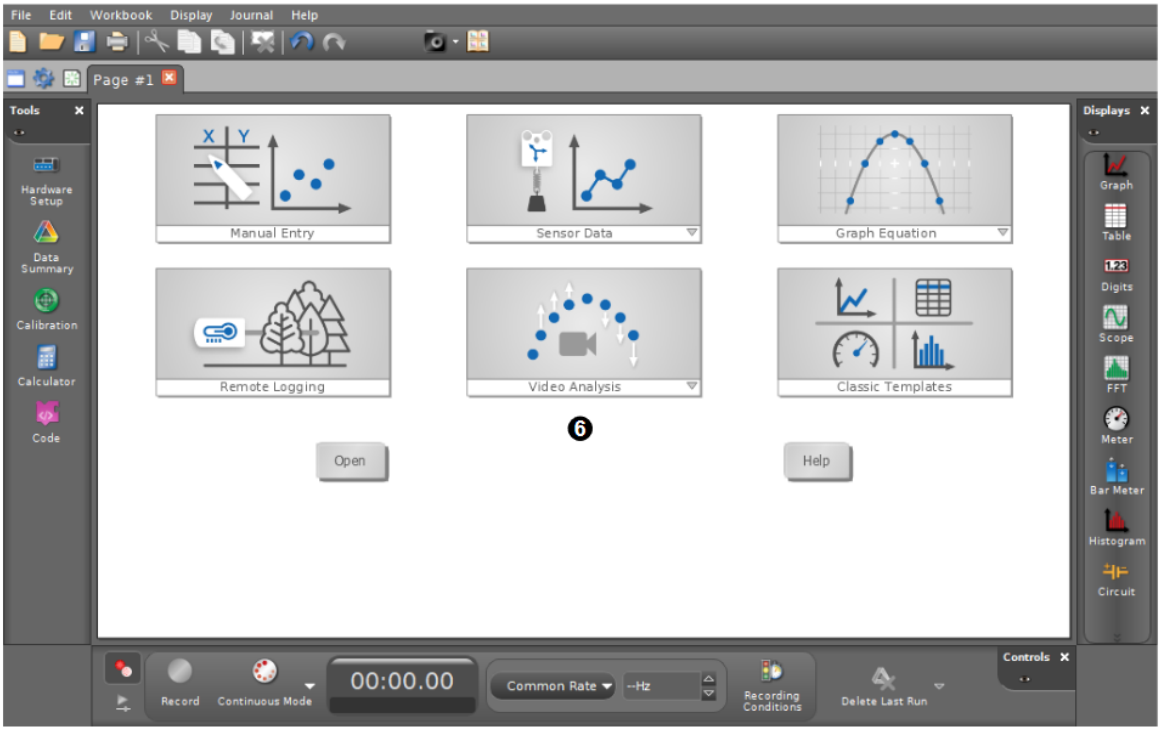


Figure 25: The Entry Screen

Capstone has all the software tools you need for data collection and analysis.

6.3.1. Data Display Features

Display features are designed to help users visualize and understand their data in a variety of ways. Here is a brief overview of each feature:

Graph: plot data on a Cartesian plane. Users can choose from a variety of graph types, such as line graphs, bar graphs, and scatter plots. They can also customize the graph's appearance by changing the axis labels, axis scales, and data point symbols.

Table: view data in a tabular format. This can be useful for viewing large amounts of data or for comparing different data sets. Users can customize the table's appearance by choosing which columns to display and by changing the column order.

Digits: The digits display shows the most recently recorded data sample as a numerical value. This can be useful for viewing data that is changing rapidly, such as the voltage of a circuit or the temperature of a room.

Meters: The meters display allows us to create custom meters to display their data. This can be useful for creating dashboards or for monitoring specific data points. Users can customize the meter's appearance by changing the meter's range, scale, and color.

Oscilloscope: view data as a waveform. This can be useful for analyzing electrical signals, sound waves, and other types of oscillating data.

FFTs: The FFT (fast Fourier transform) views the frequency spectrum of data. This can be useful for identifying the different frequency components of a signal.

Histogram: allows us to see the distribution of data. This can be useful for identifying outliers and for understanding the overall shape of the data set.

6.3.2. Data Function Features

Data function features are capabilities that can be applied to data to improve its quality, usability, and value. Here are some examples of data function features:

Data smoothing: reduces noise and fluctuations in data, making it easier to identify trends and patterns.

Data modeling: create models of your data, which can be used to simulate different scenarios and make predictions.

Recording conditions: record the conditions under which your data was collected, which can be useful for troubleshooting and analysis.

Replay data: replay your data, which can be useful for testing and debugging your models.

Blockly programming: Blockly within SPARKvue and Capstone is compatible with all PASCO sensors and interfaces as we can see in figure 26. When users combine PASCO sensors with Blockly, they are empowered to design and execute their very own sensor experiments. Users can create code that collects sensor measurements, reports data, or controls output devices such as the Smart Fan Accessory. As they execute their code, Users can visualize their data using real-time graphical displays that assist with data visualization [14].

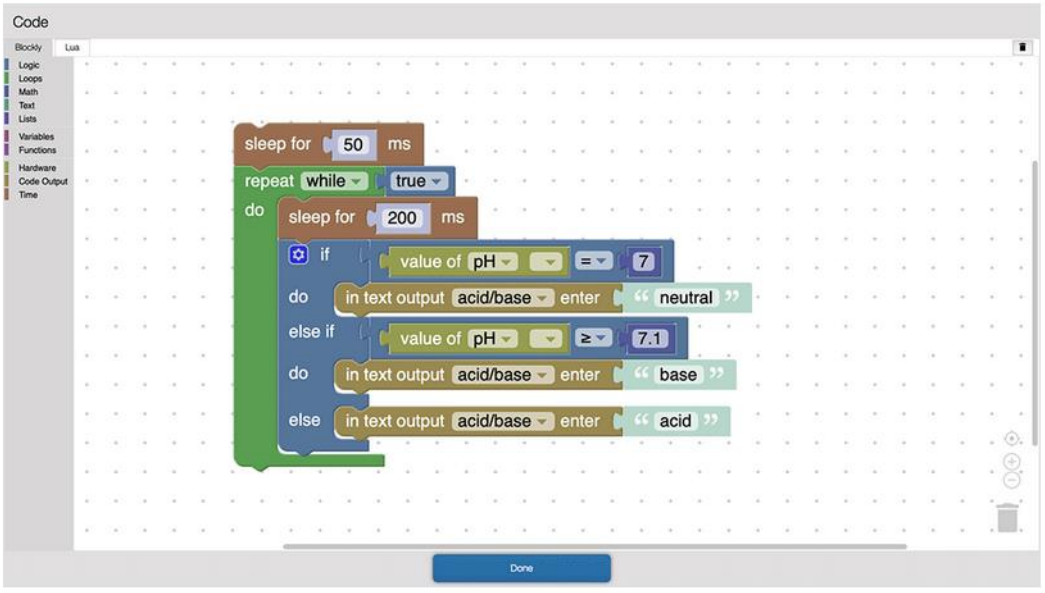


Figure 26: Instruct the sensor to identify a sample solution.

6.4.Noise mapping

6.4.1. Definition

Noise mapping is a graphic representation of sound level distribution in a particular area. This representation is typically displayed on a map of that region, and different colors are used to indicate varying noise levels.

6.4.2. Data Collection

The noise levels in the area were measured and recorded, likely using specialized equipment.

6.4.3. Software and Interpolation technique

ArcMap 10.3.1 software was used for the analysis of the noise data. ArcMap is a geographic information system (GIS) software used for spatial analysis and mapping. The values were mapped using coordinates (X and Y) along with different parameters of sound level measurements. These parameters include L_{eq}, L_{max}, and L_{min}, each representing different aspects of noise levels (e.g., maximum, minimum, equivalent) [20].

Contouring using Inverse Distance Weighted (IDW) spatial interpolation one of the most common techniques was used to capture the adjoining areas and cover the entirety of the study area. Generally, interpolation predicts cell values in a raster format using a given set of sample data. It is a valuable tool for predicting of unknown values for a given geographic point data, which in this study is noise.

The Google Maps were used to provide the real photo of the location by taken 4 points of each location at the boundary on the study area.

7. Noise

7.1 Types of noise

Noise basically can be defined as unwanted sound and may interfere with human hearing referred to in the bioacoustics literature. The noise never stops and has no tone, if the fan experiences an unexpected change such as abruptly stopping or producing a high-pitched whine, it can cause disruption or annoyance. Our hearing detects information in the source in that we hear.

Through measuring must know the type of noise so as to easily choose the parameters to operate, the material to use, and the periodicity of measurement, usually we must use our ears to

pinpoint the annoying features of the noise, making before measurement, after realizing the analysis and documenting them.

7.1.1. Continuous Noise

The noise produced by machinery operating in a continuous mode, such as blowers, pumps, and processing equipment as we shown in figure 27, can be measured using hand-held equipment (Sound Level Meter) for just a few minutes. This allows for the determination of the noise level. If tones or low frequencies are present, the frequency spectrum can be documented and analyzed further.

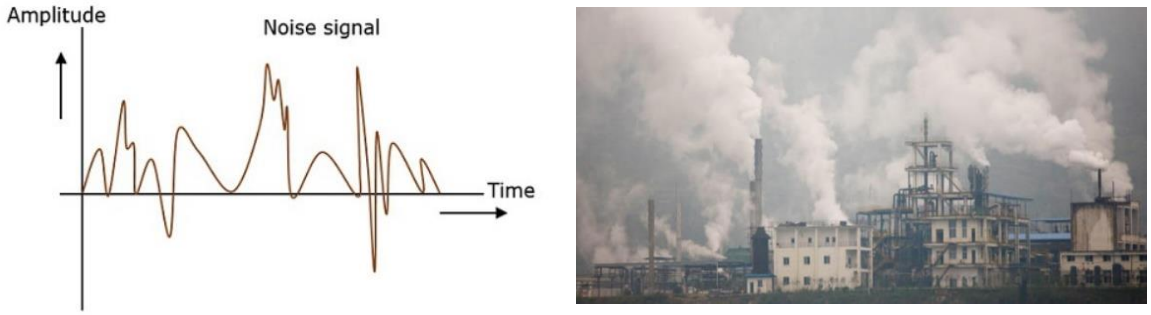


Figure 27: example of a heavy industry

7.1.2. Intermittent Noise

That noise is associated with sources that operate in cycles or single vehicles or aircraft that pass by and generate impulsive noise, characterized by rapid changes in sound pressure level. For each cycle of a machinery noise source, the noise can be measured in a similar manner to continuous noise sources. The cycle duration must be considered when measuring the noise level. When measuring the noise generated from a single passing vehicle or aircraft, the Sound Exposure Level (SEL) should be measured as it combines both the level and duration into one descriptor. In instances where a few similar events are observed, it is possible to measure the maximum sound pressure level and then calculate an average as we could see in our figure 28.

Figure 28: Intermittent examples

7.1.3. Impulsive Noise

Impulsive noise, generated from impacts or explosions, such as a pile driver, punch press, or gunshot, is distinguished from other noise sources by its brief and abrupt nature, and its ability to cause greater annoyance than its sound pressure level would suggest. The impulsiveness of noise can be quantified by the difference between a quickly responding and a slowly responding parameter (as illustrated in the figure 29). Additionally, the repetition rate (number of impulses per second, minute, hour, or day) should be documented.

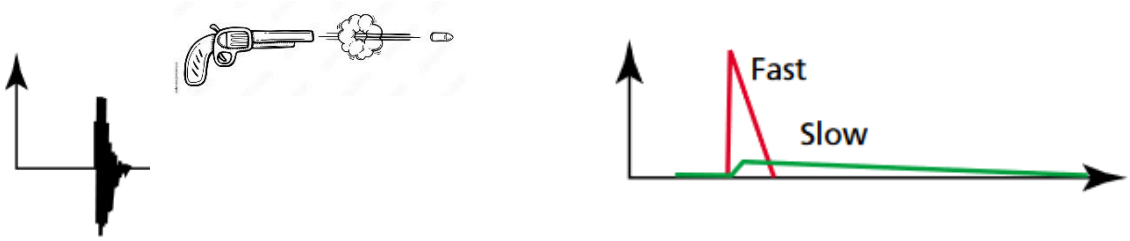


Figure 29: gunshot example

7.1.4. Tones in Noise

Machines with components that rotate, such as motors, gearboxes, fans, and pumps, are a source of annoying tones due to their unbalanced or repetitive impacts. Vibrations thus generated are transmitted through surfaces and propagate into the air, resulting in tones audible to humans. Pulsating flows of liquids and gases can also produce tones because of combustion processes or flow restrictions as we can see in figure 30. Tone identification can be done either subjectively by listening or objectively by analyzing frequency. The audibility of the tone can then be determined by comparing its level to the level of surrounding spectral components. The duration of the tone should also be noted for further analysis.

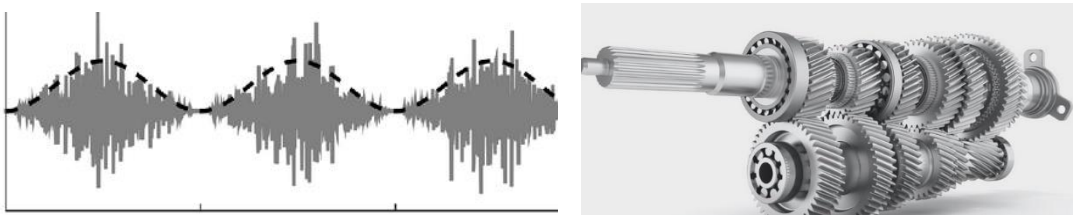


Figure 30: gearbox example.

7.1.5. Low-Frequency Noise

Low-frequency noise, which is typically emitted by large diesel engines found in trains, ships, and power plants, furthermore Low-frequency noise includes natural sources such as wind, wind effects on structures, and thunder. it has an acoustic energy present within the frequency range of 8 to 100 Hz. Due to the difficulty of muffling this noise and its tendency to spread in all directions, it can be heard at great distances. To evaluate the audibility of low-frequency components in the noise, the noise spectrum is measured and compared to the threshold of hearing. Infrasound, on the other hand, is a type of sound that has a spectrum with significant components below 20 Hz. Unlike audible sound, infrasound is not perceived as sound, but rather as a pressure sensation. Currently, the assessment of infrasound is still in the experimental stage, and there are no established international standards for its evaluation as shown in the figure 31.

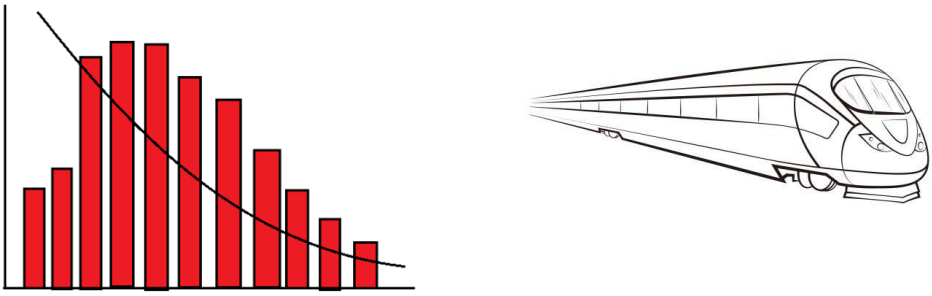


Figure 31: train example

7.2. Colors of noise

Noise signal is a sound that is intense and therefore it has the same features as sound. Any kind of filtered noise signal is called ’colored noise’ which implies to impure white noise. Noise colors are characterized by color type because of the frequency of that noise. For example, the sound wave of pink noise has the same frequency as light waves that create pink color [21].

7.2.1. White noise

White noise is a signal or process that shares its name with white light due to its resemblance. It exhibits a frequency spectrum that appears flat when graphed linearly with respect to frequency (e.g., in Hz). Essentially, the signal possesses equal power within any given bandwidth (power spectral density) when that bandwidth is measured in Hz. For instance, in the case of white noise in audio, the amount of sound power in the frequency range of 40 Hz to 60 Hz is equivalent to the power within the range of 400 Hz to 420 Hz, as both intervals are 20 Hz wide. It is worth noting that spectra are often displayed with a logarithmic frequency axis, rather than a linear one. Consequently, when a logarithmic scale is used, the equal physical widths on the displayed plot do not represent the same bandwidth. Higher frequencies cover more Hz with the same physical width compared to lower frequencies. Therefore, a white noise spectrum, sampled equally in the logarithm of frequency (X axis), will exhibit an upward slope at higher frequencies instead of remaining flat as shown in the figure 32. Nevertheless, it is common practice to calculate spectra using linearly-spaced frequency samples but present them on a logarithmic frequency axis. This may lead to confusion and misunderstandings if the distinction between equally spaced linear frequency samples and equally spaced logarithmic frequency samples is not kept in mind.

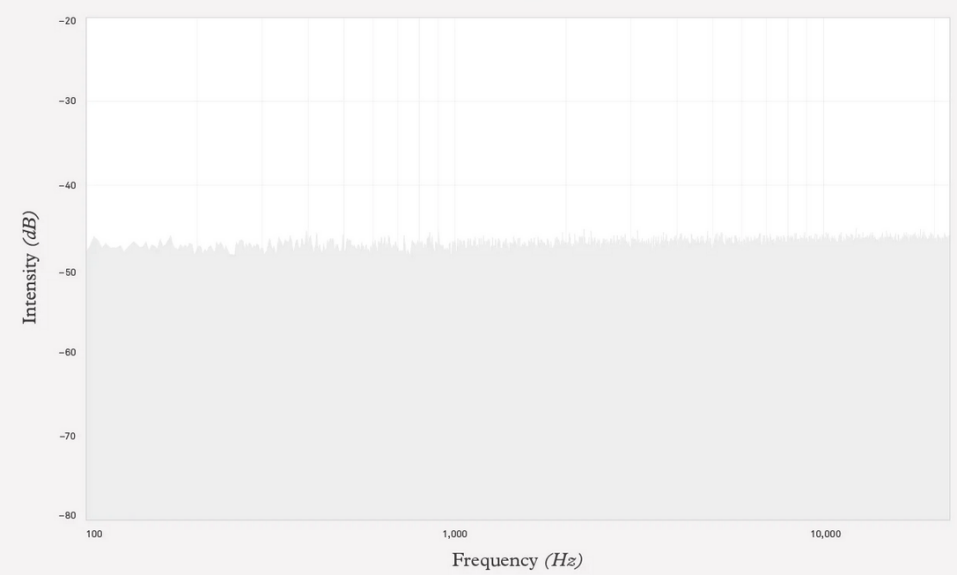


Figure 32:white noise

7.2.2. Pink noise

Pink noise exhibits a linear frequency spectrum on a logarithmic scale, with equal power in proportionally wide frequency bands. Consequently, it possesses equal power in the range from 40 to 60 Hz as it does in the band from 4000 to 6000 Hz. This characteristic aligns with human perception, where the doubling of frequency, known as an octave, is perceived uniformly irrespective of the actual frequency. Therefore, each octave contains an equal amount of energy. As a result, pink noise serves as a reference signal in audio engineering. Its spectral power density decreases by 3.01 dB per octave, following a 1/f proportionality, compared to white noise. Pink noise is often referred to as "1/f noise."

It is worth noting that due to the infinite number of logarithmic bands at both the low-frequency (DC) and high-frequency ends of the spectrum, any finite energy spectrum will contain less energy than pink noise at these ends. Pink noise is the sole power-law spectral density that possesses this property. Steeper power-law spectra, if integrated to the high-frequency end, become finite, while flatter power-law spectra become finite if integrated to the DC, low-frequency limit as illustrated in the figure 33.

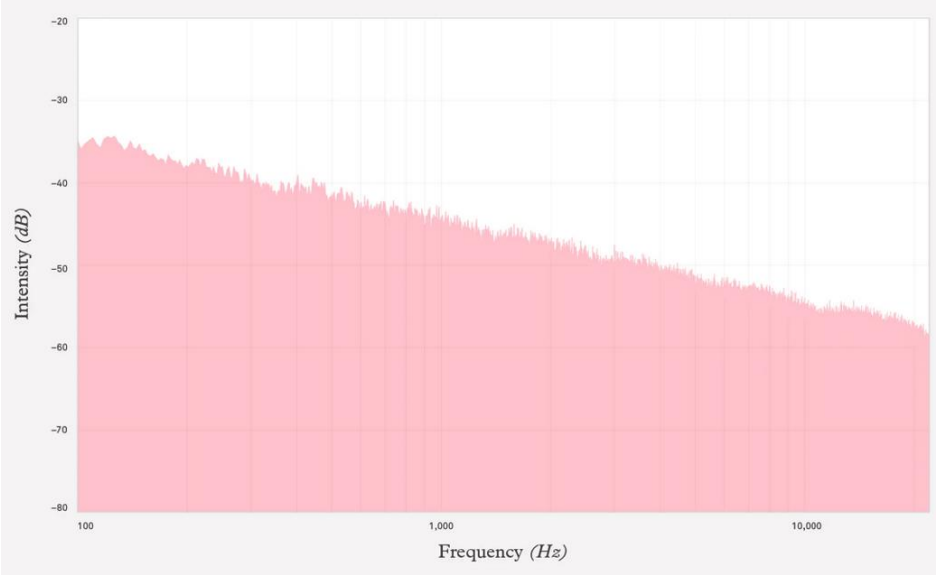


Figure 33: pink noise spectrum

7.2.3. Brownian noise

Brownian noise, also known as Brown noise, is a type of noise characterized by a power density that decreases by 6.02 dB per octave as the frequency increases. This decrease in power density follows a proportional relationship to $1/f^2$, excluding the zero frequency (DC).

To generate Brownian noise, one can integrate white noise over time. The term "Brown" in its name does not refer to a power spectrum resembling brown; rather, it originates from Brownian motion, which is also known as a "random walk" or "drunkard's walk." In terms of power spectrum shape, "red noise" describes the characteristic, with pink noise falling between red and white, as seen in the figure 34.

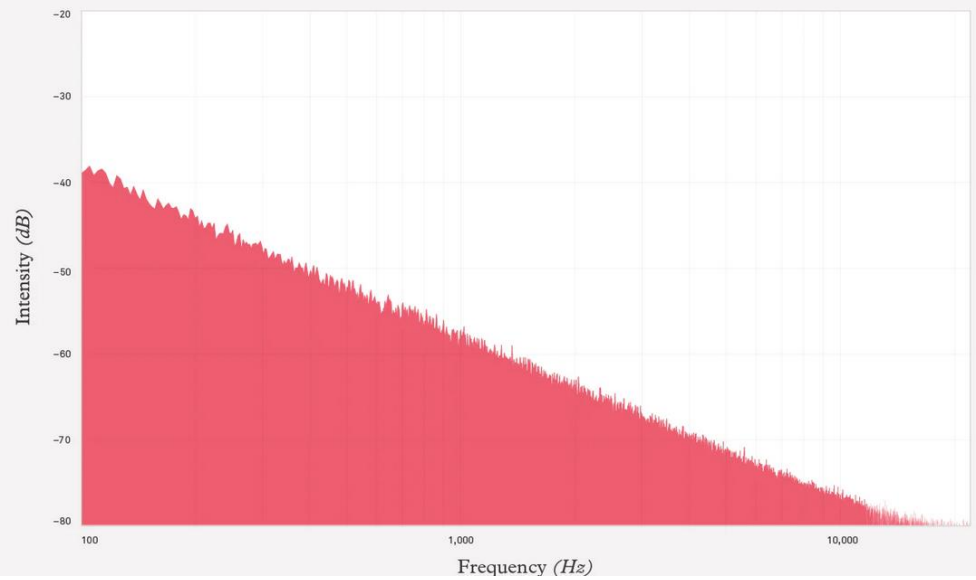


Figure 34: brown noise spectrum

7.2.4. Blue noise

Blue noise is also known as azure noise. Blue noise's power density increases by $10\log_{10} 2 = 3.02\ dB$ per octave with growing frequency (density proportional to f) over a finite frequency range. In computer graphics, it is used to describe noise that lacks low-frequency components and concentrated spikes in energy, making it suitable for dithering. Blue noise is like the arrangement of retinal cells, contributing to enhanced visual resolution.

Cherenkov radiation, an example of natural blue noise, demonstrates almost perfect characteristics with its power density increasing linearly with frequency in regions where the refractive index of the medium remains constant. This radiation appears as a vibrant blue color due to these properties, and its frequency range is limited by the refractive index of materials, as indicated in the figure 35.

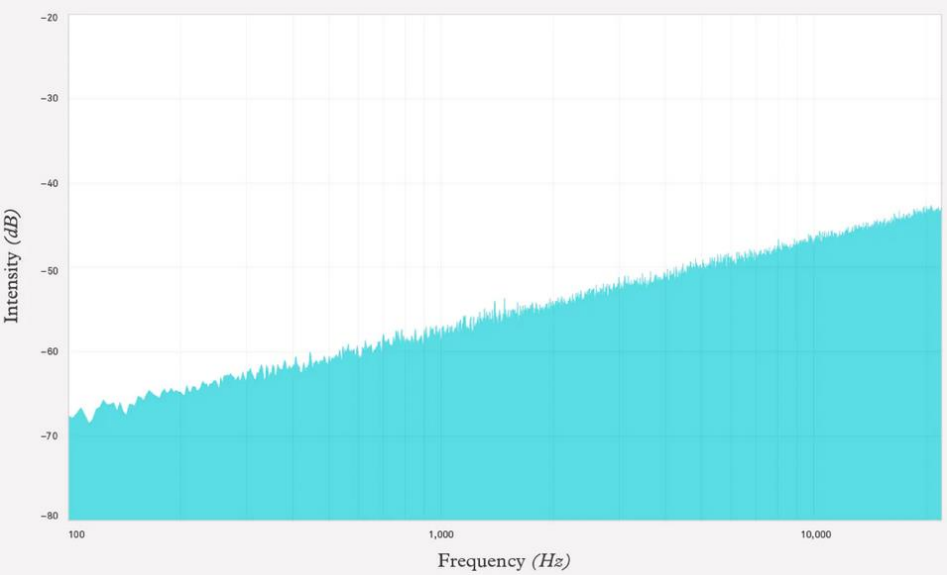


Figure 35: blue noise spectrum

7.2.5. Violet noise

Violet noise is known as purple noise; Violet noise power density increases by 6.02 dB per octave proportional to frequency. The spectral analysis makes that GPS acceleration errors seem to be violet noise processes. They are dominated by high-frequency noise." (Density proportional to f^2) over a finite frequency range.

Violet noise is sometimes referred to as differentiated white noise because it is derived from the differentiation of a white noise signal. The human ear is less sensitive to high-frequency hiss, and it is relatively easy to electronically differentiate white noise by applying a first-order high-pass filter. This property led to the early use of violet noise as a dither signal in digital audio systems.

In hydrophone measurements, the acoustic thermal noise of water exhibits a violet spectrum, meaning it becomes increasingly prominent at higher frequencies as shown in the figure 36. Classical statistical mechanics predicts a positive slope of 6.02 dB $Octave^{-1}$ for the increase in thermal noise with frequency. Alternatively, it can be expressed as a rate of 20 dB $\text{decade}^{-}1$.

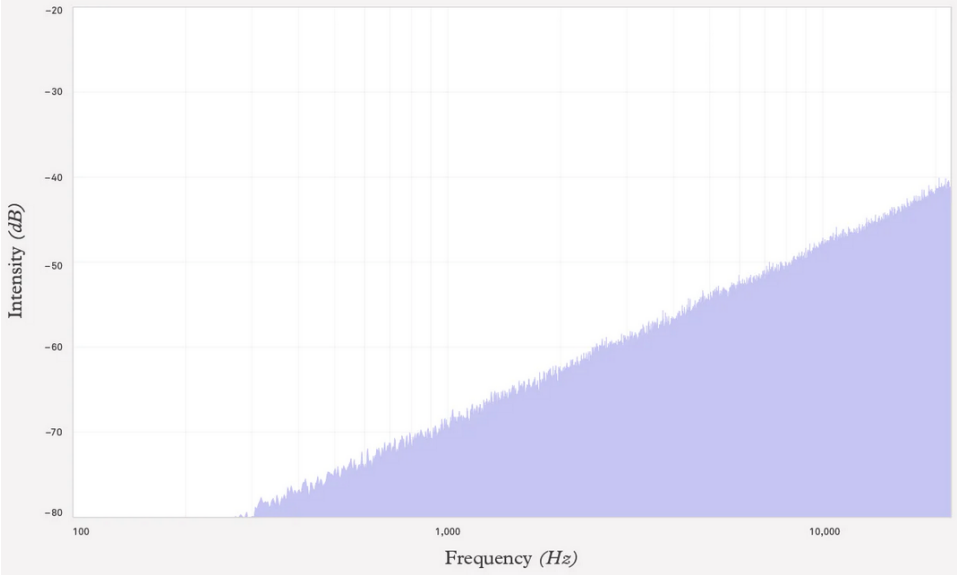


Figure 36: violet noise spectrum

There are also many colors used without precise definitions (or as synonyms for formally defined colors), sometimes with multiple definitions.

Chapter 3: Colored Noise Signal Identification Using Supervised Learning Algorithm

1. Supervised Learning

Supervised Learning is one of the machine learning types[22], it is defined by the utilization of labeled datasets to train algorithms in order to predict outcomes accurately or classify data such as between noise signals. Many pattern recognition systems can be partitioned into discrete stages, including sensing, segmentation, feature extraction, classification, and post-processing. Through the utilization of input-output samples, commonly referred to as a training dataset, we could make more accurate predictions around the original data set, and the resulting classifier is then applied to label unknown instances effectively. The aim of the learning process is to reveal the nature of the target concept. Firstly, prepare sample data with an input data matrix column and row representing variables and observations respectively [23]. To analyze and train the supervised learning model, different classification algorithms such as Decision tree, Discriminant analysis (Linear and Quadratic), and Nearest Neighbor were applied [24].

Using a machine learning approach that involves extracting features from sound signals, it is possible to make predict about on the color of the noise [25].

2. Supervised Learning method

2.1. Feature Extraction

During the pre-processing phase, an algorithm efficiently extracts key features from both the training dataset and new data in the time and frequency domains [26]. From the time domain, we've computed two critical statistics, namely the mean and variance of the noise signal, using the equations represented as (1) and (2):

$$\mu = \frac{1}{N}\sum_{n=1}^{N} A_n \qquad (1)$$

$$V = \frac{1}{N-1}\sum_{n=1}^{N} |A_n - \mu|^2 \qquad (2)$$

Where:

- A the vector of the noise signal.
- N the scalar observation of noise signal (samples).
- μ and V values for the input noise signal.

We have also the maximum and minimum values as the key features of the noise signal in the time domain. Figure 37 shown the oscillogram of pink noise.

In the frequency domain, the power spectral density (PSD) of a noise signal is estimated by taking the Fourier transform of its biased autocorrelation sequence. For an input signal sampled at 1000 Hz, the PSD estimate is defined as:

$$\hat{P}(f) = \frac{T_s}{N}\Sigma_{n=1}^{N} |A_n e^{-j2\pi fn}|^2 \qquad -\frac{1}{2T_s} < f < \frac{1}{2T_s} \qquad (3)$$

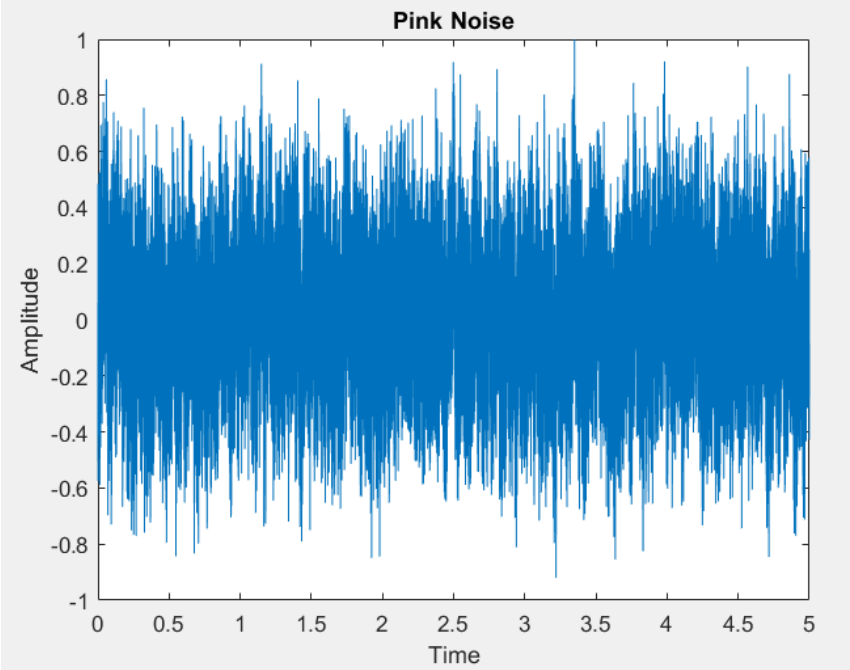


Figure 37: A sample pink noise in time domain

.

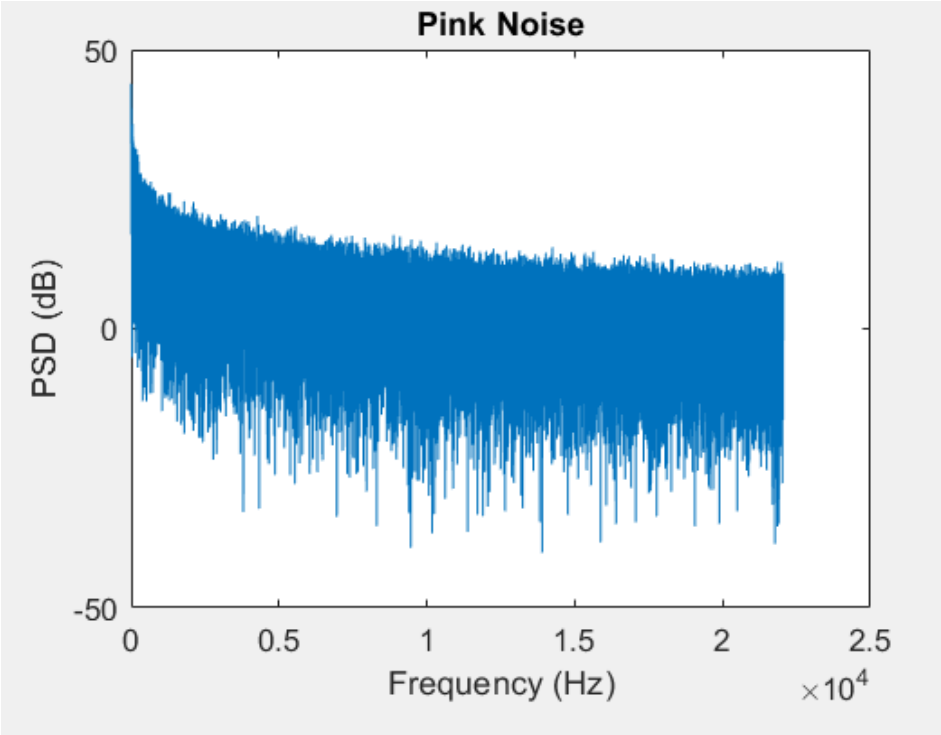


Figure 38: A sample pink noise signal in frequency domain.

The PSD estimate of a signal is a real-valued function that represents the distribution of power over frequency in the signal. The PSD estimate is defined over the frequency range of $0\ to\ fs/2$, where f_s is the sampling frequency of the signal, and T_s is the sampling time of the input signal. It contains N samples, where N is the length of the signal [27].

The most important features of a noise signal in the frequency domain are its magnitude, median frequency, bandwidth, and power. Figure 38 shown the power spectrum of pink noise.

- **Magnitude:** The magnitude of the PSD estimate at a given frequency represents the power of the noise at that frequency.
- **Median frequency:** The median frequency of the noise signal is the frequency at which half of the power in the spectrum is below and half is above.

- **Bandwidth:** The bandwidth of the noise signal is the frequency range that contains a certain percentage of the total power in the signal. It is typically defined as the frequency range that contains 99% of the total power.
- **Power:** The power of the noise signal is the sum of the absolute squares of its time-domain samples divided by the signal length. It is a measure of the overall energy in the noise signal.

2.2. Classification Algorithm

Machine learning systems usually search for similarities between the features and specified classes to classify the data training set during the earning process. To analyze the data set classification four classification learning algorithms were applied such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Decision Tree, and Nearest Neighbor.

Discriminant analysis classification methods can be derived from probabilistic models that assume different classes generate data according to different Gaussian distributions. During the training phase, the discriminant classification algorithm estimates the parameters of a Gaussian distribution for each class, including the prior probabilities (π_i), means (μ_i), and covariance matrices(Σ_i). The algorithm then constructs weighted classifiers and estimates the class mean and variance for weighted and unweighted datasets.

$$\pi_k = \frac{N_k}{N} \qquad (4)$$

Where:

- N_k is the number of observations in class k.
- N is the total number of observations.

Suppose *M* is an *N*-by-*K* class membership matrix:

M_{nk} if observation n is from class k

$M_{nk} = 0$ otherwise.

The estimate of the class mean for data $\widehat{\mu}$:

$$\widehat{\mu_k} = \frac{1}{\sum_{n=1}^{N} M_{nk}} \sum_{n=1}^{N_k} M_{nk} x_n \quad (5)$$

Where:

- x_n is the observation n of training set.
- and N_k is the number of observations in class k.

-The unbiased estimate of the covariance matrix:

$$\hat{\Sigma} = \frac{\sum_{k=1}^{K} \sum_{n=1}^{N_k} M_{nk} (x_n - \widehat{\mu_k})(x_n - \widehat{\mu_k})^T}{N-k} \quad (6)$$

Where:

- k is the number of classes.

Classifiers like LDA do not use posterior probability or cost matrix to fit the training data. They simply model the class-conditional distribution of the data $p(x|k)$ as a multivariate Gaussian distribution with density:

$$p(x \mid k) = \frac{1}{(2\pi)^n |\hat{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\widehat{\mu_k})^T \Sigma_k^{-1} (x-\widehat{\mu_k})} \quad (7)$$

Where:

- x is the data point to be classified.
- n is the dimensionality of the data.

$\hat{P}(k \mid x)$ is the posterior probability estimate of observation x belonging to class k.

$$\hat{P}(k \mid x) = \frac{p(x \mid k)\pi_k}{p(x)} \quad (8)$$

Where:

- π_k is the prior probability.
- $p(x)$ is the normalization constant, which is the sum of the posterior probabilities for all classes.

The normalization constant is necessary to ensure that the sum of the posterior probabilities for all classes is equal to 1. In linear Discriminant Analysis (LDA), it is assumed that all classes share a common covariance matrix for their Gaussian distributions, with only the mean vectors vary between classes. the prediction using Bayes rule that minimizes the expected classification cost is given by the following formula:

$$\hat{y} = \arg\min \sum_{k=1}^{K} \hat{P}(k \mid x)\, C(y|k) \qquad \text{y = 1,...,K} \qquad (9)$$

Where:

$C(y|k)$ is the cost of classifying an observation as class l while the true class is k and it is a $K - by - K$ matrix and each element is either 0 (if the observation is in the true class) or 1 (if the observation is not classified in the true class).

3. Implementation

Data analysis was achieved using MATLAB software environment for several key tasks, including preprocessing (feature extraction) and algorithm development for classification. The trained classifier was then used to predict new data.

The sample noise signals have been generated using a MATLAB noise signal generator [28]. The generator can produce different types of colored noise, including pink, red, blue, and violet noise. We considered 5 colors as shown in Figure 39. The different colors have different spectra, and each class has a unique spectrum. White noise and pink noise are used more as reference signals and have very close spectrums. Black noise was not considered as it is total silence and cannot be discerned by the human ear.

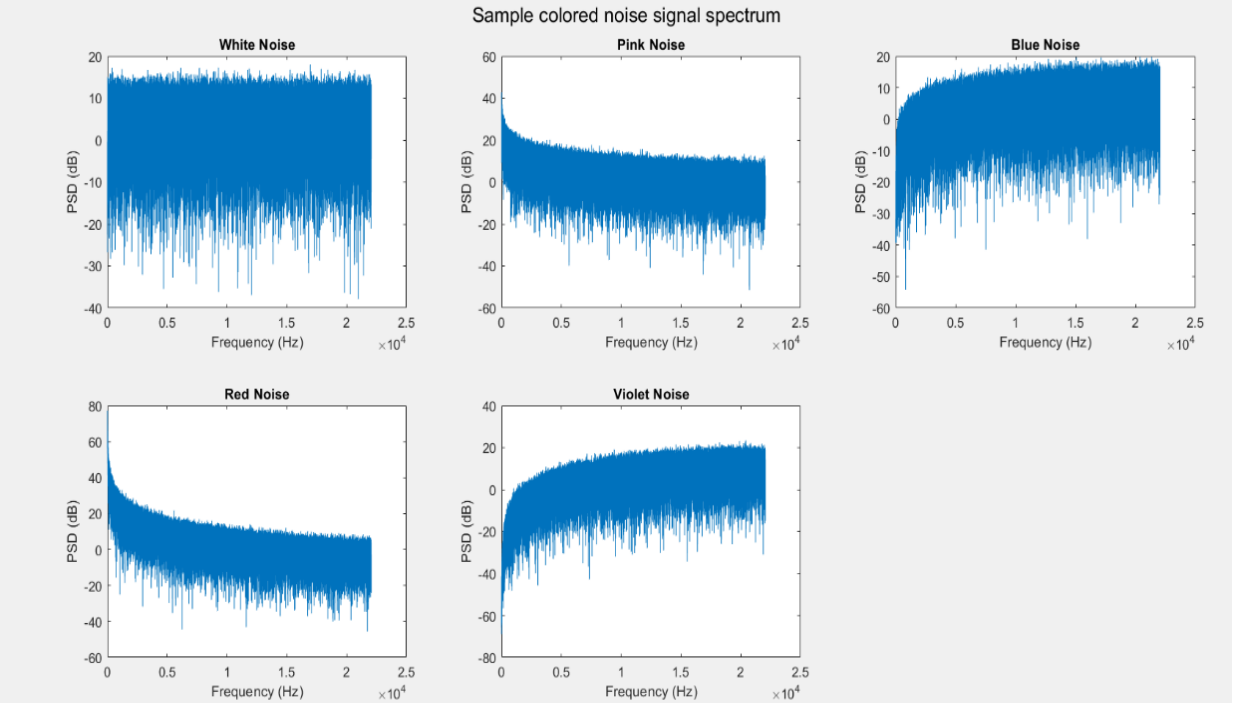


Figure 39: Sample colored noise signal spectrum

The input dataset for the classifier comprises 500 observations, each associated with a set of features extracted from both the frequency and time domains.

In the frequency domain, these features encompass the magnitude, median frequency, bandwidth, and band power derived from the power spectral density (PSD) of each signal. In the time domain, the features include the minimum, maximum, mean, and variance values calculated from each sample of the noise signal.

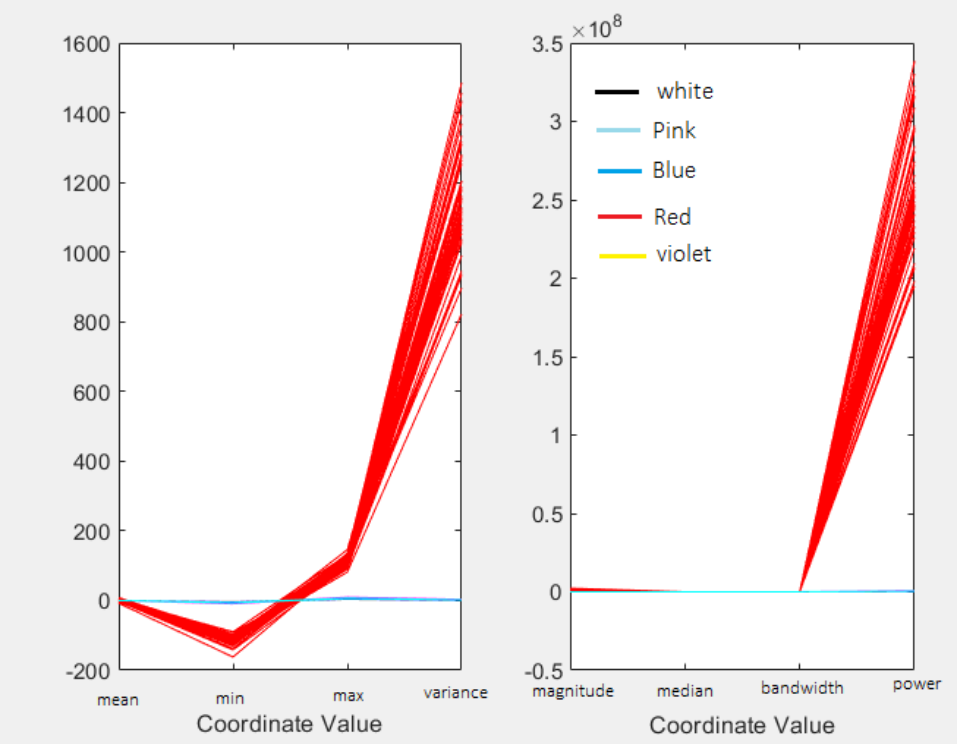


Figure 40: Sample input data features in time and frequency domain.

3.1 Learning Phase

Through extracting the features of signals and building the input training set and corresponding output responses, we utilized these data sets to train the classification model. Choosing the right classification algorithm is a balancing act and the difficult part of the problem. There exists an inherent trade-off between algorithm complexity and how restrictive the assumptions imposed on the data. For example, decision trees, linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA) are all relatively simple algorithms. However, they are also relatively restrictive in terms of the assumptions they make about the data. If the data doesn't match these assumptions, the algorithms may not perform well. On the other hand, bagged decision trees and support vector machines are both relatively complex algorithms. However, they are also less restrictive in terms of the assumptions they make about the data. This means that they can perform well on a wider variety of data sets. However, the resulting models can be extremely complex, making them difficult to understand and interpret. In

general, it is best to choose a simpler algorithm, as these are easier to interpret and explain to others. However, if the data is complex or noisy, a more complex algorithm may be necessary to achieve good performance as outlined in table 4.

Table 4: Classification method characteristics

Classifier	Characteristics			
	Multi class	***Prediction speed***	***Interpretability***	***Memory usage***
Decision Tree	Yes	Fast	Easy	Small
LDA	Yes	Fast	Easy	Small
QDA	Yes	Fast	Easy	Large
Nearest Neighbour	Yes	Medium	Hard	Medium

The classifier has two different inputs:

- **Sample input matrix (X)**: This is a $500 - by - 8\ matrix$, where each row represents a sample signal and each column represents a feature. The features are extracted from the time domain and frequency domain of the signal.
- **Known response matrix (Y)**: This is a $500 - row\ matrix$, where each row contains the known class label for the corresponding sample signal. These Inputs are given to the classifier has the time domain and frequency domain features for 5 different classes as Blue, Pink, Red, Violet, and White.
- **Q-Q Plot**

  A Q-Q plot graphically tool shows whether an empirical distribution is close to a theoretical distribution. When these distributions are identical, the Q-Q plot follows a straight 45-degree line. However, when they differ, the Q-Q plot deviates from this line.
- **Q-Q Plot for LDA**

  In the process of classifying (fitting), we computed the sample mean and covariance for each class based on the observations and the empirical covariance matrix for each class. Figure 41 displays a Q-Q for the linear discriminant. Designed to assess the validity of the Gaussian mixture assumption.

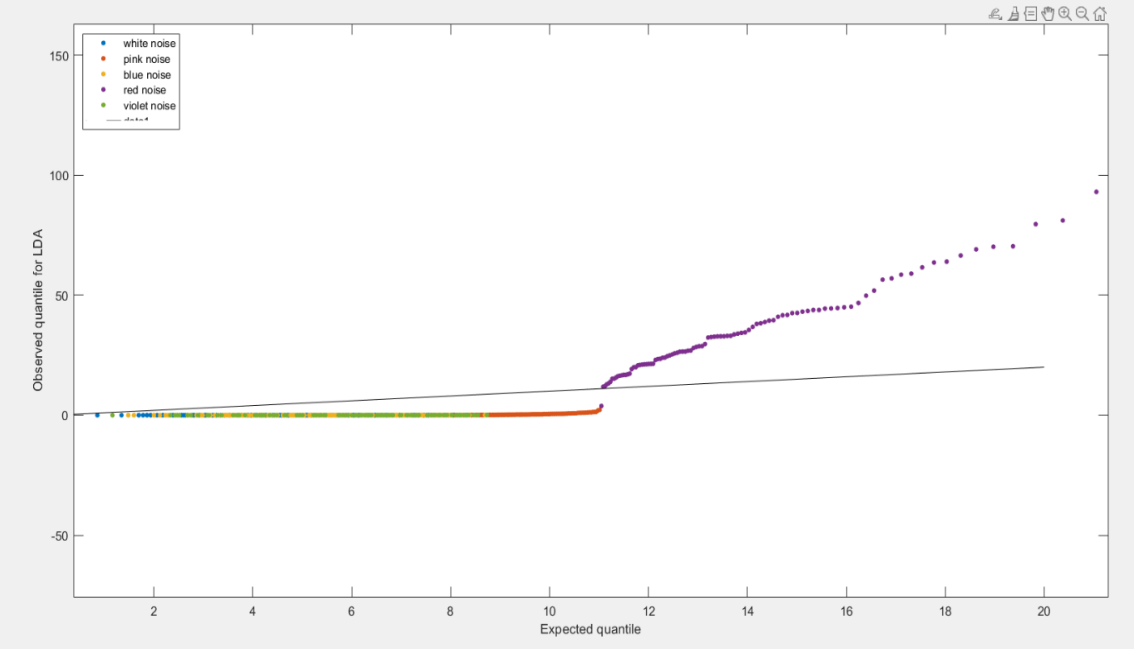


Figure 41: Q-Q for the linear discriminant

The expected and observed quantiles agree somewhat. The plot shows that the data has heavier tails than a normal distribution. The plot shows 15 possible outliers at the top from class 'Red Noise'. This means that the data is not well-modeled by a single covariance matrix, which is the assumption that LDA makes.

- **Q-Q Plot for QDA**

Figure 42 displays a Q-Q for the Quadratic discriminant. designed to assess the validity of the Gaussian mixture assumption.

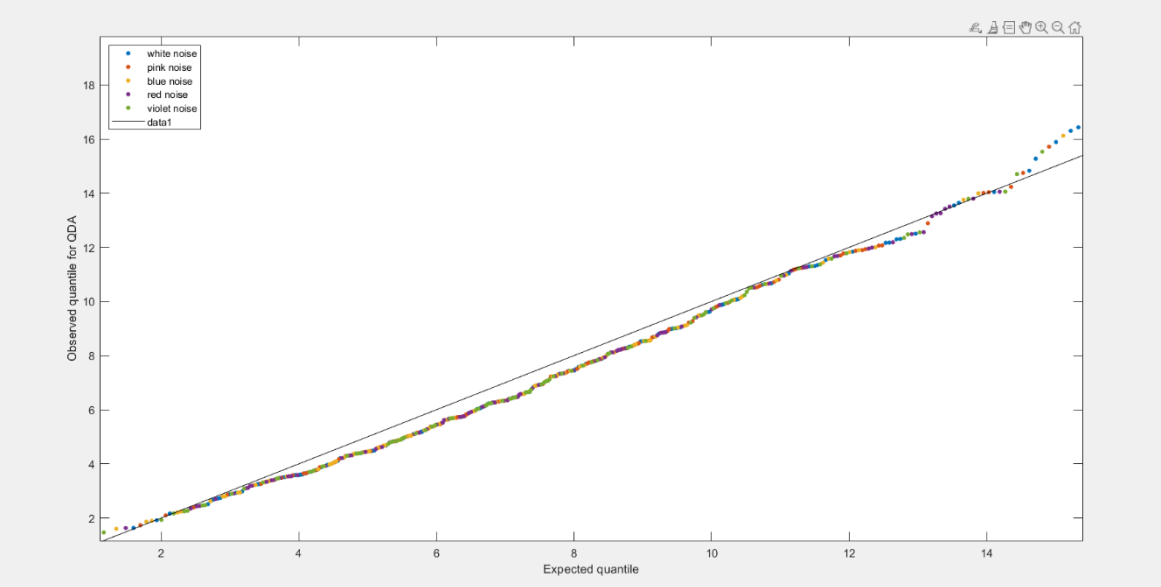


Figure 42: The Q-Q plot for the quadratic discriminant

The Q-Q plot for the quadratic discriminant shows a better agreement between the observed and expected quantiles. The plot shows only three possible outliers from class 'White Noise'. Colored noise signals data is better modeled as a Gaussian mixture with covariance matrices that are not required to be equal across classes.

3.2 Verification of Classifier

To validate the classification accuracy of a trained classifier, we can use different functions such as misclassification error and cross-validation error.

Misclassification Error is a common metric refers to difference between the response training data which are the true predictions of the known training set and the predictions that classifier makes of the response based on the input training data. Misclassification error can also be used to estimate the model's predictive error on unknown new data.

Confusion matrix is an easy way of visualizing predicted class versus true class that shows how many times misclassification occurs during learning phase. The diagonal of the

Matrix represents the correctly classified inputs and off-diagonal elements show the number of incorrect predictions.

The two figures below, show the Misclassification Error and the Confusion Matrix for different classifiers used to classify 500 sound signals for 5 classes of noise color (100 from each class that is given sequentially to the classifier in learning phase).

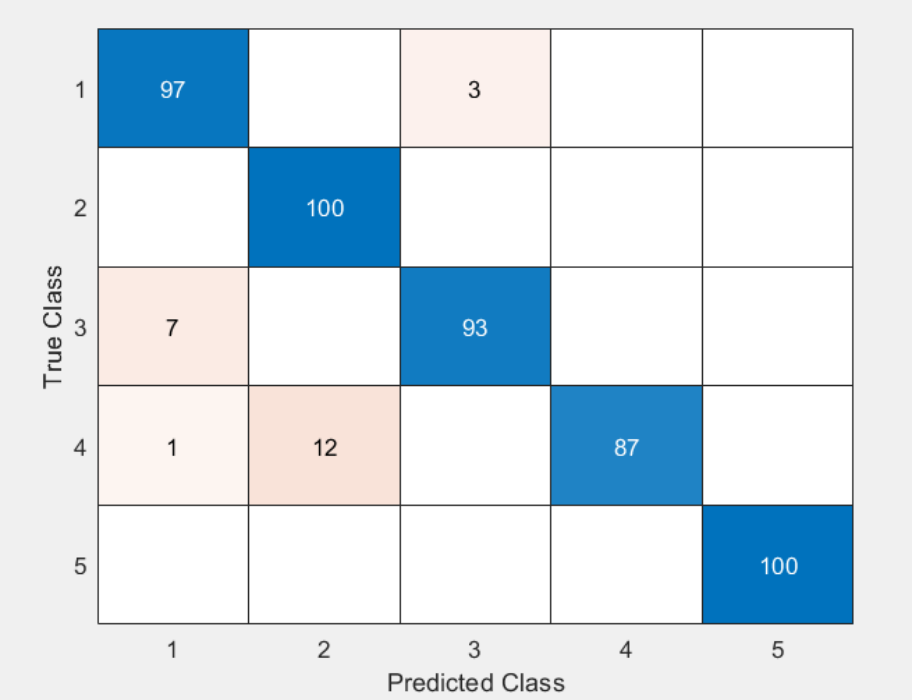


Figure 43: CLASSIFICATION VERIFICATION WITH CONFUSION MATRIX LDA

For each class number represent a noise color:

Class 1 = ‘White Noise’

Class 2 = ' Pink Noise’

Class 3 = ' Blue Noise’

Class 4 = ' Red Noise’

Class 5 = ' Grey Noise

Of the 500 training observations, 4% or 23 observations are misclassified by the linear discriminant function.

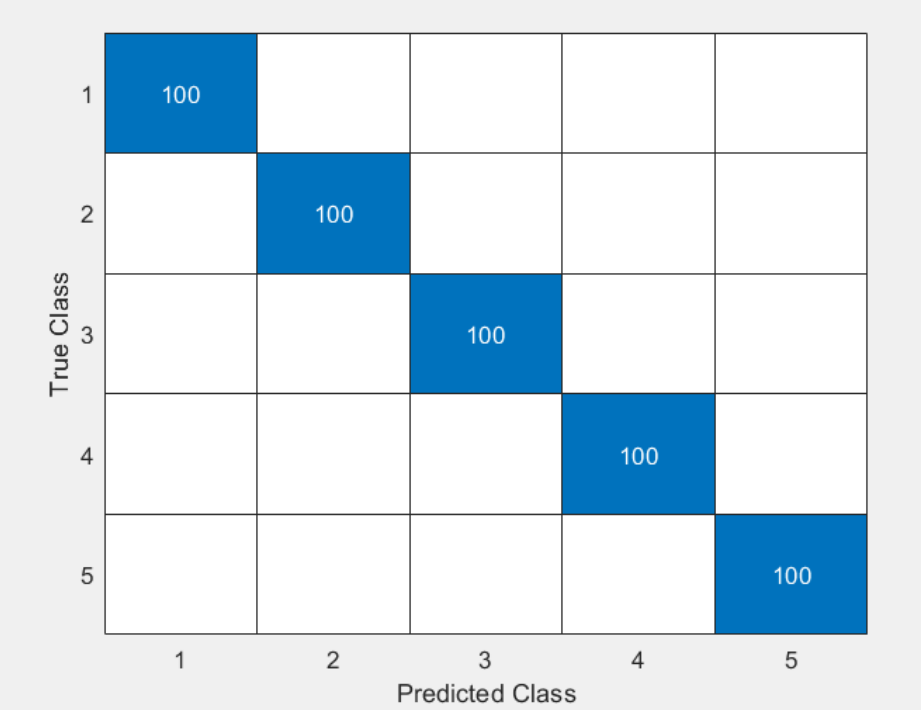


Figure 44: CLASSIFICATION VERIFICATION WITH CONFUSION MATRIX QDA

Of the 500 training observations, 0% or 0 observations are misclassified by the quadratic discriminant function.

To evaluate the performance and compare of a classifier at different threshold settings, we used Receiver Operating Characteristic (ROC) curve.

ROC curve is a graphical representation that illustrates the balance between the true positive rate and the false positive rate, also known as sensitivity and specificity, across various thresholds of the classifier's output. ROC curves are a valuable tool for evaluating the performance of classification models. They help in identifying the optimal threshold that maximizes classification accuracy, which corresponds to regions on the curve where both sensitivity and specificity are high.

Figures 45, 46 show the ROC curve for different classifiers. The filled circle markers indicate the optimal operating points. The legend displays the class name and AUC value for each curve.

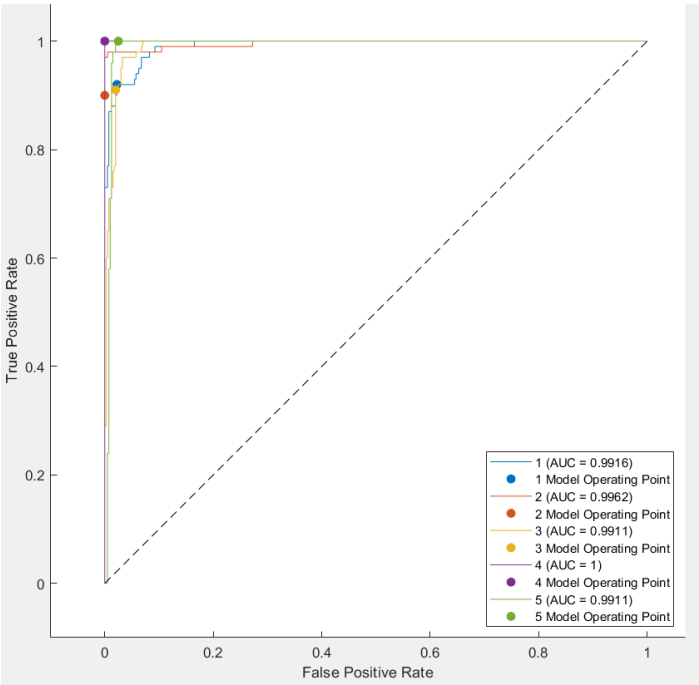


Figure 45: ROC curve by linear discriminant analysis classifier

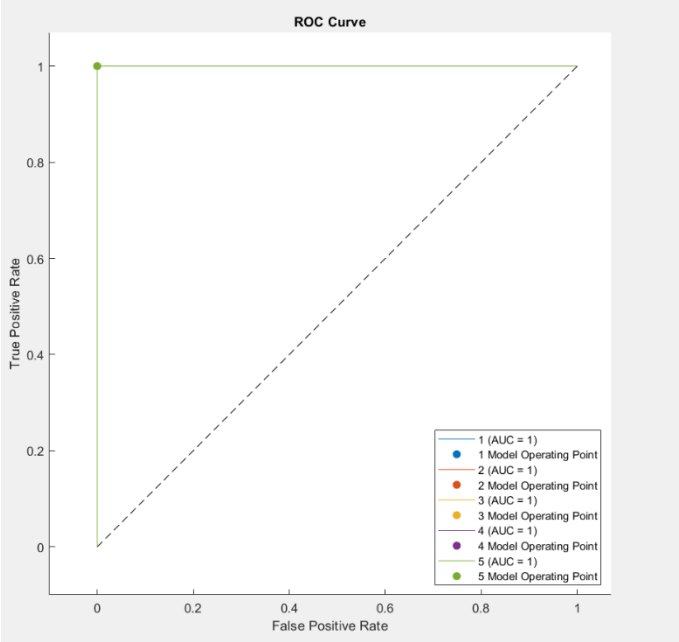


Figure 46: ROC curve by quadratic discriminant analysis classifier

Through trained our classifiers, then we can use it to see predictions about unknown new dataset that contain features of time domain and frequency domain of different noise signals. For each classifier we gave a new set of input data of 50, 100, 250, 500, 1000, and 2000 observations. To measure which one of classifiers give a better prediction of input data sets of colored noise. To evaluate the accuracy of each classifier is computed by comparing the number of correctly estimated signal to the total number of signals provided as input to each classifier. as shown in figure 47,48.

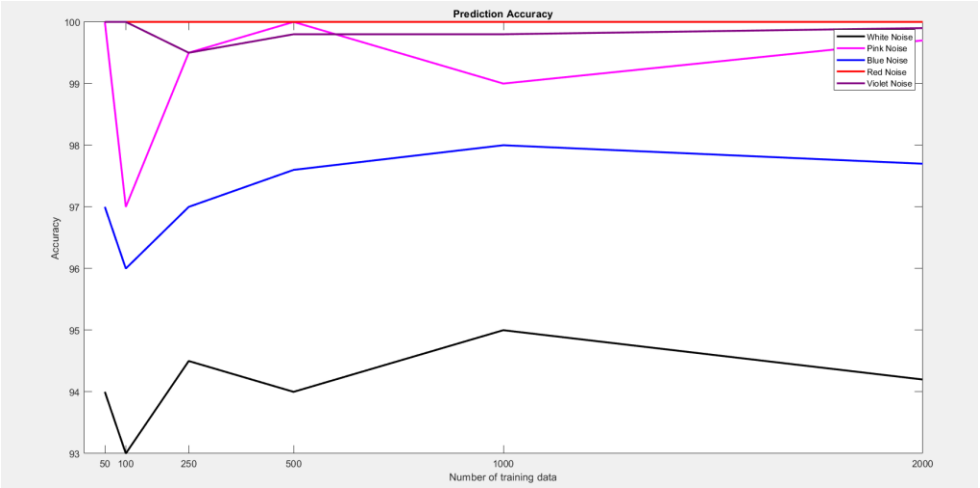


Figure 47: Prediction Accuracy by LDA

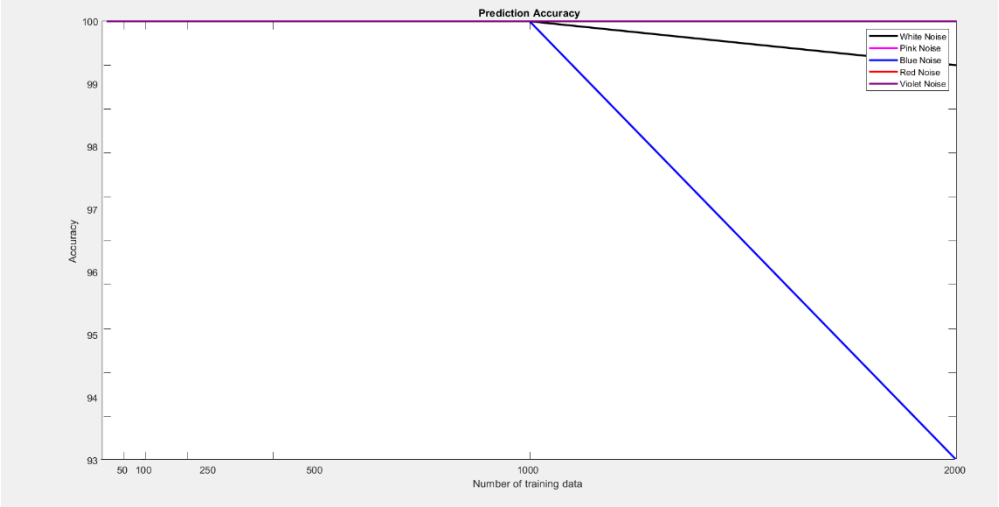


Figure 48: Figure 47: Prediction Accuracy by LDA

Quadratic Discriminant Analysis model with the average accuracy of 96.5% is more accurate than Linear Discriminant Analysis classification method used for classifying colored noise signals.

Chapter 4: RESULTS AND DISCUSSION

1. Characterization of Environmental Noise Pollution through Noise Measurement and Mapping

We conducted noise level measurements at a location known as **'Quinta Do Carmo'** and compared the results to the noise limits recommended by the World Health Organization (WHO) and the European Environment Agency defines High environmental (i.e., outdoor) noise levels are defined in the 7th EAP as noise levels for L_{den} above $55\,dB$ and for L_{night} above $50\,dB$ [29]. Figure 49 shows the equivalent sound pressure level (L_{eq}) these measurements were taken at 10 specific locations during the day, and they reflect the noise generated by various movements and activities nearby.

Table 5: GPS coordinate and noise descriptors of each point in this study

No.	Cars	Motorcycle	Medium Lorries	Agricultural machinery	Pedestrian	PCU
1	3	0	2	2	3	9
2	2	0	1	0	1	3
3	1	0	1	1	1	5
4	3	1	3	2	4	13
5	1	1	0	0	0	1
6	1	0	2	0	0	5
7	3	1	1	1	1	7
8	0	0	0	0	0	0
9	2	0	0	0	0	2
10	2	0	0	2	0	7

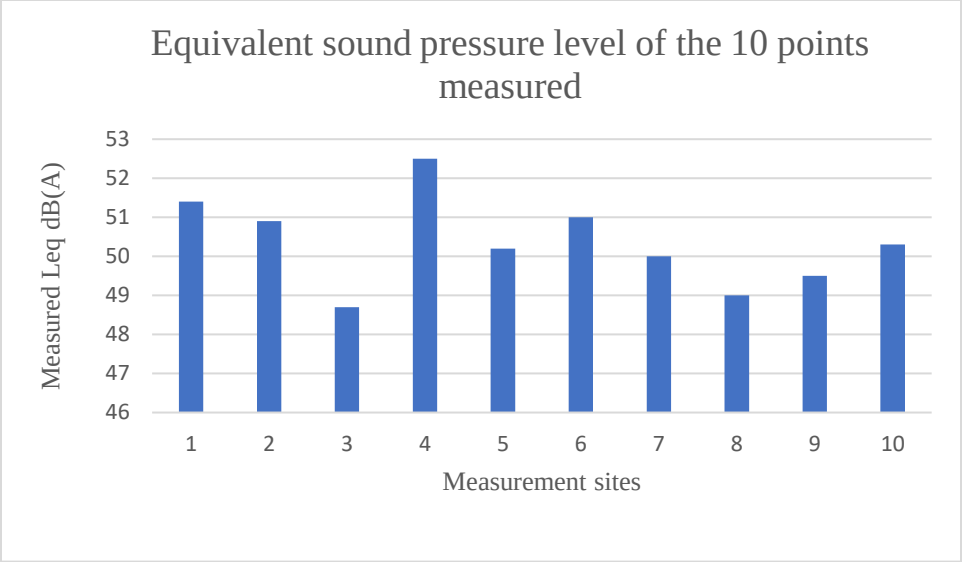


Figure 49: The equivalent sound pressure noise level

The measurement indicated 10 measured points which did not exceed the maximum permissible sound level, 55 dB (A) of the L_{eq}. The maximum sound level recorded was 66.3 dB(A) at point 4 and followed by 63.9 dB(A) at point 6 (Table 5). These measured high noise levels are attributed to the traffic movements and other activities nearby the points, such as the warehouse works, tractors, and planters where noisy equipment was in operation. Moreover, the road located near to point 4 contributed to increased sound levels, due to the elevated number of pedestrians passing. The noise limit above 55 dB (A) is generally considered to be an anxiety level and above 65 dB(A) is a noise level that affects the quality of life if the noise is continuous in the region [30].

Figure 50 until Figure 52 show the noise mapping of Quinta Do Carmo, identifying the three parameters for each of the sampling points throughout the area. The measured parameters are L_{eq} , L_{max}, and L_{min}. Figure 50 shows the noise mapping of the Quinta Do Carmo for the L_{eq} measurement, which the highest equivalent continuous level is 52.5 dB(A) with the constant noise level over a given time period. There is no existence noise from the warehouse works. This noise level would contribute greatly to diminishing people's quality of life. Unwanted sound of sufficient intensity and duration can cause temporary and permanent hearing loss [31]. It can also interfere with speech communication, the transmission of other auditory signals, disturb sleep,

act as a general source of annoyance or disturbance, and interfere with the performance of complicated tasks and the opportunity for privacy.

Exposure of people to daytime noise levels above 65 dB(A) can cause severe health problems. Thus, the noise level should be well controlled in order to ensure that people's lives can be safe without causing any health issues or damaging their hearings. Fortunately, the collected data indicated that area is not subjected to sound pressure levels exceeding the limit established by guidelines for the educational areas above the limit of 55 dB(A).

Figure 50: Noise mapping of Quinta Do Carmo for L_{eq}

Figure 51: Noise mapping of Quinta Do Carmo for L_{max}

Figure 52: Noise mapping of Quinta Do Carmo for L_{min}

Table 5 shows the raw data and converted to passenger car unit (PCU) for the number of vehicles and pedestrians for each sampling point at the Quinta Do Carmo during the measurement period. Generally, traffic doesn't cause an increase in noise levels, as shown in Figure 43, were we notice higher noise levels at the sampling points. It is worth mentioning that the number of pedestrians also not affect to higher recorded values of noise levels. For instance,

at sampling point 4, the number of people passed that point reached 4 during that period, and the measured L_{eq} was 52.5 dB(A), compared to only 49 dB(A) at sampling point 8, where no pedestrians existed at that point.

The influence of traffic and machinery on the measurement of noise levels was in accordance with several studies conducted in different cities around the world[32] [20] . There is study conducted in Kolhapur, India, where the observed results indicated that high L_{eq} measure.

2. Colored Noise Signal Identification Results

In our pursuit of characterizing environmental noise pollution through noise measurement and mapping, we were heartened to discover that our landscape had been deemed safe. With this knowledge as our foundation, our research ventured into the realm of examining potential ways to discern similarities among the five distinct noise colors that we had generated synthetically, forming a comprehensive soundscape.

To address this intriguing challenge, we employed the ever-reliable approach of Supervised Learning, an established method in the field of machine learning. Our focus was on harnessing the power of linear and quadratic discriminant analysis, both proven and efficient models in their own right. These models were meticulously detailed and explained in Chapter Three of our study, shedding light on their prowess in this context. Through a series of rigorous measurements, paved the way for the application of our classification methodology. Our primary objective was to identify potential similarities and patterns within this soundscape, a task of paramount importance in our overarching goal of comprehending the intricacies of environmental noise pollution, we present two graphs: an oscillogram and a power spectrum, illustrating our collected data as figures 53,54.

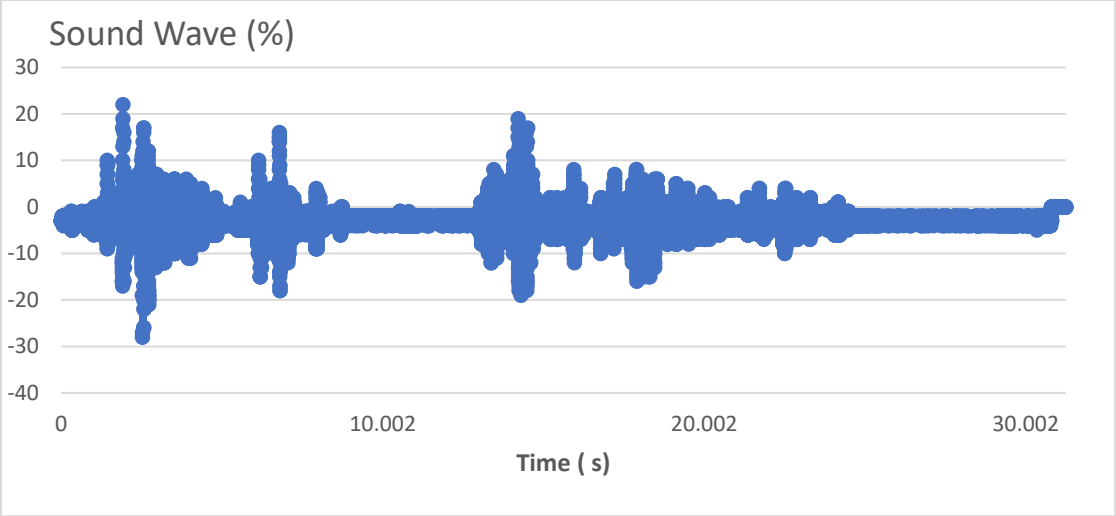


Figure 53: oscillogram GPS coordinate Nu 3.

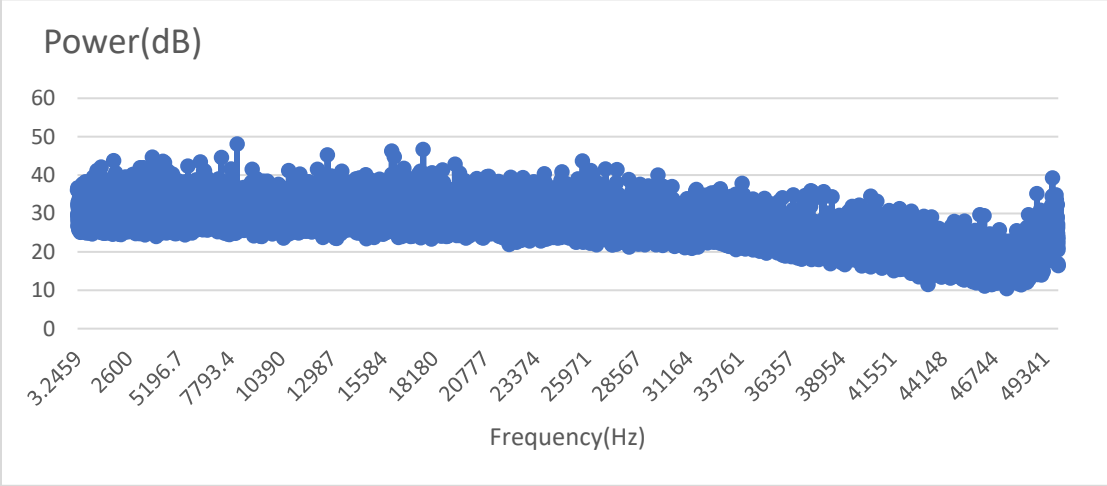


Figure 54: Power spectrum GPS coordinate Nu 3.

Signal Parameters:

- Sample rate = 1000 Hz
- Time = 5 seconds

The results obtained from our LDA and QDA classifications, through GPS of each point in this study.

Table 6: percentage of each noise color by LDA

LDA(%)	White Noise	Pink Noise	Blue Noise	Red Noise	Violet Noise
1	5	71	9	6	9
2	26	22	23	6	22
3	25	22	24	6	21
4	23	31	24	7	19
5	4	65	6	5	20
6	0	98	0	1	1
7	0	97	1	1	1
8	0	98	0	1	1
9	56	14	12	13	5
10	4	88	4	3	1

2.1. **LDA** Discussion

- Points 1, 6, 7, 8, and 10 have shown a consistent probability of over 70% for the presence of pink noise as shown in table 6. This suggests a strong correlation between these data points and the occurrence of pink noise in our environmental soundscape.
- Among these data points, the probabilities for other noise colors are relatively balanced, with each color having an approximate 20% likelihood of occurrence. This indicates that, apart from pink noise, the data exhibits a certain level of diversity in terms of noise colors, with no dominant alternative.
- Point 9, in particular, stands out, as it displays a significant 56% probability of white noise occurrence, suggesting that this specific data point is more closely associated with white noise compared to other noise colors.

Table 7: percentage of each noise color by QDA

QDA(%)	White Noise	Pink Noise	Blue Noise	Red Noise	Violet Noise
1	3	80	1	15	1
2	7	30	4	56	3
3	3	25	5	60	7
4	4	35	3	46	12
5	2	70	2	25	1
6	0	98	0	1	1
7	0	97	1	1	1
8	0	99	0	1	0
9	0	96	0	3	1
10	4	88	4	3	1

2.2. **QDA** Discussion

- The first data point exhibits a substantial 80% probability of pink noise as shown in table 7, indicating a strong correlation between this data and the prevalence of pink noise in our soundscape.
- Data point 5 also showcases a high 70% probability of pink noise, reinforcing the consistency of pink noise presence in our study area.
- Data points 6 through 10 consistently show more than a 70% probability of pink noise, further underscoring its dominance in these locations.
- However, data points 2, 3, and 4 present a different soundscape pattern. They exhibit probabilities of over 45% for red noise and between 20% to 30% for pink noise, suggesting a unique acoustic characteristic in these specific areas.

QDA analysis reaffirms the prevalence of pink noise in several data points, like the LDA results. However, data points 2, 3, and 4 deviate from this trend, showing a distinct association with red noise and a secondary association with pink noise. This unique soundscape characteristic in these areas warrants further investigation to better understand the underlying factors contributing to this noise composition.

Conclusion

Our investigation into environmental noise pollution within Quinta do Carmo, Évora, Portugal, sought to comprehensively assess the acoustic landscape across ten strategically chosen points in the area, as depicted in the accompanying noise mapping figures. The significance of such a study lies in its ability to gauge the impact of noise pollution on the surrounding environment. Our findings revealed that, while the noise levels within the studied area remained within permissible limits, discernible fluctuations were observed. Notably, a peak measurement of 66.3 dB(A) emphasized the dynamic nature of the acoustic environment. Furthermore, our examination of activities involving noisy equipment and machinery in engineering laboratories indicated their minimal contribution to elevated noise levels due to their infrequent occurrence. Returning to the realm of signal identification, we employed supervised machine learning algorithms, specifically Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), to classify a diverse set of sample noise signals. Leveraging features extracted from both time and frequency domains and utilizing 500 observations, our trained models demonstrated proficiency in predicting the color class for new datasets, showcasing the potential of supervised machine learning in environmental noise analysis [33]. The success in classifying noise signals through LDA and QDA underscores the promise of employing advanced computational techniques in environmental acoustics. Notably, the prevalence of pink noise and the variability in noise types within the soundscape highlight the intricate nature of acoustic environments. The distinctive characteristics observed at specific data points emphasize the imperative for future research to unravel the underlying factors contributing to these variations, providing valuable insights for understanding and mitigating noise-related issues within the studied landscape. In essence, this study contributes not only to the assessment of noise pollution but also advances the application of cutting-edge technologies, specifically supervised machine learning, in unraveling the complexities of environmental acoustics. The implications of our findings extend beyond the confines of this study, urging further exploration and research to foster a deeper comprehension of the intricacies associated with noise in human-engineered landscapes.

References:

[1] J. Kadis, *The Science of Sound Recording*, 1st ed. Routledge, 2012. doi: 10.4324/9780240823645.

[2] M. P. Norton and S. J. Drew, "VIBRATION GENERATED SOUND | Fundamentals," in *Encyclopedia of Vibration*, Elsevier, 2001, pp. 1443–1455. doi: 10.1006/rwvb.2001.0207.

[3] I. Nurjannah, A. Harijanto, and B. Supriadi, "Sound Intensity Measuring Instrument Based on Arduino Board with Data Logger System," *Int. J. Adv. Eng. Res. Sci.*, vol. 4, no. 9, pp. 27–35, 2017, doi: 10.22161/ijaers.4.9.7.

[4] B. C. Pijanowski *et al.*, "Soundscape Ecology: The Science of Sound in the Landscape," *BioScience*, vol. 61, no. 3, pp. 203–216, Mar. 2011, doi: 10.1525/bio.2011.61.3.6.

[5] M. Russ, *Sound synthesis and sampling*, 3. ed. Amsterdam: Elsevier, Focal Press, 2009.

[6] F. Roche, "Music sound synthesis using machine learning: Towards a perceptually relevant control space".

[7] P. Manning, *Institut de Recherche et Coordination Acoustique/Musique*, vol. 1. Oxford University Press, 2001. doi: 10.1093/gmo/9781561592630.article.42130.

[8] Hyungseob Han and Sangjin Cho, "A new pitch tracking algorithm for the spectral modeling synthesis," in *Ifost*, Ulaanbaatar, Mongolia: IEEE, Jun. 2013, pp. 364–366. doi: 10.1109/IFOST.2013.6616987.

[9] P. Djoharian, "MATERIAL DESIGN IN PHYSICAL MODELING SOUND SYNTHESIS," 1999.

[10] Z.-G. Huang and S.-J. Shiu, "Design and analysis of diaphragms in dynamic microphones," *Adv. Mech. Eng.*, vol. 7, no. 7, p. 168781401559574, Jul. 2015, doi: 10.1177/1687814015595748.

[11] J. Esteves, L. Rufer, and G. Rehder, "Capacitive Microphone fabricated with CMOS-MEMS Surface-Micromachining Technology," 2011.

[12] M. Hiipakka, "Measurement Apparatus and Modelling Techniques of Ear Canal Acoustics".

[13] R. Groschup and C. Grosse, "MEMS Microphone Array Sensor for Air-Coupled Impact-Echo," *Sensors*, vol. 15, no. 7, pp. 14932–14945, Jun. 2015, doi: 10.3390/s150714932.

[14] H. Lu, Y. Xiaoyu, W. Haodong, L. Jin, M. Xuejiao, and Z. Caihong, "Research on Application of Digital Signal Processing Technology in Communication," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 799, no. 1, p. 012026, Mar. 2020, doi: 10.1088/1757-899X/799/1/012026.

[15] X. Luo and Z. Zhang, "Data recovery with sub-Nyquist sampling: fundamental limit and a detection algorithm," *Front. Inf. Technol. Electron. Eng.*, vol. 22, no. 2, pp. 232–243, Feb. 2021, doi: 10.1631/FITEE.1900320.

[16] M. T. Heideman, D. H. Johnson, and C. S. Burrus, "Gauss and the history of the fast Fourier transform," *Arch. Hist. Exact Sci.*, vol. 34, no. 3, pp. 265–277, 1985, doi: 10.1007/BF00348431.

[17] P. Heckbert, "Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm".

[18] F. E. Clark, "From Soundwave to Soundscape: A Guide to Acoustic Research in Captive Animal Environments," *Front. Vet. Sci.*, vol. 9, 2022.

[19] O. F. S. Khasawneh, H. Halim, S. N. Abdullah, S. A. Razali, H. R. F. Algburi, and A. H. Salleh, "Characterization of Environmental Noise Pollution Based on Noise Measurement and Mapping at USM Engineering Campus," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 920, no. 1, p. 012004, Sep. 2020, doi: 10.1088/1757-899X/920/1/012004.

[20] P. H. T. Zannin, "Characterization of environmental noise based on noise measurements, noise mapping and interviews: A case study at a university campus in Brazil," 2013.

[21] K. Guo, Y. Wu, and H. Zhang, "The Effects of Color Noises on Attention," in *Proceedings of the 2022 International Conference on Science Education and Art Appreciation (SEAA 2022)*, Z. Zhan, F. P. Chew, and M. T. Anthony, Eds., Paris: Atlantis Press SARL, 2023, pp. 576–583. doi: 10.2991/978-2-494069-05-3_71.

[22] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
[23] J. Zhou, L. Zheng, Y. Zhong, S. Hao, and M. Wang, “Positive Sample Propagation along the Audio-Visual Event Line.” arXiv, Apr. 05, 2021. Accessed: Oct. 28, 2023. [Online]. Available: http://arxiv.org/abs/2104.00239
[24] E. V. Carrera, “Machine Learning for Signal Processing,” 2016, doi: 10.13140/RG.2.1.1399.4641.
[25] H. Malik, U. Bashir, and A. Ahmad, “Multi-classification neural network model for detection of abnormal heartbeat audio signals,” *Biomed. Eng. Adv.*, vol. 4, p. 100048, Dec. 2022, doi: 10.1016/j.bea.2022.100048.
[26] M. Nourbagheri and M. Zohdy, “Coloured Noise Signal Identification using Supervised Learning Algorithm,” vol. 05, no. 04.
[27] T. Bryant and M. Zohdy, “Noise Signal Identification by Modified Self- Organizing Maps,” vol. 03, no. 01.
[28] H. Zhivomirov, “A Method for Colored Noise Generation,” vol. 15, no. 1.
[29] *Environmental noise guidelines for European Region*. Copenhagen: WHO Regional Office for Europe, 2018.
[30] J. Alves, L. Silva, and P. Remoaldo, “The Influence of Low-Frequency Noise Pollution on the Quality of Life and Place in Sustainable Cities: A Case Study from Northern Portugal,” *Sustainability*, vol. 7, no. 10, pp. 13920–13946, Oct. 2015, doi: 10.3390/su71013920.
[31] B. Song *et al.*, “Music Literacy and Soundscape Perception: A Study Based on the Soundwalk Method of Soundscapes,” *Int. J. Environ. Res. Public. Health*, vol. 19, no. 14, p. 8471, Jul. 2022, doi: 10.3390/ijerph19148471.
[32] K. Fatima, R. Rampal, M. Muslim, and K. Hussain, “Evaluating the noise exposure levels of different schools in noise-scapes of Kargil town: Ladakh, India,” *Noise Vib. Worldw.*, vol. 54, no. 4–5, pp. 215–223, Apr. 2023, doi: 10.1177/09574565231161647.
[33] M. Hodges and M. Zohdy, “Intelligent Hearing Assistance Using Self-Organizing Feature Maps,” *Trans. Mach. Learn. Artif. Intell.*, vol. 2, no. 6, Dec. 2014, doi: 10.14738/tmlai.26.760.