CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

Department of Measurement

Bachelor Thesis

Acoustic Event Detection

Detekce akustických událostí

Jamal Aslanov

Study Program: Electrical Engineering and Computer Science

Supervisor: Ing. Jakub Svatoš, Ph.D.

Prague 2019

BACHELOR‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Aslanov Jamal** Personal ID number: **464281**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Measurement**

Study program: **Electrical Engineering and Computer Science**

II. Bachelor’s thesis details

Bachelor’s thesis title in English:

Acoustic Events Detection

Bachelor’s thesis title in Czech:

Detekce akustických událostí

Guidelines:

The goal of this bachelor thesis is to study current techniques of gunshot detection. On the basis of the studied materials propose methods for detection of gunshots. Consider the time domain and the frequency domain techniques. Implement these methods in one of the commonly used software.

Bibliography / sources:

[1] R. C. Maher, "Modeling and Signal Processing of Acoustic Gunshot Recordings," 2006 IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop, Teton National Park, WY, pp. 257-261, 2006.
[2] R. C. Maher, "Acoustical Characterization of Gunshots," 2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics, Washington, DC, USA, pp. 1-5, 2007.
[3] C. Clavel, T. Ehrette and G. Richard, "Events Detection for an Audio-Based Surveillance System," 2005 IEEE Internationalonference on Multimedia and Expo, Amsterdam, pp. 1306-1309, 2005.

Name and workplace of bachelor’s thesis supervisor:

Ing. Jakub Svatoš, Ph.D., Department of Measurement, FEE

Name and workplace of second bachelor’s thesis supervisor or consultant:

Date of bachelor’s thesis assignment: **20.06.2019** Deadline for bachelor thesis submission: **07.01.2020**

Assignment valid until:
by the end of winter semester 2020/2021

Ing. Jakub Svatoš, Ph.D. Supervisor’s signature	Head of department’s signature	prof. Ing. Pavel Ripka, CSc. Dean’s signature

III. Assignment receipt

The student acknowledges that the bachelor’s thesis is an individual work. The student must produce his thesis without the as sistance of others, with the exception of provided consultations. Within the bachelor’s thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt Student’s signature

Declaration:

I hereby declare that this thesis is the result of my own work and that I have clearly stated all information sources used in the thesis according to “Methodological Instruction of Ethical Principle in the Preparation of University Thesis”.

In Prague, 04.12.2019 Signature …………………......

Acknowledgements

I would like to express my gratitude that everyone involved in the development of my Bachelor research, especially my supervisor Ing. Jakub Svatoš, Ph.D. for his continuous guidance and his valuable support during project period. Besides, I would like to thank my colleague Ulvi Yusifli and Ruhani Dadashov for their continuous advice and assistance for their help in researching.

Abstract

This bachelor thesis is dedicated to the detections of gunshot signals in areas of interest and its possible improvement. The thesis contains two main methods that are divided into several algorithms. These methods analyse the current technique of gunshot detection and reveal some similar or different characteristics of signals. This thesis includes the FFT algorithm based on DFT that considered on the frequency domain. Implementations of the PSD, ZCR, Peak-Valley and MFCC methods are discussed and applied as a next step and analyse the possible way for gunshot recognitions.

Keywords

AED-acoustic event detection, muzzle blast, projectile, shockwave, GDS-gunshot detection system, ZCR-zero crossing rate, peak, valley,

Abstrakt

Tato bakalářská práce se věnuje detekci výstřelů v zájmových oblastech a jejich možnému zlepšení. Práce obsahuje 2 metody, které jsou rozděleny do několika algoritmů. Tyto metody jsou založené na současné technice detekce výstřelu a detekují podobnosti nebo odlišnosti výstřelů a falešných akustických signálů. Tato práce se zabývá algoritmy FFT, založeném na DFT ve frekvenční doméně dat, metodami PSD, ZCR, Peak-Valley a MFCC, které jsou diskutovány v dalších krocích a analyzují možný způsob rozpoznávání výstřelů.

Klíčová slova

AED - detekce akustických událostí, výbuch tlamy, projektil, rázová vlna, GDS-detekční systém střely, ZCR-rychlost křížení nuly, vrchol, údolí,

List of Abbreviations

AED - Acoustic Event Detection

FT – Fourier transforms

LT – Laplace transforms

GDS – Gunshot Detection System

GPS – Global Positioning System

MPE – Multi-path Propagation Effect

FFT- Fast Fourier Transform

DFT – Discrete Fourier Transform

IDFT – Inverse Discrete Fourier Transform

MFCC – Mel Frequency Cepstral Coefficient

FIR – Finite Impulse Response

LPF – Low Pass Filter

PSD – Power Spectral Density

ZCR – Zero Crossing Rate

DCT – Discrete Cosine Transform

Table of Contents

1. Introduction

Nowadays, there are many kinds of sensors that refer to various application fields. Acoustic signal processing-Acoustic event detection (AED) is one of the famous and most demanded areas. AED has widely used technology for today's globalized and developing scientific community. Due to the expectations for the near future, signal processing attracted the attention of numerous scientists. Its fundamental working principle separated into several sections: detection, classification, and localization of acoustic events. The main idea behind the AED sensors is based on the extraction of vital information from the signal data and classify its necessity due to the given issue. Sometimes it is challenging to detect important values from uncertain, inadequate measured audio data. As a solution to event detection, its primary technique is to use the help of well-known mathematical approaches: such as Fourier transforms (FT), Laplace transforms (LT) and wavelet transforms, etc. But still, in noisy places, it leads to many problems to classify complex audio signals. Signal processing requires more time to convert stored acoustic data from the time/frequency domain. There are still new ideas, and methods are discussed among the researchers to uncover suitable ways.

Variety at trading and innovative applications for AED improves the development of gunshot detection systems. One of the required and applied areas of AED is according to the safety and detection of various hazardous events nearby-gunshot. For this reason, the importance and growth of AED in recent years have increased significantly.

2. Gunshot Detection System

2.1 Basic Structure and working principle

Gunshot detecting system (GDS) or gunshot detector is modern developing sensors that detect hazardous and dangerous events in noisy areas. An essential object of the GDS is to collect acoustic data around the environment and classify it as alert(shot) signals and standard signals. According to this classification, GDS assists in Police, Law enforcement to intervene in the going process on time and decrease the further casualties. However, crime rates, terror attacks, violations of society have been increasing, and keeping the safety of people in public areas such as big cities, crowded squares, open malls, universities, etc. is much more attention needed. Gunshot detectors have a fundamental principle of alerting the accident, dangerous events that happen in a specific area, and store audio data concerning the incident place through microphones. In the second part, the detector will classify the signal and display it as an urgent or alerting signal. Later it will send some alarm to emergency police departments.

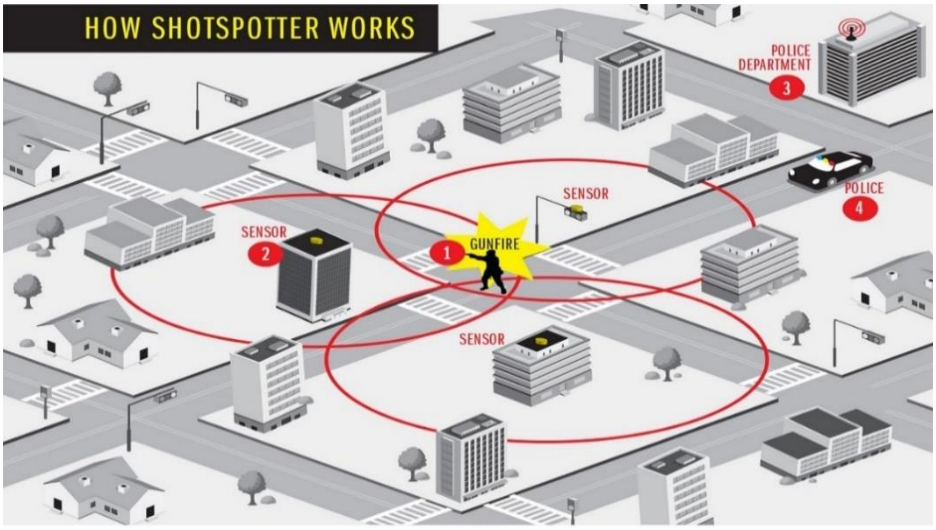


Fig1: *Shot detection procedure of sensor [1]*

Detection procedure on ig 1:

1- On the first step, microphones of sensors which is usually placed on the top of the buildings, collect audio data around the environment.

2- All the audio data collected by several sensors, unusual (gunfire) audio signals have been recorded for verifications of all sensors that attached to the same particular area. Afterward, the potential incident location has been detected by the GPS application of sensors.

3- Collected data (information) sent by the sensor has been analyzed and checked by the police department or emergency center with the help of security cameras, and all this procedure takes a few minutes to verify the incident.

4- At the last stage of the process, the police department shares valuable information about the gunshot, incident place to the closest police squads, and so police officers can reach to crime area as prepared and pretend further damage on time.

2.2 History

At the beginning of the 1990s, the United States (US) was one of the biggest countries on account of the rate of crime in cities among the world countries. It was mostly related to the gun reserves and allowance for the utilization of weapons by residents. According to this reason, the main center of GDS started to establish its applications in North America. Several cities of the US were equipped by GDS, such as Chicago, Los Angeles, San Francisco, Boston, etc. during the 2000s.

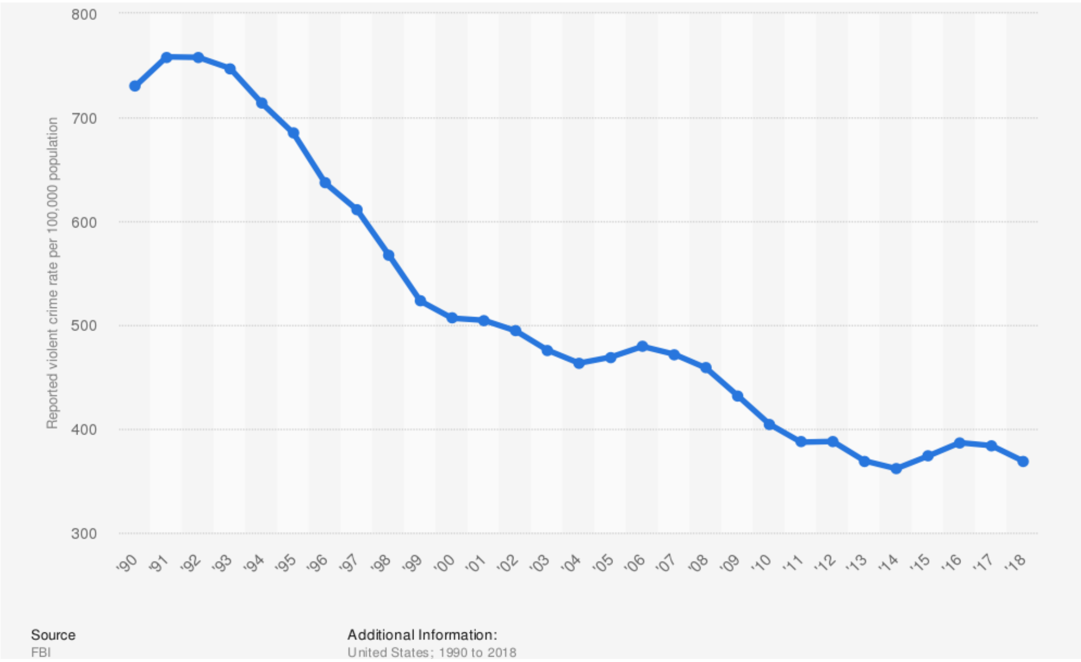


Fig 2: *Chart of crime rate in the US between 1990-2018 [2]*

Besides its public application, GDS is quite common in the military field. The primary purpose of the utilization of Gunshot detectors for the military had been started much more before than its public usage. At the beginning 20th century, ShotSpotter is used to determine the location of aircraft and submarines. The starting of WWI assisted in developing different shapes & variations of gunfire locators, and it continued with the occurrence of WWII. GDS plays an enormous role in the fate of WWI and the defense of London. As aircraft detection, it has been used to determine the location of zeppelins in lousy weather conditions. During WWII, it is the leading role to detect artillery range. GDS technology has been used by Germany, UK, Japan, during both world wars. Later GDS has been used in the Korean, Vietnam war by US forces.

Fig 3: *GDS during WWII, the Swedish army(left) and the Finnish army(right) [3]*

Currently, GDS has been used by economically stable countries. US, Germany, UK, and several world leader countries make use of GDS on their military operations and alongside this, improve and test the quality of the sensors. US armed forces will assist to their military forces in Afghanistan with GDS. It is estimated to be equipped around 10 thousand US soldiers with GDS sensors. The primary purpose of GDS is to protect soldiers and increase their security on the incident, hotline areas while giving soldiers important information about the insurgent forces in their small screens. Nowadays, the sniper area localization is the main point of the military. Due to this reason, The Sniper location determination system is also famous for today's researches.

2.3 Economic approach and development

As an economic aspect, investment for GDS has a significant influence on the world industry.GDS plays an essential role in the World Economy and Global Industry. Due to weapon utilization and gun reserves, the US stays in first position comparing to the rest of the world. This difference makes the US number one importer of the GDS system in the world.

Besides North America, Europe is also allocating space for investigation of GDS. Several developed European countries step on the researching and developing GDS. There is the Czech Republic among these countries. The Czech Republic keeps its interest in the improvement of GDS utilizations, and it is estimated to be applied in the near future. Nowadays, investigations of Czech researchers on signal processing and investment of the government on GDS are developing. Besides the US and the Czech Republic, the Global industry of GDS segmented into regions as below.

- North America (US, Canada, Mexico)
- Europe (UK, Germany, Russia, France)
- South America (Brazil, Argentina)
- Asia/Pacific (China, India, Japan, South Korea)
- Middle East & Africa (Egypt, Saudi Arabia, UAE, South Africa)

3. Acoustical characterizations

The first step of analysing and classifying gunshot detection is about to understand some essential characteristics. There are principles and significance that cause during the physical process such as reflection, diffraction, diffusion, and intake of the acoustic sounds from the solid, nearby objects. These factors can include some valuable information that is used to improve the detection capability of the sensor. Alongside this, the caliber and shape of the bullet cause different effects on the recognition of gunshot.

Pure signal of a gunshot on time domain.

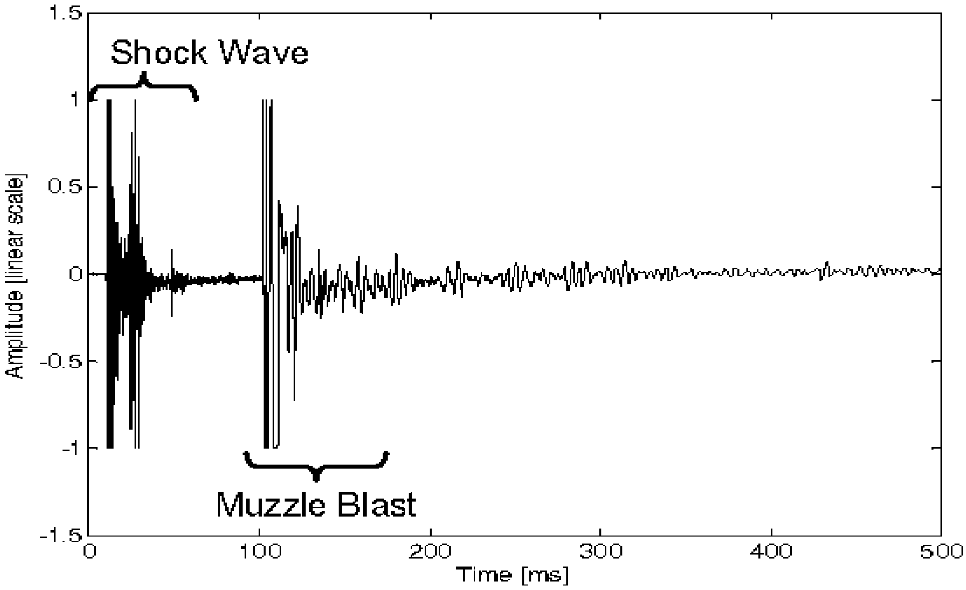


Fig 4: *Gunshot recording on time domain [4]*

3.1 Sources of acoustic data

3.1.1 Muzzle blast

When the projectile comes out of the barrel of the gun, it generates an explosive charge. This charge causes hot, high-pressured gases to expand as acoustic energy from the center of the barrel, and the outcome of this process is known as **muzzle blast**. During the procedure of muzzle blast, the majority parts of acoustic energy travel in the direction of where the barrel of the gun pointed**.** This acoustic event takes approximately 3-5 milliseconds and spreads in the air with the speed of sound (c=343 m/s at 20 °C). When sound radiates in the sky, it is contacted and reflected by various solid objects, the surface of the ground and in the between, the condition of air and more other factors that can influence the sound of a gunshot. If there is a small distance between the microphone and gunshot detector and the location of the explosion (gun), the muzzle blast of the shot will be recorded directly as a clear acoustic signal. If the distance is longer, a reached signal will display some propagation effect.

One of the factors that determine the speed of the muzzle blast is temperature. From the formula (1) it is clear, temperature degree is directly proportional with the speed of velocity:

$$c = c_0\sqrt{1 + \frac{T}{273}} \qquad (1)$$

c_0 (331 m/s) is the instant velocity of sound for 0℃, and due to calculation, it is estimated to be c = 351 m/s for 35℃. One of the components of the muzzle blast is its potential gainable pressure during acoustic radiation. So acoustic energy propagates in all directions, but its reachable peak pressure of sound is its maximum in the trajectory of the bullet. During the process, there is an attenuation in the rear acoustic sound pressure, and it is respected due to the position of the shooter's body and the characteristics of the gun. Respectively to this characteristic, the formula of angular-peak overpressure is the function of azimuth angle-φ based on line-of-fire.

The source of sound pressure decreases with the exponential function while time passing, consequently muzzle blast is going to be very sensitive towards the background noise. Muzzle blast sensitivity based on the current flow characteristics of the source. Besides, the muzzle blast can be characterized by the equation of Friedlander. Friedlander equation based on the time dependence of pressure with variables such as atmospheric pressure and peak overpressure created by muzzle explosion.

$$p(t) = P_0 + P_S(1 - \frac{t}{T_0})e^{-bt/T_0} \quad (2)$$

T_0- is the positive phase duration, b is the rate of exponential decay. When Fourier transform applied Friedlander equation (2), it is the final analytic spectrum will be given as:

$$P(w) = \frac{P_S((b-1)+jwT_0}{T_0(\frac{b}{T_0}+jw)^2} \quad (3)$$

w- is angular frequency, *j* is an imaginary unit. Magnitude spectrum of Friedlander final equation is given:

$$|P(w)| = \frac{P_S}{T_0}\frac{\sqrt{(b-1)^2+{T_0}^2w^2}}{\frac{b^2}{{T_0}^2}+w^2} \quad (4)$$

Magnitude spectrum at formula (4) has a peak (w_{max}):

$$w_{max} = \frac{1}{T_0}\sqrt{-b^2+4b-2} \quad (5)$$

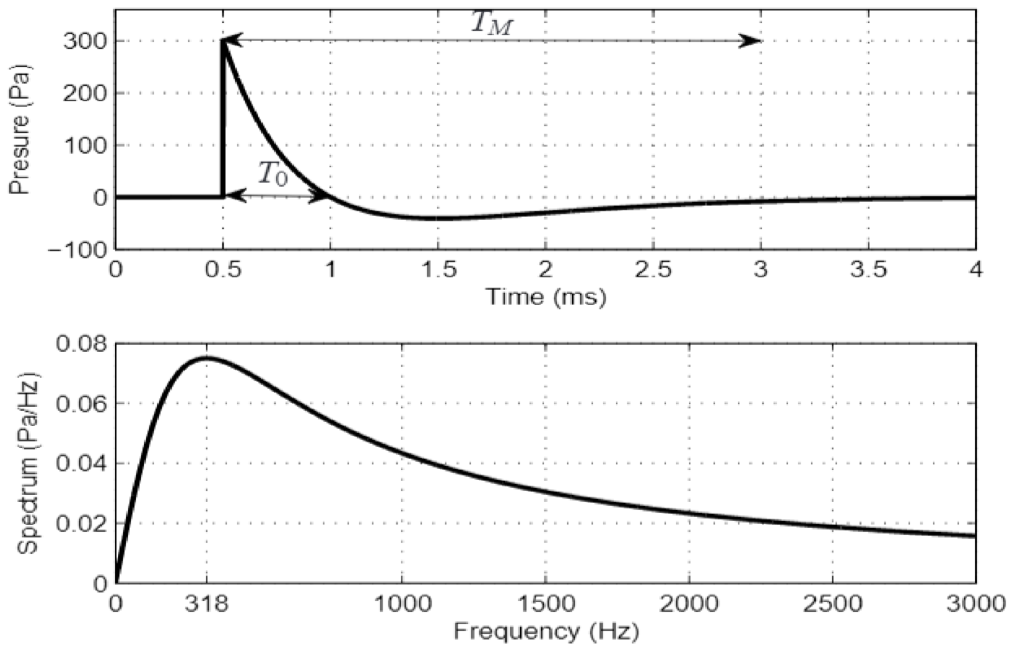


Fig 5: *Time and Frequency domain of typical small muzzle blast based on the Friedlander's equation [4]*

Fig 5 describes Friedlander's model for the frequency and ideal-time spectrum of the muzzle blast. In this example, the positive phase duration was assigned to 0.5 ms, the rate of exponential delay

b was set to 1, and the peak overpressure was assigned to 300 Pa (144 dB). So obtained result (T_M) in blast duration is 3 ms, which is related to the ideal small muzzle blast duration.

Muzzle blast is not such a truthful source for the researching of the signal. Because Muzzle blast is mostly propagated in the direction where the barrel of a firearm pointed, and it is defined as weaker at all other courses. So, it means, if there are solid surfaces, obstacles between barrel and sensor, it will decrease the efficiency of muzzle blast for determination of vital information by the microphone of GDS sensors. Some suppressors could supply some firearms and weapons. The primary role of the suppressor is to decrease the Muzzle blast effect, and due to this reason, it makes harder the GDS to detect the location of the gunshot. Thereby muzzle blast detection system on GDS should consider the utilization of suppressors. Muzzle blast is usually applied for the recognition of sniper location as military purpose.

3.1.2 Supersonic Projectile: Shockwave

Another source of investigation for the acoustic event is a supersonic projectile that bullet travels with the speed of sound. This acoustic motion appears on the bullet with the supersonic speed that generates a shockwave effect, which is propagating on the conic fashion behind the bullet trajectory (Fig5). Shockwave is based on the combinations of two different types of shockwaves: compression and expansion shock. This combination of two shockwaves is referred to as **ballistic shockwave**. Ballistic shockwave is based on the theory of a sudden rise of acoustic pressure.

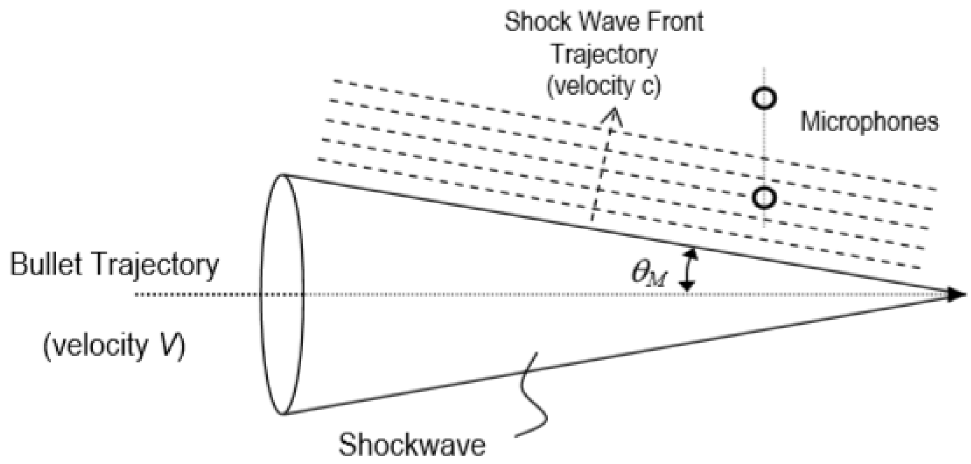


Fig 6: *Scheme of Supersonic bullet wave [5]*

In the supersonic projectile process, there is an inner angle between the bullet trajectory and shockwave propagation. This angle is defined as θ_M:

$$\theta_M = \arcsin\left(\frac{1}{M}\right) \tag{6}$$

M- is Mach number, which is defined with the ratio between sound and bullet velocity. The central role of the utilization of Mach number is in some disciplines such as aerodynamics:

$$M = \frac{V}{c} \tag{7}$$

$$V - velocity\ of\ bullet$$

$$c - velocity\ of\ sound$$

From the formulas (6) and (7), it is observed that the inner angle depends on the ratio of Mach number. While the velocity of the bullet increasing, the value of the Mach number increases, and the inner angle decreases. It creates a perpendicular angle between the propagation of shockwave and the trajectory of the bullet. In the inverse case, the distribution of shockwave is almost parallel to the bullet's path. If the velocity of the bullet is faster than the speed of sound, the inner angle is approaching 90°. When the Mach number defined between the interval 0.8 and 1.2, this regime is known as the transonic regime. The transonic regime is between subsonic and supersonic regimes. When Mach number is smaller than 0.8 is called subsonic, and $1.2 < M < 5$ is known as a supersonic regime. Besides, there are new regimes that have been invented by NASA: Low and High hypersonic regimes. Low hypersonic regimes ($5<M<10$) considered for the aircraft, which have a higher speed than the velocity of sound. A high hypersonic regime ($M>10$) is recognized for much higher rates, so that accounts for atmospheric re-entry speed.

Shockwave propagation direction can be defined with the assist of several GDS sensors. It is needed to determine the time arrival relations between collected data for each sensor. Besides, the determination of shockwave propagation vector from the trajectory of the bullet, bullet velocity is necessary. (2) When the bullet moves with the speed that is higher than the velocity of sound, the entropy of the air raised due to the two main components of ballistic shock.

When bullet reaches higher speed than the velocity of sound at air, there are two principal shocks that causes growth at the entropy of environment: two parts of ballistic shockwave-(1) an expansive shock wave which propagates at subsonic velocity relative to the materials, at supersonic speed corresponding to the content behind of it and compressive shock wave that spreads at supersonic velocity relative to the air , subsonic velocity relative to the sky behind of it. This case is very vital information for ballistic shockwaves.

3.1.3 Mechanical actions

Alongside sources such as muzzle blast and supersonic projectile, there are some other factors (source) that relate to the determination of gunshot. These kinds of effects depend on the mechanical, physical status of firearms, and throwing (ejection) the spent cartridges. This includes the mechanical sounds such as trigger of the gun, hammer mechanism, reloading and supplement of new ammunition for a firearm with an automated or automatic system. This kind of action is considered as quiet sounds with a comparison muzzle blast and supersonic projectile. Thus, the detection of such a variety of sources is much more complicated and is only available if the microphone is located closer to the weapons.

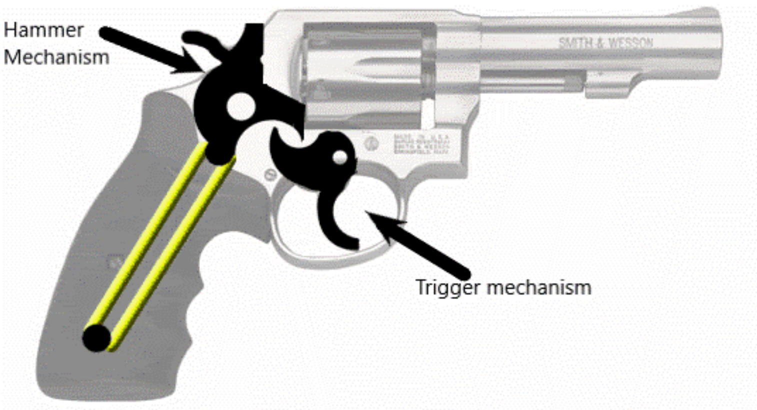


Fig 7: *Smith & Wesson revolver 67*

3.1.4 Surface vibrations

During the process of the gunshot, propagation of acoustic waves could be vibrated through the solid materials, grounds, if there are some. Sound of bomb explosions, various firearms can spread from the ground ten times faster and more prominent than air from the source of the sound. So, this kind of vibrations can be measurable with GDS. Propagation of the sound on the solid material such as rock, metallic obstacles, is five times faster than the distribution of noise on the air. For this reason, it can be considered as a suitable and productive technique for the detection of various hazardous events.

$$V = \sqrt{\frac{C_{ij}}{\rho}} \tag{8}$$

At the formula (8) it has been discovered the velocity of sound for different environment and C_{ij}- Is the elastic property (which is variable due to solids, liquids, and gases) and ρ- is the density of the environment. It is exhibited sound velocity for some materials following the table (Fig 7).

Materials	*Sound Velocity (m/s)*
Rubber	*60 m/s*
Air at 40 ^oC	*355 m/s*
Air at 20 ^oC	*343 m/s*
Lead	*1210 m/s*
Gold	*3240 m/s*

Fig 8: *Table of sound's propagations on various materials*

3.2 Multi-path propagation effect

At a multi-path propagation environment, all the stored signals are formed of direct-path, which is a realizable path if the region (route) between the receiver and source of the signal is unobstructed. Additionally, it is calculated the N-scaled delayed copies of primary direct-signal. The scaling factor depends on the frequency and determined by the surface reflection characteristics and traveling distance. In open areas, MPE poses fewer problems. It is related to the absence of objects on the neighborhood area of the Gunshot detectors. It has been proven that there always exists shockwave ground reflection of the signal.

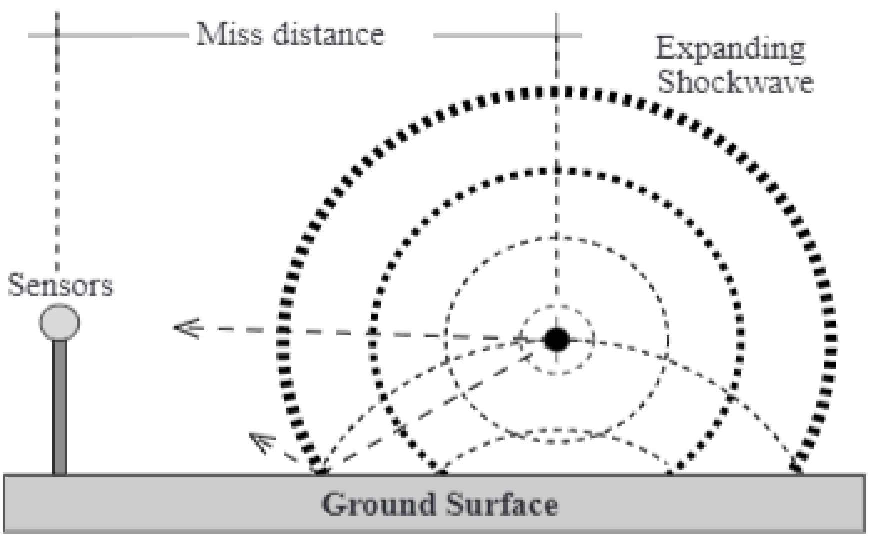


Fig 9: *Shockwave ground reflection of the signal [4]*

Alongside ground reflection, there are two main types of reflections of acoustic signals. Regular reflection's theory based on the two-shock configuration and three-shock structure (Fig10a) and Mach reflection (Fig10b). Nowadays, shock wave reflection distributed as two primary types of reflections. Regular reflections and irregular reflections. Irregular reflection consists of Mach reflection and includes some subtypes.

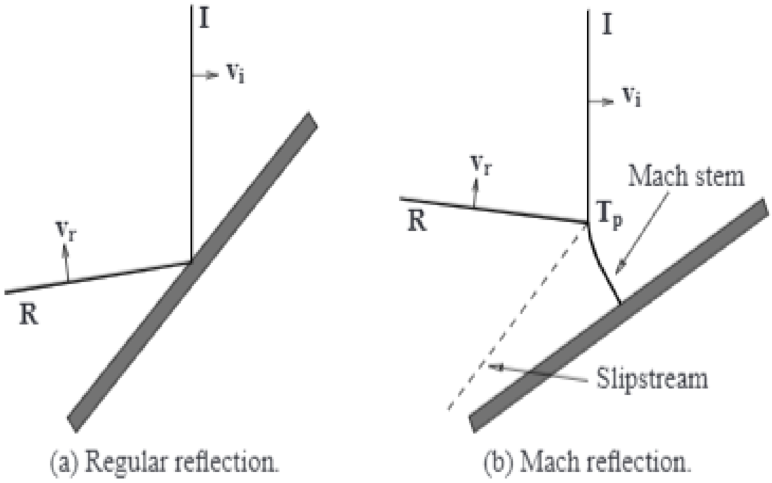


Fig 10: ***I***- *incident shock,* ***R****-reflective shock,* v_i *and* v_r*- the velocity of the incident and reflective shock.* T_p*- The triple point where three shockwaves met [4]*

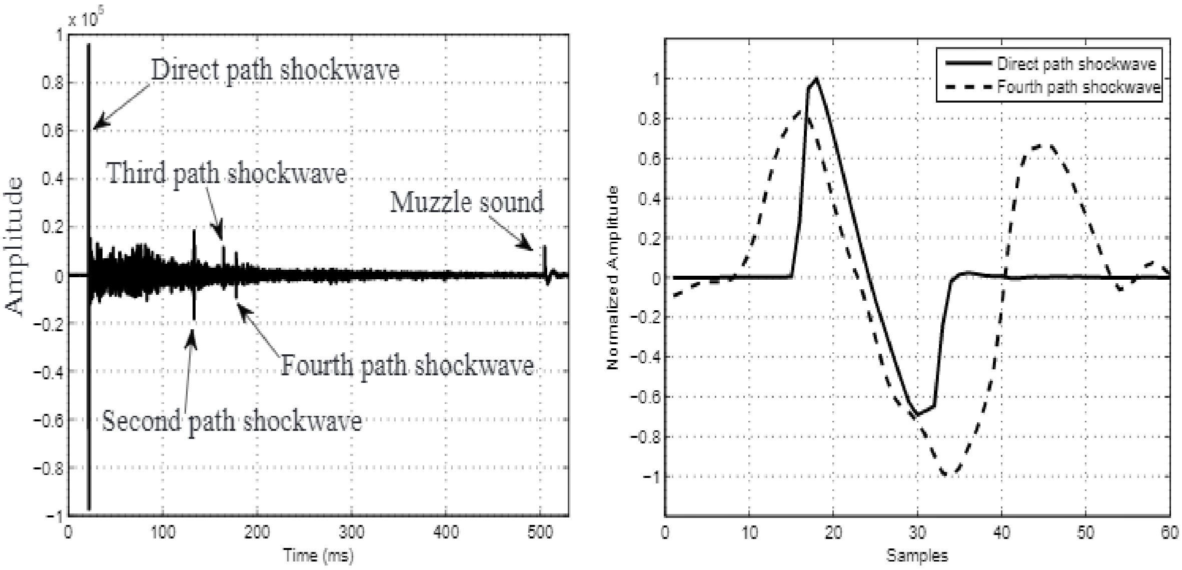


Fig 11: *Recording of gunshot due to reflections on the multi-path environment(left) and direct/fourth path reflection of signals(right). [4]*

In Fig 10, the gunshot signal is recorded by GDS in the multi-path environment. The bullet used on the experiment has 12.7 mm length from the distance of the sensor is approximately 62 meters. In Fig 11, the main shockwave has a higher amplitude comparing to its reflections, which are followed to the direct path signal before the muzzle blast. In this experiment, the exponential factor has been used to scale for better visualization. On the right side of Fig 10, it is discovered the comparison of direct and fourth reflections of shockwave reflection. There is some various version of factor that has been used to avoid muzzle blast and shockwave for determination of them.

In this part of the project, we will take account of some acoustic characteristics of signal processing for our thesis. Application of MATLAB and detection of gunshot require more information based on the muzzle blast and shockwave of strange events. Some other specifications that are mainly related to AED such as ground reflections, surface vibrations, mechanical actions (trigger and hammer mechanism) will not be compulsory for further investigation of gunshot detection.

4. Methods

In AED, the acoustic characteristic of a signal is one of the factors that must get an account for analyzing signals. Besides its characteristic, an algorithm that used to detect gunshots remains one of the fundamental bases of the GDS. Due to algorithm differences, several methods are used to identify shots such as the Zero Phase Technique, Cross-correlation, FFT, etc. In this part of the thesis, there are two various methods chosen for the experiment of GDS. Fast Fourier Transform (FFT) and Mel Frequency Cepstral method (MFC).

4.1 Fast Fourier Transform (FFT)

4.1.1 DFT, Basics of FFT.

DFT is a form of Fourier transform that converts finite sequences of data from the original (time, space) domain into the frequency domain and vice versa (IDFT). Formula (9) and (10) are the base explanation of DFT and IDFT.

$$X(k) = \sum_{n=0}^{N-1} x(n) * e^{-j2\pi k(\frac{n}{N})} \tag{9}$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) * e^{j2\pi k(\frac{n}{N})} \tag{10}$$

DFT is used in many applications, and it can be directly implemented. But computing DFT directly from the definition is not convenient, and it will cause some problems for the speed of the process. Therefore, FFT has been developed for the utilization of DFT, and it is one of the widely used algorithms that has been applied on signal theory to computes DFT of sequences. FFT is involved in many fields of science and is used significantly on the radar, sonar, and software based on radio applications. GDS is one of these fields that uses the FFT algorithm to detect hazardous actions. Due to its extensive range, FFT has been used in many applications such as engineering, mathematics, music, science, and s.

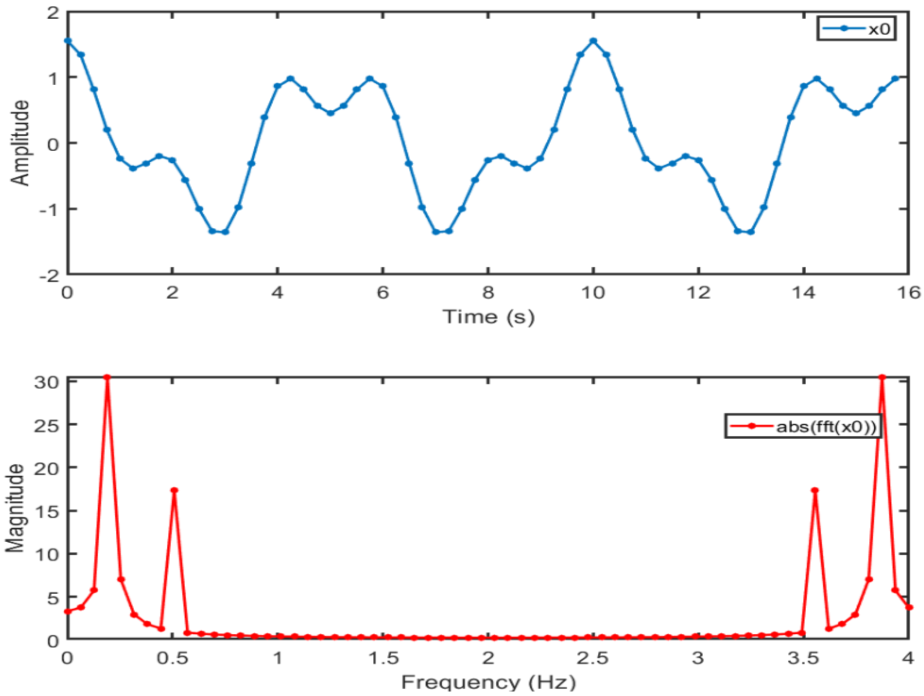


Fig 12: *Conversion of signal from a time domain into a frequency domain [6]*

FFT's central principle of computing of DFT is to select 2-point and 4-point DFT and combines them to generate 8-point,16-point, etc. Another difference between DFT and FFT algorithm is when we compute the DFT it will take O (N^2) multiplications and additions.

4.2 MEL Frequency Cepstral (MFC)

4.2.1 Mel-frequency scale

The human auditory system is not considered for the pitch linearly. Pitch is linearly increasing with the frequency. It causes several problems and is not needed in some applications. Due to this behaving of the pitch, Mel-scale is developed.

MFC is usually based on the detection of speech characterizations. Apart from this, MFC is used on other applications and analyzing signals generated by different scientific fields such as Machine engines, gunshot detections. Based on our topic, it has similar aspects for GDS applications. MFC is widely used for the detection and classification of gunshots and still one of the techniques that are developing in GDS aspects.

Mel-scale is one of the components of the MFC technique. Mel-scale is a kind of perceptual scale of pitch that is based on the equal distance one another for listeners. Formula (10) below can be described as one of the approximated formulas for converting the signals from the frequency domain to the mel-frequency domain. Cause there are several ways of revealing the formula for converting pitch.

$$F_{mel} = \frac{1000}{\log{(2)}} * (1 + \frac{F_{hz}}{1000}) \tag{10}$$

F_{mel}-measured frequency on Mel-scale in mels, F_{Hz} is the normal frequency recorded in Hz.

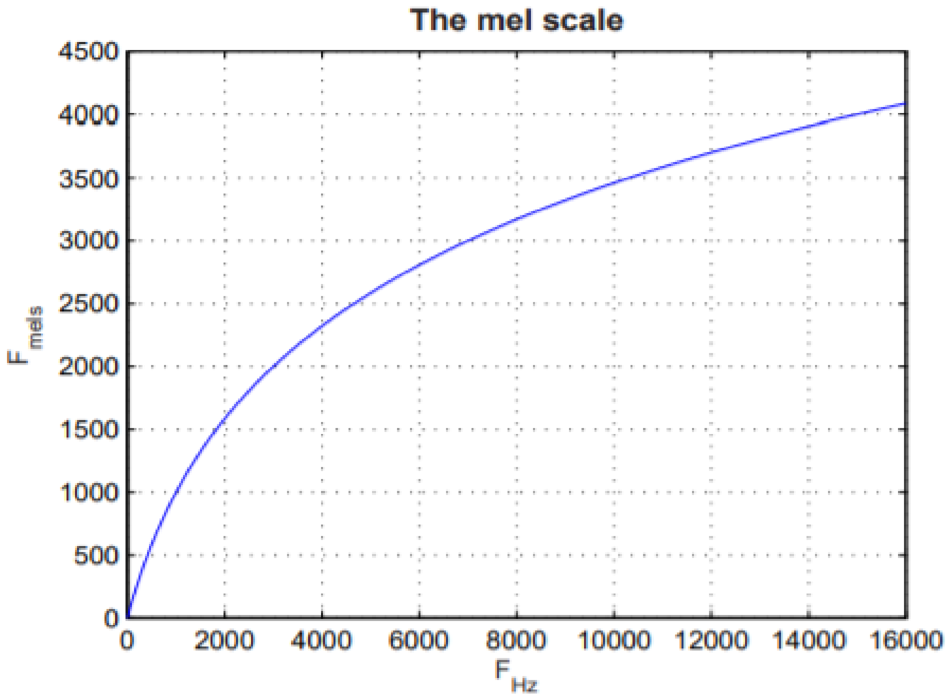


Fig 13: *Mel-scale & frequency scale relationship [7]*

Mel-frequency's main application based on the characteristic of the human auditory system on a linear scale. From the **Fig12**, the pitch is linear in the frequency between 0-1000 Hz. Higher frequencies than 1000 Hz, its scale considered as logarithmic.

4.2.2 Mel frequency cepstral coefficient (MFCC)

The essential point about speech is the human articulatory system generates the sounds. During the procedure of speech, it has been filtered by the organs that are part of the articulatory system, such as vocal tract, teeth, and tongue. Depends on the shape of them, characteristic of sound varies. The shape of the vocal tract displays itself in the window (envelope) of the short time power spectrum. The purpose of MFCC is to the identification of automatic speech.

The utilization of the filterbank usually derives MFCC. It came to light that the human auditory system has a connection between the energy in the critical band of a specific frequency. Depends on the value of the rate, a bandwidth of critical band varies. It is linear below 1 kHz and is logarithmic function above the 1 kHz. The critical band is defined as such a band filter that is regulated around the center of frequency. The calculation for the total energy of MFCC for each critical band is interpreted on formula (11).

$$Y(i) = \sum_{k=0}^{N/2} \log(s|n|) * H_i(k * \frac{2\pi}{N'}) \quad (11)$$

Y(i) is the total energy in the critical band, s(n) is a DFT signal which has been calculated according to MFCC. $H_i()$ is a critical band filter for the I coefficient and N' is the number of points that have been used in filter for short-term DFT signal.

We calculate each critical band separately as formula (12).

$$\tilde{Y}(k) = \begin{cases} Y(i), \ k = k_i \\ 0, \ other\ k \in [0, N' - 1] \end{cases} \quad (12)$$

After this procedure continues with the deriving from the IDFT of formula (12) and the final cepstrum can be expressed as on equation (13).

$$c_s(n) = \frac{1}{N'} \sum_{k=0}^{N'-1} \tilde{Y}(k) e^{jk\left(\frac{2\pi}{N'}\right)n} \quad (13)$$

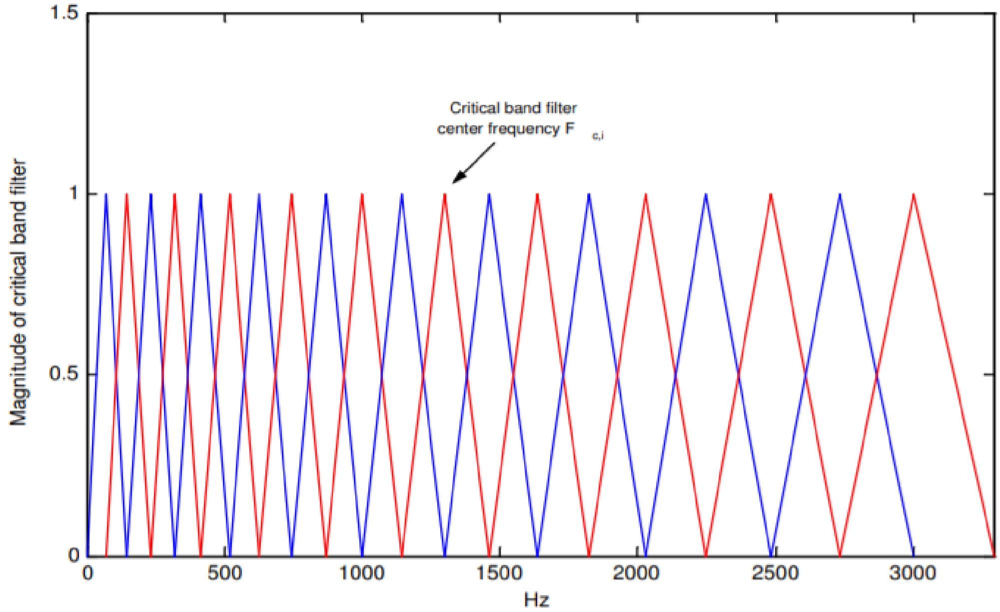


Fig 14: *Mel scale filterbank of 20 filters, center of frequency is considered on each peak of the critical band [7]*

4.3 Computation of FFT methods

In this part of the thesis, we will discuss the various methods that have been used with the application of FFT on MATLAB. At first, we recorded several signals, and these signals distributed into two groups: true and false alarm signals. True signals contain real gunshot signals, and, in our experiment, we used three different pistols types (caliber): 9 mm (Luger short gun), .22 inches and 7.62 mm (Tokarev short gun). The main point of choosing the three types of guns is to examine the characteristics of gunshot signals deeply and discover similarities. Alongside with gunshots, we used several signals that caused by household applications. We used the glass breaking, plastic bubble snap, door slam, book slam, and hand clap signals to differentiate the true and false alarm signals. We used different samples of signals according to each type. The main reason is to decrease the uncertainty of methods and approach the signal characterizations. Totally, we have 40 signal samples from each 8 types of data: 4 samples of 0.22 inch, 4 samples of 7.62 mm Tokarev, 4 samples of 9 mm Luger, 4 samples of book slams, 2 samples of handclap, 6 samples of door slam, 8 samples of glass breaking and 8 samples of plastic bubble. As the representation of spectrums, we chose one of the samples from each signal to display on our thesis.

Primarily, signals were computed and revealed the exact time spent on recording for each signal. In our case, it is crucial to know the sampling frequency (F_s), which is equal to 44100 Hz. It is computed the total time of the signal with the help of the ratio between the number of samples (N_s) and F_s (44100 Hz) on formula (14).

$$t_{total} = \frac{N_s}{Fs} \tag{14}$$

Signals	*Number of Samples*	*Time [Second]*
Plastic bubble	*184321*	*4.2*
Door slam	*222208*	*5.1*
Book slam	*196097*	*4.5*
Handclap	*87552*	*2*
Glass breaking	*185857*	*4.3*
Luger-9 mm	*497153*	*12*
Tokarev-7.62 mm	*190792*	*4,4*
.22 inch	*390000*	*9*

Fig15: *Table of the duration of each signal*

From the obtained data above, it is evident that the all-time for each signal measurement is different. Usually, not all the parts of the signal are necessary for further investigations. In this time, we can neglect some unused parts (samples) of signal and equalize all the signals into the same sample (it is convenient to balance all signals into 2 seconds). For this, we must define the sample in which the function reaches its maximum. We determined the maximum amplitude of the signal on MATLAB with a function of *max.* we used two parameters for defining the maximum point, its value *M*-maximum and its *I*-index. An index is necessary for identifying the vital parts of the signal and removing the unwanted parts of the signal from both (left & right) approach. (which part of the signal to keep).

[M, I] =max (Signal)

First border = I – Samples

Last Border = I + Samples

I- is the index of the sample, which obtained the maximum *M*-amplitude. We put the border to the *I*-index and define the first and last samples as the border of the total new designed signal. In this way, we removed the not essential parts of the signal and minimized them as much as possible. As a result, we decrease all the signals into 44000 samples, which is roughly equal to 1 second.

In signal processing, filters are used to neglect some specified range of frequency from the signals. Mainly filters are used to design some signals which are allowed at for specific frequency and frequency bands. There are many types of filter. For signal processing, digital filters are used. In our case, we will use the Low Pass Filter (LPF) to remove the unwanted high frequencies.

Some spectrum of the various filters:

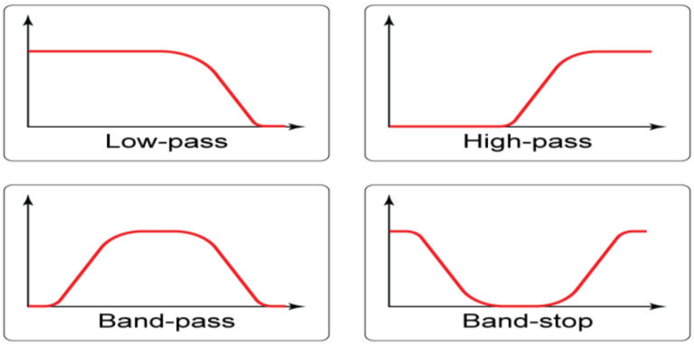


Fig16: *Different version of filter [8]*

Low Pass Filter allows the frequencies that are lower than selected cut-off frequencies. As its working principle, low frequencies go through the system with less resistance, while higher frequencies encounter much bigger resistance. LPF is constructed by using some fundamental components of physics: resistors, capacitors, or inductors. LPF which is contained the components: the capacitor and resistor are known as RC while the filter maintains inductor, and a resistor is called RL filter. LPF can be called a high-cut filter or treble-cut filter in the acoustic field.

The procedure of designing is displayed below:

Filter designer (MATLAB)

1. (FIR)-Design method is chosen as a window: type-Blackman Harris.

2. Filter order has been selected in an interval between 100-200 (150). Between [Page 12, Fig5]. As a result, if we take a bigger order than 200, it causes a delay in ‘real time.’ We chose the 150 as filter order.

3. Sampling frequency, it should match with signals frequency which is 44100 Hz

4. Cut off frequency is determined as 2500 Hz (due to the relation between spectrum & frequency on Fig5. Frequency higher than 2500 Hz has linear dependence)

A low-pass filter is designed as on the Fig16:

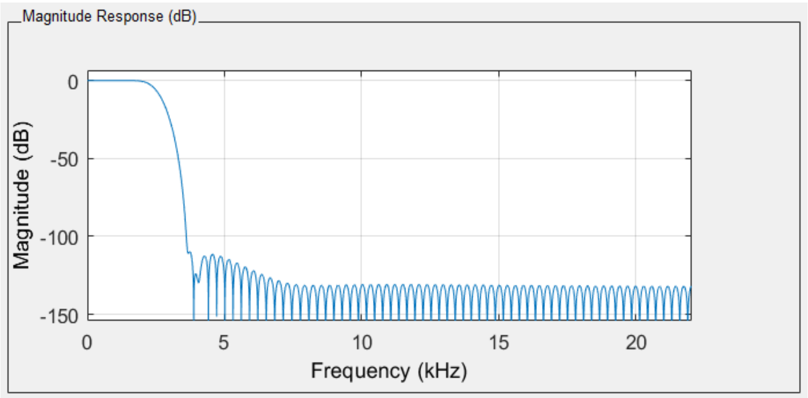


Fig17: *Designed Blackman-Harris Low pass filter*

Afterward, we applied our low-pass filter into each signal. Therefore, we removed the saturation and unnecessary frequency bands from each data. We converted the domain from samples to time, and the outcome of all filtered signals has been described in Fig18 and Fig 19.

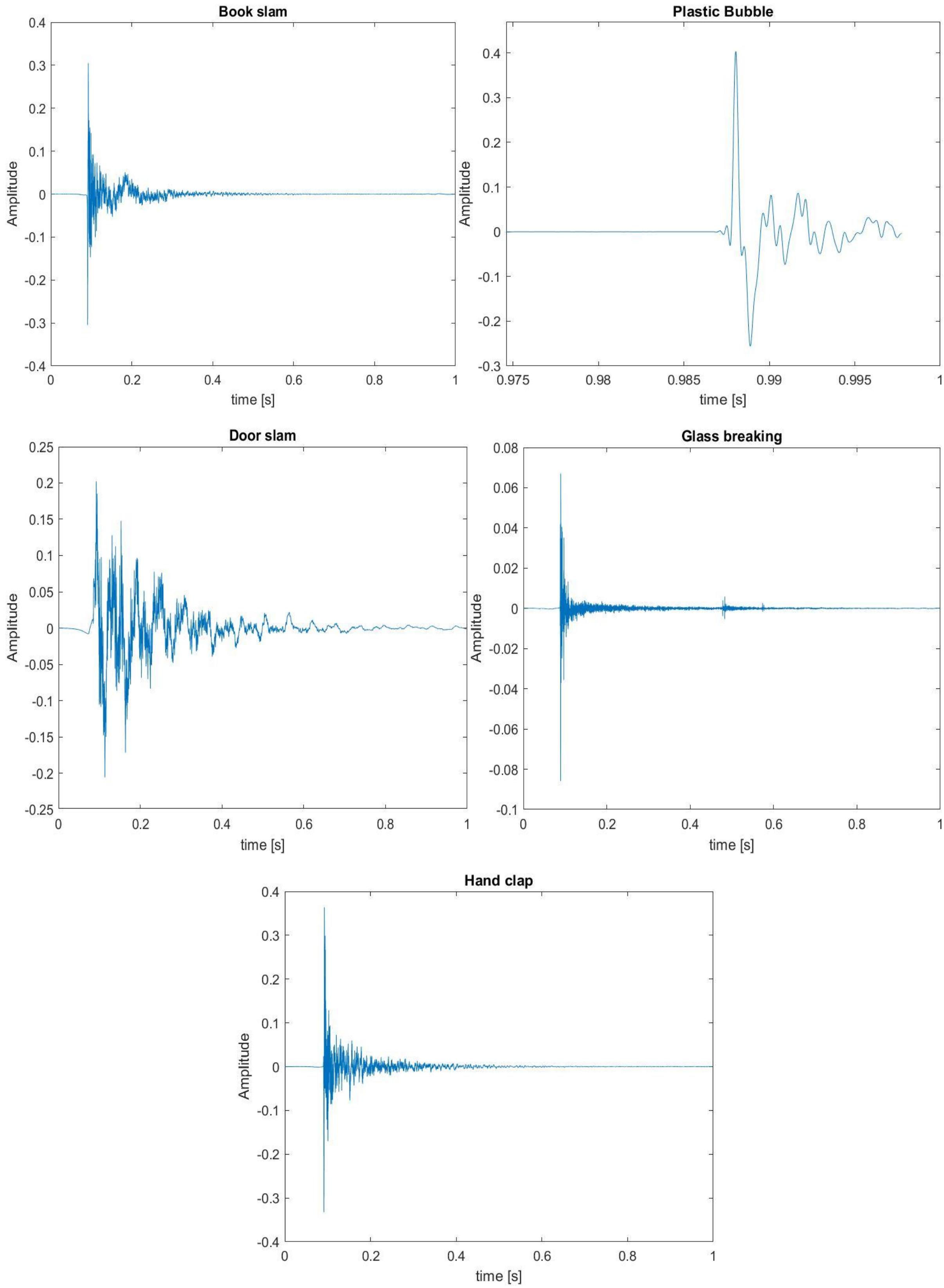


Fig 18a-e: *Spectrum of none-gunshot signals (Book slam, Plastic Bubble, Door slam, Glass breaking, Handclap)*

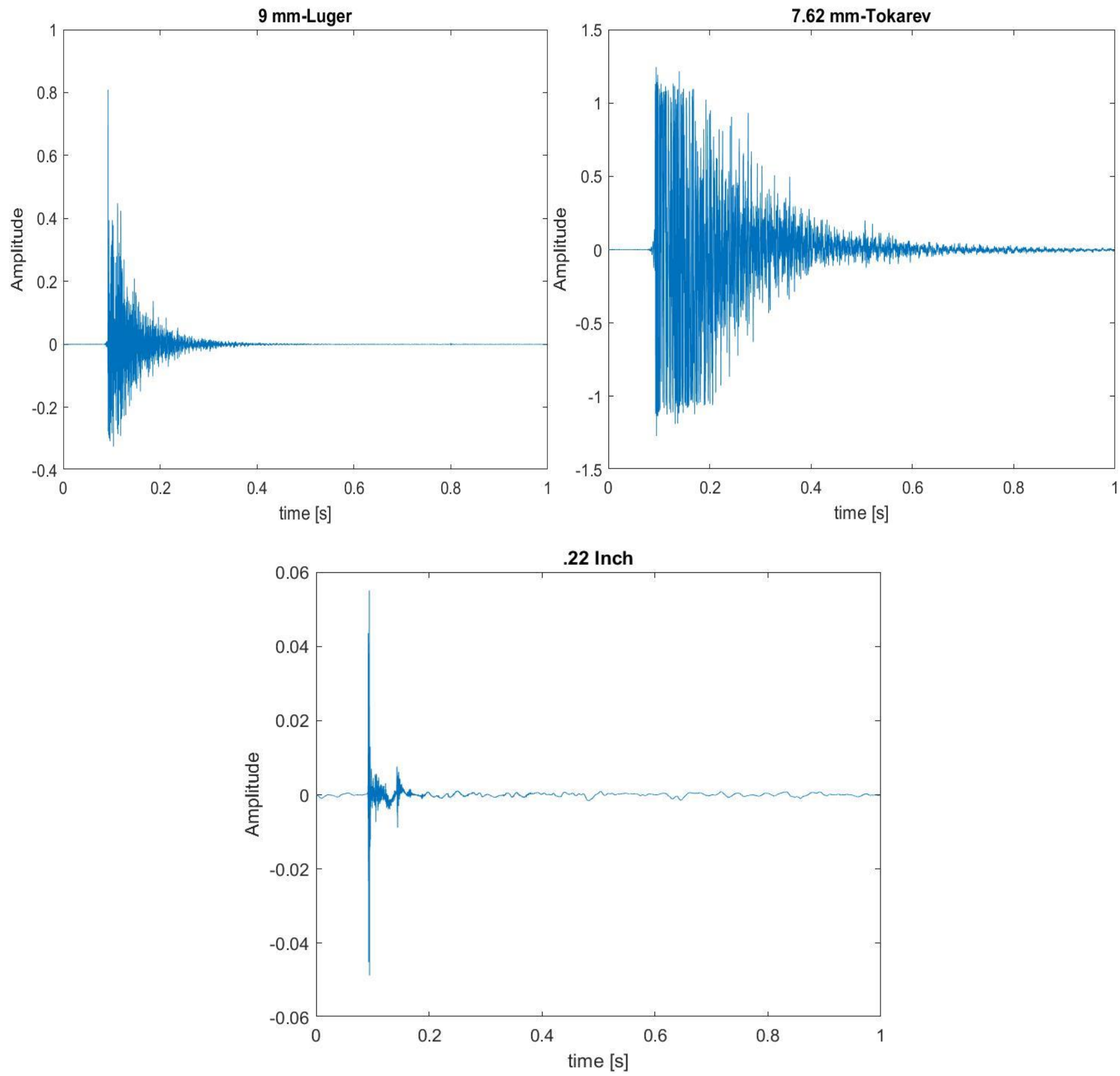


Fig 19a-b: *Spectrum of gunshot signals (9 mm-Luger, 7.62 mm-Tokarev, .22 Inch)*

Filtered signals are prepared for further processes. As the first step, it is required to convert signals into frequency domain because we must examine the frequency interval of signals. According to this reason we will use transforms to analyze signals on the frequency domain. We will apply the Fast Fourier transform.

In this part of the thesis, it will be discussed and analyzed different methods for signals. Zero crossing method, Peak-Valley difference, frequencies that contain the first 100 maximum amplitude and PSD will be checked on the further investigations.

4.3.1 Simple FFT method

The first method for signal analyzing is FFT. It is a common way of characterizing the signals. we filtered all signals and displayed them on amplitude dependence on the frequency domain with FFT algorithm. Due to the long range of frequency domain, we cut the parts of frequencies till 4 kHz. Frequencies are bigger than 4kHz, amplitude close to 0. In the second step, the FFT method displayed the various frequencies that all signal contains. As seen from Fig19, Gunshot signals have more intensity of frequencies on the middle rate while none-gunshot signals contain small frequencies.

As of the last procedure, we should normalize amplitudes for each signal. Because, when a microphone of the sensor has measured the signals, every signal had various distance from the actual sensor, and it can affect the result of amplitudes. Normalize function sets all signal amplitudes in the same scale. This way, we could approach the characterizations and analyze them better. We normalized the amplitude scale in interval 0-1 with the help of command: *Normalize (Signal, "range").*

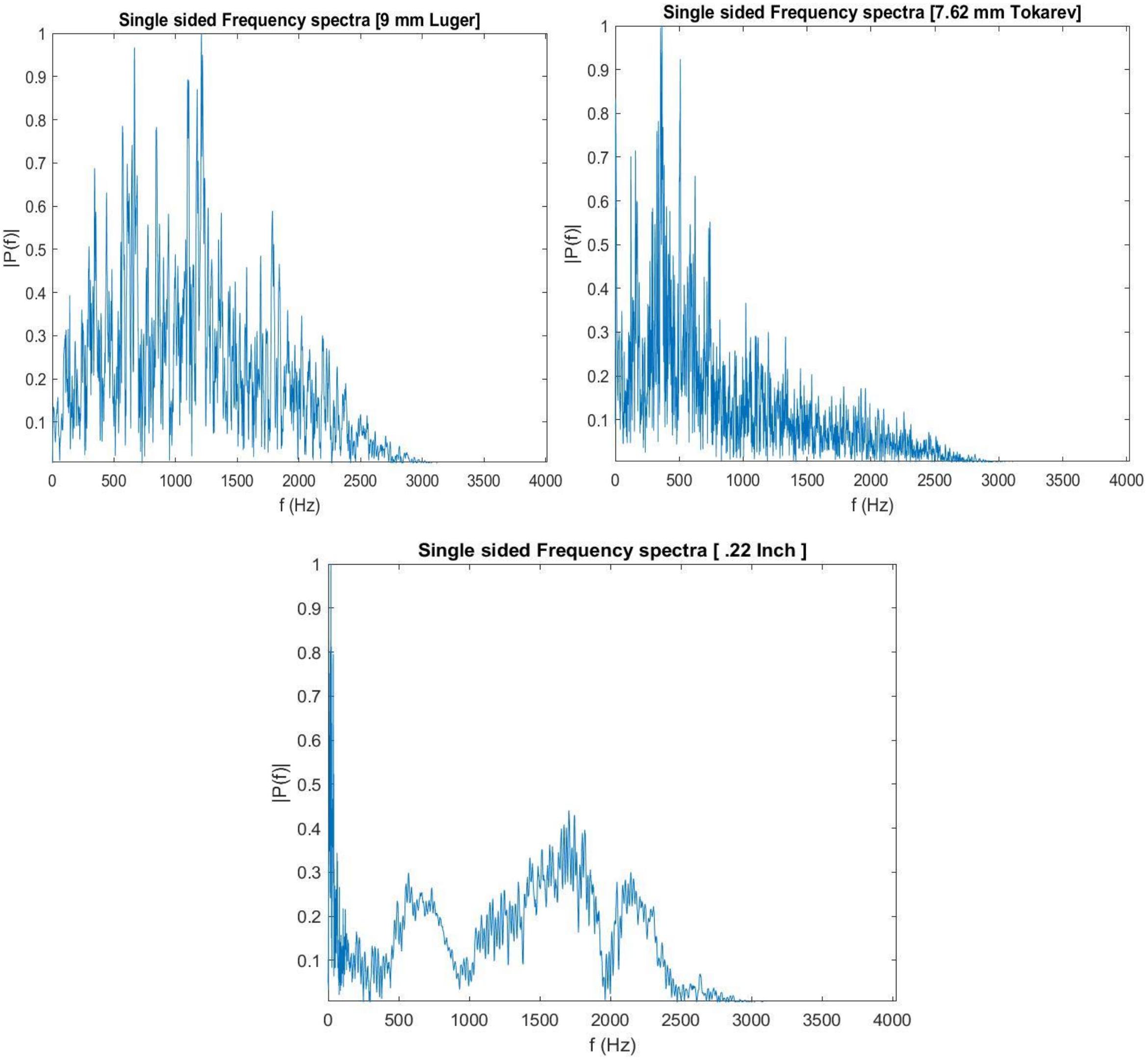


Fig 20a-c: *Normalized amplitude: Frequency spectrum of gunshot signals (9 mm-Luger, 7.62 mm-Tokarev, .22 Inch)*

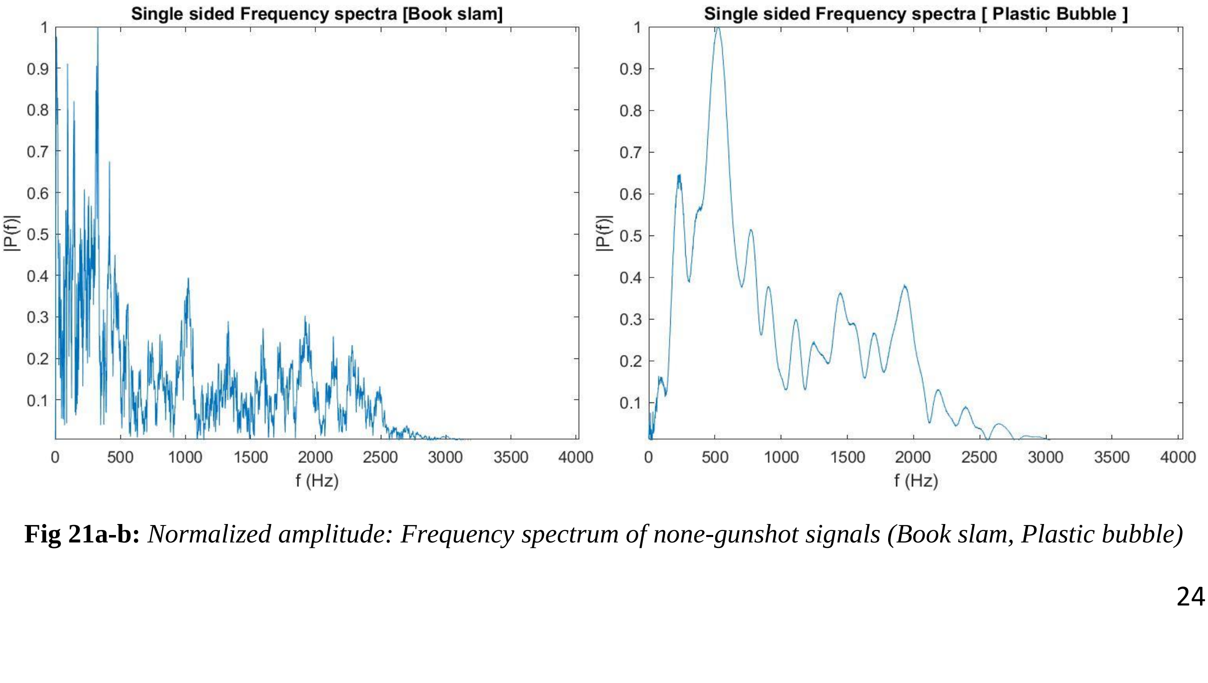


Fig 21a-b: *Normalized amplitude: Frequency spectrum of none-gunshot signals (Book slam, Plastic bubble)*

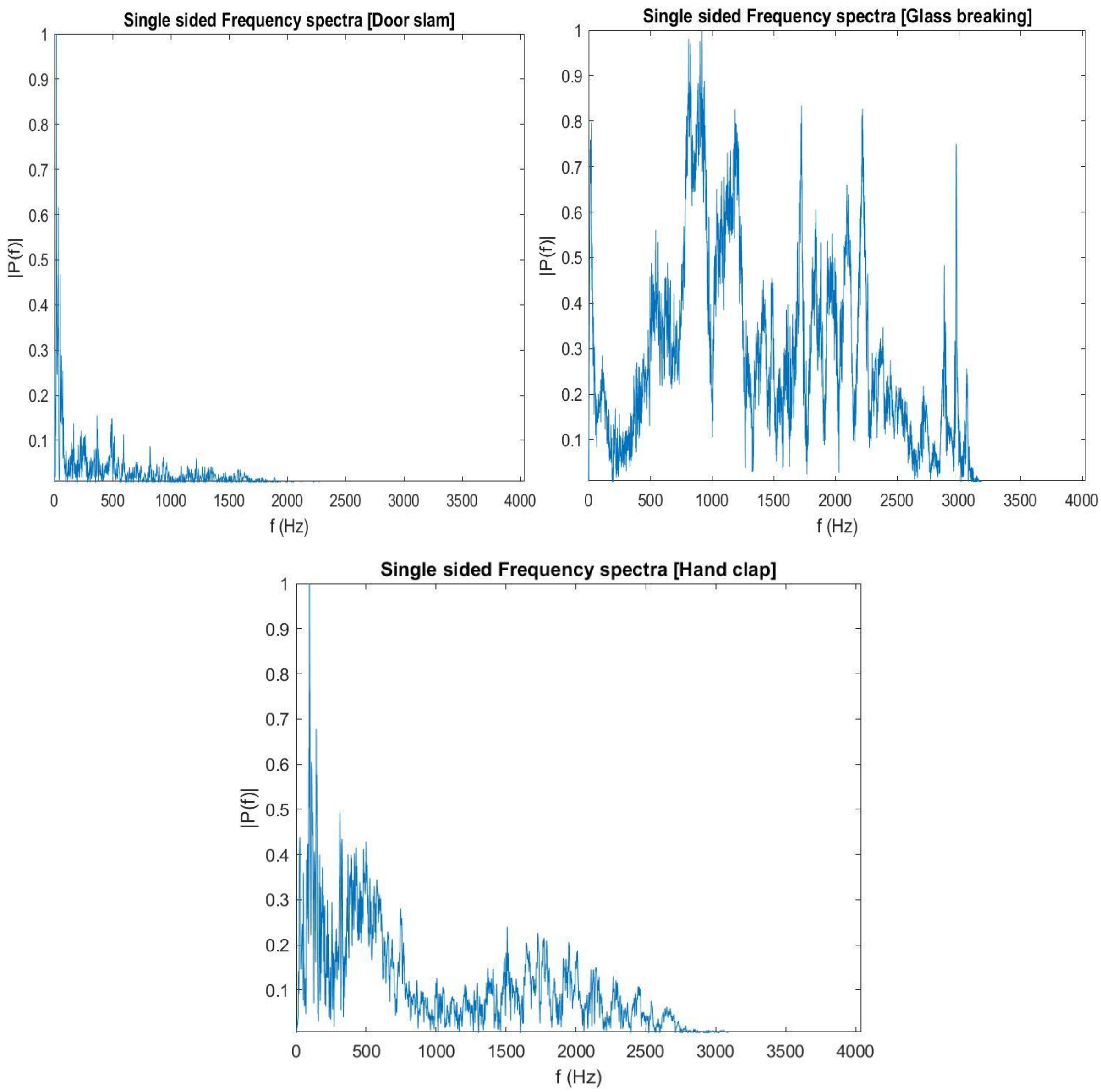


Fig 21d-e: *Normalized amplitude: Frequency spectrum of none-gunshot signals (door slam, glass breaking, handclap)*

The results of the FFT application are displayed in Fig 20 and 21. It is exposed to the outcome that the goal of the thesis is not provided as estimated. it is hard to discover the distinctions from the frequency spectrum of the signals. If considered as the first step, the FFT method is not enough to compute the similarities and differences between the signals due to the sampled frequency of the power spectrum. It is required to use some more process on the signal. According to the result, it is required to use further methods afterward for the detection of the gunshot characterizations of the signal.

4.3.2 Power Spectral Density (using FFT)

Power Spectral Density of periodic and random signals is the main factor that is used in signal processing. As a primary purpose of spectral estimation, the speech recognition problem uses spectrum analysis as a preliminary measurement to perform speech bandwidth reduction and further acoustic processing. Power Spectral Density (PSD) is a dependence between the signal's power and related frequency. PSD is used to describe broadband signals. During the application of PSD, attentive normalization should provide that the power calculated over time should be the same as the one that would be acquired from the integration of the spectrum.

Some of the signals meet the problem that their future behaving (variation) can't be predicted precisely. The only solution is to do some probabilistic experiments about the changes in the signal. The mathematical gadget that used to define these types of signals is **random sequences**. It consists of several possible realizations that each of them has its own specific associated probability for occurrence. In conclusion, researchers can only observe one realization among all the possibilities. Later it could be expected that the deterministic explanations of preceding sections could be kept as the unmodified to the present case. Nevertheless, according to the random signals viewed as DTFT sequences, they do not have finite energy, it is not feasible. Besides that, random signals have limited average power, and because of that, it can be analyzed by power spectral density. There are two main versions of the description of PSD.

1. The first way is to define the PSD as DTFT of the covariance sequences. PSD's usual formula can be expressed as (15).

$$\emptyset(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-i\omega k} \tag{15}$$

$r(k)$-is the autocovariance sequences, as an inverse transform $r(k)$ reestablished from the formula 16.

$$r(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \emptyset(\omega)\, e^{i\omega k} \tag{16}$$

when *k* is defined to 0, formula (16) can be written as:

$$r(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \emptyset(\omega) d\omega \tag{17}$$

Since *r* (0) = $E\{|y(t)|^2\}$ indicated the average power of *y(t)*, *the* formula shows as $\emptyset(\omega)$ can be defined as PSD.

2. The second definition can be displayed like formula 18.

$$\emptyset(\omega) = \lim_{N\to\infty} E\{\tfrac{1}{N} | \sum_{k=1}^{N} y(t) e^{-i\omega t} |\}^2\} \tag{18}$$

This equation is equivalent to the formula (15). It is under mild assumptions. The equivalence between formula (15) and (18) expressed below (19).

$$\begin{aligned} \lim_{N\to\infty} E\left\{\frac{1}{N} | \sum_{k=1}^{N} y(t) e^{-i\omega t} |\}^2\right\} &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} E\{y(t) y^*(s)\}\, e^{-i\omega(t-s)} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} (N - |\tau|) r(\tau) e^{-i\omega t} \\ &= \sum_{\tau=-\infty}^{\infty} r(\tau) e^{-i\omega t} - \lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} (\tau) r(\tau) e^{-i\omega t} = \emptyset(\omega) \end{aligned} \tag{19}$$

The procedure of PSD estimating using the FFT algorithm:

1. The first step, FFT is applied to find out the amplitude spectrum of the signals: Y=FFT(X)

2. Afterward, we computed the square of the absolute value of the amplitude spectrum, and due to density, we divide the whole spectrum with the multiplication of samples and sampling frequency. It gives us the power spectral density.

3. For a decibel scale: we applied 10log10 multiplication to PSD.

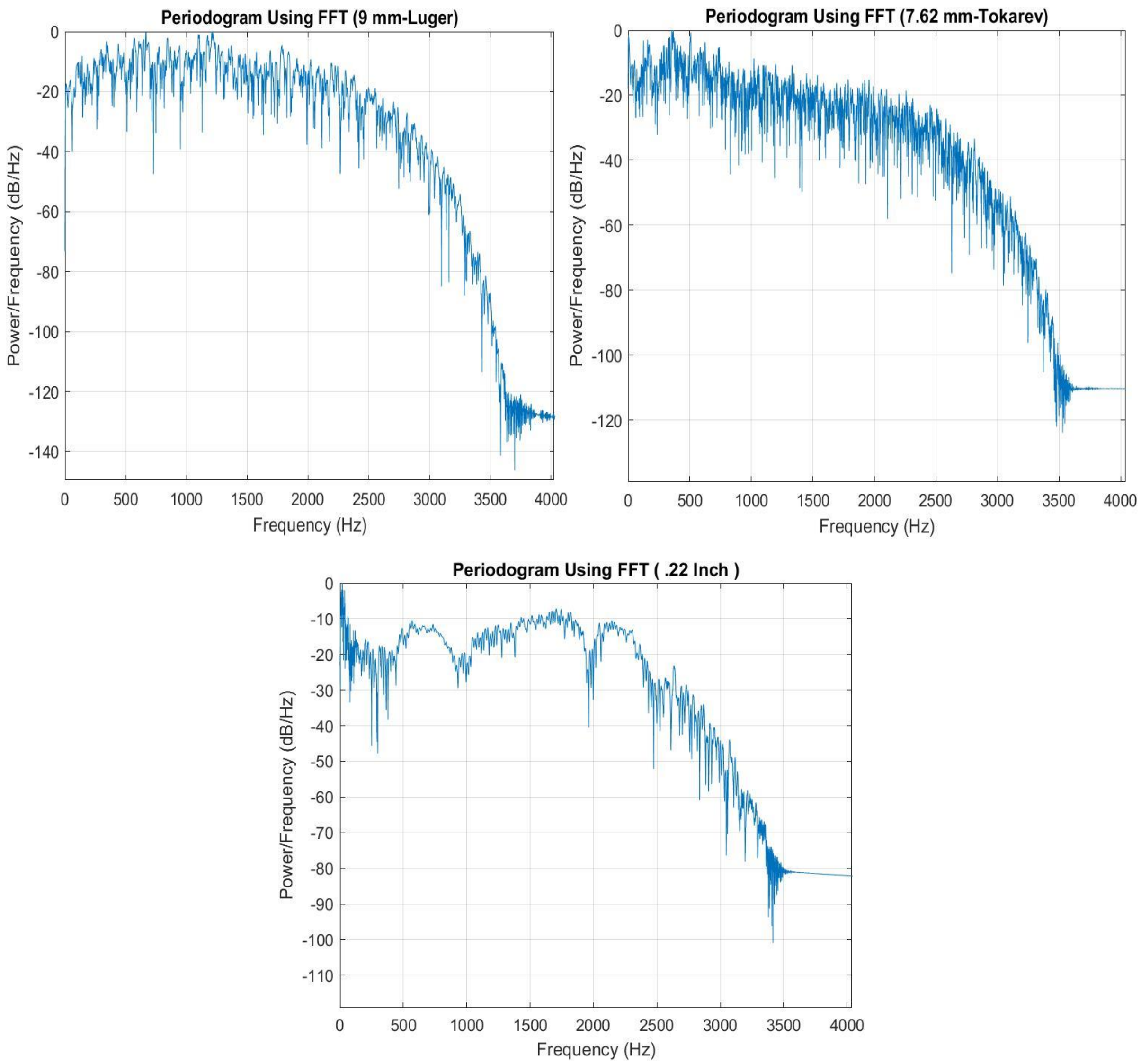


Fig 22a-c: *Normalized Periodogram spectrum of gunshot signals (9 mm-Luger, 7.62 mm-Tokarev, .22 Inch)*

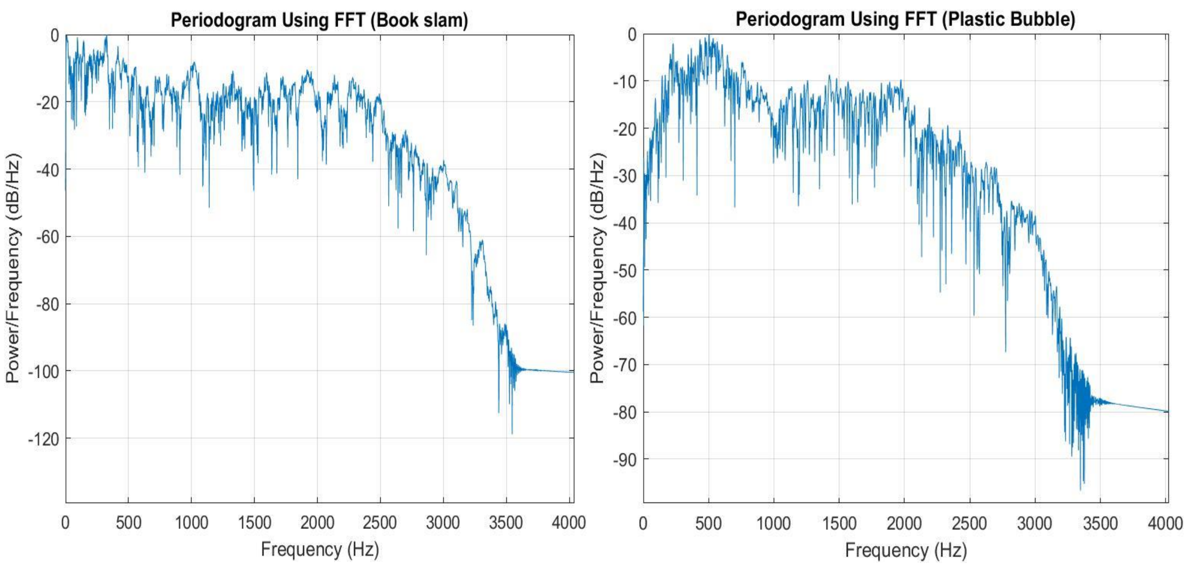


Fig23a-b: *Normalized Periodogram of none-gunshot signals (Book slam, Plastic Bubble)*

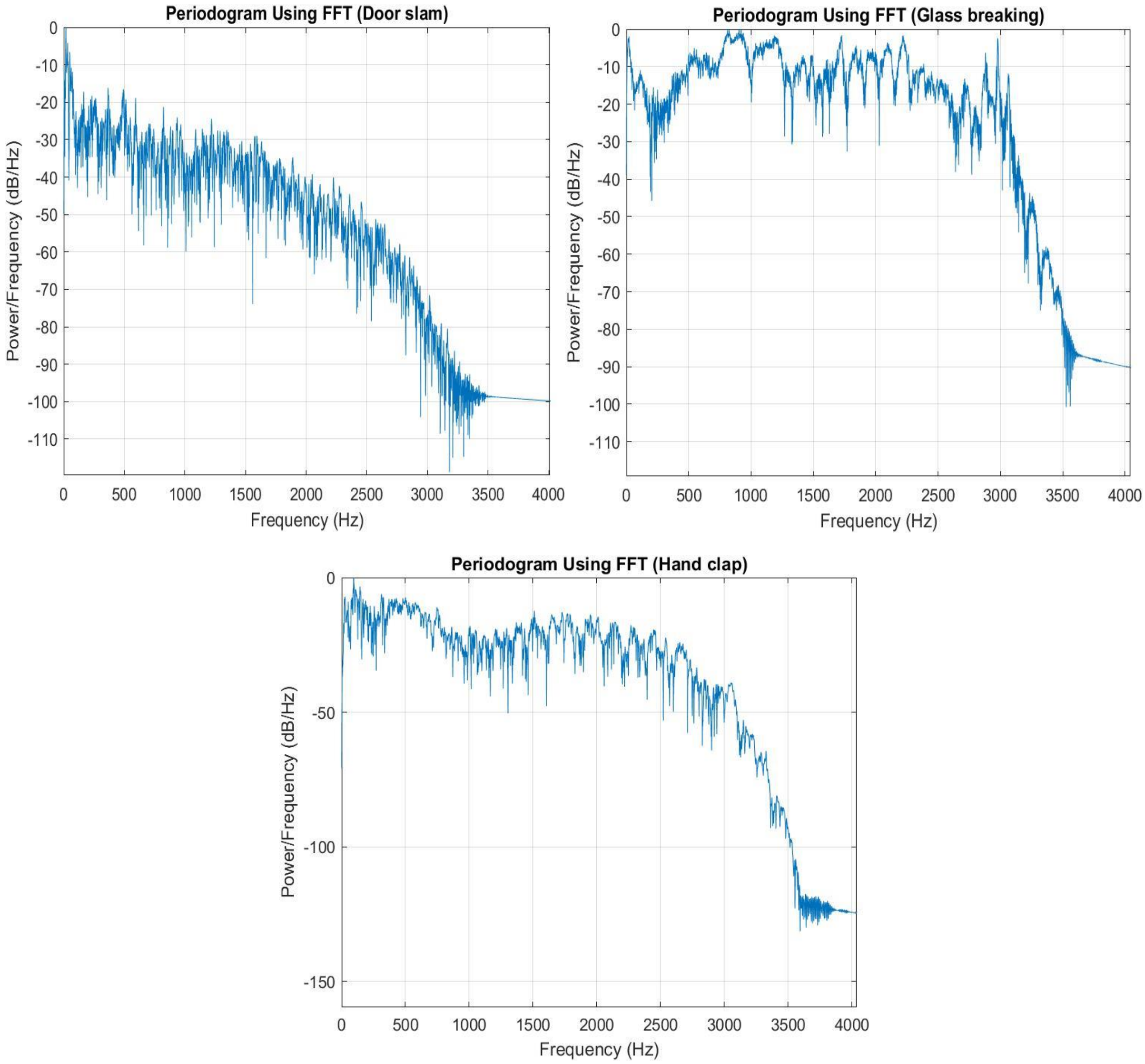


Fig23c-e: *Normalized Periodogram of none-gunshot signals (Door slam, Glass breaking, Handclap)*

4.3.3 Zero Crossing method.

A zero-crossing method is one of the techniques that is applied in signal analyzes. Its essential application considered for the measurement of frequency or period of periodic signals. On acoustic signal detection, this method used to find out the characterizations of the signal's type. The zero-crossing method is a well-known method on MATLAB, and in our case, it was used to determine the number of the crossing of zero points for each signal. It is estimated to reveal several resemblances between the same kind of signals (gunshot signals).

Threshold points are a set of points that create a line to determine the rate of certain signals crossing the rate of this selected line along the graph. Pre-detection low pass filtering or bandpass filtering assist in restricting the bandwidth to frequencies close to the frequency of the signal being measuring.

Zero crossing rate (ZCR) is the main algorithm that is used for the detection of zero-crossing samples at signals. In this part of the signal, it approaches the linearity. Close to the zero crossings, the signals are described as a linear line, which has two parameters: angular parameter a and linear parameter b. With the help of these two parameters, it is possible to express the system equation that consists of two samples between zero-crossing lines (20).

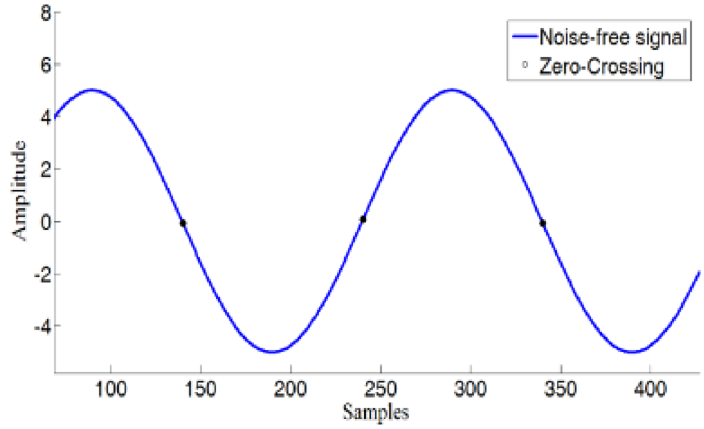


Fig24: *ZCR of Simple cosine wave signal [24]*

$$x_n = a * n + b$$

$$x_{n+1} = a * (n + 1) + b \qquad (20)$$

x_n –first sample and x_{n+1} –second sample between zero-crossing points. So, from the formula (20), it could be written the equation for parameters: (21).

$$a = x_{n+1} - x_n$$

$$b = (n + 1) * x_n - n * x_{n+1} \qquad (21)$$

The primary condition that compensates that there is a zero-crossing point between two close samples is interpreted as expression (22):

$$x_n \leq 0 < \ x_{n+1} \qquad (22)$$

So, we applied the ZCR method to our signals on frequency spectra. As starting, we considered a threshold point, and then we considered the counter due to crossing of threshold point. Considering the threshold point is important while we can arrange some conclusion due to various points that are considered as threshold point and all signals data will be different due to this point. We decided to take 3 main threshold points: 0.1 we choose the point from the below, 0.9 we choose the point from above average, and 0.5 we choose the point from the middle of the spectrum. Here is the result of ZCR for 3 different threshold points:

	ZCR Threshold (0.1)	*ZCR Threshold (0.5)*	*ZCR Threshold (0.9)*
7.62 mm Tokarev-1	*155*	*42*	*8*
7.62 mm Tokarev-2	*160*	*60*	*4*
7.62 mm Tokarev-3	*169*	*36*	*2*
7.62 mm Tokarev-4	*157*	*44*	*4*
9 mm Luger-1	*156*	*46*	*6*
9 mm Luger-2	*148*	*60*	*8*
9 mm Luger-3	*144*	*48*	*4*
9 mm Luger-4	*148*	*48*	*4*
0.22 Inch-1	*100*	*20*	*2*
0.22 Inch-2	*120*	*18*	*2*
0.22 Inch-3	*72*	*23*	*6*
0.22 Inch 4	*106*	*14*	*6*

Fig25a-*ZCR rate of gunshot signals*

	ZCR Threshold (0.1)	*ZCR Threshold (0.5)*	*ZCR Threshold (0.9)*
Book slam-1	*97*	*18*	*3*
Book slam-2	*81*	*8*	*1*
Book slam-3	*84*	*10*	*5*
Book slam-4	*100*	*8*	*1*
Door slam-1	*24*	*2*	*1*
Door slam-2	*32*	*2*	*1*
Door slam-3	*18*	*2*	*1*
Door slam-4	*53*	*1*	*1*
Glass breaking-1	*122*	*103*	*4*
Glass breaking-2	*106*	*59*	*6*
Glass breaking-3	*108*	*100*	*6*
Glass breaking-4	*112*	*73*	*7*
Plastic bubble-1	*216*	*44*	*8*
Plastic bubble-2	*120*	*31*	*4*
Plastic bubble-3	*101*	*24*	*2*
Plastic bubble-4	*81*	*20*	*6*
Handclap-1	*81*	*5*	*1*
Handclap-2	*66*	*9*	*1*

Fig25b-*ZCR rate of none-gunshot signals*

We could arrange the average rate of zero crossings according to the tables above. as a final result, it is considered that average ZCR of gunshots signals varies: 5.62 mm-Tokarev (160-0.1, 45-0.5 5-0.9), 9 mm luger (149-.01, 50-0.5, 5-0.9) and .22 inch (100-0.1, 18-0.5, 4-0.9)

4.3.4 Peak-Valley method.

The peak-valley method is a well-known algorithm that is used on the recognition of human speech. It uses a geometrical approach due to the localizations of the local maximums and minimums of the signal. Besides the local maximums and minimums, all other points that are not included to locals will be neglected during the process of the algorithm.

$$X_p = x_{i-1} < x_i > x_{i+1}$$

$$X_v = x_{i-1} > x_i < x_{i+1} \tag{23}$$

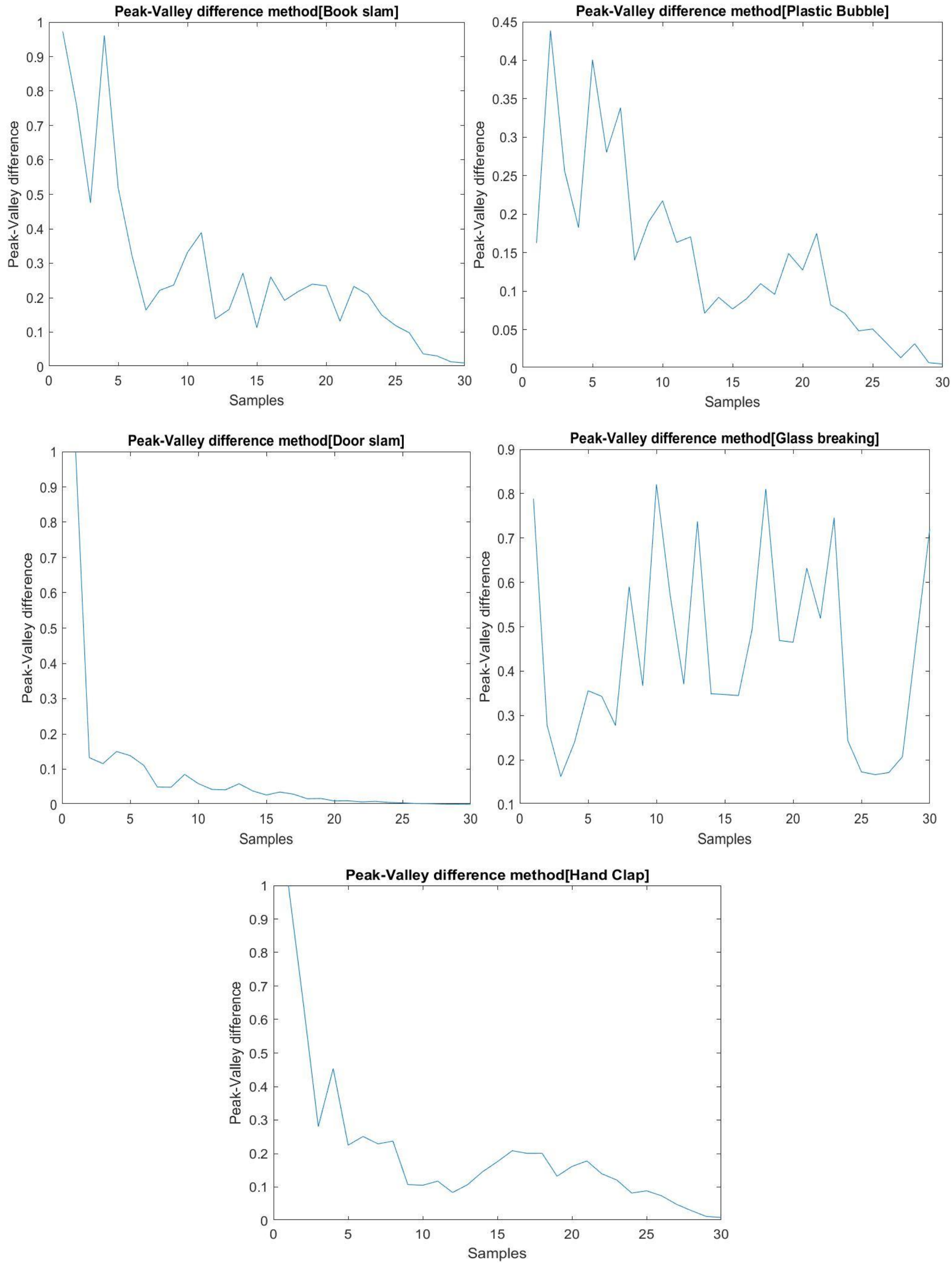
From the supervision, we discovered that the signal's peaks and valleys are synchronous with the filtered versions of the signal. There is a relation between both either on peaks or valleys. The peak-valley method computes the two costs of the amplitudes by applying the summation method: $s[m]$ m- is representing the locations of the local extreme point of the filtered signal. We can calculate the cost estimation of the signals $s[\cdot]$ from formula 24.

$$C_{peak} = \frac{1}{N_{peak}} \sum_{n=1}^{N_{peak}} s[Pos_{peak}[n]]$$

$$C_{valley} = \frac{1}{N_{valley}} \sum_{n=1}^{N_{valley}} s[Pos_{valley}[n]] \tag{24}$$

C-cost estimates on the peak (valley), *N*- total number of peaks (valley), *Pos*-positions of n-th peak (valley)

In our task, we apply the Peak-Valley algorithm to all signals at the frequency spectrum. Due to the length of the signal, we distributed into pieces that each part contains 100, 200, 500 samples. The variation on samples made to analyze signals from varying aspects. Due to the selected frequency length of signal equal to 3000, we created the Peak-Valley difference vector with 30 components. As of the last part, we plotted the final spectrum of Peak-Valley differences on samples for each signal (according to 100 samples).



26a-e: *Peak-Valley difference of none-gunshot signals (Book slam, Plastic bubble, Door slam, Glass breaking, Handclap)*

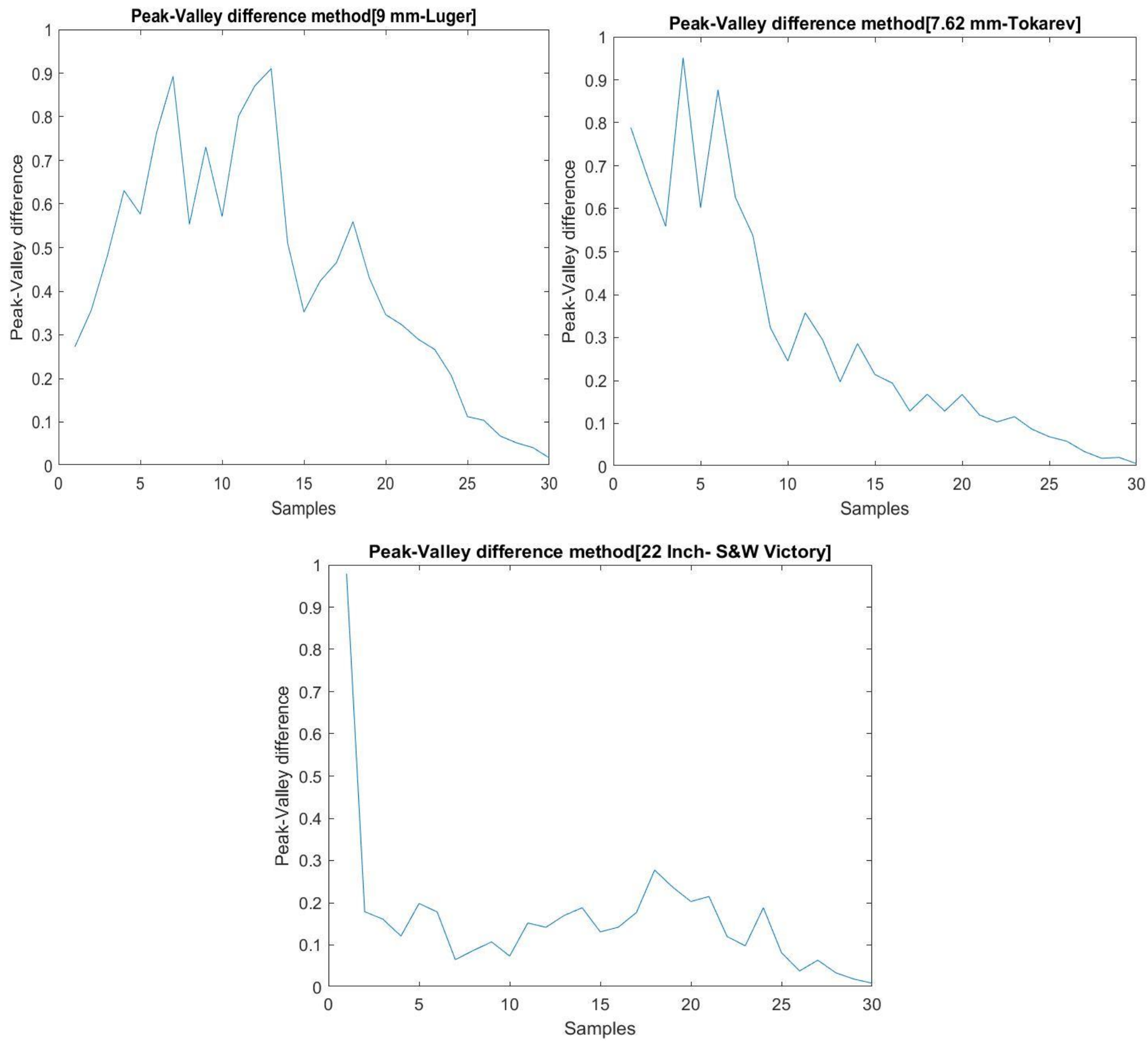


Fig27a-c: *Peak-Valley difference of gunshot signals (9-mm Luger, 7.62 mm-Tokarev, .22 Inch)*

4.3.5 Frequency interval of highest amplitudes estimating with FFT

This method is specific for the frequency spectrum of the signals on the frequency domain. There is each power amplitude according to every frequency. The main task is here to store 100 specific frequencies according to 100 maximum amplitudes. It is estimated to reveal some way that we could arrange some points where the difference between a gunshot and none-gunshot signals. The procedure of MATLAB code.

- In the first step, we assigned 8 obtained and filtered data-name Signal. we created the vector of frequency to collect all the frequencies due to the first 100 maximum amplitudes for those signals.
- Later, we applied FFT to all data and obtain the frequency spectrum of signals:

  FilteredSignal=FFT(Signal)
- We found out the certain frequencies for each data with the assist of max function at MATLAB. "For" loop helped to build a vector that contains 100 maximum values of amplitudes, and according to each amplitude, it matches the certain frequency, and we store all the frequencies into one vector.
- As the final part, our algorithm sorted the frequencies from the highest to lowest on the vector. With the help of $F_{highest} - F_{lowest}$ subtraction method, we learn where the frequency interval lies.

The result is displayed on the Fig28:

	F_{max}	F_{min}	$F_{interval}$
7.62 mm Tokarev-1	743	2	741
7.62 mm Tokarev-2	734	3	731
7.62 mm Tokarev-3	739	2	737
7.62 mm Tokarev-4	691	2	689
9 mm Luger-1	1237	343	894
9 mm Luger-2	1351	9	1342
9 mm Luger-3	917	5	912
9 mm Luger-4	1292	9	1283
0.22 Inch-1	1825	7	1818
0.22 Inch-2	1950	9	1941
0.22 Inch-3	1893	14	1879
0.22 Inch 4	1763	4	1759

Fig 28a: *Table of the frequency interval for gunshot signals*

	F_{max}	F_{min}	$F_{interval}$
Book slam-1	420	5	415
Book slam-2	396	27	369
Book slam-3	252	5	247
Book slam-4	410	6	404
Door slam-1	595	4	591
Door slam-2	500	4	496
Door slam-3	510	4	506
Door slam-4	414	11	403
Glass breaking-1	2978	23	2955
Glass breaking-2	1569	1164	405
Glass breaking-3	2372	923	1449
Glass breaking-4	2524	1902	622
Plastic bubble-1	783	207	576
Plastic bubble-2	874	143	731
Plastic bubble-3	598	143	455
Plastic bubble-4	760	169	591
Handclap-1	509	486	23
Handclap-2	488	465	23

Fig 28b: *Table of the frequency interval for none-gunshot signals*

4.3.6 .22-Inch pistol

As a result, it is estimated that gunshots have ZCR between 40-100 intervals with the threshold considered as 0.5, while gunshot .22 Inch behaved different characteristics comparing to the other 2 firearms for ZCR method. Besides, the result of .22 inch wasn't satisfied according to the methods of Peak-Valley difference and Frequency interval. All the obtained data from the measurement of three methods let us conclude that the characteristic of .22-inch short gun varies comparing to rest two firearms. It can be considered as .22-inch gun is not related as the professional weapon while its utilization is useful for some sports championship, and its caliber is considered for "Starting Pistols." "Starting pistols" can have different characterizations comparing to the rest of firearms, and its behaving can relate to these aspects.

4.4 Computation of MFCC methods

In the second part of the computation, the MFCC method is discussed. Comparing with FFT, the MFCC method is usually applied to speech recognition and several steps are set for the process. FFT is one of the steps that is used on the MFCC method. In this project, we will check the possibilities for recognitions of gunshot and household processes. For the MFCC algorithm, we use FFT, Mel scale Filterbank and a discrete cosine transform (DCT). MFCC procedure is described in the below.

1. Divide the signals into frames.
2. Take the FFT of all frames of signal and calculate the power spectrum.
3. Multiplications of Filterbank with power spectrums.
4. Log the values of energy stored in each filter.
5. Apply the DCT algorithm to the logged energies to get a matrix of MFCC coefficients dependence on the frame.

As the first step, we should divide the main signals into short frames. Audio signals are constantly changing in time intervals. In a short time, interval or frame, we assume that signals are regular. According to this statement, we use frame windows to divide the signal into short frames. It is important to choose the right time interval for a frame. If the frame is too short, it can be complicated to analyze signals while the signal length is longer, there is a possibility of changes at certain intervals.

In our case it is accepted to choose the frame length between the 20-40 ms time interval. If we consider the total samples of the signal are 44000 samples, and sampling frequency selected as 44100 Hz, it will be approximately 1 second as total. We chose a short frame as 25 ms. It means each frame sample length can be calculated *44000*0.025=1100* samples for each frame. For overlapping, we chose overlapping frame time as 10 ms which is equal to 440 samples. This is the preprocessing part of the MFFC. At the outcome, we obtained the matrices of 98 frames with 1100 samples with the overlapped frame as 10 ms.

On the next step we used DFT. We will choose the length of DFT filter as K=512, our windowing frame length equal to N=1100. We obtained the complex spectrum of signals using the formula 25:

$$S_i(k) = \sum_{n=1}^{N} S_i(n) * h(n) * e^{(-j*2*\pi*k*n)/N} \qquad 1 < k < K \tag{25}$$

Here $h(n)$-N sample long hamming window, $S_i(n)$-our framed time domain signal, $S_i(k)$- is our final complex spectrum. On the next step we calculated the periodogram estimates of the power spectrum. We took the absolute value of complex spectrum, square the outcome and divide the result by sample length according to formula 26.

$$P_i(k) = \frac{1}{N}|S_i(k)|^2 \tag{26}$$

On the next step, we created the MEL scale filterbank. One of the main points of the filter bank to choose how many filters we want to apply for our power spectrum and calculate the stored energy. The ideal range varies between the 26-40 filters. In our case, we chose the 26 filters which each of them is a vector that contains 257 coefficients. Vectors are mostly zero but are none-zero in certain parts which are defined by the formula given below. Our MEL filterbank is discovered on the Fig30.

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \le k < f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) < k \le f(m+1) \\ 0, & k > f(m+1) \\ 1, & k = f(m) \end{cases}$$

(27)

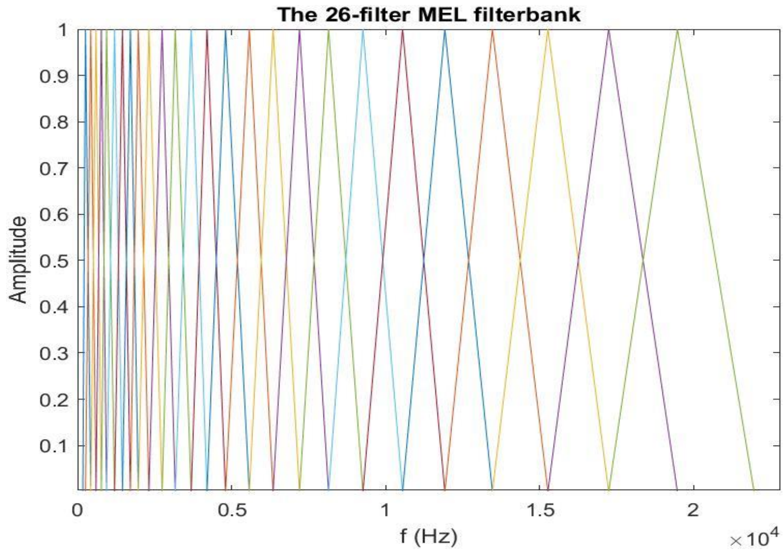


Fig 29: *Mel Filterbank*

The application of filterbank is to calculate the stored energy on each part of the filter according to the power spectrum of frames. We multiplied filters by the frames of the power spectrum, then we summed the energies at each filterbank and got a vector of filterbank-frame.

Afterwards, we multiplied each component of the vector with the logarithmic function and applied DCT. As a result we neglected the second half of the filters and kept the energy stored on the first 13 filters. This part is identified as MFCC. It is possible to compare the results of 7.62 mm Tokarev signals MFCC according to the different energy on different filterbank. Final MFCC coefficients displayed in Fig31.

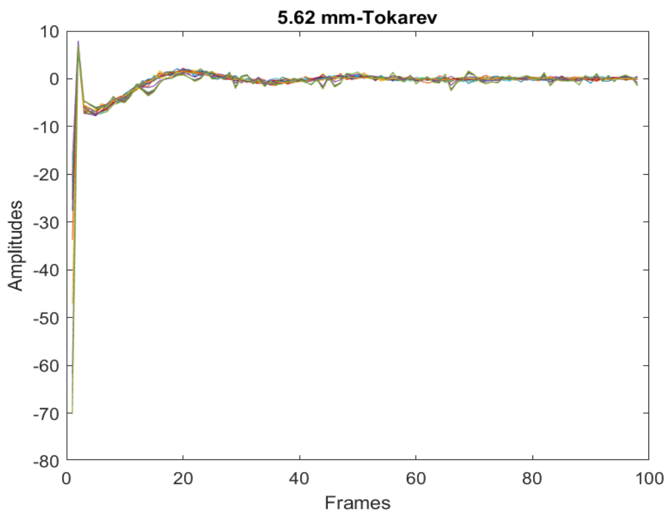


Fig30a: *MFCC coefficient of 5.62 mm-Tokarev*

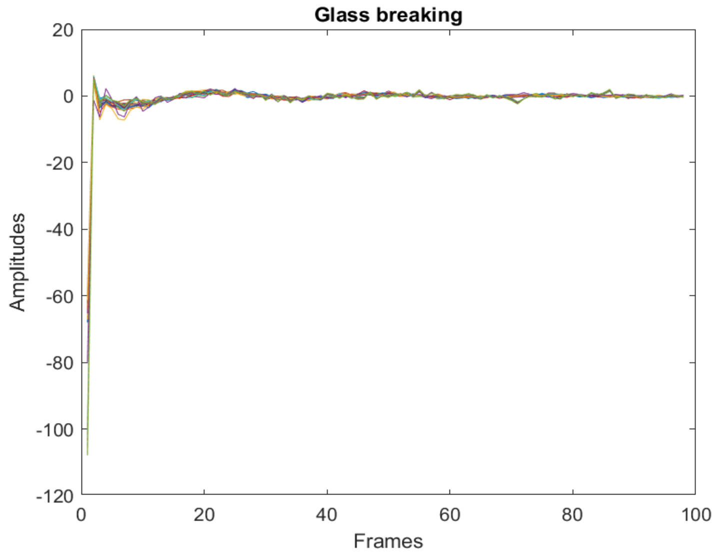


Fig30b: *MFCC coefficient of Glass breaking*

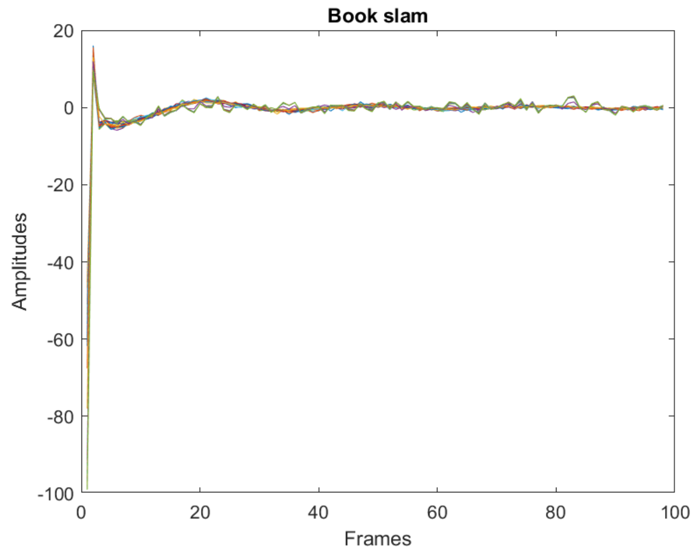


Fig30c: *MFCC coefficient of Book slam*

5. Conclusion

To conclude, the main purpose of the thesis is presented by 4 objectives:

- Study the essential factors of gunshots (muzzle blast, shockwave).
- Analizing the applications of the FFT and MFCC algorithms on gunshot recognitions and general review of the signal characteristics dependence on the frequency change.
- Development and utilization of PSD, ZCR, Peak-Valley and Frequency interval (FFT) methods on the distinguishing of the gunshot & none-gunshot signals.
- Close approach to MFCC process: generating the MEL filterbank and applications of the filter to stored signals for gunshot detection.

During the bachelor project, 5 methods belonging to 2 different algorithms have been discussed and applied to the recorded 40 samples from 8 different types of signals (3 gunshots and 5 none-gunshot). As a result, there are several expected and unexpected factors appeared during the measurement. Each way stands to the basis of the FFT method and parses the characteristics of signals in the frequency domain. However, some methods are not valuable to differentiate and didn't give clear information for characterizing signals.

FFT

A method such as PSD, Peak-Valley and ZCR methods were available to detect some similarities between gunshots and distinguish the characteristics of these type signals with a none-gunshots signal while frequency interval and MFCC methods were not successful during the project. Besides detections, during the procedure of techniques, some vital information about signal identifications have appeared. PSD and Peak-Valley methods gave us a clear review from the graphs that characteristics of gunshot signals can be distinguished from the false alarm signals.

The outcome of PSD allows us to observe the relation of the ratio between the power spectrum and its related frequency. Each signal's power spectral density is computed on the base of an algorithm which is displayed on the fifth step of FFT (Appendix). As the first approach, Periodogram of the 5.62 mm (Tokarev) and 9 mm (Luger) revealed the specific result comparing to the rest of the signals. In these graphs, there is functional conformity between both gunshot signals. Uunfortunately, graph of 0.22 inch varied than the rest of the gunshot according to the type of pistol.

ZCR method is divided into 3 different sections according to applications. 3 different threshold point is examined during measurement. Results of the threshold points: 0.1 and 0.9 weren't so convenient to distinguish the signals while point 0.5 gave us a legible report about the distinctive of gunshot signal. Gunshot signals threshold crossing rate varies in the interval 40-60 while most of the false alarms remain under the value of 30. The specific signal that contains higher frequency such as the glass breaking rate, was higher than gunshot signals. ZCR method can be considered as successful way for the detection of gunshots signals.

Peak-Valley method's result was displayed in Fig 26 and 27. The result of this way can be considered reasonable. Higher differences in gunshot signals vary on lower and mid frequencies while the none-gunshot data, it was expected to be on the small frequency. Glass breaking signal has a higher difference in every frequency interval.

In the last method of FFT, gunshot signals have higher amplitude intervals on the middle frequencies while none-gunshot signals, it was computed to be on the small frequencies. The frequency interval of glass breaking, and plastic bubbles are the same as gunshots. This method is not a convenient way of detecting gunshots for now.

MFCC

MFCC method has many parameters. Size and number of frames overlap times and number of the chosen filter. According to these parameters, the result of the MFCC coefficient can be varied. In our case we used 98 frames with the filterbank contains 26 filters. Unfortunately obtained graphs from the MFCC were not enough to determine gunshot signals as an initial step. However, according to the different selected parameters, the result of MFCC can be different and recognitions of gunshots are possible.

All these methods based on a simple algorithm of FFT can be implemented in the future for acoustic event devices to detect and classify the acoustic events.

6. Appendix

FFT

First part- Opening all observed data

```
load('Tokarev_762.mat', 'Tokarev_1_fs_44100senzor_1')
load('Luger_9mm_8.mat', 'Luger_9mm_8_fs_44100senzor_1')
load('handclap.mat', 'ruka_1_Audio_wav_fs_44100')
load('glass.mat', 'sklenicka_typ1_1_Audio_wav_fs_44100')
load('doorslam.mat','dvere_1_Audio_wav_fs_44100')
load('022inch.mat','vystrel_1_1')
load('book.mat', 'Mlaceni_skript_1_Audio_wav_fs_44100')
load('bubble.mat', 'Bouchaci_bublina_1_Audio_wav_fs_44100')  %we loaded original signals

vystrel_1_1=vystrel_1_1.';
ruka_1_Audio_wav_fs_44100=ruka_1_Audio_wav_fs_44100.';
Luger_9mm_8_fs_44100senzor_1=Luger_9mm_8_fs_44100senzor_1.';
Tokarev_1_fs_44100senzor_1=Tokarev_1_fs_44100senzor_1.';
sklenicka_typ1_1_Audio_wav_fs_44100=sklenicka_typ1_1_Audio_wav_fs_44100.';
dvere_1_Audio_wav_fs_44100=dvere_1_Audio_wav_fs_44100.';
Mlaceni_skript_1_Audio_wav_fs_44100=Mlaceni_skript_1_Audio_wav_fs_44100.';
Bouchaci_bublina_1_Audio_wav_fs_44100=Bouchaci_bublina_1_Audio_wav_fs_44100.'; %converted signals row to column
```

Second part-Minimize the total signal

```
% in this part, we found out the maximum amplitude point of each signal and
% later we chose the certain interval as start and stop points.
[~,I]= max(ruka_1_Audio_wav_fs_44100);
start=I-4000;
stop= I + 40099;

handk =ruka_1_Audio_wav_fs_44100(start:stop);
doork = dvere_1_Audio_wav_fs_44100(start:stop);
glassk = sklenicka_typ1_1_Audio_wav_fs_44100(start:stop);
inchk = vystrel_1_1(start:stop);
lugerk = Luger_9mm_8_fs_44100senzor_1(start:stop);
tokarevk = Tokarev_1_fs_44100senzor_1(start:stop);
bubblek= Bouchaci_bublina_1_Audio_wav_fs_44100(start:stop);
bookk= Mlaceni_skript_1_Audio_wav_fs_44100(start:stop);
```

Third part-Low Pass Filter

```
Fs = 44100;  % Sampling Frequency
N    = 150;      % Order
Fc   = 2500;     % Cutoff Frequency
flag = 'scale';  % Sampling Flag
%Create the window vector for the design algorithm.
win = blackmanharris(N+1);
b  = fir1(N, Fc/(Fs/2), 'low', win, flag);
Hd = dfilt.dffir(b);
% we apply our Blackman-Harris filter into signals and got the filtered
% datas
luger = filter(b,1,lugerk);
inch = filter(b,1,inchk);
tokarev = filter(b,1,tokarevk);
door = filter(b,1,doork);
glass = filter(b,1,glassk);
hand = filter(b,1,handk);
book = filter(b,1,bookk);
bubble= filter(b,1,bubblek);
Fs= 44100;
samples=length(bubble);
t=0:1/Fs:samples/Fs-1/Fs;
plot(t,bubble)
xlabel('time [s]');
ylabel('Amplitude');
title('Plastic Bubble')
```

Fourth part-FFT method

we applied FFT algorithm into our filtered datas

```
Fs=44100;
L=length(bubble);
Y=fft(bubble);
P2=abs(Y/L);
P1=P2(1:L/2+1);
P1(2:end-1)=2*P1(2:end-1);
f = Fs*(0:(L/2))/L;
P1=normalize(P1,'range');
plot(f,P1)
title('Single sided Frequency spectra [Signal name]')
xlabel('f (Hz)')
ylabel('|P(f)|')
```

Fifth part-PSD method

```
Fs = 44100;
t = 0:1/Fs:1-1/Fs;
x=bubble;
N = length(x);
xdft = fft(x);
xdft = xdft(1:N/2+1);
psdx = (1/(Fs*N)) * abs(xdft).^2;
psdx(2:end-1) = 2*psdx(2:end-1);
freq = 0:Fs/length(x):Fs/2;
psdx=normalize(psdx,'range'); % normalized our signal to put the all datas value in the same range
plot(freq,10*log10(psdx))
grid on
title('Periodogram Using FFT (Signal name)')
xlabel('Frequency (Hz)')
ylabel('Power/Frequency (dB/Hz)')
```

Sixth part-ZCR method

```
Fs=44100;
L=length(inch);
Y=fft(glass);

P2=abs(Y/L);
P1=P2(1:L/2+1);
P1(2:end-1)=2*P1(2:end-1);
f = Fs*(0:(L/2))/L;
P1=normalize(P1,'range');
plot(f,P1)
title('Single sided Frequency spectra [Signal name]')
xlabel('f (Hz)')
ylabel('|P(f)|')
%
% Positive-slope zero-crossing detector
z = and((P1 > 0.1), not(circshift((P1 > 0.1), 1)));  z(1) = 0;
%
% Find the locations of the zero-crossing points
crossing_points = find(z);
```

Seventh part-Peak Valley method

```
Fs=44100;
L=length(bubble);
Y=fft(door);
P2=abs(Y/L);
P1=P2(1:L/2+1);
P1(2:end-1)=2*P1(2:end-1);
f = Fs*(0:(L/2))/L;
P1=normalize(P1,'range');
P1=P1(1:3000);
differencevector=[1:30];
a=1;
b=100;  % interval of 100 samples and found out the peaks and valley inside of these interval
d=1;
for c=1:30
    M=max(P1(a:b));
    N=min(P1(a:b));
    Difference=M-N;
    differencevector(d)=[Difference];
    a=a+100;
    b=b+100;
    d=d+1;
end
plot(differencevector)
title('Peak-Valley difference method[Signal name]')
xlabel('Samples')
ylabel('Peak-Valley difference')
```

Eighth part- 100 maximum amplitude

```
Signal=bubble;
maxfrequencyvector=[1:100]; %vector of first 100 maximum amplitudes
L=length(Signal);
FilteredSignal=fft(Signal);
Fs=44100;
Powerspectrum2=abs(FilteredSignal/L);
Powerspectrum1=Powerspectrum2(1:L/2+1);
Powerspectrum1(2:end-1)=2*Powerspectrum1(2:end-1);
[M,I]=max(Powerspectrum1);
for c=1:100 %counter
     [M,I]=max(Powerspectrum1); % obtaining the each highest frequency
     maxfrequencyvector(c)=I; % adding the each highest frequency into samples vector
     Powerspectrum1(I)=[0]; % deleting the already added frequency from the main spectrum
end
biggestfrequency=max(maxfrequencyvector);  % the highest frequency
smallestfrequency=min(maxfrequencyvector); % the lowest frequency
interval=biggestfrequency-smallestfrequency; % releasing the final frequency interval where all
                                             % frequencies included
```

MFCC

First step- Frame signals

```
signal=luger;
framesample=1100;
frameSignal=zeros(98,framesample);
frameoverlap=440;
a=1;
b=1100;
for c=1:98
    frameSignal(c,:)=signal(a:b);
    a=a+440;
    b=b+440;
end

fsignal=signal;
```

Second step-DFT and Power Spectrum

```
signal=fglass;
Ysignal=zeros(98,257);

w=hamming(1100);
w=w.';
signalsumframe=0;
for c=1:98
    for k=1:257
        for n=1:1100
            signalsumframe=signalsumframe + signal(c,n)*w(n)*exp((-1i*2*pi*k*n)/1100);
        end
    Ysignal(c,k)=signalsumframe;
    signalsumframe=0;
    end
end

Powerspectrum=abs(Ysignal).^2/1100;

psignal=Powerspectrum;
```

Third step- multiplication of powerspectrum with mel filter

```
multi=zeros(26,98);
ps=pinch;
s=0;
for c=1:26
    for j=1:98
        for k=1:257
        s=s+ps(j,k)*H(c,k);

        end
        multi(c,j)=s;
        s=0;
    end

end
msignal=multi;
```

Fourth step- Log the energies

```
logsignal=log10(msignal);
```

Fifth step- DCT

```
mfccofsignal=dct(logbook);
```

7. References

[1] https://time.com/4951192/shots-fired-shotspotter/

[2] https://www.statista.com/statistics/191219/reported-violent-crime-rate-in-the-usa-since-1990/

[3] https://www.theatlantic.com/photo/2013/05/finland-in-world-war-ii/100519/

[4] Antonio L. L. Ramos, On acoustic gunshot localization systems

[5] https://link.springer.com/chapter/10.1007/978-3-319-70831-7_3

[6] https://www.mathworks.com/help/fixedpoint/examples/convert-fast-fourier-transform-fft-to-fixed-point.html

[7] http://kom.aau.dk/group/04gr742/pdf/MFCC_worksheet.pdf

[8] https://www.sciencedirect.com/topics/engineering/low-pass-filters

[9] Candy J, Signal Processing in Acoustics: Science or Science Fiction, Acoustic today (2008) DOI: 10.1121/1.2994726

[10] Clavel C, Ehrette T, Richard G, Events detection for an audio-based surveillance system, ICME 2005, DOI: 10.1109/ICME.2005.1521669

[11] Maher R, Acoustical characterization of gunshots, 2007

[12] Davila, Vianna, (August 17, 2017) [August 16, 2017]. "San Antonio police cut pricey gunshot detection system." San Antonio Express-News

[13] Bragg, William Lawrence. "Personal Reminiscences" (PDF). Archived from the original (PDF) on 2007-09-30. Retrieved 2006-05-14

[14] Rawlinson, Alfred (1923), Rawlinson, The Defence of London, Andrew Melrose, London & New York, pp. 110–114

[15] https://www.zdnet.com/article/us-armys-gunshot-detector-may-save-lives/

[16] https://www.securityinfowatch.com/alarms-monitoring/alarm-systems-intrusion-detection/gunshot-detection/article/21081663/tech-deep-dive-gunshot-detection

[17] https://www.marketsandmarkets.com/Market-Reports/gunshot-detection-systems-market-58054729.html

[18] Loeffler C, Detecting Gunshots Using Wearable Accelerometers, DOI: 10.1371/0106664

[19] Maher R, Modelling and signal processing of acoustic gunshot recordings, DOI: 10.1109/DSPWS.2006.265386

[20] Jr. J Freire I, Gunshot detection in noisy environments, DOI: 10.14209/sbrt.2010.92

[21] Heckbert P, Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm, Notes Computer Graphics (1995) 3(2) 15-463

[22] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[23] Luciano C. Caldas, Luiz A. Baccala, Power spectral density (PSD) estimation: MATLAB algorithm implementation for array signal processing and test validation

[23] https://www.sciencedirect.com/topics/engineering/low-pass-filters

[24] http://www.learningaboutelectronics.com/Articles/Low-pass-filter.php

[25] Wall R, Simple Methods for Detecting Zero Crossing, DOI: 10.1109/IECON.2003.1280634

[26] Mendonca T, Pinto M, Duque C, Least squares optimization of zero crossing technique for frequency estimation of power system grid distorted sinusoidal signals, EEE INDUSCON 2014, DOI: 10.1109/INDUSCON.2014.7059443

[27] Petre Stoica, Randolph Moses, Spectral Analysis of Signal, ISBN 0-13-113956-8

[28] Narayanam R, An Efficient Peak Valley Detection Based VAD Algorithm for Robust Detection of Speech Auditory Brainstem Responses, DOI: 10.5121/csit.2013.3636

[29] Chen J, Kao Y, Pitch Marking Based on an Adaptable Filter and a Peak-Valley Estimation Method, Computational Linguistics and Chinese Language Processing (2001) 6(2) 1-12

[30] https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd

[31] Leiming Chen, Lianzhang Zhu, Emotion Recognition from Chinese Speech for Smart Affective Services Using a Combination of SVM and DBN, DOI: 10.3390/s17071694

[32] https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html