2017

# Sound analysis prototype to enhance physical security in academic institutions

Kevin Ochieng' Omondi
*Faculty of Information Technology (FIT)*
*Strathmore University*

# Sound Analysis Prototype to Enhance Physical Security in Academic Institutions

**OMONDI KEVIN OCHIENG'**

**065233**

**Submitted in partial fulfilment of the requirements of the Degree of Master of Science in Information Technology at Strathmore University**

**Faculty of Information Technology**

**Strathmore University**

**Nairobi, Kenya**

**June, 2017**

## Declaration and approval

I declare that the design and the implementation of this research, sound analysis prototype to enhance physical security is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references and that this work has not been submitted before for any other degree at any other institution.



Omondi Kevin Ochieng'

…….……

Date: 8th June 2017

The thesis of Omondi Kevin Ochieng' has been reviewed and approved by the following:

Dr. Humphrey Njogu

Senior Lecturer, Faculty of Information Technology

Strathmore University

Dr. Joseph Orero (PhD)

Dean, Faculty of Information Technology

Strathmore University

Professor Ruth Kiraka

Dean, School of Graduate Studies

Strathmore University

## Abstract

Competence in the provision of security to the civilians in Kenya has generally deteriorated and hence negatively affecting the public trust accorded to security agencies. Indeed, the police to civilian ratio is low and this has affected the institutions of learning as they have become new attack grounds for the terrorists. Institutions of learning have suffered the worst since they are expected to be accountable of their own security in many cases. As a result, many institutions of learning use available security agencies, most of which employ outdated and less efficient means of implementing security. Examples of commonly used physical security techniques include the use of security guards, perimeter walls, some places use turnstiles as well as CCTVs. The inefficiencies that comes along with these security measures has still however exposed these institutions to great dangers of insecurity.

This study proposes the use of sound classification to enhance physical security. The solution relies on the integration of the possible solutions of the artificial neural networks (ANN) in sound classification to detect sound variations in the leaning institutions. It is expected that decisions made through classification assist security personnel on the ground to tighten the physical security. The solution offers automatic analysis of the recorded sound from the environment, compares it to the stored dataset which has urban sounds and the score labels displayed on the output screens for the security personnel to help them enhance the available physical security.

The usage of scientific research methodology through experimentation ensured that the sounds were captured, the dataset sounds were collected and trained for comparison to take place and finally results validated to prove the theory. The system proved an accuracy percentage of 78%, and the efficiency, user friendliness and reliability were al passed.



**Keywords: physical security, security personnel, artificial neural networks, sounds**

**Acknowledgement**

Due appreciation are directed to my project supervisor, Dr. Humphrey Njogu for providing guidance through the scope of the project. I would also like to acknowledge my coursework lecturers for providing adequate knowledge in the relevant fields of Information Technology (IT) though the two year course and my MSc. IT class of 2017 for the encouragement and the general academic support. I also acknowledge my family members for their tireless support and finally to the almighty God for the great experience and the opportunity he has granted me through the entire journey.

## Table of Contents

## List of Figures

## List of Tables

## List of Equations

## List of Abbreviations

| | | |
|---|---|---|
| ANN | - | Artificial Neural Networks |
| CCTV | - | Closed – circuit Television |
| CCD | - | Charge Coupled Device |
| DVR | - | Digital Video Recorder |
| HVAC | - | Heating, Ventilation, and Air Conditioning |
| IP | - | Internet protocol |
| IT | - | Information Technology |
| MAS | - | A Multi-Agent System |
| NLP | - | Natural Language Processing |
| KSIA | - | Kenya Security Industry Association |
| PSIRA | - | Private Security Industry Regulatory Authority |
| SAPS | - | South Africa Police Service |
| WSNs | - | Wireless Sensor Networks |

# Chapter 1 Introduction

## 1.1 Background of the study

Sound, usually produced when objects or particles vibrate, Hollis (2017), and is always perceived through different frequencies that distinguish the different types of sounds produced. There exists a lot of sounds produced in our environment on a daily basis depending on the activities that we do. Most of these sounds may symbolize a source of threat but no one uses them to detect and act on the threats that they pose. The current security situation in Kenya today is not favoring physical security of the citizens as the police to civilian ratio is at 1: 1150, a report by Kenya National Commission on human rights (2014) and this is a sign that the population has grown too large for the police or the responsible security stakeholders to handle. This therefore calls for better ways, which is being driven by the use of IT to help the people stay safe.

Sometimes, there are too many sound recorders that are available but barely any of them is being put into usage. In a busy city like Nairobi there is so much sounds being produced that can be termed as sound pollution Karue, Kinyua, & Njau, (2014) and most of the time, none of it is being analyzed so as to produce a usable decision with regards to security. This study narrows down to the academic institutions that we currently have in Kenya, most of them are used to the friendly environment where not so much noise is being produced. Recently, having these academic institutions being the focus of the terrorists in Kenya, a research by Pate , Jensen, & Miller (2015), better security can be enhanced by using sound classification to analyze the environmental conditions in these institutions such that the production of unfamiliar sounds can be used to predict the threats that are around us for instance, if the academic institutions are used to the sounds of students playing or moving around, then a sound produced by firing a gun or a grenade explosion in the same environment should be taken seriously, analyzed as it is most likely a threat in that environment.

With this great potential of using sound to predict dangers in a confined environment, then sound analysis seems to be a very promising way to detect them and come up with a legit verdict as to what the danger might be and what its causes are. Some solutions that are trying to use sound to monitor the safety of a place, for instance, mobile-based security agency sound monitor and alert system, a model developed by Machanje (2014), has limitations such as being costly to the low end users as they may not be able to afford the online charges from the service providers. This in

makes the proposed methods passive and slow in solving real-time problems of insecurity and disaster management.

Through the use of artificial intelligence (AI) to analyze the sounds that are being produced in the leaning institution's environment, this research aims at improving the physical security that is already in place by producing a reliable decision trough sound classification that will help the security personnel on the ground know the cause of the insecurity at hand.

## 1.2 Problem statement

A lot of sounds are produced through the different activities that take place in the environment for example gunshots, use of pangas and sometimes unordered noise from individual's activities such as random movements. Despite the existence of many sound recorders that we have, these sounds are hardly used as many have not realized the importance of sound analysis. Relating the sounds being produced to their sources can help to map the type of activities happening at the source hence the importance of sound interpretation.

Currently, the available methods of enhancing physical security do not consider the use of sound as one of the ingredients to consider in decision making yet, if sound is mapped back to the source, then the activities happening at the source can be determined. If the activities are malicious, then a security enhancing decision can be implemented immediately. Currently, some of the existing solutions using sound technology to enhance physical security are very costly to implement and are not tailored to solve the local problems such as high costs of security which are faced by the Kenyan institutions of learning.

Having all these challenges in place, this research aims at developing a sound based solution that monitors all these sounds being produced in the environment, analyzes and classifies them through the use of artificial neural networks to predict the activities happening at the source hence helping the security personnel reinforce physical security by making the right decision.

## 1.3 Objectives of the study

i. To identify the common causes of sound related physical security breaches in academic institutions,
ii. To review the existing sound based solutions for different physical security breaches such as trespass and vandalism,

iii. To design and develop an intelligent sound solution to address physical security breach,
iv. To validate the effectiveness of the system.

## 1.4 Research questions

i. What are the common causes of sound related physical security breaches?
ii. What are the current existing sound based systems trying to enhance physical security?
iii. How will the application be designed and developed, to address physical security breach?
iv. How will the effectiveness of the system be validated?

## 1.5 Project scope and limitations

This study focuses on solving the gaps in physical security faced by the educational institutions in case of an attack or a breach, this has been done by capturing the sounds from the environment that are generated by the attackers or by the random victims seeking safety in case a terror attack occurs. The research has therefore investigated and implemented an automatic monitoring alert system that captures the sounds from the environment, analyses them by comparing them to the system's dataset to generate a decision that helps the security personnel on the ground to react to the situation, this helps to enhance the physical security that is already in place. Any additional functionality will have to fall under the vicinity of these technologies in order to be incorporated in the implementation of the study.

Some of the limitations that were faced by the study included;

i. Limited resources in terms of computing that could analyze the dataset as fast as possible, the dataset has a total of 8742 sound that needs to be analyzed through supervised learning in order to make a verdict from the sound input whether it's a dangerous sound or not.
ii. The programming of the system which had to be a combination of procedural languages such as C++ and a modern language such as python for analyzing the dataset.

## 1.6 Justification

The need to enhance physical security is a major milestone that most of the institutions in Kenya today Pate , Jensen, & Miller (2015) would like to achieve, this is because lives are lost from the major attacks that occur from time to time an example from the Garissa University attack that occurred in Kenya in 2015. Many institution still rely on the help of the security agencies that are currently available in the country such as G4S, Senaca Security Group, RADAR Security Group and KK Security, Idriss, Jendly, Karn, & Mulone (2010) which is still not enough to reinforce the tight rules and precautions needed to enhance the physical security that has already been implemented. The challenges faced by these security groups such as acting on command, slow reaction time and doing their operations on a routine and predictable manner has caused major dissatisfaction and lack of confidence by the people who seek their services. This security challenges has been observed both locally and internationally (Kevin Strom, et al., 2010).

Apart from the challenges caused by these security groups, the means of communication that is used by them such as walkie-talkies, mobile phones, use of whistles and shouting at one another is still not adequate in enhancing the physical security in case of a security breach as they are slow and unreliable.

The proposed sound analytical tool used to classify the sounds produced in the environment whether dangerous or not aims to solve the problems experienced in these scenario. The application captures sounds of different frequencies and automatically analyzes and detects the abnormal sound variations then alerts the security personnel on the ground to enhance the adequate action that will improve the physical security that is already in place. This can also help them to seek immediate backup from external sources in case there is need to.

# Chapter 2 Literature Review

## 2.1 Overview

This chapter has been broken down into four sub sections excluding the introductory and the summary sub sections. The first segment narrates the current security scenario in the country, specifically in the academic institutions. The second segment delves into the challenges faced by the academic institutions and what causes there challenges, it also focuses on the different sound technologies that already exists and how they have been used. The third segment, which features the majority of this chapter, reviews applications and systems done in the past through various research programs locally and in other countries. This provides a sneak peek into some of the technologies and algorithms implemented in the application at hand. The last segment of this chapter explains the current state of knowledge based on the literature review. It brings up the gaps that exist in the current security scenario, especially to educational institutions, thus giving an opportunity for introducing a smarter solution for the same.

## 2.2 Physical Security

Security is the degree of resistance to, or protection from, harm. It applies to any vulnerable and valuable asset, such as a person, dwelling, community, nation, or organization George , Jean , & Tim (2009). Security can either be physical or logical. Physical security is concerned with access control while logical security is concerned with virus detection and network intrusion. Organizations place the management of the physical and logical security under different departments with little or no interaction at all. Separating the two types of security can compromise the security of the organization assets since physical security provide first line of defense and the logical security secures data and information. Logical security cannot be assured when there is lapse in physical security which has always been the case, management has always been wrong about physical security for instance, viewing it as an unsophisticated where physical security has always been considered as an afterthought, something to be tackled if need be after the technical issues have been resolved.

Physical security systems prevent intruders from accessing the organizational facilities by detecting intrusions, or facilitating their capture once detected. Such systems include video surveillance, sensor-base, barriers and the security guards. Physical security involves mitigation of certain risk levels just like logical security. Determination of the risk level involves carrying

critical assessment of the value of the assets as conceived by the users and the probability of compromise. The value and usefulness of an asset to users and organization as a whole is determined by considering its criticality and the ease with which it can be replaced (Gary, 2004). The protection accorded to an asset depends on the risk level. This in turn determines the security procedures, physical protection and terrorism counter measures that must be applied to secure the assets.

Most organizations such as Universities have critical infrastructures that require absolute security. Securing is the act of detecting and isolating any problem such as theft, high temperatures, unauthorized accesses etc. Detection entails announcing the existence of a problem whereas isolation involves determining the nature and the actual point where the problem occurred (Alkhateed, Maghayreh, Tubishat, & Aljawarner, 2010). According (Department of the Army, 2001), threats to physical assets such as damage and theft of computer hardware and other information systems are perpetrated by unauthorized users, employees and public and private sponsored groups with an intention of destroying information and data.

### 2.3 Security in Kenyan institutions

Kenyan institutions in general have suffered more than their share since the start of the attacks by the Al-Shabaab militia in the year 2011, recently, academic institutions have become the center of target as it has the large number of students who are vulnerable. Between 1970 and 2014, more than 3,800 terrorist attacks targeting educational institutions took place in 111 countries (Alkhateed, Maghayreh, Tubishat, & Aljawarner, 2010). These attacks comprised 2.7 percent of all terrorist attacks worldwide during this time period. Although attacks on educational targets have the capacity to be highly lethal, this is certainly atypical. In fact, the average lethality of attacks on educational targets was 0.9 deaths per attack, compared to 2.14 deaths per attack on average for all other types of targets combined (Cox, Orsborn, & Sisk, 2013).

The lethality and the deadliness of the Garissa University attack where 147 people were killed was extremely unusual and noteworthy. This attack, and the December 2014 attack on the Army Public School in Peshawar where more than 150 killed as well, are among the three most deadly terrorist attacks on educational targets on record since 1970 (Pate , Jensen, & Miller, 2015).

Nearly 70 percent of all terrorist attacks on educational targets between 1970 and 2014 (2,637 attacks) caused no deaths, compared to approximately 50 percent of attacks on other types of targets. Many attacks against schools and universities took place when the buildings were unoccupied and targeted the facility rather than individuals. This produced a considerably lower likelihood that the attack caused any casualties. Between 2004 and 2014, the percentage of non-lethal attacks against education targets actually increased to 72.5 percent, while attacks against other types of targets were more likely to be lethal than they had been previously (Pate , Jensen, & Miller, 2015). Figure 2.1 shows the range of targets on different institutions in Kenya between the years 1970 to 2003 and in comparison to the years between 2004 and 2013. It is therefore evident that the educational institutions are in so much danger as the numbers have increased exponentially.

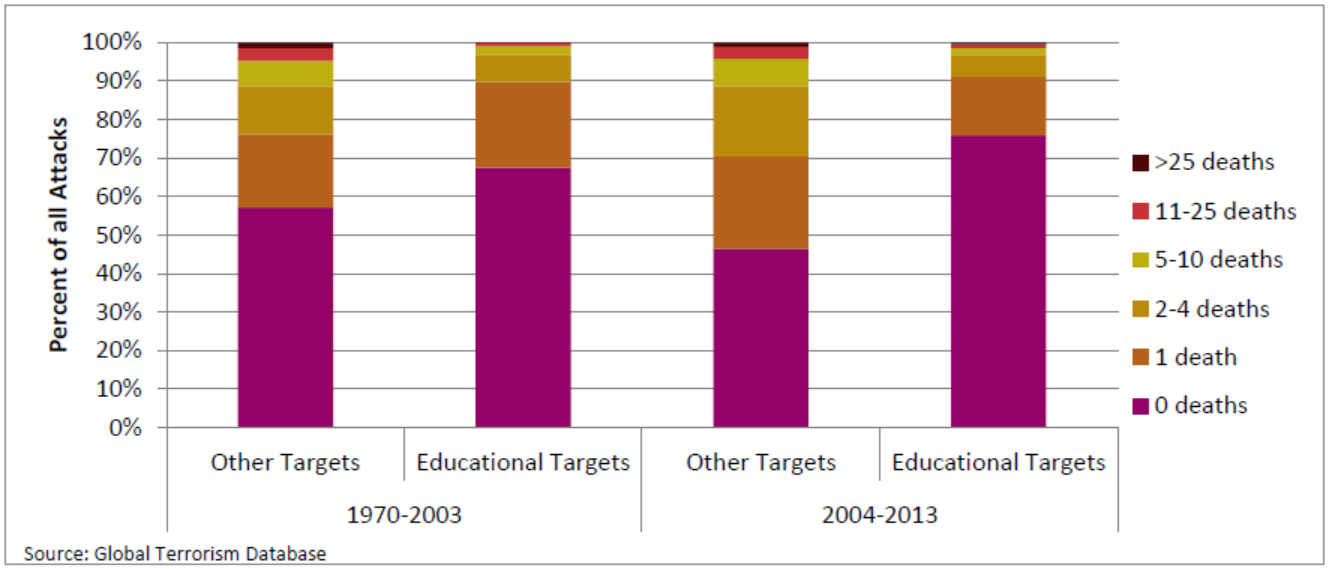



Figure 2.1: Comparison of the number of security risk over the years (Source: Global terrorism database)

It is therefore evident from the review that the academic institutions have a milestone to beat when it comes to securing themselves against the terrorist. This motivates the research to seek a smatter way of using the ignored environmental attributes such as sounds being produced to enhance physical security.

## 2.4 Approaches to improve security in educational institutions

To better the situation, the Kenyan institutions, specifically the educational institutions have tried to deploy a few measures to enhance the situation just to reduce the incidences that occur over the given periods of time.

### 2.4.1 Security guards

Security guards are people who are paid to protect property, assets, or people. They act to protect property by maintaining a high visibility presence to deter illegal and inappropriate actions, observing either directly, through patrols, or by watching alarm systems or video cameras for signs of crime, fire or disorder; then taking action and reporting any incidents to their client and emergency services as appropriate. Private security provided by guards is essential to protect intellectual property, people and sensitive data based in the various information systems and any other relevant information for the given organizations.

Organizations outsource security services to security companies or hire their own security guards. Others hire their own security personnel and at the same time outsource the services to enhance security of their assets (Kevin Strom, et al., 2010).

The use of human security guards and bouncers has become a popular method of protecting organizations critical assets. A report by (Department of the Army, 2001) postulates that human guards respond quickly to threats because of their ability to differentiate between real and false threats. However, the cost of hiring human guards is high and also considering that guards are placed between a criminal and their targets thus guarding is perceived to be a dangerous occupation.

A study conducted by Julie, Uwe, Jonathan & Kendall (2010), found that human security systems are difficult to preempt and can adapt to new circumstances. However, the rest period for the security personnel, preferences based on gender, race, age and their susceptibility to prejudices determines their level of performance.

(Waiguru, Kamunju, & Singo, 2004), noted that government's inability to protect its citizens in conjunction with increased crime rate lead to hiring of private security in organizations and homes in Kenya. Failure by state to guarantee security to its citizens warrants individuals to organize their

own security. Majority if the poor depend on informal protection such as vigilante groups and communal surveillance because they could not afford to hire private guards. However, the use of these informal groups' results into even a bigger challenge since some of them transforms to become extortionist and met violence against their rivals. The organizations and the rich make use of private security since they could afford the multinational security companies and small security firms enjoys the market share of private security in Kenya, however, the former offers high standard services than the smaller firms. Although there were over 2,000 security companies operating in Kenya, only 21 companies were registered members of the Kenya Security Industry Association (KSIA) hence hindering application of standard to regulate the industry. (Wakefield, 2005), noted the government's failure to regulate the private security industry as critical omission since it compromises security standards provided to its citizens.

Many local private security firms disregards service standards, work ethics and labor laws since they not guided by any legal framework. The companies hire untrained or inadequately trained and illiterate personnel with low self-esteem and lack motivation to deliver quality services. According to Wairagu (2009), employees of such companies are holders of secondary school certificate or less and aged between 18 and 30years. The employees lack insurance cover which leaves their dependents vulnerable in case of injuries or death. This leaves employees of this industry disillusioned and eventually colluding with criminals to steal from their employers and clients. For example, a survey conducted by Wairagu (2009), noted that G4S (a private security firm that is also based in Kenya) personnel had been robbed of cash on transit from one place to another. It was suspected that security personnel connived to steal the cash in all the three instances.

### 2.4.2 Perimeter and barrier security systems

Perimeter security is a set of physical security and programmatic security policies that provide levels of protection against remote malicious activity. Ancient people used physical security such as weapons, lake or cliff dwellings, walls and gates for perimeter security for example the Great Wall of China. Perimeter security controls are a necessary step to accomplishing a secure environment to maintain information security systems and protect data that is created or maintained by an organization.

Successful perimeter protection efforts require integrated entities which are capable of communicating seamlessly and provide real time monitoring and alerts in order to ensure quick response to crime events all the time, irrespective of the weather conditions. Organizations use perimeter security and guards who patrols the whole compound all the time.

ANSI-HSSP as per Standards (2007), refer to Perimeter Security System as the system of people, technologies, geophysical features, processes and operations employed to secure a particular security interest normally a potential target from unauthorized access, particularly premeditated attacks intended to injure, damage, destroy, or impede the normal operations of the security interest.

Perimeter security provides the first line of defense where the intruder is detected before compromising the security of the organization assets. It uses barriers for delaying intruders from achieving their goals which sometimes might be malicious. Barriers also protect the organization against stray animals such as stray dogs and runaway game animals thus reducing the probabilities of false alarm. Perimeter security cannot guarantee 100% success just like other systems because they are not accurate. This means that organizations require an integrated approach to security to reduce their vulnerabilities to attacks.

A research by Standards (2007), indicated that the yard stick to measure the effectiveness of the perimeter security is its ability to deter attacks, detect and profile threats to a given target with an aim of mitigating, capturing and/or destroying the intruders and their intentions. Deployment of benches along the perimeter fence and at close proximity to buildings as well as flower pots arranged close to majority of the buildings provides additional security. Researcher Jean, P, from University of Minnesota Jean (2005), noted that pedestrian benches are one of the most effective security control used in conjunction with other furniture, bollards and flower pots. When deployed in organization along the perimeter barrier they portray a welcome gesture for visitors and entice people to sit and relax. Benches add an extra layer of defense because no intruder would like to be seen committing crime by people sitting on the benches since the availability of an evidence would always frame them to the authorities. Perimeter barriers deter vehicles and projectiles from penetrating the organization premise by providing high anti-crash or anti-blast capabilities.

### 2.4.3 Closed-Circuit Television Cameras and Alarms

CCTV cameras are electronic systems that maintain surveillance on real time basis. CCTV technology was first used in the 1940s to monitor the testing of V2 missiles. The CCTV system allowed officials to monitor the testing at a close range without danger, watching out for defects and other problems that might have otherwise gone undetected. In 1949 the first commercial closed-circuit television system became available in the United States. By the 1960s, officials in the UK began installing CCTV systems in public places to monitor crowds during rallies and appearances of public figures (Woodhouse, 2016).

The term surveillance and monitoring are fancy terms used to describe the act of looking for intruders in the restricted areas. Anomalous condition is identified through the use of human or electronic means. Once identified an appropriate alarm is raised so that the appropriate action may be taken. The actual sighting or evidence of an intruder is an example of anomalous conditions. For example, a guard coming upon a hole cut in a perimeter fence has just found an anomalous condition (Ustun & Smith, 2010).

A study by Titus, (2016), reveals that video cameras provide real-time images of the activities across the organization. They are deployed to detect intrusion or anomalous conditions and enable verification of a threat or a false alarm by the security personnel. Cameras deter intruders from penetrating the organization since they provide real-time images. Organizations use unmonitored cameras to reconstruct events after intrusion has occurred.

Julie, Uwe, Jonathan, & Kendall, (2010), note that static sensor networks such as CCTV cameras, standard burglar alarms etc. record events, store and replay them when needed. The advantages of these security components are that they are always active and their response to situations is always predictable. Static sensors are easily replaceable when damaged and when used in systems with centralized data storage, evidence is not compromised. However, criminals can circumvent these sensors because they can be obstructed and cannot maneuver through obstacles.

A review and analysis of Public Area CCTV and Crime Prevention report by Welsh & Farrington (2009), suggest that CCTV caused a modest 16 % decrease in crime in experimental areas compared with control areas, the warning that a given area is under the surveillance of CCTV has also enabled to increase the human compliance to the security measures that have been put in place.

This overall result was largely driven by the effectiveness of CCTV schemes in car parks, which caused a 51% decrease in crime, again as a result of compliance. The use of static sensor network is limited since an individual cannot monitor many sensors at the same time and the volume of information generated by those sensors. A study by Julie (2010), estimated that a single operator can effectively monitor approximately 16 sensors feeds, with the detection rate falling from 83% for a four camera system, to 64% for a 16 camera system. The study also revealed that short range sensors are more effective when mounted on a robot, because the sensor moved to the actual target. In order to deal with the challenges of the camera surveillance an integrated approach to security require to be adopted which is believed to be more reliable.

## 2.5 Challenges of security in educational institutions

The results shown in Figure 2.1 is a result of the existing gaps in the current security structures that already exists in the Kenyan institutions. Some of the gaps incudes the over reliance of the institutions in the security guards. In fact, collaboration between private security and public security has been a big challenge in trying to maintain law and order. According to Ohlhausen (2004), lack of information sharing among the public and private law enforcement agencies coupled with mistrust and misinformation hinders effective collaboration to enforce security. Police officers feel that private security personnel have infiltrated their territory and are generally illiterate and nonprofessionals who are ineffective in their duties (Idriss, Jendly, Karn, & Mulone, 2010). On the other hand private security personnel believe that police do not understand the role they play to compliment crime prevention.

In most of the institutions, during the routine patrols, the way the security checklist is being marked at the time the guards are doing their regular patrols can be easily predicted by a trespasser since it is always structured, procedural and is always repeated the same way every single day. Trespassers may find it easy to make their way into the restricted areas with simple techniques such as piggy tailing and tailgating since most of the times, due to lack of enough training, the security guards are usually naïve and misinformed. Some of the physical security measures that have been put in place are not enough to assist the security personnel to tell whether there is a genuine security breach or not hence leading to ignorance of these situations most of the time.

## 2.6 Sound technology and security

Sound can be recognized in different classes, to find out which acoustic environments are critical in everyday life and thus most important to be recognized, it seems a good idea to ask hearing impaired persons. This has been done for example in a study from Fedtke (1991). Subjects with a moderate hearing loss were asked to judge how important it is to hear well in 52 different situations in the area of home life, work, culture, leisure time and traffic. The situations judged most important can be roughly divided into four classes:

i. Speech (dialogue, lectures, theater, cinema, phone calls, television)
ii. Speech in noise (cocktail party situation, announcement at train station or airport, speech in a car)
iii. Alarm signals (ringing phone, doorbell)
iv. Nature (chirping birds).

The first three of these classes contain essential information for the people. The class ‘nature’, however, shows their desire for a certain listening comfort. In this context, it is a bit astonishing that ‘music’ is not named as an important sound class as well. This is probably due to the sounds that were presented in the study; no musical signals had to be judged apart from ‘singing in a theatre’, which was regarded as being quite important. Thus, it is assumed that music sounds also belong to the more important situations.

Haubold et al. (1993) developed a new hearing instrument fitting procedure based on natural acoustical patterns. They propose to use eight different classes for the fitting, which included ’speech’, ’speech in noise’, ’noise’, ’warning signals’, ’nature’, and ’music’.

The classes ‘speech’ and ‘speech in noise’ are again situations that are apparently important for communication. It is of course also important that warning signals such as car horns, phone or door bells etc. can be heard. These sounds can be very short. The concept of an automatic program switch in the hearing instrument, however, will probably be that it reacts to events that remain stationary over a longer period of time, in the order of ten seconds or longer. The class ‘alarm’ is therefore a special situation which will be omitted in a first approach, assuming also that the hearing instrument will anyway amplify such sounds by default.

Using these classes, sound can be used to detect the type of threat it poses depending on the environment that it is. For example, in academic institutions, some sounds like those that are produced by firing a gun is not expected and thus the class can be used to group it as a threat in that environment.

## 2.7 State of Art in sound Classification and existing Algorithms for sound classifications

The general structure of a sound classification system can be described with a block diagram, as it is shown in Figure 2.2. From the sound data, a number of characteristic features are extracted, which are then classified with some sort of pattern classifier. An optional post processing step may correct possible classification outliers and control the transient behavior of the algorithm. The output of the algorithms are the recognized sound classes.

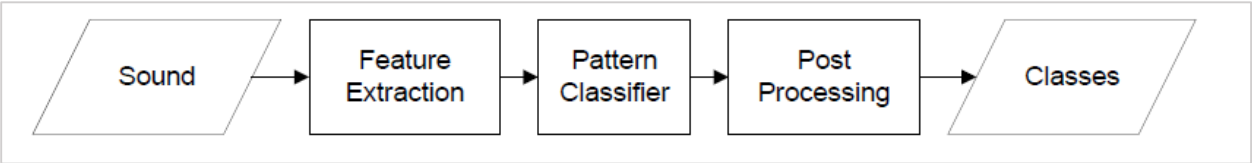



Figure 2.2: General block diagram of a sound classification system. (Source: State of the art in sound classification, Haubold et al. (1993))

Sounds have been used to aide hearing in hearing instruments before and different classifications have been used. Currently, five known methods for sound classification in hearing instruments have been found and three of them are already exploited in commercial hearing instruments, the analysis of the amplitude statistics by Ludvigsen (1993), the classification based on temporal fluctuations and spectral form by Kates (1995) and further developed by Phonak (1999), and the analysis of the modulation spectrum (Ostendorf et al., 1997). The two other algorithms are also designed for hearing instruments, but not exploited so far (Feldbusch, 1998, and Nordqvist, 2000).

The feature extraction blocks of the three already exploited approaches will be evaluated and compared in this paper. It will be shown that they are related in that most of the features described in these algorithms represent the amplitude modulations in the signal, and that this enables the discrimination of speech signals from other sounds very well. A more detailed classification of the acoustic environment is however hardly possible with these approaches.

### 2.7.1 Feature Extraction statistics

Ludvigsen (1993) proposes to automatically control the amplification and/or the frequency response of a hearing instrument by investigating the continuity of the input signal; that is, by discriminating impulse-like and continuous signals. He does this by investigating the amplitude statistics of the signal.

Ludvigsen states that the amplitude histogram of more or less continuous signals, like background noise and certain kinds of music, shows a narrow and symmetrical distribution, whereas the distribution is broad and asymmetric for speech or knocking noises. The examples in figure 2.3 and figure 2.4 show the amplitude histogram of speech, party noise and speech in party noise. The histograms were built over thirty seconds of the envelope of each signal.

Due to the pauses in the speech signal, its level varies very much over time, resulting in a broad and asymmetrical amplitude histogram. The level of the party noise is much more constant, that is, the amplitude histogram has a narrow and symmetrical form. The speech in party noise signal is a bit broader, but still symmetric; the two modes are not typical for speech in noise sounds.

In addition to the histograms, some percentiles are also drawn in the figures. The 30 % percentile, for example, shows the level below which the envelope is 30 % of the time. The asymmetrical distribution in the speech signal results in a much larger distance between the10 % and the 50 % percentile than between the 50 % and 90 % percentile, or, in other words, the 50 % percentile is far away from the arithmetical mean of the 10 % and 90 % percentile. For the noise and the speech in noise signals, the 50 % percentile is more or less in the middle of the 10 % and the 90 % percentile, representing the symmetrical distribution.

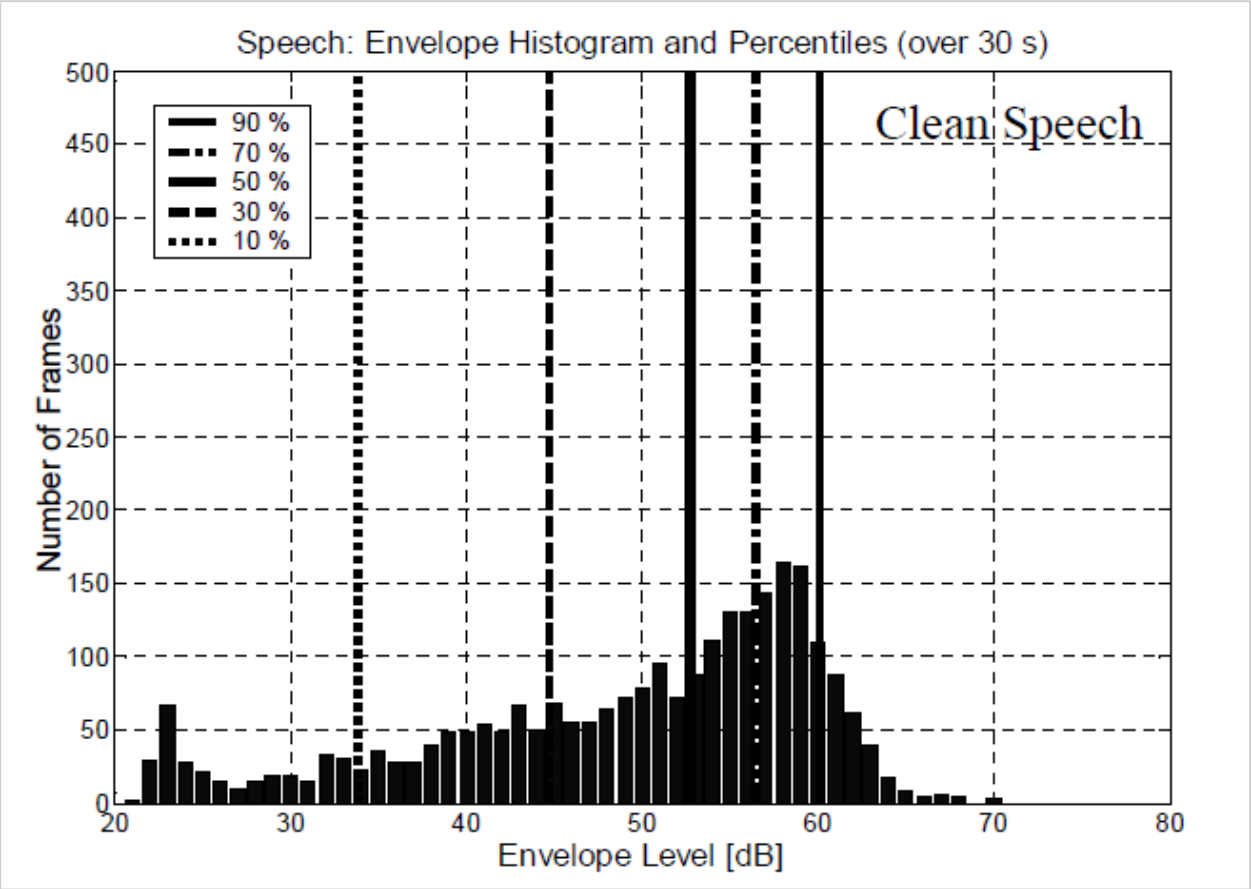



Figure 2.3: Amplitude envelope histogram of clean speech acquired over thirty seconds. (Source: Amplitude statistics research)

Due to the pauses in the speech signal, the histogram is very broad and asymmetric. In addition, some percentiles are plotted. The distance between the 10 % and 50 % percentiles (or 30 % and 50 %) is much larger than the one between the 50 % and the 90 % percentile (or 50 % and 70%).

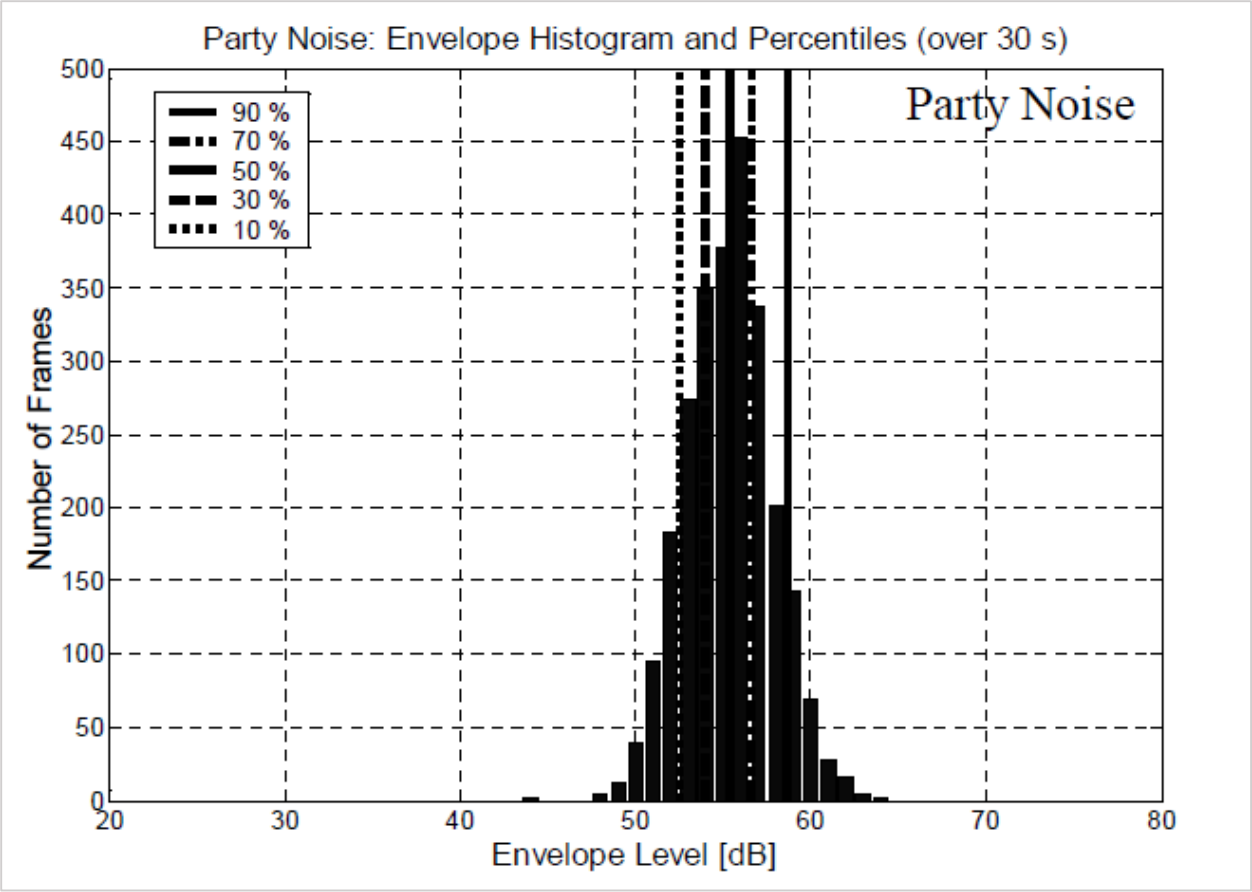



Figure 2.4: Amplitude envelope histogram of party noise acquired over thirty seconds. (Source: Amplitude statistics research)

### 2.7.2 Modulation Frequency Analysis

Ostendorf (1997) investigated modulation spectra of different signals and confirmed that they show systematic differences. The goal was to distinguish speech, speech and noise, and noise signals to allow the automatic control of the compressor settings of a hearing instrument.

The modulations of a signal, described by the signal envelope, are characterized by the modulation frequencies and the corresponding modulation depths. The modulation frequency denotes the velocity of the modulations, and the modulation depth denotes the strength of the modulation. It has been shown that different signal classes exhibit different characteristics in their modulation frequency spectrum. The envelope of speech for example is determined by the phonemes, the syllables, the words, and the sentences. Normally we articulate about 12 phonemes, 5 syllables, and 2.5 words per second. To formulate sentences, several seconds are required. Thus, speech has

modulation frequencies of approximately 12 Hz (phonemes), 5 Hz (syllables), 2.5 Hz (words), and < 1 Hz (sentences). Due to the speech pauses, the modulation depth of speech is large (Holube, 1998). The maximum in the modulation spectrum of clean speech is in the area of 2 to 8 Hz. Note that this corresponds very well to psychoacoustic findings: the maximum sensitivity to fluctuation strength occurs at 4 Hz and indicates the excellent correlation between the speech and the auditory system (Zwicker and Fastl, 1990). By way of contrast, noise shows often weaker but faster modulations and has therefore its maximum at higher modulation frequencies. Hence, modulation frequencies and the corresponding modulation depths represent a powerful feature for the perception and discrimination of sounds.

The modulation spectrum was first calculated as shown in Figure 2.5. The envelope of the signal is scaled to its root mean square and Fourier transformed. The third spectrum of the absolute values of the FFT bins is then calculated.

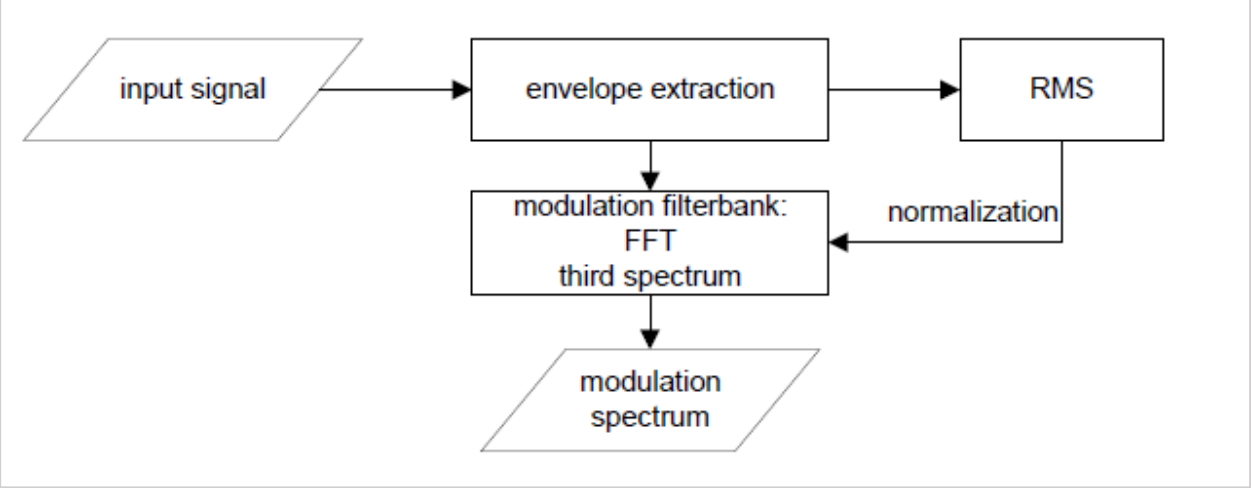



Figure 2.5: Block diagram for computing the modulation spectrum (Source: modulation frequency analysis research, (Ostendorf 1997)).

The envelope of the signal is scaled to its RMS and Fourier transformed. The absolute values of the FFT bins are combined to the third spectrum.

### 2.7.3 Noise Classification with Neural Networks by Feldbusch

An algorithm for automatic switching between different hearing programs was presented by Feldbusch (1998). The main classes to be identified are speech, babble noise and traffic noise. Further classes are speech in babble noise, speech in traffic noise, music, nature, and possibly alarm signals.

A large set of features is calculated both in the time domain and in the frequency domain. In the time domain, the zero-crossing rate, the maximum of change of the zero-crossing rate and some derivatives are computed, in the frequency domain, either Fourier or wavelet coefficients are taken as features, together with some features that are extracted from each of the coefficients, such as the mean and maximum of each coefficient in a certain time window. Feldbusch states that for practical applications, the features should be independent of the signal level.

This results in over 150 different features that are fed into a neural network. Several network topologies have been tried; a so called Time Delayed Neural Network (TDNN) was used to analyze the temporal structure in the features. The classification with the TDNN was quite poor, however. Feldbusch assumes that some of the features express already the temporal dynamics of the signal, and that the temporal changes of the other features do not contribute much to the classification.

The best results were obtained with a neural network with one hidden layer. The classification of the main classes was quite good, for music and nature however bad. Feldbusch states that these classes may contain very different signals, which makes it difficult to cluster the signals into a class. The large number of features might make it difficult to train the network properly; a pre selection of the best features is currently in progress (Feldbusch, 2001). Feldbusch recommends to apply a post processing stage at the output of the network to make the system more inert and robust.

### 2.7.4 Noise Classification with HMMs by Nordqvist

Nordqvist (2000) designed another automatic classifier for different listening situations, which shall enable the hearing instrument to switch between different filters, look-up tables or other settings, such as directional microphone, noise reduction and feedback suppression. The classes are speech, babble noise, traffic noise, subway noise and outdoor noise.

The algorithm is based on hidden Markov models (HMM). As features, LPC coefficients are used, which are vector quantized before being fed into the HMM. A post processing block controls the switching from class to class (that is, between the HMMs), in order to allow different switching delays for different class transitions.

The long-term classification error is close to zero. At first sight, it seems however quite easy to recognize two of the classes: Subway noise contains very low frequencies, and outdoor noise probably just means that the average signal level is low. Anyway, the use of a HMM as classifier appears to be a good way of identifying the temporal structure that lies in the features. It would be interesting to know how well other noises or musical signals can be recognized with this approach.

### 2.7.5 Environmental Noises and Alarm Signals

Goldhor (1993) calculated cepstral coefficients for 23 familiar environmental sounds, such as door bell, ringing phone, car engine, vacuum cleaner, running water, closing door, etc. He then performed a cluster analysis and found that only the cepstral coefficients representing low frequency spectral and temporal variations were required in order to obtain accurate classification. Obviously, cepstral coefficients are useful to separate transient noises with quite different spectral and temporal variations. This might not be the case for more stationary sounds, like speech in noise and music.

Gaunard (1998) and Couvreur (1998) present a method for classification of five types of noise events: Car, truck, moped, aircraft, and train. These noises all have a transient nature, as the recording was made with the vehicles passing by. The best performance was achieved with LPC coefficients and a five-state left-right HMM. For this type of noise events, a HMM seems indeed the best solution, as it is able to model the temporal structure, that is the different phases in the transient noises.

Oberle, Kaelin (1995) and Oberle (1999) tried to identify four different alarm signals: Car horns, streetcar bells, streetcar rings, and phone rings. Cepstral LPC coefficients and the energy were taken as features. An ergodic HMM with four states outperformed a minimum distance classifier and a neural network. Again, the sounds have a transient nature and are quite short, with a silent phase at the beginning and the end of the alarm signal.

It would be interesting to see how a HMM is suited to model more stationary noises and sounds of other classes. There, the temporal structure has quite a different character; it is more the fluctuations within the sound itself that contributes to the structure than the sound appearing and fading out again.

## 2.8 Review of the existing solutions for sound classification and security

Sound classification has been used to solve problems in many areas such as in the medical area, military applications and now in the security sector. Some of the applications have been review in this section stating their advantages and their short comings.

### 2.8.1 Sound classification for event detection

This system shows the application of sound classification into the medical field, where patients are being telemonitored in the hospitals. This system, proposed by Phuong and Dat (2013), uses the results of their study in order to build an automatic alarm system using sound analysis which can be used for medical telemonitoring. The system raises the alarm if it detects abnormal sounds coming from abnormal situations in a patient or an aged person's room. Their system consists of building three sound corpora and developing a sound analysis algorithm consisting of three sound discriminators. The obtained results of above 95% events correct recognition rate are promising for the first studies.

The system proposes that the problem of detecting abnormal sounds could be solved by classifying sounds. From Figure 2.6, sounds are firstly separated into speech and nonspeech. At this step, there should be a speech/nonspeech discriminator. Secondly, in their turns, speech sounds are divided into normal speech and abnormal speech; and nonspeech sounds are split into normal nonspeech and abnormal nonspeech. At this step, it requires two other discriminators: normal/abnormal speech, and normal/abnormal nonspeech. In short, it is obvious that those discriminators should be the main components of our sound analysis algorithm.

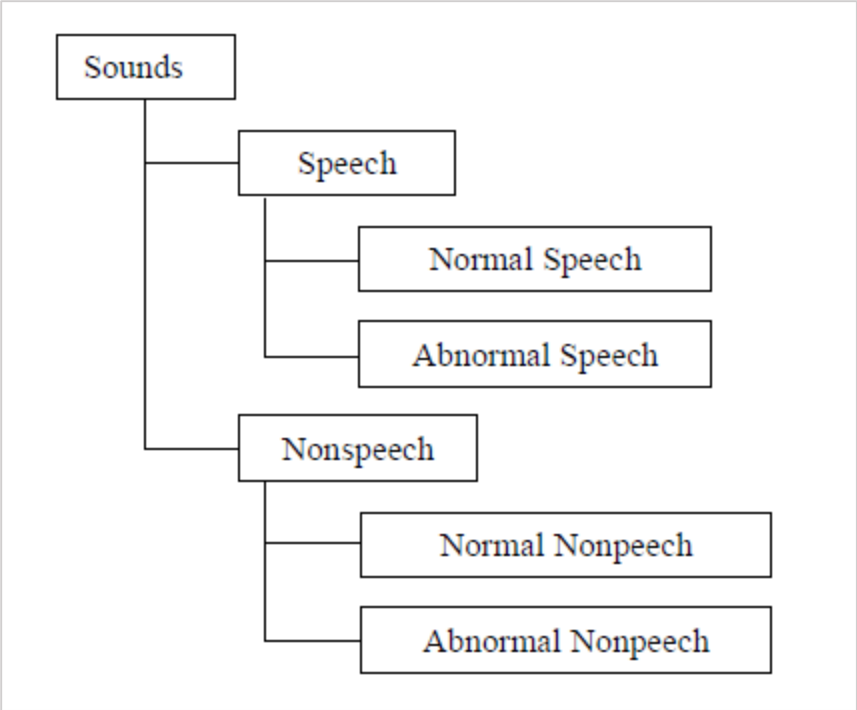



Figure 2.6: The sound classification used in the system.

Their work involved the use of study and developing the three discriminators, which is needed to build four sound corpora: normal speech, abnormal speech, normal nonspeech, and abnormal nonspeech. They used two approached to build the corpora;

i. Collecting sound corpus from existed sound corpora or sources.
ii. Recording a new sound corpus.

Four corpora are all mono, sampled at 16 kHz, and quantized at 16bit.

**The Abnormal Speech**

Abnormal speech is the sound coming from patient's throat when he/she is in bad health or in a serious situation, e.g. fall, faintness, sick, etc. Firstly, we think of recording this corpus in a hospital. But in Vietnamese hospitals, setting up and operating a sound recording system in a patient's room is nearly impossible, due to the violation of privacy and the bad condition of those hospitals. Then raises another way: building the corpus in a professional studio. But it is found that that work is a very hard one, because it is difficult to imitate those abnormal sounds. At last, we come to the following final solution. Signals of this corpus are collected from internet, films, and

sound effect CDs. It consists of 445 signals (cough, cry, gasp, groan, moan, hiccup, scream, vomit) with a total duration of about 21 minutes, an illustration is as shown in table 2.1

Table 2.1: The abnormal speech corpus

| | **Number of Signals** | **Duration (s)** |
|---|---|---|
| Cough | 51 | 116 |
| Cry | 58 | 234 |
| Gasp | 16 | 14 |
| Groan | 50 | 103 |
| Hiccup | 12 | 9 |
| Moan | 51 | 248 |
| Cream | 191 | 495 |
| Vomit | 16 | 47 |
| **TOTAL** | **445** | **1266** |

**The normal speech**

Normal nonspeech is the sound coming from things in patient's room when he/she is in good health, or sounds of normal life. Those audio signals could originate from, for instance, doors (closed), doorbell, chairs (dragged), drawers (opened and closed), liquid (pour in and out), glasses, cups, dishes, bowls, thermos flask, etc. the study proposes that the patient's room of the 108 Hospital in Hanoi, Vietnam, had been considered as the pattern room. They collect things from the patient's room, such as bed, chair, table, cupboard, glasses, cups, etc., then record their sounds or sounds from their clashes. The corpus was recorded in the recording studio. Two high quality microphones were used together to capture sounds. The first one was positioned 40cm away from objects, and the other 1.5m. In the study, they also recorded sounds of objects in bathroom, such

as shower, faucet, flushing of toilet, etc. To enrich this corpus, they also collected those sounds from internet and sound effect CDs. Finally, they obtained a database containing 194 signals (about 29 minutes in total), which was summed up as seen in table 2.2.

Table 2.2: The normal speech corpus

| | **Number of Signals** | **Duration (s)** |
|---|---|---|
| Appliances | 17 | 81 |
| Bathroom | 40 | 86 |
| Doors and Drawers | 25 | 23 |
| Kitchen | 48 | 62 |
| Liquid | 18 | 58 |
| Others | 46 | 1606 |
| **TOTAL** | **445** | **1266** |

Finally, the system detects the abnormal sounds by classifying sounds into normal (speech and nonspeech) and abnormal (speech and nonspeech) sounds hence used to report an emergency.

### 2.8.2 Utilization of Audio Source localization in security systems

Nowadays very dynamical development in electronics and computer science enables applying of the sound localization systems in areas where it was impossible due to technical and economic aspects several years ago. These areas include applications in security, teleconferencing and robotic systems where information is coded in audio signal source position. The findings of, Dostálek, Vašek, Křesálek, & Navrátil (2015) suggests a research that deals with the utilization of audio source localization in security systems especially dedicated for additional securing of larger objects like squares or military basis for instance. They suggest that an intruder usually makes some noise which can be picked by the microphone array which can then be used to redirect the cameras to pick the location of the intruder.

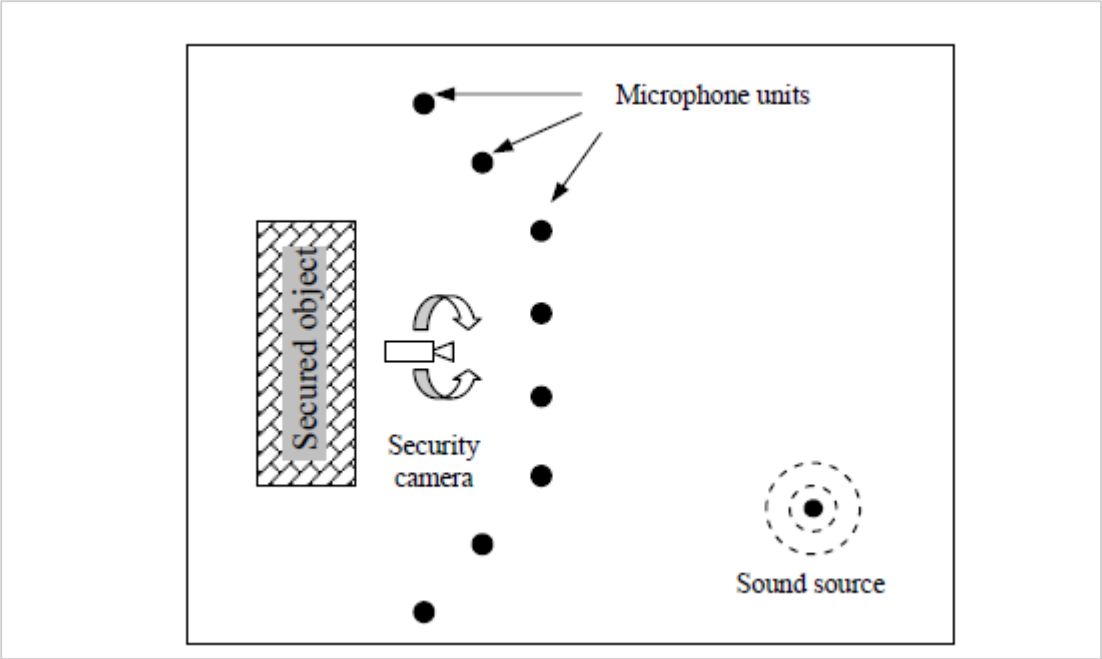



Figure 2.7: Security system structure. (Source: Dostálek, Vašek, Křesálek, & Navrátil (2015))

Figure 2.7 shows the system's structure and how the sounds triggered will assist in diverting the cameras to point into the problem which is the intruder. It is evident that where the secured object are placed microphone units and connected with evaluation unit which is installed in the security station. Audio event occurred outside the microphone array will trigger audio localization process resulting in determination of audio source azimuth. The paper also suggests that the evaluation unit can automatically point connected security camera to event position and inform operator. Structure of the whole localization system depends on the size of secured area. The researchers also proposed that for large objects it is better to use decentralized structure with more local localization units due to less requirements to system implementation.

### 2.8.3 Mobile-Based Security Agency Sound Monitor and Alert System

In this system, Machanje (2014), proposes a Mobile-Based Security Agency Monitor and Alert System that relies on the integration of the vital solutions of Natural Language Processing (NLP) in voice and sound recognition, GPS location services, and message broadcasting to detect sound variations in the environment and alert security agents at the user end. He states that an automatic analysis of the recorded sound is responsible for determining the need for a notification on the admin end, the one who controls the system. He also suggests that the abnormal sound variations

based on the pitch measured in decibels on the user end are responsible for alert notifications on the admin end that triggers immediate response by security guards on the ground. This proposed model uses android and is built to fit into the home environment which has been personalized to an individual's mobile phone. Figure 2.8 shows the system model proposed by Machanje to help in improving security in the victims' homes.

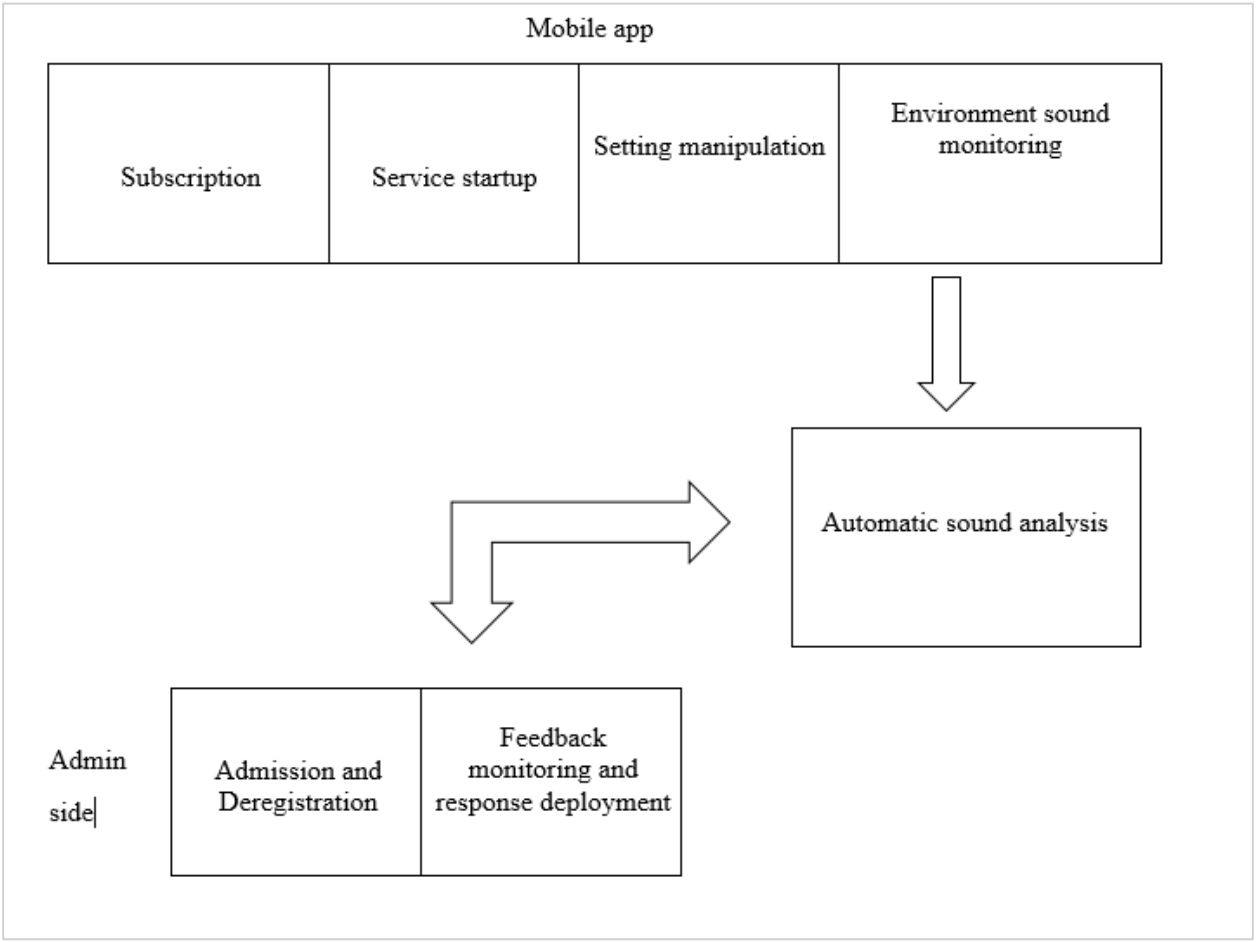



Figure 2.8: Mobile-Based Security Agency sound and alert system architecture. (Source: Mobile-Based Security research, (Machanje, 2014)).

## 2.9 Conclusions

The reviews from other papers and works done before that has been carried out contain significant knowledge on the area pertaining the proposal and implementation of this project. The research findings has contributed to determine the methodologies and technology choices that has been employed in the scope of this research. For example, the utilization of audio source localization in

security systems, the use of the microphones as actuators to get the sounds from the environment has been borrowed in this study to get the sound signals as well. This is cheaper in terms of implementation and productivity hence being borrowed in the proposed system. The main challenge with this system is that it has been tailored to fit into the army environments and has not been tailored to fit into the local scenario such as the academic institutions.

The use of sound input in decibels as used in the mobile-based security agency sound monitor and alert system has been employed in the system to input the sounds into the system for classification. This has been made easier with the use of artificial neural networks which has been made possible through the use of supervised learning. However in the mobile-based security agency sound monitor and alert system some of the challenges comes out clearly for example;

i. It is not easy to use ones phone in case of a panic attack as everyone would run to save their lives instead of alerting the system application in the phones
ii. The technology moves fast in the current technological age and this would want the system to be upgraded after a very short time. For instance, in 2014 when the system was developed, the android version that existed as the latest was android version 4 kit Kat. Currently there has been procedural upgrades to version 6.0 Marshmallow and 7 Nougat which has outlived some mobile applications that existed before.

The proposed system, sound analysis to enhance physical security has been tailored to use the good features of the previous works and to improve the disadvantages that exists in those systems. The system is cheap and readily exists in the confined areas in the institutions hence avoids the use of human effort to trigger the requests to the server. This will in return help the security personnel in the ground to enhance the physical security that is already in place.

### 2.10 Conceptual framework

This model showed in Figure 2.9 shows the workings of the system and how the layout has been implemented. From the communication of the sensor that has been simulated by the microphones, the sounds collected are sent to the server through a secure network. The server through supervised learning learns and clusters them to make a valid decision that helps the security personnel to make a move that enhances physical security that is already in place.

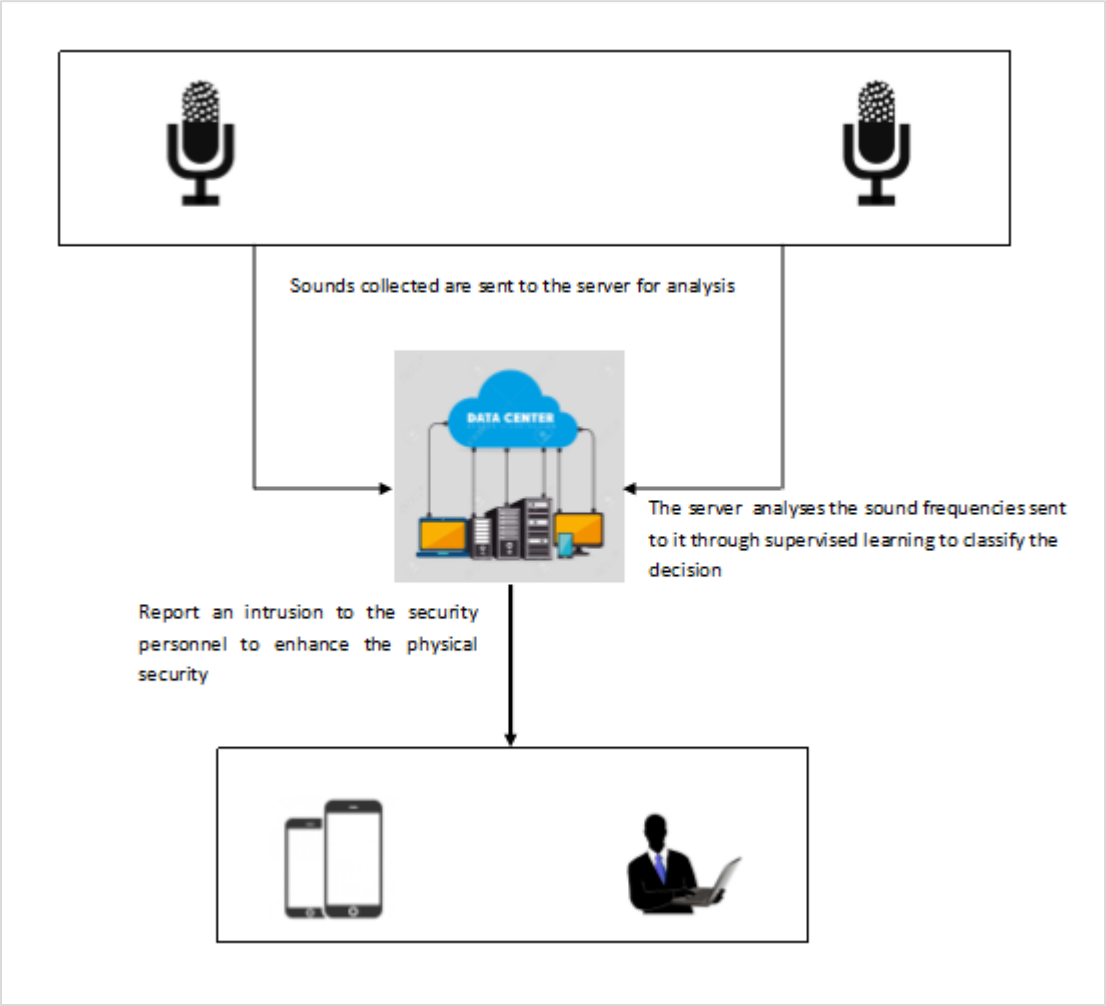



Figure 2.9: Conceptual framework for sound analysis

# Chapter 3 Research Methodology

## 3.1 Overview

This chapter presents the research methodology used in this research. It outlines the research design, approach and strategy, the research methods used, a discussion of the methodology, the methodology used to develop the system and the system architectural design.

## 3.2 Research design

A research design is a plan according to which one obtains research participants and collects information from them (Mugenda & Mugenda, 2003). The proposed system has used a scientific approach to achieve the results displayed as output since it involves doing experiments on the system to determine the outcome. The scientific research method involves the investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws (Dobbins, 2004).

The use of the scientific approach is relevant since it reduces the ideas into a small, discrete set to tests, such as the variables that comprise hypotheses and research questions with a set of rules and guidelines. These set of rules are then tested, analyzed and tested to prove the proposed facts.

## 3.3 System development methodology

This research has adopted the incremental model of system development. This model ensures that the software is made in increments, that is, initially the software is delivered with the basic requirements for example module or class delivery as the version management takes place. This enables the appropriate use of the limited time and the number of the developers available. A research done by Prakriti & Sharma (2013) suggest that the incremental model uses less resources when trying to maximize the productivity of the development process, this has also motivated the choice of his system development methodology. Figure 3.1 shows the different development stages that has been incorporated in the study for the development process.

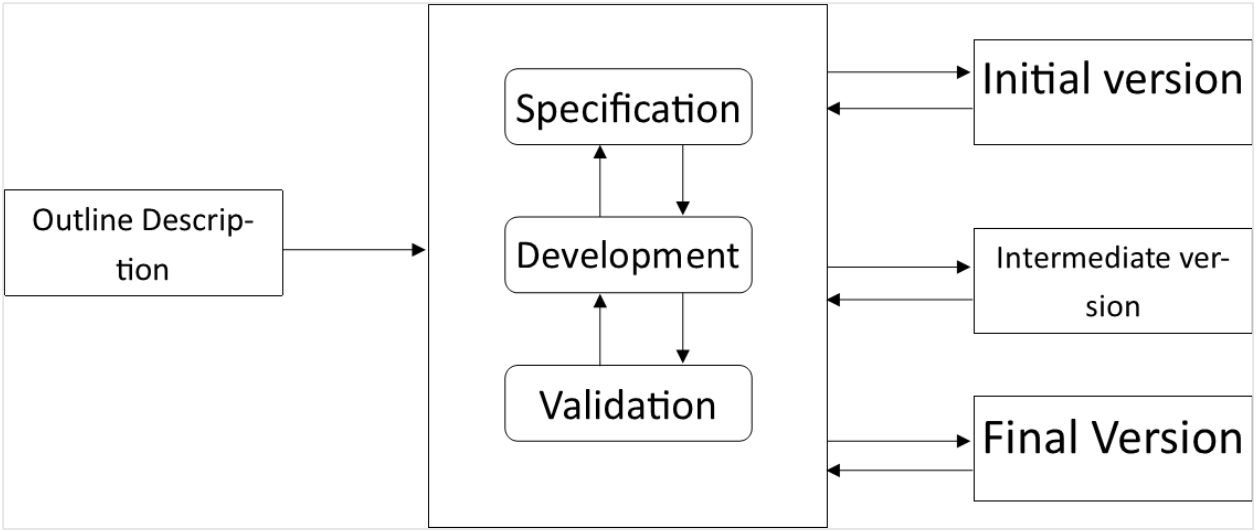



Figure 3.1: Incremental system development methodology

With the ability to initially develop an initial implementation, expose it to user comment for acceptance, and evolve the development until the full development is fully achieved will enable the system to get better and better every time it is released hence having the full functional system in the final release. The choice of this methodology has been influenced by the advantages such as the cheaper cost of accommodating changes in case of any, it is much easier to get the feedback from the users since there is rapid testing at each release.

### 3.3.1 Outline Description

Outline description shows first step of the system development process, at this point, a sketch of parts or the whole system is discussed. In the proposed system, the requirements were initialized at this point and discussed in detail to measure the weight so as to come up with the hierarchy of priority during development

### 3.3.2 Specification

Expansion of the requirements gathered at the description outline were expanded on at this point, the researcher ensured that they were revised on to avoid ambiguity and conflicting requirements. At this stage, the priority set was revised to make sure that during system development, the development curve runs smoothly from the beginning to the end saving on the time allocated for the project to be completed

### 3.3.3 Development

Active programming happens at this stage, in the proposed system, all the libraries were set, all the dependencies were set and most of the code was written to bring into reality, all the specifications set at the specifications stage, preliminary system runs was done to make sure that the code was working and preliminary results were achieved. Test classes were set to further make the results real as they helped the developer stay on the expectations scope.

### 3.3.4 Validation

The developed code was tested to make sure that the results being returned is as expected. The validation involved adding the constraints in the proposed system making precise and accurate results.

### 3.3.5 Initial version

The initial system version or prototype of the system is developed at this point, the basic system requirement are implemented to set the floor for the improvements to take place. In the proposed system, the work framework was set to make it possible to use the existing open source python libraries. This was done by installing python as a service in the hosting machines Linux operating system

### 3.3.6 Intermediate version

The initial version developed was tested and recommendations that came up were reviewed at this point, this was done in a procedural manner to make the changes coincide into the already built system's structure. In the proposed model, Github was used to control the version control and errors were solved in an organized manner.

### 3.3.7 Final version

All the fixes have been worked on and the final design and considerations have been arrived, the product can now be launched for use at the end of this stage. The proposed system was launched at the end of this stage and the performance measures were analyzed to keep making the system better in efficiency and accuracy.

### 3.4 Data collection instruments

During the research secondary data sources were used. The experimentations allowed the researcher to find the best datasets which was urban sounds to use as analyzing different perspectives of data was helpful.

Availability of the secondary sources most of which have been discussed in chapter two enabled the researcher to compare different algorithms and procedures used before to extract sound and their applications to security. Additionally, these sources also assisted the developer realize the best open source software to use to develop the system's prototype.

### 3.5 Data analysis and Presentation

The proposed model uses the training data as a base comparison to the data being input into the system, using artificial neural networks (ANN), for supervised learning, the researcher was able to draw conclusions from the results which were presented as arrays from the system as recommended by the library used (Google Developers, 2017).

### 3.6 Ethical Considerations

Prior to the development of the system and any other deductions made during the research, as a recommendation by Creswell (2003) in his report, the researcher made sure that all the relevant parties involved were consulted and their consent was taken into account. This ensured that no privacy issues and any other ethical aspects were overridden thus enhancing the legitimacy and the validity of the research.

### 3.7 Research Quality and Reliability

The reliability of the sources of information from the experiments, the location of study, the research instruments, and any other concerned research aspect was guaranteed. This was based on the approval of the same from the academic circles concerned with supervision, and other sources of knowledgeable information featuring the same methods in previously successful research studies as discussed in chapter two.

## 3.8 Summary

Having looked at the various research design methodologies discussed and used in this study, there is little doubt that any areas were left not done in a bid to find a solution of an integration of the vital solutions of Artificial Neural Networks (ANN) in sound classification. The results shows the output after the processing was done and this verifies that the research was thorough.

# Chapter 4 System design and architecture

## 4.1 Overview

This section of the study contains the details of the design of the proposed solution by incorporating the system requirements as per the design methodology in the previous chapter. The requirements of the system has been selected and analyzed through the study of the previous works done in a similar way. The design diagrams though the use of Unified Modelling Language (UML) has been drawn detailed information for each design diagram put down. The design diagrams and structures put down for the design purpose includes the use case diagram with detailed follow-up use case descriptions, systems sequence diagram, data flow diagram, entity relational diagram (ERD) and class diagram. The use of these diagrams shows how the system modules interact with each other and the data flow from one entity to another.

## 4.2 Requirement analysis

Based on the literature review and the research methodology alongside the initial study objectives, this section encompasses the task that the system needs or the conditions to be met in order to make it work as initially proposed in the study. This section comes in hand for the success of the development project. The systems requirements in the proposed systems are actionable, measurable, testable, related to the identified business needs and opportunities. These requirements have been classified into either functional, non – functional, performance requirements or design requirements.

### 4.2.1 Functional requirements

Functional requirements explain what has to be done by identifying the necessary task, action or activity that must be accomplished. Functional requirements analysis has been used in the proposed system as the top level functions for functional analysis. Some of the functional requirements in the system includes;

i. Start service and stop service – this functional requirement defines the runtime and the idle state of the system that it starts to listen and stops to listen to the sounds from the environment
ii. Sound listening and recording – this functional requirement is a main objective of the system, it listens to the sound variations from the environment which is the area

confined by the educational institution then record the sounds in preparation for analysis in the application's model.

iii. Sound automatic analysis – this functional requirement makes sure that the sound captured from the recorder is sent to the systems model for analysis. Upon reception, the artificial neural network learns from the stored dataset in comparison to the input and a verdict is issued as an output. This is done though supervised learning.

iv. Storage of missing sounds in .wav or .wave format – this functional requirement learns from the dataset and makes a decision based on the available library. When the library is limited such that the sound cannot be classified, the system saves the sound as a new sound and this can help in forming a new dataset hence making future verdicts even more accurate.

v. Verdict processing and output – this functional requirement makes sure that the processed data is classified appropriately and that the verdict is relayed to the security personnel so that they take an appropriate action in terms of enhancing physical security.

### 4.2.2 Non – functional requirements

Non-functional requirements used in this research have specified the criteria that has been used to judge the operation of the system, rather than its specific behaviours. Some of the non – functional requirements that has been applied to the proposed system includes;

i. System availability and accessibility – the system has to be up and available to execute the commands or the information being sent to it at any times. This ensures that maximum amounts of data will be collected since the ready state will most of the time be available.

ii. Feedback – the system needs to provide necessary feedback that the security personnel can use to make a valid verdict at the end of the sound analysis. This makes sure that the system has met its objectives

iii. Performance and reliability – the system needs to meet its operations at a 100% efficiency in its monitoring, this ensures that the systems ready state, processing state and finally at the point of delivery it is able to handle the input, processing and the output efficiently for decision making.

### 4.2.3 Performance Requirements

Performance requirements shows the extent to which a mission or function must be executed in order to achieve goals. This is usually measured in terms of quality, quantity, coverage, timeliness or readiness. In the proposed system, the performance requirements has been interactively developed across all the identified functions based its life factors, and characterized in terms of the degree of certainty in their estimate, the degree of criticality to the systems success, and their relationship to other requirements. Some of the performance requirements used in the proposed system includes;

i. Response time – the systems response time to a new request which in this case is to record a new sound is as soon as the previous verdict will have been availed to the security personnel. As for the sound analysis, the dataset has been subdivided into sections to make the analysis faster as compared to putting up all the datasets into one section, this causes delay and hence making the sound analysis model slower.
ii. Workload – the proposed system is capable of analyzing one sound at a time, this is so because when sound is recorded it is saved as a sound wave and then sent for analysis. The comparison is done across 8042 sound waves that makes up the dataset and this is done through the supervised learning technique.
iii. Scalability and platform considerations – the proposed model is quite heavy and is faster when run on a Linux platform as compared to Microsoft's Windows operating system. Linux make the system light weight and this enhances productivity. The scalability of the proposed system is relying on the use of a Linux or UNIX operating system as the platforms are light weight and are capable of making the system run faster.

## 4.3 System Architecture

This is a representation of the proposed system in a chart manner. It shows how the systems components interact from the point of sound input, its analysis and finally the output. The input processes in the proposed system includes the capturing of the sound from the environment, this is done by a sound capturing hardware such as through the use of a microphone.

The recorded sound is them sent to a temporary location having been saved in the .wav or in .wave format to allow analysis and comparison between the recorded frequencies and the saved frequencies in the dataset. This is done through artificial neural networks. Finally the analysis is then sent for decision making and the decision that is in the form of an output message is then sent to the security personnel. At the end of the input to output process, a new process starts as the system goes back to its original listening/ready state.

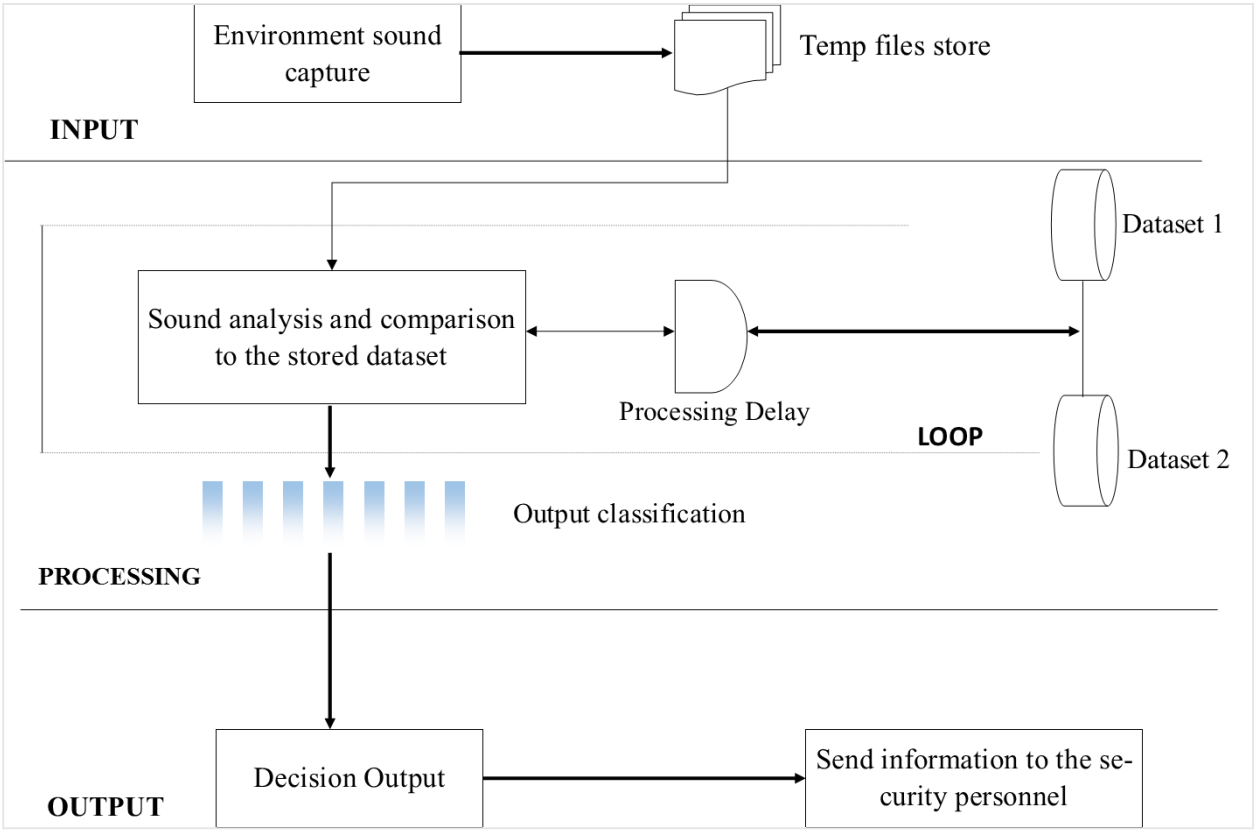



Figure 4.1: System Architecture

Figure 4.1 shows the step by step operations of the system form input to output, then finally the loop rolls back to its initial ready listening state.

### 4.3.1 Data Input

In the proposed system, the sound input is done at this level, from the literature, it is evident that most of the time, during a physical security breach, noise comes out as part of the attackers activity, this includes activities such as gunshots, a grenade blast and sometimes the victims running for

their lives, this results to noise from the resulting commotions. The sound input is therefore done from a client which is the sound capturing devices placed allover in a distributed way, this includes the microphones that can capture sounds based on their different frequencies.

Upon capturing of this sounds from the environment, the client has to save the file temporarily so as to be sent to the processor, this is done by the use of a Microsoft language which is C++. The files are saved in a .wave or a .wav format so as to allow processing to take place.

### 4.3.2 Data Processing

The stored .wave or .wav sounds form the temporary files are then sent to the model that does the classification, this has the artificial neural network (ANN) that has 10 nodes to give the verdict from the analysis. It has a delay in the processing since the classification involves reading from the datasets that are available in the system which are 10 in total. The segregation of the datasets makes sure that similar sounds are placed together and thus making the processing work much easier.

The different frequencies are then extracted from the .wave files and compared against the stored datasets. Since .wave or the .wav files are not compressed, it is much easier for the system to perform supervised learning. The system tries to get the accuracy of the sound that was captured against the sounds that is stored in the dataset. Once the classification is done, the nodes can then be ready to display the output in the final section.

### 4.3.3 Classification Output

This is the final stage of the proposed system's sound analysis. The output from the nodes is generated and sent to the security personnel. The message format is in the text format, currently just displayed on the systems screen. This will therefore be very important for the security personnel to make a quick decision so as to enhance the physical security that is already in place.

## 4.4 Process Design

This section outlays the set of input resources which has been used to transform the input into the meaningful output and services. The system has used, firebase, different languages and the use of a cloud platform to make sure that the system processes and gives the relevant output to the security

personnel on the ground. A detailed information of this section has been discussed in chapter five under software development environment.

## 4.5 System Design

The collection of the different objectives as from the analysis of the literature and other similar systems were merged with the ideas that the developer had in mind to synergize an application design with desirable functionalities to fulfil its objectives. The following design diagrams has been used to give an insight into the actual implementation of the system. This includes the different system's modules.

### 4.5.1 Class Diagram

In the proposed system, the class diagram has been used to describe its structure showing it classes, their attributes, operations and the relationships between them. This quickly provides a snippet of how the object oriented entities are related. Figure 4.3 shows the proposed system's class diagram.

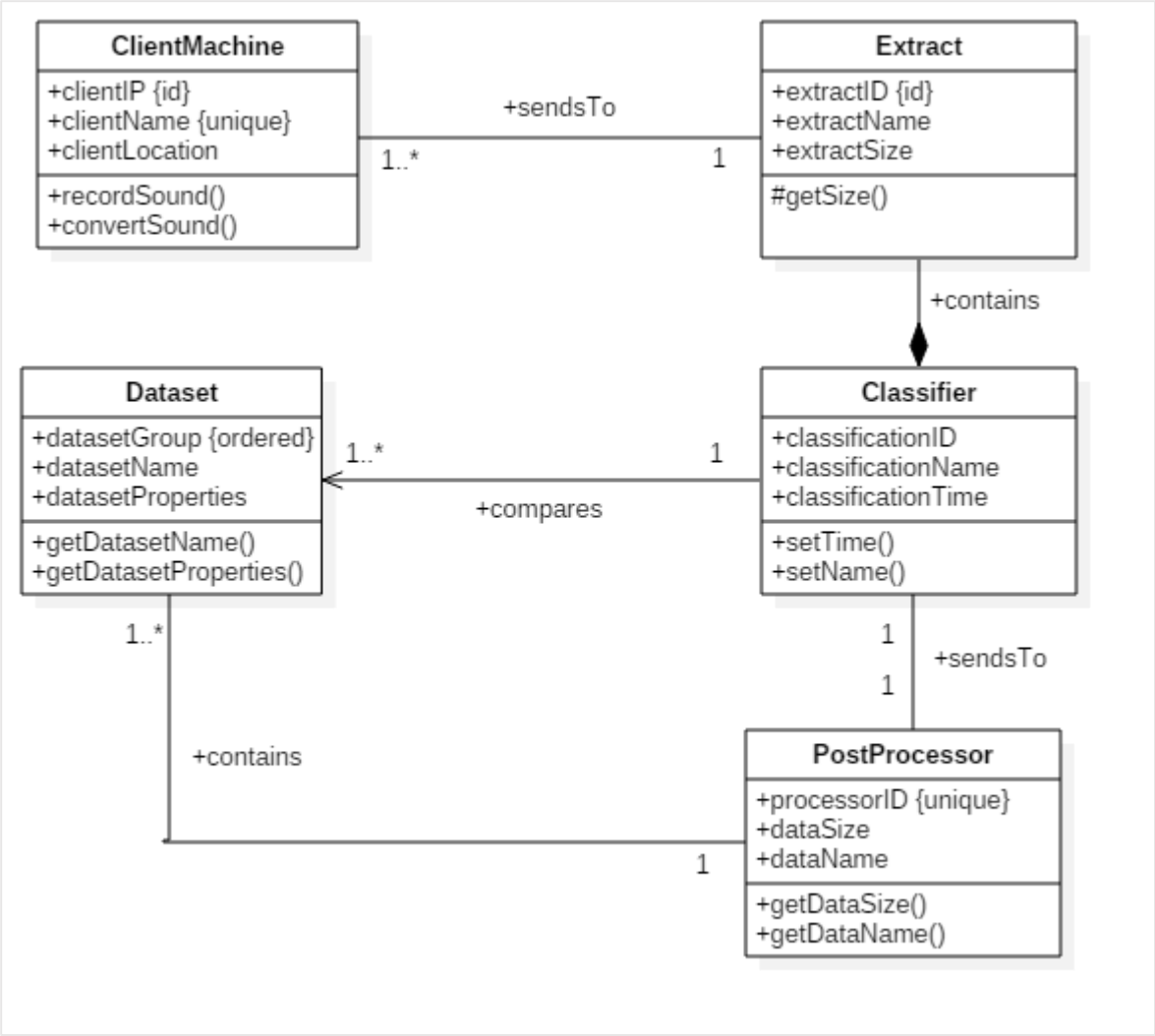



Figure 4.2: Class Diagram

### 4.5.2 Data Flow Diagrams

Data Flow Diagram (DFD) provides a visual representation of the flow of information that is data within a system. In the proposed system, this diagram has been used to show the information provided by and delivered to someone who takes part in the system process, the information needed to be stored and accessed.

**Context Diagram**

Context diagram shows only the top level characteristics of the system data flow. It demonstrates only one visible process node that represents the functions of a complete system in regards to how

it interacts with external entities. It therefore show briefly the boundaries of the proposed system thus viewing the system as a black box. Figure 4.3 shows how the external entities interact with the system.

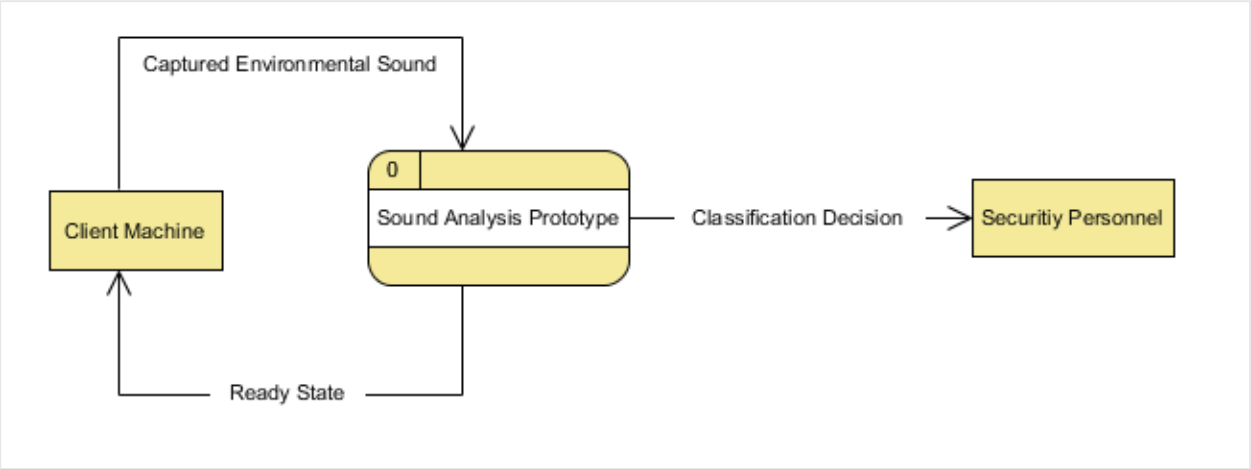



Figure 4.3: Data flow diagram

The entities are interacting with the system as a black box, the client machine that is in a distributed environment captures the environmental sounds and sends it to the server. The server does all the classification and sends the messages from the classification results to the security personnel, this is done through a screen display which is part of the proposed system model.

**Level 0 Diagram**

The different entities initially in the system as a black box has been broken down into various processes which interact with the external entities. The data stores represents the storages done in the system during processing. Some of the data stores are temporary as they involve temporary storage of information before post processing or display. Figure 4.4 shows how processes and entities interact in the level 0 diagram

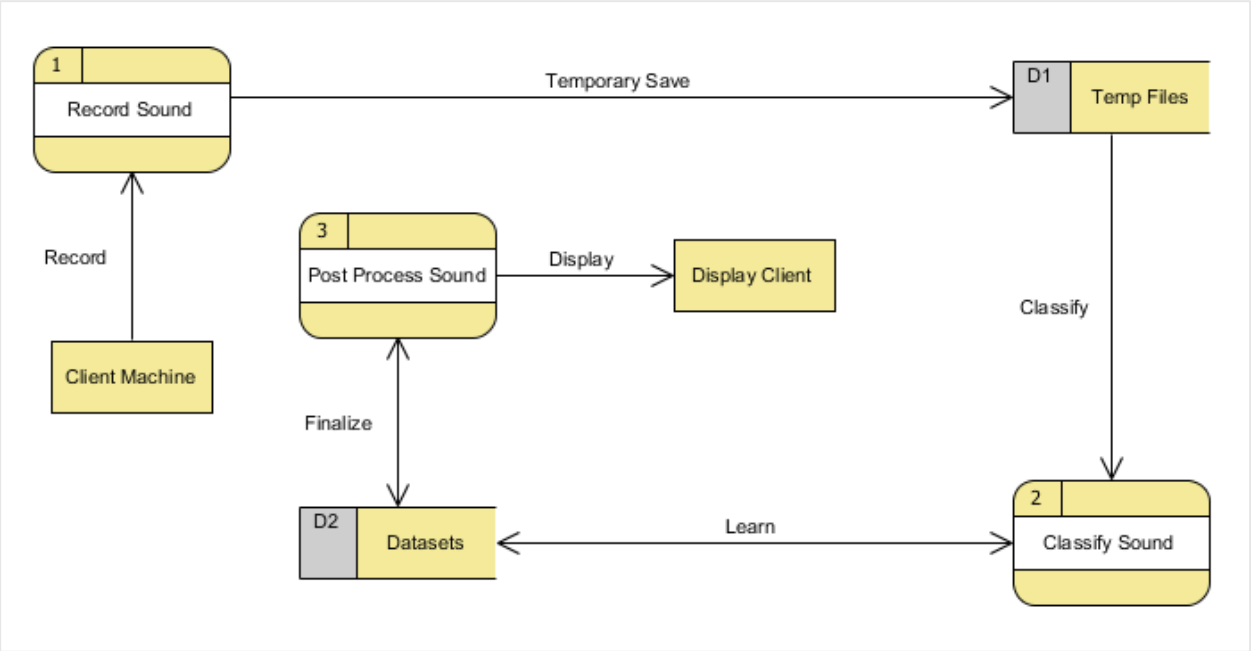



Figure 4.4: Level 0 Diagram

### 4.5.3 Use Case Diagram

In the proposed model, different cases have been simulated diagrammatically as shown in Figure 4.5. The entities involved included the independent system classes such as the classier function that is in the system but interacts with the system components.

Some of the assumptions that has been made in the use case diagram includes;

i. The display and the client machine does the actions independently even though in the system, the machine responsible for both actions is one.
ii. The client machine auto controls itself to initiate the listening, the wait and loops till the output then then back to the start again

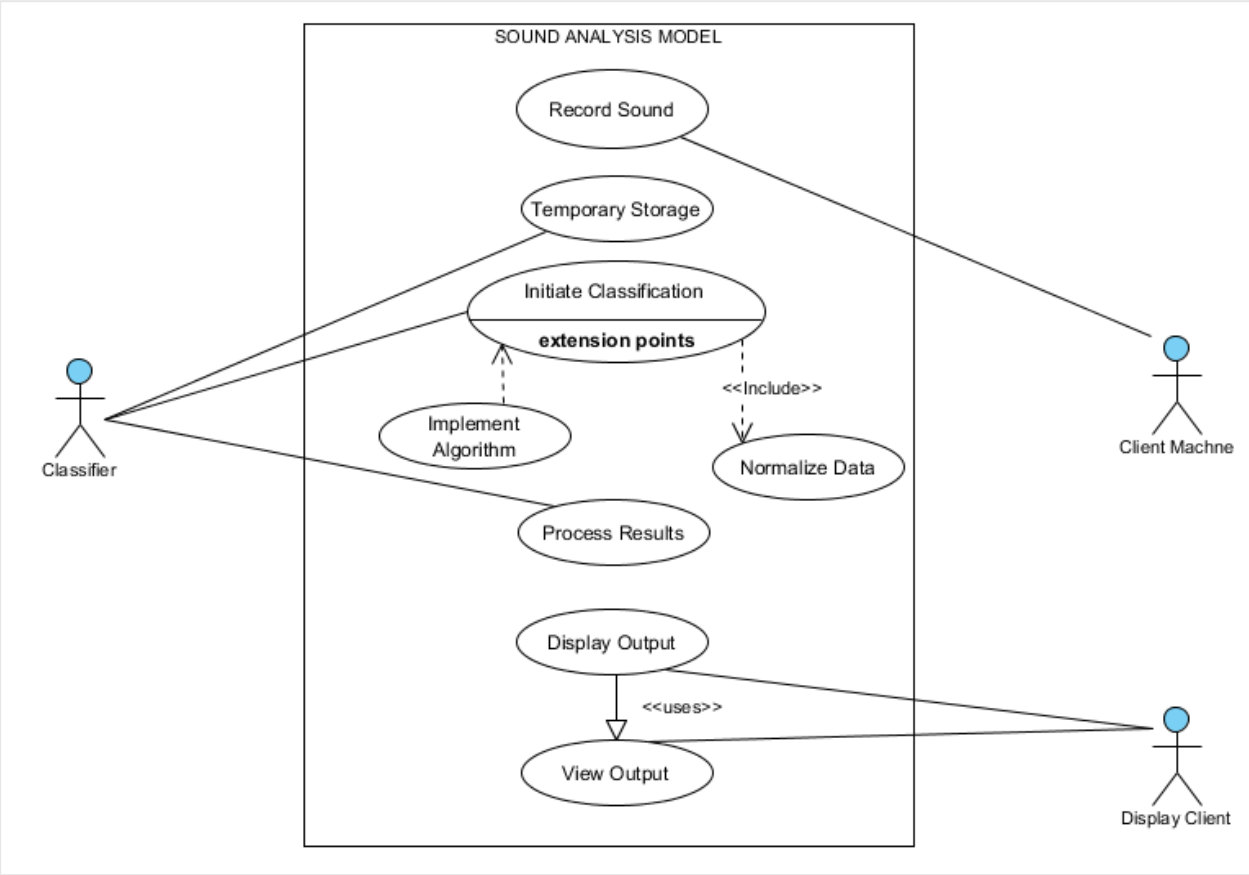



Figure 4.5: Use Case Diagram

### 4.5.4 Sequence Diagram

Sequence diagrams illustrates how the input is done, how the processing is done till the output without focusing on how the system does it. This focuses on how the system's processes interact to make sure that the data processing is achieved and a desirable output is reached.

Figure 4.6 shows the proposed system's sequence diagram focusing on how the system's processes interact and an output is reached.

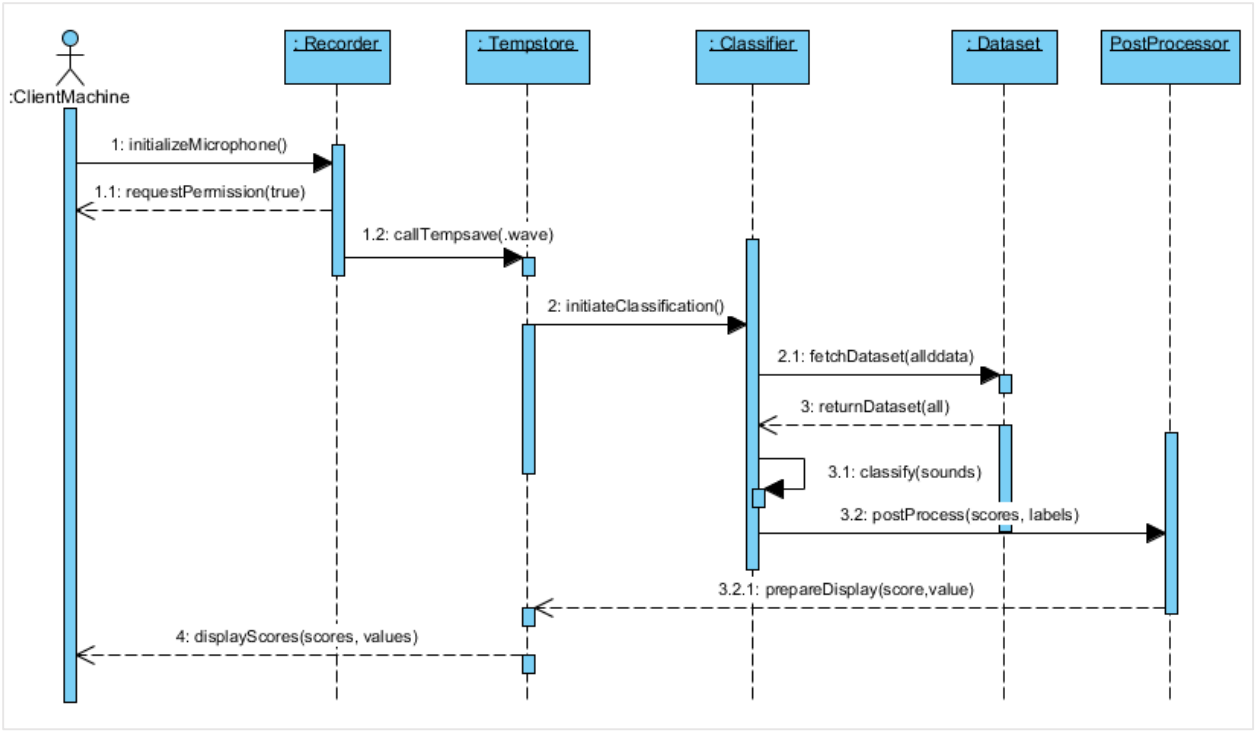



Figure 4.6: Sequence Diagram

## 4.6 Network Design

The proposed model runs on a client server architecture in which each computer or process in the network is either a client or a server. As proposed by Schuster, Jablonski, Kirsche and Bussler (2010), the client server architecture enables the faster workflow and processes management thus boosting the performance in the proposed model. Figure 4.7 shows the client server model used in the proposed system to enhance productivity.

Figure 4.7: Client Server Architecture in the proposed system

The clients sends the requests to the compiler which is situated in the server, the compiler does the analysis through supervised learning as it compares the datasets to the input sounds then finally the decision results saved into the database which uses the firebase and thus enables streaming of the results to the security personnel.

The proposed model uses http2, which tends to improve the data accuracy and efficiency in its transfer over the network since its speed is fast as it enables multiplexing and header compression. HTTP/2 comes in hand as it encrypts the data as compared to before when HTTP was in use which had machine readable format of the information. This also increases the security as one it can't be breeched through the use of telnet. This protocol also has additional features such as compression, multiplexing, server push and priority (Saxc´, Oprescu, & Chen, 2017). Figure 4.8 shows a model that motivated the use of http/2 over the older version of http/1.1.

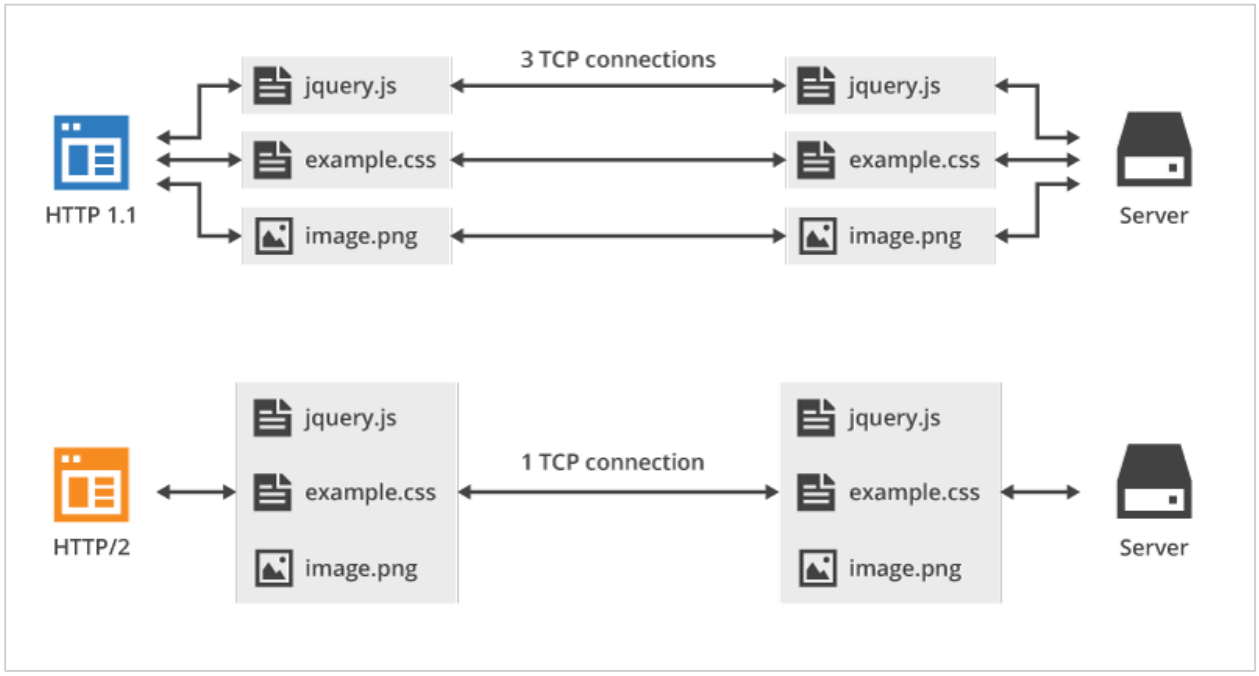



Figure 4.8: HTTP/2 Multiplexing and header compression

### 4.7 Security Design

The proposed model has implemented the used of http/2 protocol which enhances the systems security, as discussed in the previous section, 4.6 which covers the network design, http/2 allows the data to be sent on a secure form since it's in a non-human readable format. There is no applied encryption algorithm that has been applied in the system hence making it less secure in case of a

man in the middle attack but this can be added in the incremental model as development proceeds to future works.

### 4.8 Wireframe Design

Wireframe design known as the schematic paging or blueprint focuses on the visual guide representing the skeletal framework of a system. Wireframes are created for the purpose of arranging elements to best accomplish to a particular purpose which in this case is the systems UI design (User Interface). In proposed model, figure 4.9 focuses on the UI elements of the proposed system focusing on its hierarchical interaction with the user interface

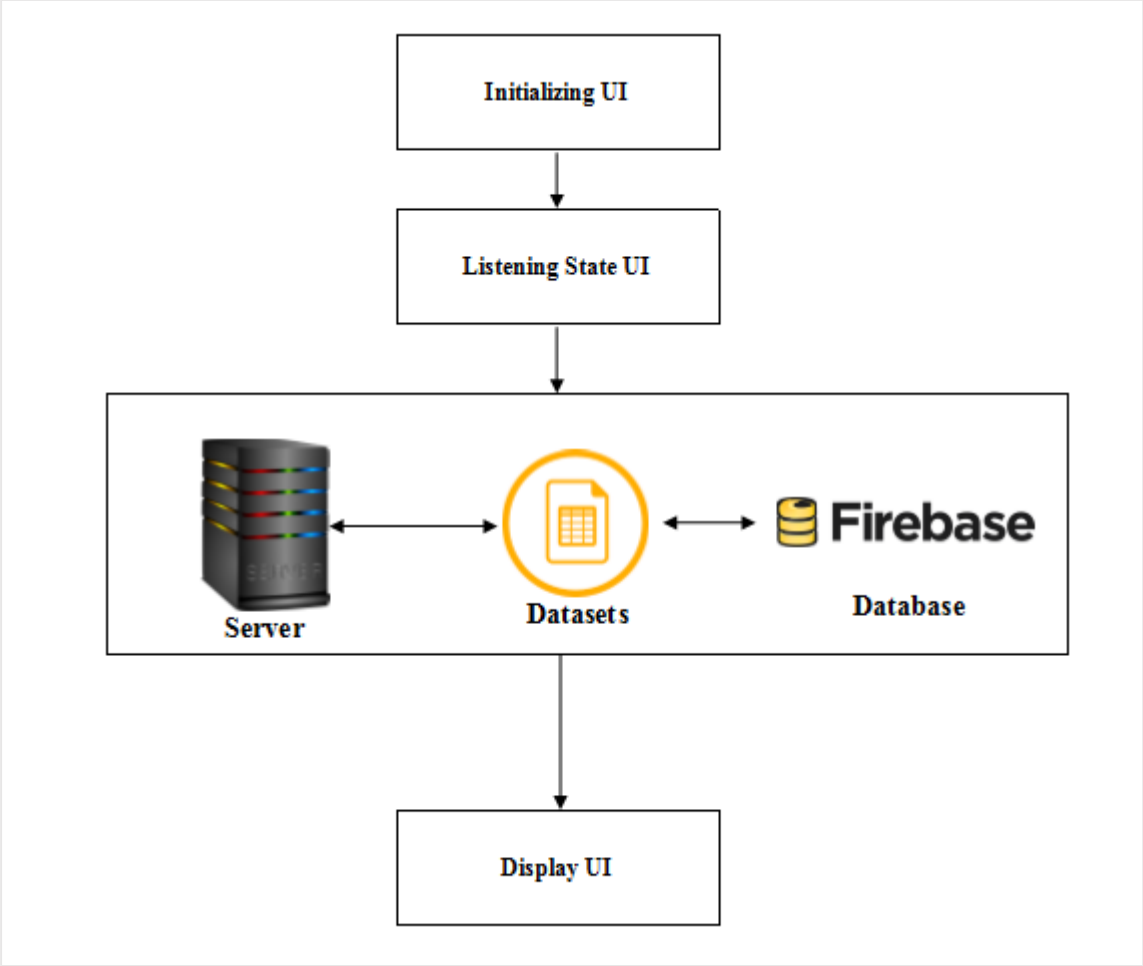



Figure 4.9: UI wireframe of the proposed model

# Chapter 5 Implementation and Testing

## 5.1 Overview

This chapter contend the implementation of the actual prototype as proposed. The functionalities incorporated in the system includes the systems requirements as from the literature. The section also strictly adheres to the design diagrams as proposed and described earlier in chapter four. To appreciate the feel of the system implementation, this section also includes screenshots of the various application which is mainly in the client side of the distributed system.

This section also shows how the classification model works and the detailed information about its functionalities. A brief acceptance questionnaire results from the security personnel and other few individuals in the university. Finally, this section demonstrates the proposed system's performance measures which includes the scalability, response time and the systems accuracy, this is supported with detailed data form the systems testing.

## 5.2 Description of the testing Environment

The testing environment entails both the hardware and the software, this describes the hardware specifications and the software specifications that comprises the systems running environment and how it enhances the system's performance. This can also assist in making sure that the components are key variables in the system's characteristics such as being distributed and scalability.

### 5.2.1 Hardware Specifications

The systems runs on a test computer with a RAM (Random Access Memory) of 6 GB, this enhances the systems performance since the dataset is quite heavy, the test data has 8742 .wave sound files and is totaling to a capacity of 7.1 GB. For better performance, the RAM can be enhanced making the system a lightweight on the hardware. The hard disk space used in the test environment is a minimum of 50 GB, this makes sure that there is enough storage for the dataset files which might grow in size to while trying to get more data to compare against the input.

A high processing speed of 2.7GHz and above is recommended and this has been made to run on a corei5 hardware, this make it easier for the processors to handle different threads or processes that results during processing. To make the system even better, the local server can be boosted with mode processors hence making a computer with corei7 a more suitable model.

For data capturing, the system uses the inbuilt computer microphones which captures the environmental sounds in a distributed way and sends them to the server for processing. For better performance, better microphones that capture clear sounds even with the presence of minimal frequencies are recommended, this will enhance the clarity of the sounds being sent to the server for processing hence facilitating clear and accurate classification hence clear results.

### 5.2.2 Software Specifications

The proposed system runs on a Linux operating system with an Ubuntu distribution, this enhances the systems performance, Linux's use of processes as opposed to threads make the proposed system a lightweight when it comes to the classification which consumes a lot of memory as it is a loop of many sound waves while comparing the input data against the dataset.

Firebase real-time database has been used to maintain the client server architecture communication as it relays the data in real-time making it available even when the network is off. In Firebase, Data is stored as JSON and synchronized in real-time to every connected client. When you build cross-platform apps with our iOS, Android, and JavaScript SDKs, all of your clients share one real-time Database instance and automatically receive updates with the latest data

In the proposed model, instead of typical HTTP requests, the Firebase Real-time Database uses data synchronization—every time data changes, any connected device receives that update within milliseconds. Provide collaborative and immersive experiences without thinking about networking code. (Google Developers, 2017).

The additional advantage is that Firebase apps remain responsive even when offline because the Firebase Real-time Database SDK persists your data to disk. Once connectivity is reestablished, the client device receives any changes it missed, synchronizing it with the current server state, moreover, the Firebase Real-time Database can be accessed directly from a mobile device or web browser; there's no need for an application server. Security and data validation are available through the Firebase Real-time Database Security Rules, expression-based rules that are executed when data is read or written.

Figure 5.1 shows how the application is fetching the data from Google's firebase in real-time and serving it to its clients

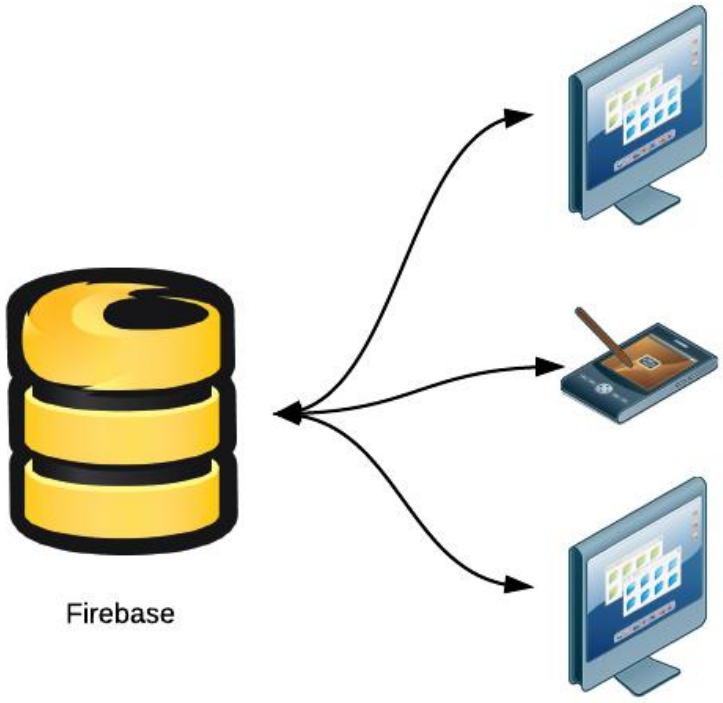



Figure 5.1: Real-time data access from the firebase database

### 5.3 Prototype development Environment

The proposed model has used a number of applications and development environment to enhance its productivity. Some of the development used includes;

i. Visual studio code for programming – this is a text editor that makes it easier to develop applications that use the Microsoft languages such as C++ or C#.

ii. Bazel build for building – this is a tool that automates software builds and tests. Which includes tasks like running compilers and linkers to produce executable programs and libraries, and assembling deployable packages for Android, iOS and other target environments. Bazel is similar to other tools like Make, Ant, Gradle, Buck, Pants and Maven. Bazel enhances speed, correctness, scalability, reliability, flexibility and repeatability.

iii. gRPC for the distributed system - gRPC is a modern, open source remote procedure call (RPC) framework that can run anywhere. It enables client and server applications to communicate transparently, and makes it easier to build connected systems. It has been used in last mile of computing in mobile and web client since it can generate libraries for iOS and Android and uses standards based HTTP/2 as transport allowing

it to easily traverse proxies and firewalls, this makes the productivity of the proposed system more accurate.

**Specified languages**

i. C++ - a general purpose object oriented programming language that has been used in the proposed system to support the distributed service. This makes sure that the client communicated with the server.
ii. Python – enables quick working by letting one quickly and integrate code more efficiently. In the proposed model, the language has been used for building the classifier and the client stub.
iii. Proto – in the proposed model, this has been used to enhance the information transfer between the client and the server. In this model, the information is inform of sound waves.
iv. JavaScript - a high-level, dynamic, untyped, and interpreted programming language which has been standardized in the ECMAScript language specification. In the proposed model, JavaScript has been used to enhance client data input and building the user interface (UI)
v. LibROSA – this is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems. In the proposed model, libROSA has been used to extract the sound waves from the client to the server.

**Specified Development packages**

The proposed model has used the package for mathematical computations. NumPy, (NumPy developers, 2017) has a powerful N-dimensional array object, sophisticated (broadcasting) functions, tools for integrating C/C++ and Fortran code, useful linear algebra, Fourier transform, and random number capabilities has also been used in the proposed model as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined and this allows to seamlessly and speedily integrate with a wide variety of databases.

Tensorflow, an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the

multidimensional data arrays (tensors) communicated between them, its flexible architecture allows one to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API. In the proposed system, the tensor graph has been used to build the neural network and this enhances the performance of the hidden layers.

**Utilized Cloud platform**

The proposed model has also utilized the use of open source cloud platforms to enhance the productivity of the system by making it distributed. The cloud services used includes;

i. Google Cloud – this is an open source cloud service which offers platform as a service (PaaS). With Google Cloud, one can build can build, test, and deploy applications on its highly-scalable and reliable infrastructure for your web, mobile, and backend solutions.
ii. Firebase – this is a real-time no SQL database provided by Google to enable streaming of messages to other devices. Figure 5.1 shows how firebase is capable in terms of real-time data streaming

## 5.4 Model Components

This section shows the proposed system's components, supported by screenshots where applicable, explaining on how the different modules of the system works and how they affect the system's performance and output.

### 5.4.1 Sound Input Components

i. *Microphone or a sound frequency sensor* - the proposed model requires a microphone or a sound frequency sensor to identify and capture the environmental sounds as they occur or as they are caused on the designated environment.
ii. *Sound converter library* – this is a library used in the proposed system to convert the captured sound into a .wave or a .wav file which further allows processing as this can be read as a result of no compression state.
iii. *Temporary storage location* – this is where the captured sounds are stored before being sent to the processor. This makes sure that data is not lost as streaming is not possible in the proposed model.

## 5.5 Neural Network Components

The proposed system uses the back propagation neural network algorithm which consists of a number of components as discussed below. The neural network diagram below, a model proposed by Jain & Jianchang (2006) shows a sample similarity to the network used in the proposed model.

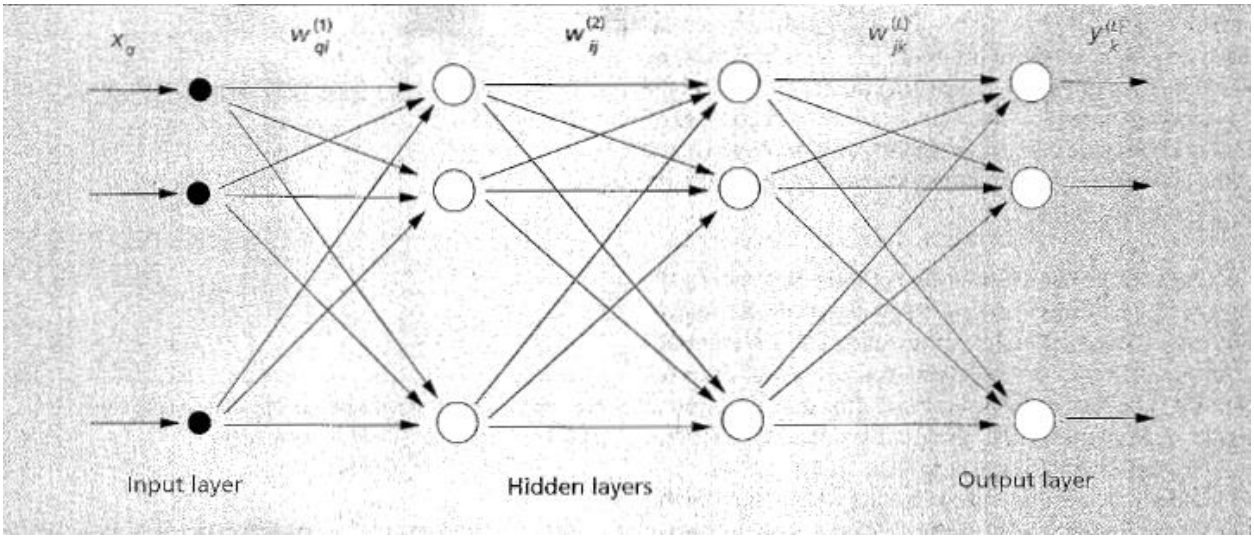



Figure 5.2: Sample three layered feed forward neural network

### 5.5.1 Input layer

In the proposed model, this is the first layer of the model, it comprises of 5 neurons for each specific attribute used to by the network to classify the sounds captured from the environment. These number of neurons determines how the input layer is structured.

### 5.5.2 Hidden layers

The proposed model has two hidden layers, hidden layer one and hidden layer 2 in the hidden layer one, hyperbolic tangent was used, a regression algorithm that computed the hyperbolic tangent. The main purpose of using two hidden layers was to make sure that the output was more accurate and defined to the security personnel. Figure 5.3 is a 2 hidden neuron simulator show how the hyperbolic tangent is able to compute having the capability to use a number of neurons.

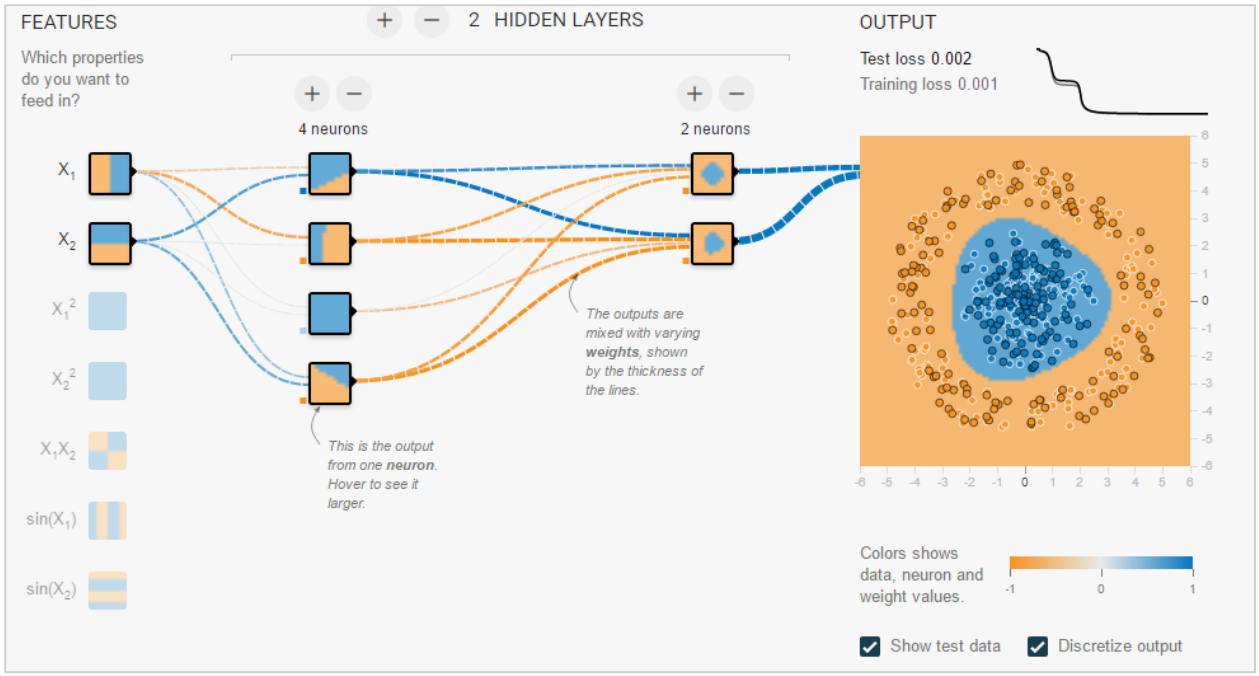



Figure 5.3: Two hidden layer neuron network simulator

In the first hidden layer, a total of 280 neurons and it enables the neural network to produce better results of the expected output given the input. This layer provides an intermediate layer for which the activation function has been implemented. In the second hidden layer, a total of 300 neurons was used to make the results from the first layer better and finer. The use of a total of 280 and 300 neurons makes sure that the problem of over fitting or under fitting is taken care of.

The neural network was implemented by attaching weight and biases to the input variables in the network. The biases helps the neural network give a more realistic results at the output level. In the proposed system, a transfer function was implemented hence the use of a sigmoid transfer function which lies between 0 and 1. As soon as the network was trained, it classified the target class as output.

A learning rate of 0.01 was used to train the network and finally a test was carried out in the proposed model to validate that the network performed as it was expected to. The critical error was not used in stopping the network but has been implemented in the input of data which distinguishes between a tensors and arrays.

### 5.5.3 Output layer

This is the last layer of the neural network and the softmax function was implemented at this level to enhance its productivity. The output is finally provided after all the inputs have been processed and a presentation of a pattern is presented to the external environment, which in this case is the security personnel. The output layer finally consists of the specific groups for which the output are assigned to. These groups includes a class which is the label and the scores which are the prediction scored showing how much the system feels that the output is more efficient and accurate.

## 5.6 System modules

The system modules that have been used in the proposed model includes the main modules and the sub modules. The main modules has the client server which is a listening interphase that can show the sound input thus providing the module with an interface for interaction to the environment for demonstration purposes. The sub modules includes the decision interface showing the security personnel the decision made by the system as a result of classification.

### 5.6.1 Main modules

At the front end, on the client side, a listening interface has been set to initiate capturing of sounds and sending them to the server. Appendix A shows the main interface at the client side. While in a listening state, the browser, which enables the client to share the collected sounds with the server, displays the microphone symbol. Initially when the button initiating listening is pressed, the microphone seeks the permission to run on the clients IP (Internet protocol). Appendix B and C shows the permission request for the microphone to run on the client's IP and the microphone in a listening state after the permission has been granted.

### 5.6.2 Sub modules

At the receiving end, the client, after analysis sends the feedback to the screen as this can help the security personnel know the classification decision. This helps them to know the right step to take in order to take control of the physical security breach. Figure 5.4 shows the result serial number, the time of completion and the verdict made from the classification.

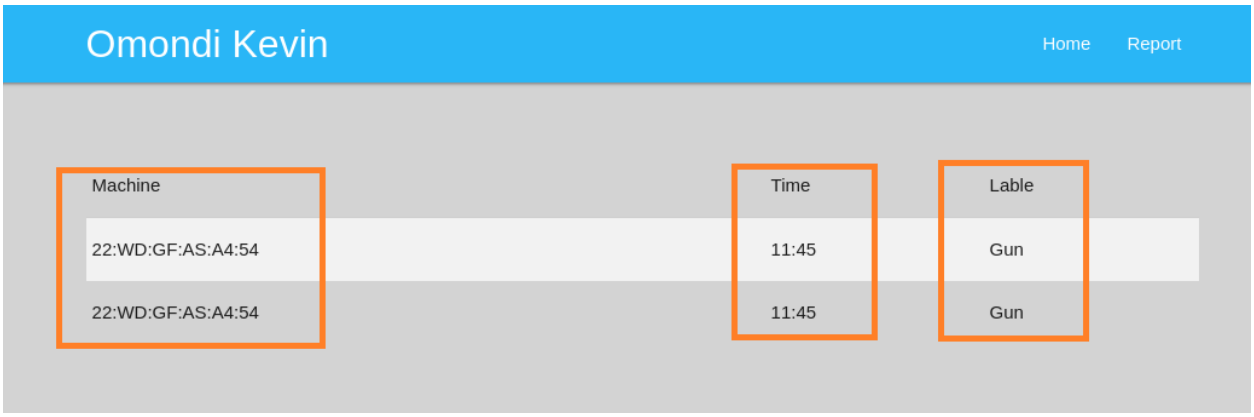



Figure 5.4: Results classification interface

The main components of the interface has been marked in rectangles, which is the machine, the time and the label.

## 5.7 Training and testing the model

A training data set was constructed from the urban sound library that was available online for sound analysis and was stored into 10 different folders to spread the load when learning starts. The input features from the sounds collected included;

i. Mel scaled power spectrogram – which is a representation of the short-term power spectrum of the input sound.
ii. Mel – frequency cepstral coefficients – in the proposed system, from the collected sounds, this is the dominant feature used for speech recognition (Logan, 2014).
iii. Spectral centroid features – this is a measure used in digital signal processing to characterize a spectrum. It indicates where the "center of mass" of the spectrum is. Perceptually, it has a robust connection with the impression of "brightness" of a sound.
iv. Tonal centroid features which supports the automatic cord extractions.

In the prosed model, the training takes 70% of the sound processing time and the testing takes 30%. In the testing, the features used includes the sound after which the label is initialized. For accuracy the results is fed into a feed dictionary which has x and y variable as shown in the equation below.

$$\boldsymbol{Equation} = \boldsymbol{soundFeed}\{\boldsymbol{x}: \boldsymbol{value}, \boldsymbol{y}: \boldsymbol{value}\}$$

Equation 5.1: Feed dictionary array equation

The *x* represents the input and its value is the score allowed, and the *y* shows the class while its value is the label given. During testing, each sound input is mapped into the equation to replace the *x* after which the *y* can be mapped from the training hence allocating the score giving the verdict. In the testing, some of the variables inspected were as shown in table 5.1

Table 5.1: Model Testing Table

| **Test Class** | **Sounds used** | **Inspection Variables** | **Priority** |
|---|---|---|---|
| Functional | Animal movement<br>Crowded area<br>Gunshot | Were all the sounds that where captured returned after processing? | High |
| Functional | Animal movement<br>Crowded area<br>Gunshot | Did the system validate the input to make sure that they were only sounds and the extension was in .wave or .wav | High |
| Reliability | Animal movement<br>Crowded area<br>Gunshot | Was there a difference in the sounds input from the sounds that were in the dataset | High |
| Response | Explosion<br>Crowded area<br>Gunshot | Did the system collect the sounds from the environment sent to the model for analysis | High |
| Actual Data | | Was the actual data present in the system for comparison to take place | Medium |

| | | | |
|---|---|---|---|
| Accuracy | Animal movement<br>Crowded area<br>Gunshot<br>Explosion | Was the output decision from the other experiments reliable enough to make a decision | High |

### 5.7.1 Model test results

The system successfully validated the user input to ensure that only acceptable sounds which are in the .wave or .wav format were accepted. Table 5.2 shows the deductions obtained from the tests done

Table 5.2: Model test results

| Test Class | Test Results | Sounds Used | Comments |
|---|---|---|---|
| Functional | Pass | Crowded area<br>Explosion | The sounds were captured form the environment and the files uploaded were .wave and .wav only, the library responsible for the conversion was accurate |
| Reliability | Pass | Crowded area<br>Gunshot<br>Explosion | There was a mild difference between the sounds collected from the environment to the ones in the dataset. This because the testing environment imitates the urban sounds |
| Response | Pass | Music<br>Crowded area | The system was able to collect the sounds from the environment and send them to the model for analysis |

| | | | |
|---|---|---|---|
| Actual Data | Pass | | The actual data was present in the dataset for comparison |
| Accuracy | Pass | Gunshot<br>Crowded area | The accuracy was above average, able to make a decision |

### 5.7.2 System Testing

The system was tested to check how the developed model performed in comparison to having the security personnel figuring out everything by themselves then applying the reinforcement. Table 5.3 shows the inspection and the results.

Table 5.3: System testing variables and results

| Test Class | Inspection Check | Priority level | Test Results |
|---|---|---|---|
| Performance | Does the system provides the verdict in the output as per the requirements | High | Pass |
| Accuracy | Did the system successfully classify most of the sounds used for the testing | High | Pass |
| Compatibility | Did the system run on the specified operating system? | Medium | Pass |
| Effectiveness | Was the verdict displayed on the output screen effective enough in making a decision to help the security personnel | High | Pass |

### 5.7.3 Acceptance testing

Key aspects determined the acceptance of the system as the features discussed in table 5.4 below were set as the verification variables

Table 5.4: Acceptance testing

| Test Class | Inspection Check | Priority | Test Results |
|---|---|---|---|
| Usability | Has the system met the user requirements? | High | Pass |
| Usability | Were the users satisfied by the systems output decision | High | Pass |
| User friendliness | Was the system's UI easy to use? | High | Pass |

# Chapter 6 Conclusions and Recommendations

## 6.1 Overview

This section focuses on the discussions of the key findings as seen in the previous chapter. It goes ahead to discuss the results explaining each of them and why they are important to the study, It also highlights the recommendations and the suggestions listing various improvements in case of a future work.

## 6.2 Discussions

The sound analysis model was implemented by using features extracted from the sounds captured at the input level. These features were used to compare their similarity against the trained model so as to make a decision that could help the security personnel on the ground to enhance physical security in case of a security breach. The model was tested for correct classification on the basis of accuracy, efficiency and precision and the error was obtained. The proposed model was declared suitable as compared to the other methods initially discussed in the research.

Physical security has faced a lot of challenges especially in academic institutions as discussed initially in the literature section. The other systems that has been used before have tried to solve the issues but were not efficient enough to help the localized environment suiting the learning institutions. For example, the utilization of audio source localization in security systems was able to solve the intruder issue but for an army based environment where cameras were also used to focus on the intruders. Based on the results obtained from the structured findings, the sound analysis prototype to enhance physical security comes in handy when security personnel needs assisting tool that helps them plan early as early intrusions can be detected.

The system model developed in this research gives more accurate classification and efficiency based on the fact that it was implemented based on the artificial neural network (ANN) algorithm. The use of machine learning algorithm, enabled the proposed model to work much faster thus the feedback to the security personnel is almost immediately. A combination of the advantages provided by the computing strengths and machine learning has enabled the model to provide accurate results while classifying the sounds to come up with the verdict to the security personnel.

## 6.3 Conclusions

A background research done at the onset of this study supported by the literature confirmed that the available statistical data shows that there is a major concern about insecurity in the country. The research turned its focus on the academic institutions available in the country today which of late has become target grounds to the terrorists (Miller, 2013). Due to the insecurity that has been experienced by these institutions of learning, most turned to the available security agencies for help but still, these agencies have not been able to fully solve their problems, as a result of the gaps that are available in the way they do their operations.

At the begging of the study, objectives were set to find a security monitoring tool which collects the sounds produced in the environment, classifies them and sends the decisions to the security personnel to take relevant action in enhancing the physical security that is already in place. The actions that the security personnel takes are meant to make physical security more reliable as through early information by the system, it is easier to reinforce backup, enable a lockdown and even manage the disaster appropriately. This was covered in chapter one during the early stages of the proposal.

To support the objectives of the study and the gaps that existed, in chapter two, which contains the literature review. The study mainly focuses on a number of topics like the current security situation in the Kenyan institutions stating the mass casualties that has been experienced in the past, the current approaches that these institutions have taken to safeguard themselves and the challenges they have faced. This chapter also reviewed already existing applications that used sound technology to make decisions, the sound algorithms used, similar systems that might have been built for security monitoring and alert purposes and finally the conceptual model of the proposed system.

A phase of the research methodology, research design and the proposed system development methodology was discussed in chapter three. In chapter four, the research focused on the proposed system requirement analysis which discussed its functional, non – functional and performance requirements. The proposed system's architecture, development process design and system design, network design, security design and the wireframe design were also discussed in this chapter.

Chapter five puts down the proposed system's implementation and testing where the focus was on the system's development environment, its model components and the different modules that exists on the client and the server side. This chapter also displays the various screenshots of these modules. Chapter six focuses on discussing the key findings from the testing results as experienced in the system, it also focuses on the tabular comparison of the results against the expected results. This chapter conveys the success of the study given the set out objectives at the beginning of the study.

## 6.4 Recommendations

The researcher recommends that;

i. For the system to be deployed in any academic institution, security personnel training on how to use the system would therefore come in handy so as to enhance the system's performance.
ii. The system can be expanded to adopt a broad input range of variables such as the distance from the sound generating item to make the analysis more accurate and hence a more accurate decision at the output level
iii. The database to be expanded to save more data to facilitate referencing in the future in case of need

## 6.5 Future Research Work

With the emerging technology for sound analysis through the use of artificial intelligence, a lot has been left unexplored especially on how other researchers all over the world can use sound to enhance any kind of security. From the results obtained from the research, a future researcher can;

i. Implement a better way of reducing the latency when it comes to collecting the sound to be sent to the server as it will improve the output efficiency.
ii. Find a faster way to send the data to the security personnel either through push messages into their mobile phones or sounding an alarm to make sure that the system is fully distributed.
iii. Incorporate a better encryption method that will fully hide the information sent from the client to the server as it will improve the data integrity.

## References


Alkhateed, F., Maghayreh, E., Tubishat, M., & Aljawarner, S. (2010). The use of location based services for very fast and precise Accidents reporting and locating . *2010 International Conference on Intelligent systems, Modelling and Simulation* (pp. 21-24). Liverpool: IEEE.

Cox, F. D., Orsborn, C. R., & Sisk, T. D. (2013). *Identity and Insecurity in Modernizing Kenya.* Nairobi.

Creswell, J. W. (2003). *Research Design, qualitative, quantitative and mixed methods approaches.* London: SAGE Publications.

Department of the Army. (2001). *Physical Security.* Washington, DC: Department of the Army, USA.

Dobbins, T. (2004). *Scientific Methods of Research.* Califonia.

Dostálek, P., Vašek, V., Křesálek, V., & Navrátil, M. (2015). *Utilization of Audio Source Localization in Security Systems.* Zlín, Czech Republic: Tomas Bata University in Zlín.

George , C., Jean , D., & Tim , K. (209). *Distributed systems concepts and Design.* New Delhi, India: Dorling Kinsersley Ltd.

Goldstein, Jonhson, E. J., & Daniel. (2003). *Do Defaults Save Live?*

Google Developers. (2017, April 3). *Firebase Documentation.* Retrieved from Firebase: https://firebase.google.com/docs/database/

Hollis, B. (2017, January 25). *Physics of Sound.* Retrieved from The Method Behind The Music: https://method-behind-the-music.com/mechanics/physics/

Idriss, M., Jendly, M., Karn, J., & Mulone, M. (2010). *CRIME PREVENTION AND COMMUNITY SAFETY: TRENDS AND PERSPECTIVES.* 465, rue Saint-Jean, Suite 803 , Montreal, Quebec, Canada, H2Y 2R6: International Centre for the Prevention of Crime (ICPC).

Jain, A. K., & Jianchang, M. (2006). *Artificial Neural Networks: A tutorial.* Nebraska, USA: Wayne State University.

Jean, P., & P, P. (2005). *InformeDesign Where Research Informs Design.* Minnesota: University of Minnesota.

Karue, J., Kinyua, A., & Njau, L. (2014). *Pollution in Kenya.* Nairobi: Centre for Nuclear Science Techniques, Nairobi University.

Kenya National Commission on human rights. (2014). *Are we under siege? The state of security in Kenya.* Nairobi: KNCHR.

Kevin Strom, P., Marcus Berzofsky, M., Bonnie Shook-Sa, M., Kelle Barrick, P., Crystal Daye, M., & Nicole Horstmann, B. (2010). *The Private Security Industry: A Review of the Definitions, Available Data Sources, and Paths Moving Forward.* Washington, DC: U.S. Departments of Justice.

Le, V., & Tran, B. (2004). Spoken and written language resources for Vietnamese. *LECRE'04*, 599-602.

Lerer, L. (2007, April 27). *Exceutive Protection.* Retrieved from Forbes: http://www.forbes.com/2007/04/27/security-ceo-compensation-tech-security-cx_ll_0430ceosecurity.html

Logan, B. (2014). *Mel Frequency Cepstral Coeficients for Music Modeling.* Massachusetts: Cambridge Research Laboratory.

Machanje, D. I. (2014, June). Mobile-Based Security Agency Sound Monitor and Alert System. *Mobile-Based Security Agency Sound Monitor and Alert System*. Nairobi, Nairobi, Kenya: Strathmore University.

Miller, E. (2013). *Al-Shabaab Attack on Westgate Mall in Kenya.* Maryland: U.S. Department of Homeland Security.

Mugenda, O., & Mugenda, A. (2003). *Research Methods Quantitative and Qualitative Approach.* Nairobi: African Centre for Technology Studies.

NumPy developers. (2017, March 4). *NumPy.* Retrieved from NumPy: http://www.numpy.org/

Ohlhausen, P. E. (2004). Building Private security/public policing partnerships to prevent and respond to terrorism and public disorder: Vital issues and Policy recommendations. *National Policy Sumit* (pp. 1-45). 8803 Prudence Drive, Annandale, VA 22003-4156 USA: Ohlhausen Research Inc.

Pate , A., Jensen, M., & Miller, E. (2015). *Al-Shabaab Attack on Garissa University in Kenya.* Baltimore : Department of Homeland Security Science and Technology Directorate's Office of University Programs.

Phuong, N. C., & Dat, T. D. (2013). Sound classification for event detection. 330-333.

Prakriti, P., & Sharma, A. (2013). A comparative study between iterative waterfall and incremental software development life cycle model for optimizing the resources using computer simulation. *2nd International Conference on Information Management in the Knowledge Economy*, 188-194.

Saxc´, H. d., Oprescu, I., & Chen, Y. (2017). Is HTTP/2 Really Faster Than HTTP/1.1? *18th IEEE Global Internet Symposium*, 299.

Schuster, H., Jablonski, S., Kirsche, T., & Bussler, C. (2010). A CliendServer Architecture for Distributed Workflow Management Systems. Nuremberg: University of Erlangen.

Standards, A. N. (2007). *Standardization for Perimeter Security*. Washington, DC: ANSI Homeland Security.

T., O. (2016, October 27). *Physical Security: Managing the Intruder*. Retrieved from http://resources.infosecinstitute.com/physical-security-managing-intruder/

Ustun, V., & Smith, J. S. (2010). Creating Realistic Human Behavior in Physical Security Systems Simulation. *Proceedings of the 19th Conference on Behavior Representation in Modeling and Simulation, Charleston, SC, 21 - 24 March 2010.* Charleston: IEEE.

Vigne, L., Palmer, T., & Hetrick, S. (2008). *Planning for Change: Security Managers' Perspectives on Future Demographic, Crime, and Technology Trends.* Washington, DC: Urban Institute/ASIS Foundation. Mendez.

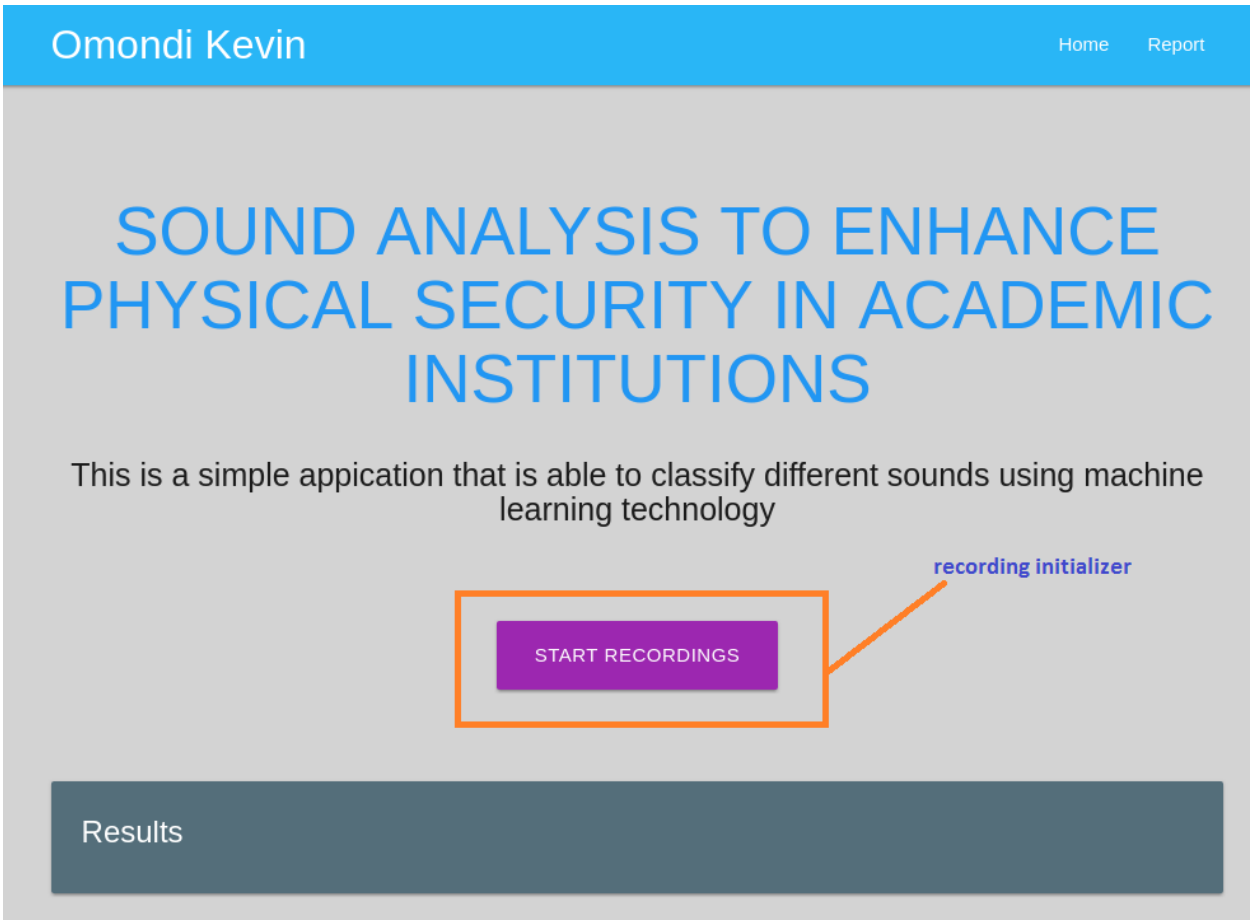
Waiguru, F., Kamunju, J., & Singo, M. (2004). *Private Security in Kenya. Nairobi.* Nairobi: Security Research and Information Center (SRIC).

Wakefield, A. (2005). The Public Surveillance Functions of Private. *Surveillance & Society 'People Watching People'*, 529-545.

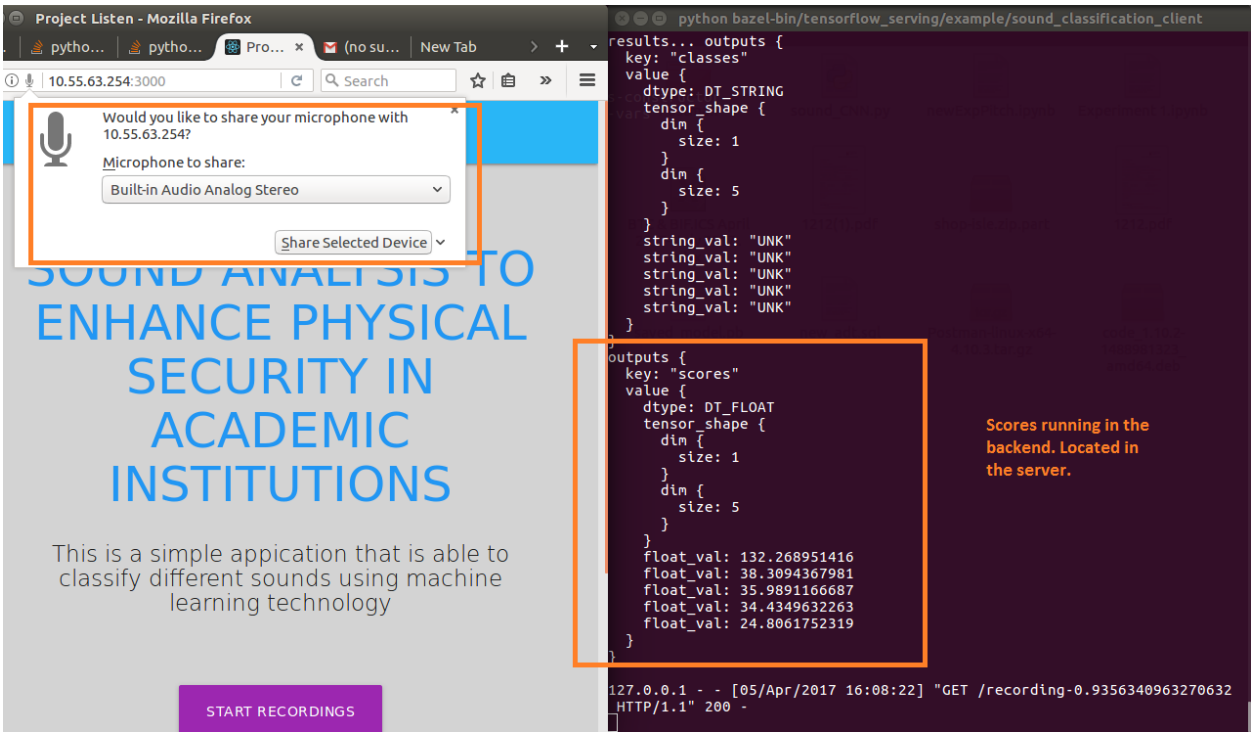
Welsh, B. C., & Farrington, D. P. (2009). Public Area CCTV and Crime Prevention: An Update Systematic Review and Meta-Analysis. *Justice Quarterly Volume:26 Issue:4*, 716-745.

Woodhouse, A. (2016, October 20). Retrieved from History of CCTV: http://www.maplin.co.uk/history-of-cctv

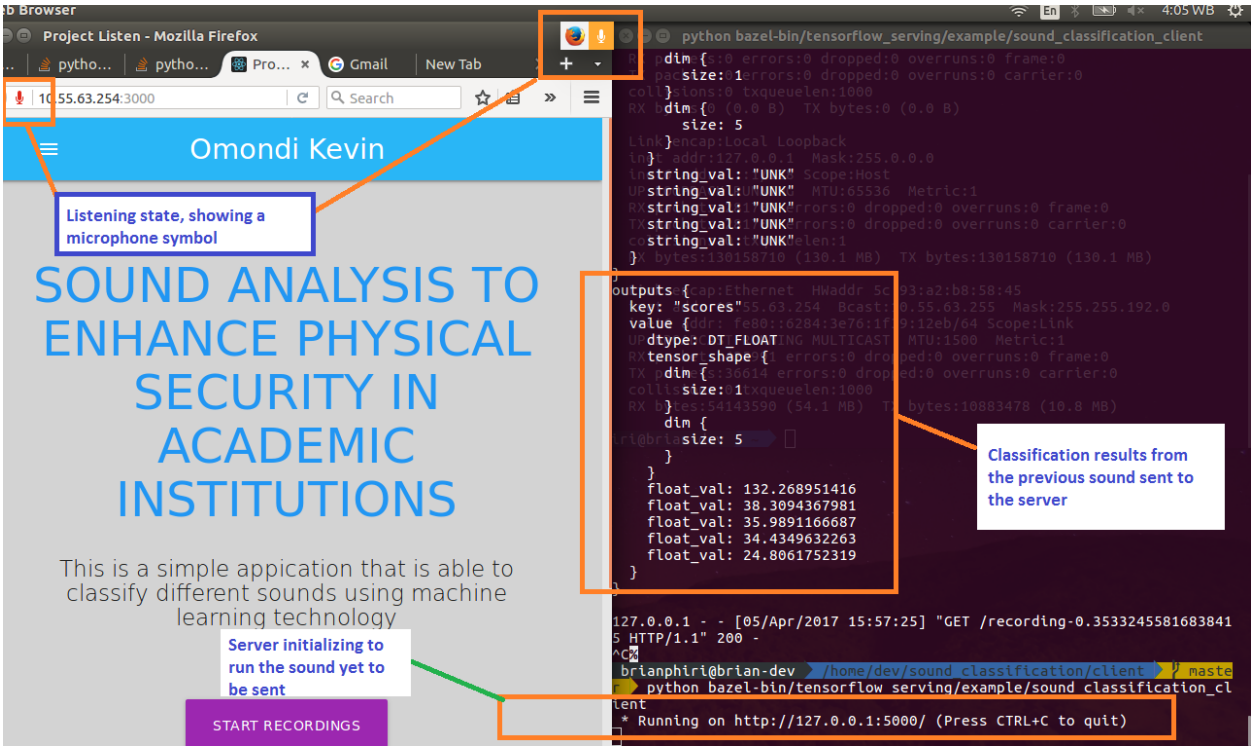
**Appendix A: Client side interface, Main module**



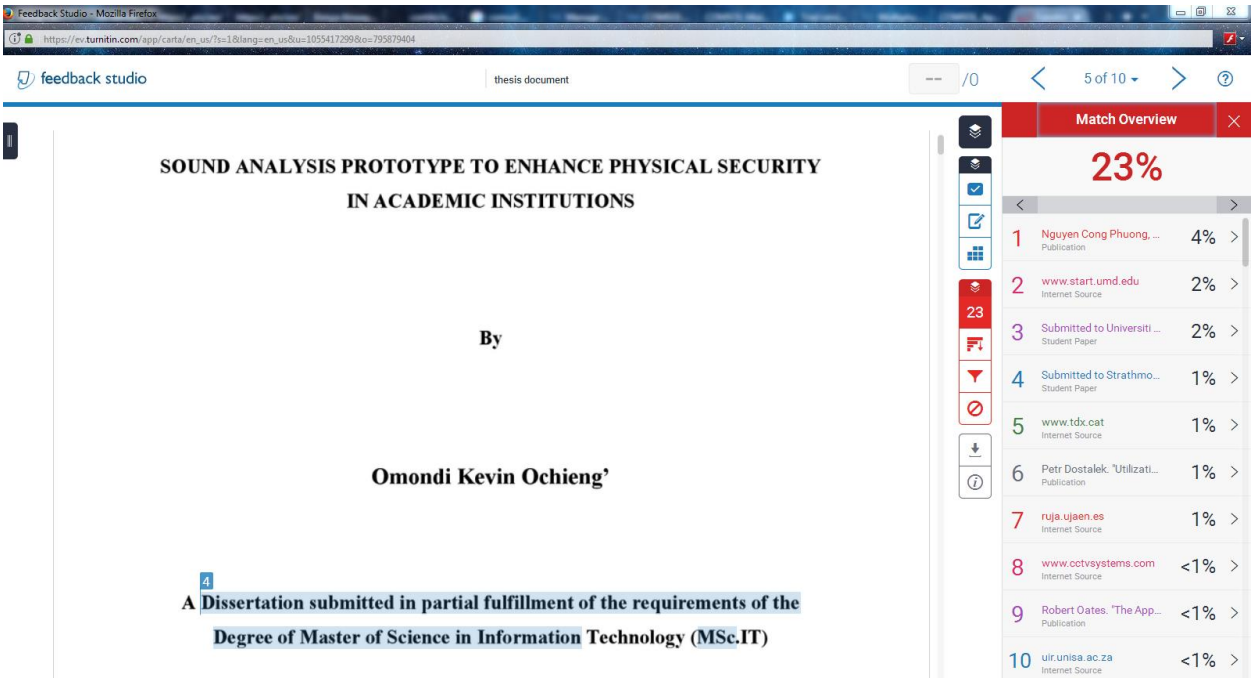
## Appendix B: Client Microphone ready to listen



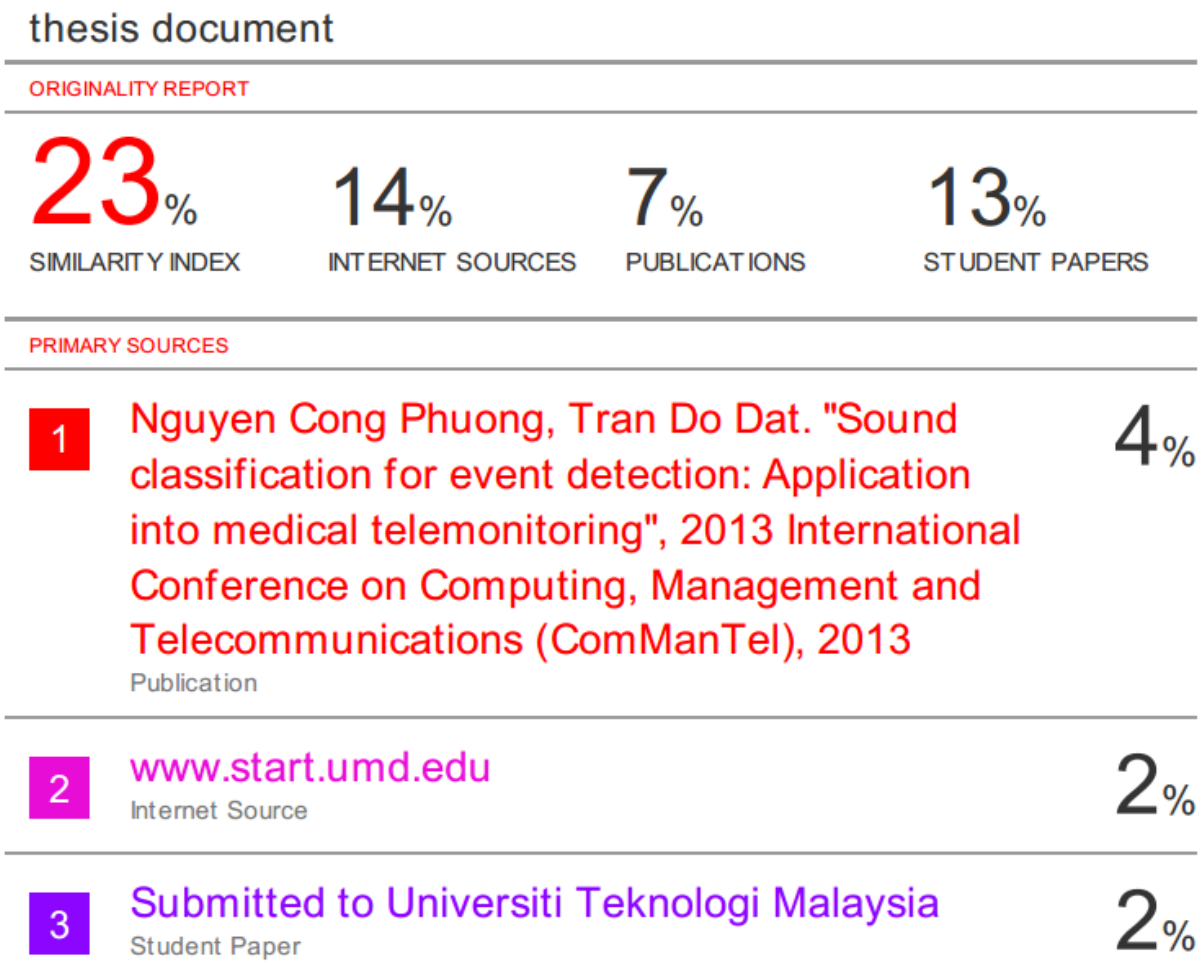

## Appendix C: Client Computer in a listening state



# Appendix D: Originality Report with name



**Appendix E: Originality Report with General percentage**



thesis document

ORIGINALITY REPORT

| 23% | 14% | 7% | 13% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Nguyen Cong Phuong, Tran Do Dat. "Sound classification for event detection: Application into medical telemonitoring", 2013 International Conference on Computing, Management and Telecommunications (ComManTel), 2013<br>Publication | 4% |
| 2 | www.start.umd.edu<br>Internet Source | 2% |
| 3 | Submitted to Universiti Teknologi Malaysia<br>Student Paper | 2% |