Scientific Research Publishing

Communications and Network, 2023, 15, 98-119
https://www.scirp.org/journal/cn
ISSN Online: 1947-3826
ISSN Print: 1949-2421

# Design and Simulation of an Audio Signal Alerting and Automatic Control System

**Winfred Adjardjah[1], John Awuah Addor[2], Wisdom Opare[1], Isaac Mensah Ayipeh[1]**

[1]Department of Electrical and Electronic Engineering, Takoradi Technical University, Takoradi, Ghana
[2]Department of Mathematics, Statistics and Actuarial Science, Takoradi Technical University, Takoradi, Ghana
Email: winfred.adjardjah@ttu.edu.gh, john.addor@ttu.edu.gh, wisdom.opare@ttu.edu.gh, isaac.ayipeh@ttu.edu.gh

**How to cite this paper:** Adjardjah, W., Addor, J.A., Opare, W. and Ayipeh, I.M. (2023) Design and Simulation of an Audio Signal Alerting and Automatic Control System. *Communications and Network*, **15**, 98-119.
https://doi.org/10.4236/cn.2023.154007

**Received:** July 8, 2023
**Accepted:** September 25, 2023
**Published:** September 28, 2023



## Abstract

A large part of our daily lives is spent with audio information. Massive obstacles are frequently presented by the colossal amounts of acoustic information and the incredibly quick processing times. This results in the need for applications and methodologies that are capable of automatically analyzing these contents. These technologies can be applied in automatic contentanalysis and emergency response systems. Breaks in manual communication usually occur in emergencies leading to accidents and equipment damage. The audio signal does a good job by sending a signal underground, which warrants action from an emergency management team at the surface. This paper, therefore, seeks to design and simulate an audio signal alerting and automatic control system using Unity Pro XL to substitute manual communication of emergencies and manual control of equipment. Sound data were trained using the neural network technique of machine learning. The metrics used are Fast Fourier transform magnitude, zero crossing rate, root mean square, and percentage error. Sounds were detected with an error of approximately 17%; thus, the system can detect sounds with an accuracy of 83%. With more data training, the system can detect sounds with minimal or no error. The paper, therefore, has critical policy implications about communication, safety, and health for underground mine.

## Keywords

Emergency Response, Emergency Management Team, Audio Signal Alerting, Automatic Control System, Uni Pro XL, Manual Communication, Fast Fourier Transform Magnitude, Zero Crossing Rate, Root Means Square

## 1. Introduction

Audio information plays a very important role in our everyday life. Audio in-

formation ranks along with video, as the two most dominant means by which humans perceive their environment. The overwhelming volumes of audio information and the sheer speed with which the information needs to be processed often present enormous challenges. This results in the need for applications and methodologies that are capable of automatically analyzing this content. These applications can then be leveraged in tasks such as automatic content analysis and emergencyresponse systems where they work in tandem with human operators to provide the needed information about the situation. Furthermore, interpreting audio information in the soundtracks of video can provide additional context for machine vision applications such as scene understanding [1] [2] [3]. Since a break in manual communication can occur in an emergency which can cause accidents and equipment damage, it is necessary to prevent this situation. The audio feature extraction and classification were performed based on convolutional neural network (CNN), which has the potential to extract more relevant sound spectrum characteristics of the audio data category [4].

The audio signal does a good job by sending a signal to the emergency control room that will give information on the type of emergency happening underground and the action to be taken by the emergency control center at the surface [5] [6]. Therefore, this research seeks to design and simulate an audio signal alerting and automatic control system using Unity Pro XL to prevent manual communication of emergencies and manual control of equipment [7]. This research aims to design and simulate an audio signal alerting and automatic control system for underground mine to improve communication and safety. To achieve this aim, the paper purposes to: detect fault and emergency conditions using sensors, integrate an alerting system with a buzzer to give a notification at the emergency control room, introduce an automatic switching system to turn on the standby pump when the duty pump is faulty and no action is taken after the alert and test the physical model of the system. The trained sound extracted from the simulation is used a as control signal to automatically send test sound level of the monitoring equipment at the underground pumps to alart the operators of an impending emergency. The location and the state of the equipment is displaced on the SCADA red if there is fault, for the necessary action to be taken. Hitherto the troubleshooting was done through manual inspection. The techniques used include employing programable logic control (PLC) and the supervisory control and data acquisition (SCADA) functionality of MATLAB to create the system's circuit design and simulation using Unity Pro XL [8] [9]. The contributions of this paper are:

1) The use of engineering tools and methods. That is the use of MATLAB, PLCs etc. to draw meaningful conclusions.

2) The created predictive model for identifying the various levels of sounds and their implications to the emergency response team.

## 2. Related Works

Nowadays, emergencies are the leading cause of fatalities. In addition to mortal-

ity, there is a general decline in health that occasionally coexists with significant social challenges, especially for individuals over the age of 60. These folks often find it very tough to escape during circumstances because of their weakness. For senior living facilities, Personal Emergency Response Systems (PERS) were developed to increase security [10]. A call button that the subscriber wears on a bracelet or neck chain, along with a two-way intercom that is connected to a phone line, make up the most popular type of PERS now in use. If help is needed, the subscriber just hits the button, which places an immediate call to a live operator via the intercom. The operator consults with the subscriber to determine the problem and choose the best course of action, which may include getting in touch with a neighbour, a member of the family, or an emergency response team. This condition frequently leads to false alerts due to accidental button pressing. This makes it more difficult for emergency personnel to react to actual circumstances. The push-button activator must be worn at all times, and subscribers must be conscious and physically able to activate it in case of emergency. This places a heavy load on subscribers under the current methods. The requirement to wear the activator causes stigma for many elderly people.

Using arrays of infrared sensors, thermal images of an occupant were produced [11]. The tracking system will be used, per the research, to notify the PERS. Regardless of the technique utilized for detection, once a PERS alarm has been triggered, the response effort must be coordinated with the user. By including the user, the PERS responds in the best way while maintaining the user's control over decisions affecting their health. Mann [12] set out to create an innovative approach to automatically connect the user to a call center when an alarm is triggered. In their investigation, they took into account the viability of applying automated dialog and artificial intelligence techniques. It improves the usability and efficiency of PERS for people over 60 in an emergency. Other research has also proven the use of Automatic Speech Recognition (ASR) with microphone arrays and speech recognition software to enable dialogue and dialog as a means of interacting with a PERS. The widespread use of the most recent ASR technology in a variety of fields, such as interactive voice response (IVR) telephone systems, office speech-to-text software, and others, is evidence of the enormous strides made in accuracy and practicality. ASR provides a simple, intuitive, and unobtrusive way to interact with the PERS by letting the user choose the best course of action in response to the detected alert, such as dismissing a false alarm or interacting directly with a call center operator [13]. The user has more control as a result.

In [14], the core problem of audio segmentation and categorization was examined. The topic at hand is how to distinguish between speech and music, the two most important types of audio. Speech and music both have incredibly varied temporal morphologies and spectral distributions, which makes it simple to classify sounds with a high degree of accuracy. When further classifying audio data, consideration may be given to sounds other than voice and music. After

the music was first recognized using the average amount of time that peaks last in a certain frequency range, pitch tracking was utilized to separate speech from music. Another area of interest for Grosche was the investigation of audio content-based music retrieval [15]. The information that verbally characterizes the audio content itself is a prerequisite for traditional retrieval algorithms. Since there are no textual descriptions accessible in this case, content-based retrieval algorithms that just use the raw audio data are required. In their contribution, they discussed a-query-by-example paradigm-based content-based retrieval strategies. In their approach, given an audio query, they get any documents from a music collection that are even vaguely linked to the audio query. They stressed that these strategies can be grouped generally according to their specificity, which is the degree of connection between the database records and the query. High specificity refers to a rigorous sense of likeness, whereas low specificity refers to one that is quite fuzzy. Additionally, they developed a theory of granularity-based classification that distinguishes between retrieval at the fragment and document levels. They used a categorization method based on specificity granularity to identify a variety of retrieval scenarios, such as audio identification, audio matching, and version identification. They concluded that they serve as a basis for user-oriented retrieval systems that incorporate a range of retrieval methods.

To greatly reduce the system's normal power consumption, Dong [16] examined the settings for automated fire alarm systems, proposed a wireless automatic fire alarm system architecture, and designed and created system module hardware, including a communication protocol. However, no sensors were used in his investigation. Smart emergency response solutions for fire threats based on the Internet of Things (IoT) have been introduced [17]. IoT links previously disconnected objects and people. First responders may connect and receive the information they need thanks to IoT. The emergency response for fire concerns was created using the standardized IoT structure. They adopted modern technology that modifies conventional forms of human-to-human or human-to-machine communication. High levels of intelligence and scalability are present in IoT.

In a nutshell, with the literature reviewed and other related materials, several researches have been done about audio content analysis and automatic emergency response However, little work has been done on the interpretation of the audio signals. This work, therefore, seeks to design an audio content interpretation system to detect whether a sound is an emergency or not.

## 3. Material and Methods

This chapter focuses on the design of an audio content interpretation device using digital signal processing (DSP) techniques and machine learning. The system will interpret the content of audio signals, whether they are emergency or not. Several audio signals will be collected from the internet, and audio sound data-

base and will be programmed through MATLAB. The features of these audio sounds will be extracted and grouped according to their classes. The scope of this work is limited to MATLAB software implementation.

### 3.1. Data Collection

The methods used to acquire data are covered in great length in this section. [18] offers a thorough data collection method for speech performance. For this experiment, audio samples that were picked at random from the internet were used. The information was compiled using two categories: human noises, such as everyday speech, shouts, and sobs, and emergency sounds, such as gunshots, earthquakes, natural disasters, and fire sounds. Male, female, and child speakers were present in the speech audio recording that was chosen. These files were all in various file types, including Windows Media Audio (WMA), Waveform (WAV), and MPEG-1 (MP3). The sounds were all converted to WAV format to be used in MATLAB programs and to guarantee that they all had the same sampling frequency.

### 3.2. Data Pre-Processing

Any type of processing done on raw data to prepare it for processing later is known as data preparation. A common first stage in the data mining process is data preparation, which transforms data into a format that can be processed for the user's purpose more quickly and efficiently. Cleaning, instance selection, normalization, transformation, feature extraction, sale, and auction are a few examples of data preparation techniques. The final training set is the result of data preparation. Pre-processing is crucial because real-world data is frequently unreliable, deficient in certain pertinent features, noisy, and contains errors or outliers. There are also inconsistencies (different names or codes) [19] [20].

Pre-processing can be done with a number several methods, including sampling, which selects a representative subset from a large population of data, transformation, which modifies raw data to create a single input, de-noising, which removes noise from data, normalization, which arranges data for easier access, and feature extraction, which isolates specific data that is significant in a given context [21]. It prepares the raw data for further processing. Users have access to a method for logical data processing for data mining [22]. "Data cleaning" is a general phrase for data processing that refers to replacing missing values, smearing noisy data, identifying or removing outliers, and correcting abnormalities in audio signals. The processes for data preparation are described below.

1) Data integration: It is done using multiple databases, data cubes, or files to integrate audio signals.

2) Data transformation: It is used to normalize and aggregate data signals. Normalization involves scaling attribute values to fall within a specified range.

3) Data reduction: The amount of signal is decreased, yet the analytical results are the same or almost the same.

4) Data discretization: It entails both data reduction and the substitution of notional qualities for numerical ones.

### 3.3. Audio Pre-Processing

Audio pre-processing, which comprises two phases, ensures the consistency of audio assets before employing them in a research from one session.

The pre-editing and standardization of raw audio processing are done before effect processing. This typically involves eliminating undesirable segments, such as a conversation between takes, coughing, sneezes, and any abnormal peaks, such as clicks, thumps, and paper rustling. The audio is then normalized to a predefined RMS level after measuring the Root Means Square (RMS) level, ensuring that all audio files have the same RMS level before any effect processing (FX processing) [23] [24].

### 3.4. Algorithm

There are many methods to represent data. Shifting data into a different field, specifically the frequency domain (amplitude of individual frequencies), is one method of showing data. The device receives sampled audio data in the form of a waveform (WAV), where values are recorded at specified intervals. Additionally, stated is the sample rate. The dataset is used to generate subsets for both training and validation. The selected attributes were extracted using the training data. On the training and validation datasets (non-emergency), emergency and regular sound labels were applied. In the training process, one stands for a routine sound, and zero for an emergency sound. [25].

In this work, the writers adopted machine learning technic by mimicking information from data, instead of depending on existing equations as models. The algorithms will modify its performance when there are additional examples there to learn from. This method was used since a lot of data was involved. In our research supervised learning techniques was employed to train a model on known input and output data to predict future outputs [26] [27]. During the training phase, the neural network algorithm in [28] was deployed to train the data to look for what its ideal output should be using binary values (0 or 1) to decide if the output is true or false.

### 3.5. Audio Feature Extraction

Feature extraction is the process of breaking down an audio signal into a collection of feature vectors that each contain a specific piece of information about the sound. The variety of audio features accessible for categorization tasks is extensive. There are two types of these features: time domain and frequency domain.

These features include Mel-Frequency Cepstral Coefficients (MFCC), frequency, pitch short-time are the main parameters investigated in this work include maximum values, minimum values, Zero Crossing Rate (ZCR), dynamic range values, Fast Fourier Transform (FFT) magnitude ratio, and Root Mean

Square Value (RMS) [29].

### 3.6. Frequency

A sound wave's frequency was based on the number of vibrational cycles that occur in a second. Cycles per second (cps) or Hertz (Hz) was used as express in [30]. High, low, or medium frequency options are available. The audible range of the audio signal is between 20 and 20,000 Hz. Infrasonic sounds are part of this frequency range, but ultrasonic sounds are those that fall outside of it [31].

### 3.7. Root Mean Square

Root mean square energy is based on all samples in a frame. It acts as an indicator of loudness, since higher the energy, louder the sound. It is however less sensitive to outliers as compared to the Amplitude Envelope. This feature has been useful in audio segmentation and music genre classification tasks.

The square root of the arithmetic mean of the squared deviations of the correctly predicted sounds from the actual sounds is known as the Root Mean Square (RMS). According to [32], the RMS is an extremely significant validation tool, where its usefulness in validating predictions was established. The distribution of RMS features is employed in this study to identify the boundaries between emergency and non-emergency sounds [33]. The method of boundaries is based on the amplitude distribution of the audio samples' dissimilarity measure. It can mathematically be calculated as:

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{i=n}\left(x_i - \overline{x}\right)^2} \tag{1}$$

where, $n$ = number of samples.

$\overline{x}$ = the mean value of the samples.

$x_i$ = the $i^{\text{th}}$ sample.

### 3.8. Fast Fourier Transform Magnitude Ratio

The Fast Fourier Transform (FFT) technique is used to compute the discrete Fourier transform. It transforms a signal from the time domain to the frequency domain by decomposing a function into a frequency domain composition. By comparing the magnitudes of the low-frequency and high-frequency FFT components, the FFT magnitude ratio calculates the difference between them [34].

In the time domain, RMS energy, zero crossing rate, and amplitude envelope are examples of extractions that can emanate from extracted waveforms of the raw audio. In the case of frequency domain: These focus on the frequency components of the audio signal. Signals are generally converted from the time domain to the frequency domain using the *Fourier Transform*. Band energy ratio, spectral centroid, and spectral flux are examples. Time-frequency representation: These features combine both the time and frequency components of the audio signal. The time-frequency representation is obtained by applying the Short-Time Fourier Transform (STFT) on the time domain waveform. Spectrogram,

mel-spectrogram, and constant-Q transform are examples.

### 3.8.1. Zero Crossing Rate

The Zero-Crossing Rate (ZCR) of an audio frame refers to the rate at which the signal alternates in signs during the frame. Differently expressed, it is defined as the frequency by which the signal changes in value; that is, alternating between negatives and positives, which is then divided by the frame's length. The equation below expresses the ZCR.

$$Z(i)=\frac{1}{W_L}\sum_{n=1}^{W_L}\left|\operatorname{sign}\left[x_i(n)\right]-\operatorname{sign}\left[x_i(n)-1\right]\right| \tag{2}$$

where $\operatorname{sgn}(\cdot)$ is the sign function expressed as

$$\operatorname{sgn}\left[x_i(n)\right]=\begin{cases}1, & x_i(n)\geq 0\\ -1, & x_i(n)<0\end{cases}$$

ZCR can be thought of as a measurement of a signal's noise level. For instance, in the situation of noisy signals, it typically displays greater values. It is also known to reflect a signal's spectrum properties, albeit somewhat coarsely. Its adoption by several applications, such as speech-music discrimination, speech detection, and music genre categorization, to name a few, has been attributed to its simplicity in computation. In speech frames, the respective ZCR values are often lower (depending, of course, on the type and context of the phoneme that is pronounced each time), but the values for the noisy regions of the signal are higher. It's noteworthy to note that speech signals can exhibit higher standard deviation for this attribute across subsequent frames than musical signals [35].

### 3.8.2. Dynamic Range

The difference between the maximum and minimum values is calculated. In terms of math, it is calculated as: [36].

$$\text{Dynamic Range} = \text{Maximum Values} - \text{Minimum Values} \tag{3}$$

**Figure 1** shows the flow chart which is developed to create a representation of the sequence of operations that are carried out to define the logic basis for the programming.

## 4. Results and Discussions

This chapter focuses on validating the audio content interpreter designed. To ensure the effectiveness of this system, many audio signals (pure human speech, sounds from fire or flames, gunshots, or explosions) were used. Six features were used to know the distinct characteristic of these sounds.

### 4.1. Selected Features Extracted

Regarding the desired attributes of the original data, these features must be illuminating. Because we want our analysis algorithms to be based on a relatively limited number of features, feature extraction can also be thought of as a data

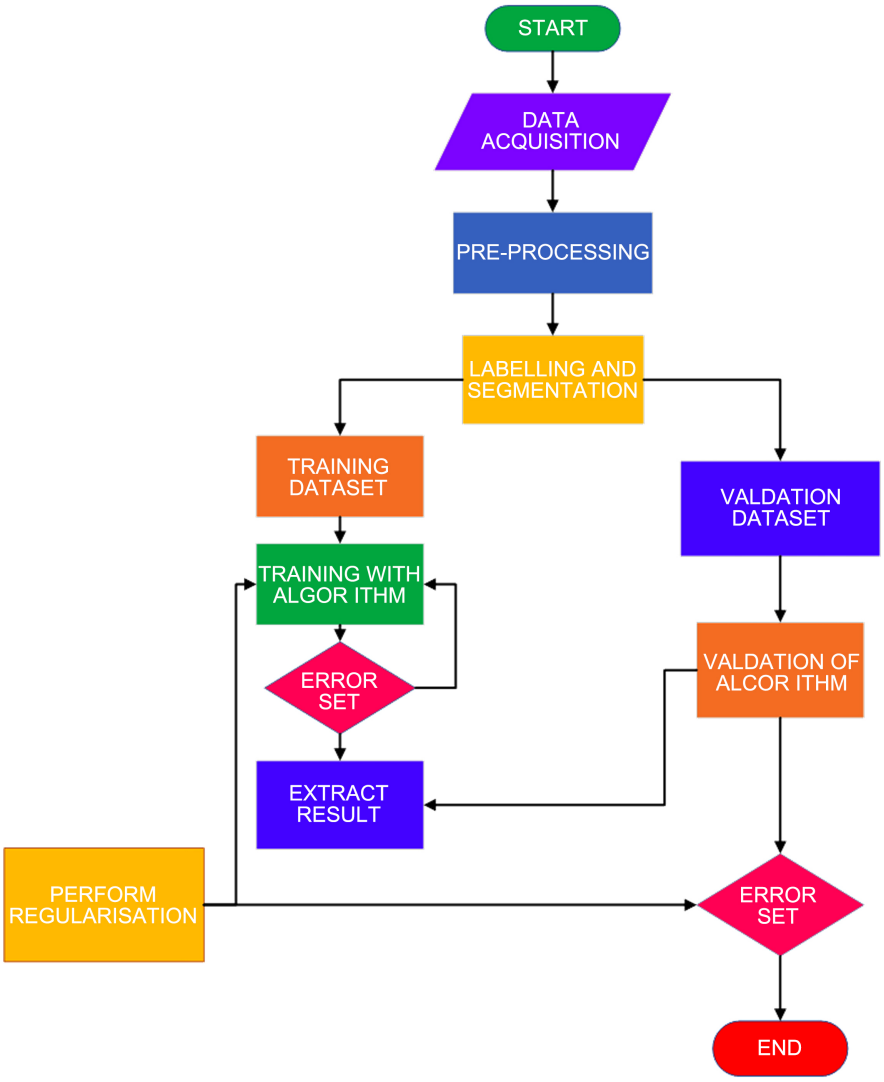


**Figure 1.** Flow chart of the system operation.

rate reduction technique. The audio signal in our situation is a large amount of original data, making it difficult to process it directly for any analysis work. Therefore, we must convert the initial data representation into one that is more appropriate by removing audio aspects that describe the characteristics of the original signals while lowering the data volume. For example, the divergence of the signal's energy when discriminating between speech and music segments is an attractive feature candidate since this feature has a physical meaning that matches well with the specific classification task [37].

#### 4.1.1. Short-Term and Mid-Term Processing

The majority of audio analysis and processing techniques involve short-term feature extraction, which divides the data into short-term frames (windows). The feature extraction stage likewise uses this method; the audio stream is divided into potentially overlapping frames, and a collection of features is computed for each frame. Each audio signal subjected to this form of processing yields a se-

quence, F, of feature vectors [38].

#### 4.1.2. Mid-Term Windowing in Audio Feature Extraction

The processing of the feature sequence on a mid-term basis is another typical technique. This method of processing divides the audio signal first into mid-term windows, and then performs the short-term processing stage for each window. In a subsequent phase, feature statistics, such as the average value of the zero-crossing rate, are computed using the feature sequence, F, which is extracted from a mid-term segment. In the end, a set of statistics that correspond to the corresponding short-term feature sequences for each mid-term segment serve as its representation. Depending on the application domain, the duration of mid-term windows often ranges from 1 to 10 seconds in practice [39]. This aids to-create the corresponding mid-term audio statistics and divide a big audio file (or audio stream) into mid-term windows. That is when the audio file being examined is particularly long in duration, loading all of its material at once may be impossible due to memory constraints. For this reason, the function shows how to read a huge content of audio file gradually using data chunks (blocks), which may be one minute long.

#### 4.1.3. Extracting Features from an Audio File

The presentation of some of the most significant and popular audio features will now be done, along with the MATLAB code and examples that go with them. In the context of the examples given, we also discuss a number of statistics that, over the long term, offer respectable discrimination capabilities among the audio classes. The feature extraction functions are to be called inside a short-term analysis process, as is the case with function Feature Extraction, and it should be noted that we presume the input to the feature extraction functions is an audio frame [40].

#### 4.1.4. Time-Domain Audio Features

Typically, the audio signal's samples are used to directly extract the time-domain audio properties. The short-term energy and short-term zero-crossing rate are typical instances. Although it is typically essential to integrate them with more complex frequency-domain characteristics, such features provide a straightforward method of analyzing audio signals [41].

$$E(i)=\sum_{n=1}^{W_L}\left|x_i(n)\right|^2 \tag{4}$$

Usually, dividing energy by $W_L$ to eliminate the dependency of the length of the frame leads to normalization of energy. Subsequently, (4) reduces to

$$E(i)=\frac{1}{W_L}\sum_{n=1}^{W_L}\left|x_i(n)\right|^2 \tag{5}$$

The power of the signal is given by (5).

#### 4.1.5. Frequency-Domain Audio Features

MATLAB's built-in function makes it simple to compute a signal's discrete fourier transform (DFT). Because it offers a practical representation of the distribution of the frequency content of sounds, or of the sound spectrum, DFT is commonly employed in audio signal analysis. The DFT of the audio signal serves as the foundation for a few often-utilized audio features. Frequency-domain (or spectral) audio characteristics are another name for these features. We first compute the DFT of the audio frames using the get DFT method, before we can compute the spectral characteristics. Each audio frame's DFT is calculated by function Feature Extraction. And numerous spectrum characteristics are computed using the DFT coefficients that are produced [41].

#### 4.1.6. MFCCs Mel-Frequency Cepstrum Coefficients (MFCCs)

In our work, MFCCs is used a sort of cepstral representation of the signal in which, as opposed to the linearly spaced method, the frequency bands are spread according to the mel-scale. MFCCs are extracted from a frame by performing the following steps:

1) A DFT computation is made.

2) The resulting spectrum is fed into a bank of L filters in a mel-scale filter. The frequency responses of the filters typically exhibit, triangular overlap. In an effort to be consistent with some psychoacoustic observations that suggest the human auditory system may discriminate nearby frequencies more easily in the low-frequency domain, the mel-scale incorporates a frequency warping effect [42] [43] [44].

The plots of the maximum values of the four sound groups; that is gunshot, fire, scream, and speech (normal sound) are presented. The maximum values of these sounds are compared. However, the minimum values show the least values from the four audio sounds. From comparison in **Figure 2** and **Figure 3**, it was noticed that gunshots had the least value.

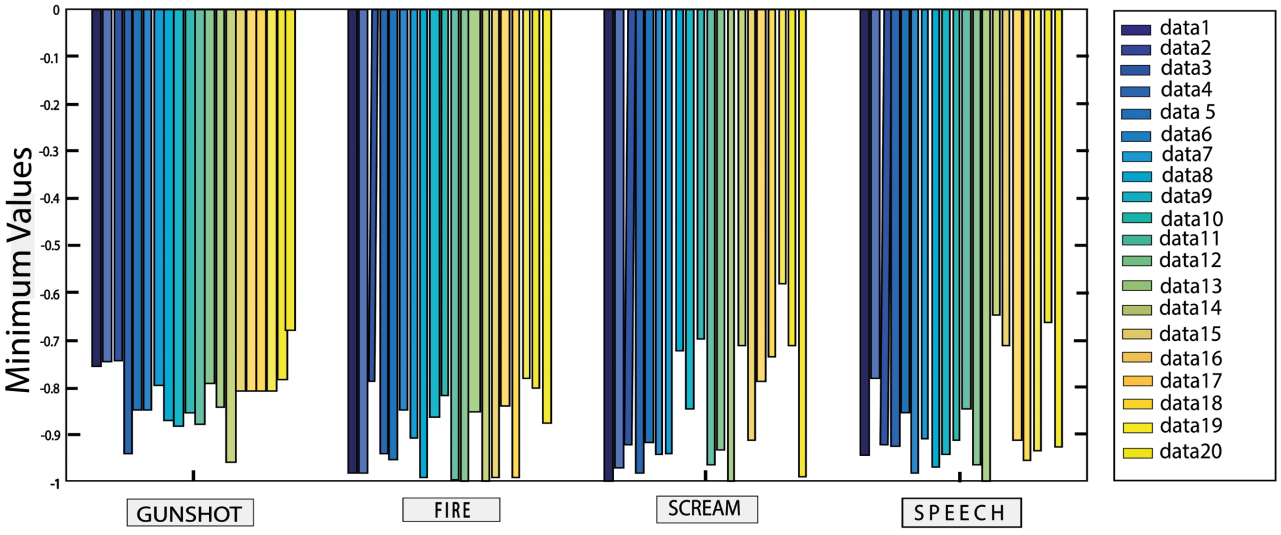


**Figure 2.** Minimum values of selectedsound types.

The minimum and maximum values of the selected soundhelped to find the average minimum sound level and average maximum sound level for all the selected sound types. The average sound level for Gunshot, Fire and scream was found which was used as the setpoint or the reference point. Example, if the average sound level for gunshot is 100 dB, fire is 70 dB and scream is 40 dB, when any signal is sent to the control room whose sound level is 100 dB and above everybody will know it is incident which involves Gunshots and if it is 70 dB and above it is incident which involves fire.

Comparing Figure 3 and Figure 4, it is seen clearly, the distinctness of the two graphs. Hence these two features were selected for the training of the model based on the maximum variation across the various classes and the minimum variance across samples in the same class.

With the dynamic range of the selected sound, it is possible to know the loudest sound peak for each sound type that will be sent as in Figure 5. For instance,

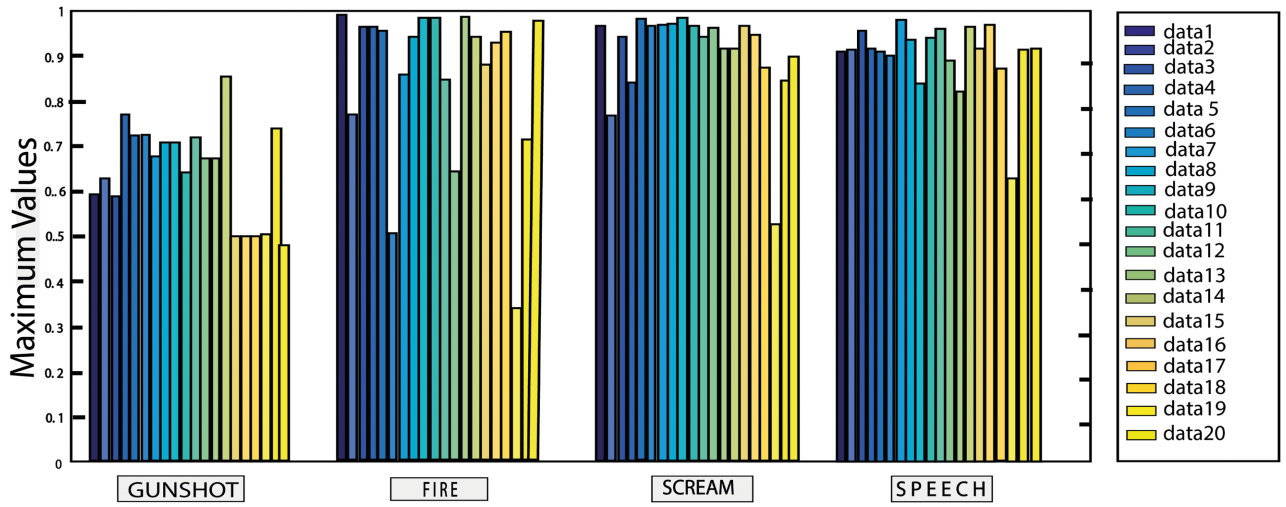


Figure 3. Maximum values of selected sound types.

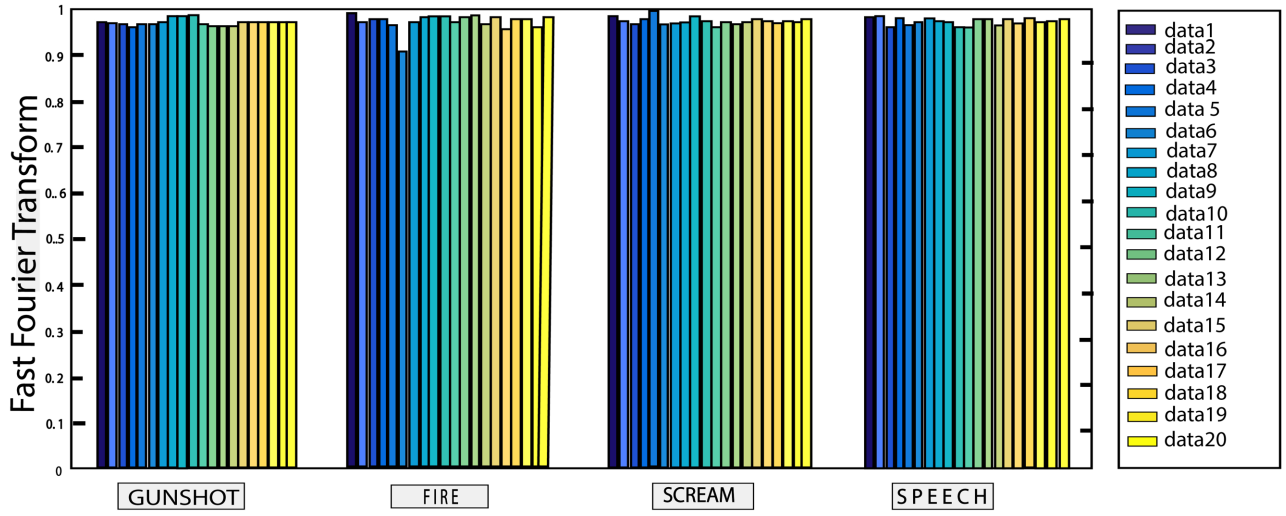


Figure 4. FFT magnitude ratios of selected sound types.

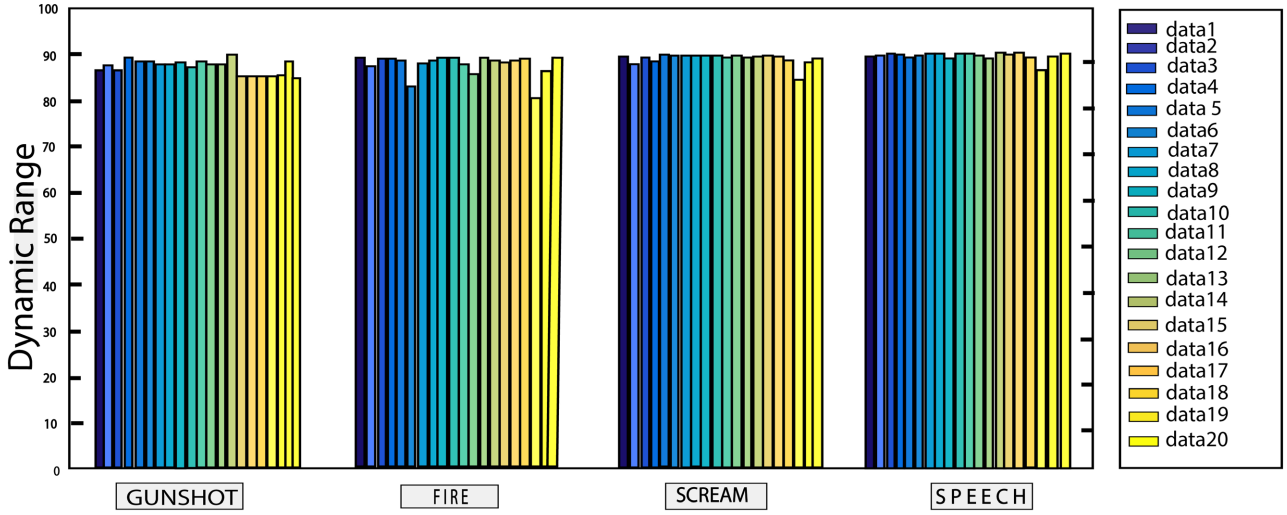


**Figure 5.** Dynamic range of selected sound types.

if the human auditory system has a dynamic range of about 90 dB and gunshot in the range of 160 to 168 dB, any signal sent from the emergency response unit which gives more than the 160 dB will definitely indicate incident involves gunshots.

#### 4.1.7. Comparison of FFT Magnitude Ratio and ZCR

The disparity of the ZCR for gunshot differs strongly from the other sounds (**Figure 6**). With regard to the RMS, the scream has the highest average amplitude, followed by gunshot, speech and fire (**Figure 7**). This means that scream contains the highest energy waveform, while fire has the lowest energy waveform. For that of FFT, the variance between each sound group is minimum (**Figure 4**), indicating that the selectedsound types share similar acoustic (or similar temporal and frequency) features.

### 4.2. Simulation of the System

This study makes use of two datasets. The system is trained using the first dataset, and the classification model is tested using the second dataset. Selected audio sound kinds are taught in this procedure so that the system can tell one sound type from another. It is then tested to check the validity and the errors generated during the running of the process. A neural network was trained using the graphs in **Figure 4** and **Figure 6** (ZCR and FFT) that had the most distinguishable features out of all the features that were plotted. This training phase is carried out to enable the system to evaluate the relative weights of the chosen features and the respective contributions of each to the sound classification. It is trained to classify emergency sounds as zero (0) and normal sounds as one (1).

### 4.3. Analysis of Extracted Features

From the above feature plots, it is seen that the FFT magnitude ratio has the

lowest variance among samples of the same class. Also, the ZCR shows the widest distinctions across the various classes. Therefore, these two features were selected for the training of the neural network. The training data for emergency sounds were 150 files and that of normal sounds were 200 files (Table 1).

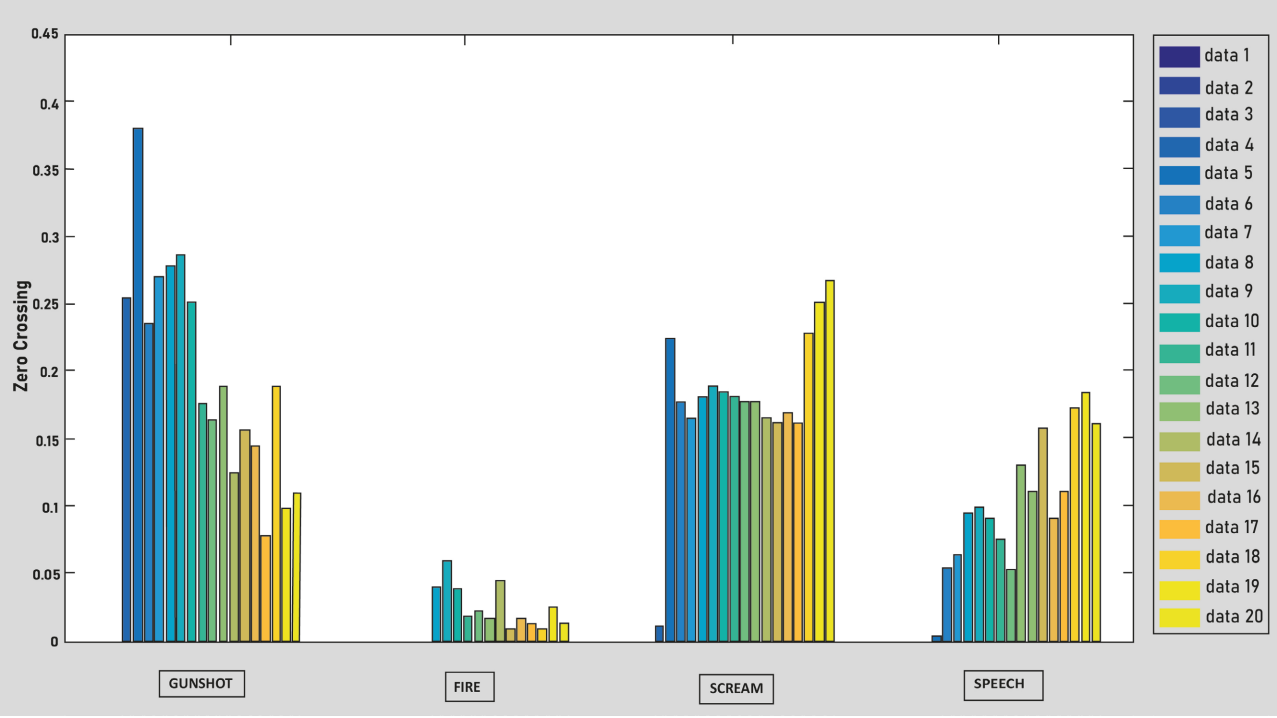


Figure 6. ZCR values of selected sound types.

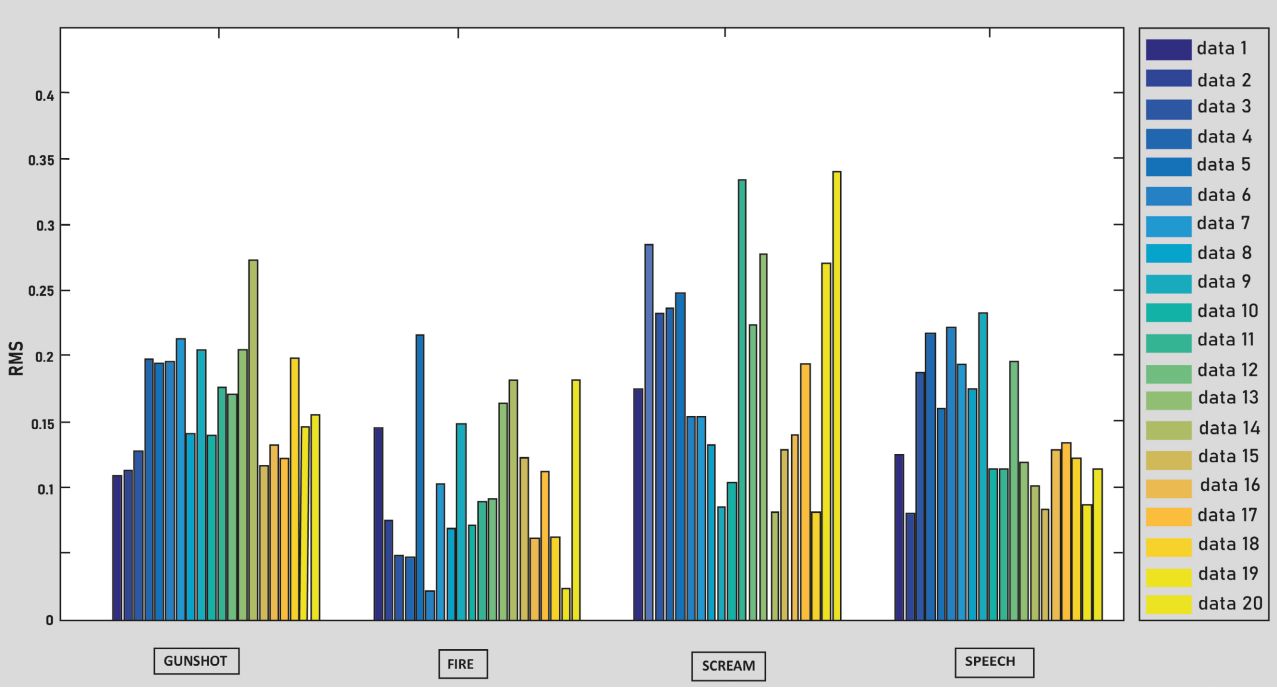


Figure 7. RMS values of selected sound types.

**Table 1.** Segmentation of collected audio files.

| Audio Class | Number of Audio Files |
|---|---|
| Emergency | 150 |
| Normal Sound | 200 |

### Validation of the System

Validation was done to determine if the system can perform on test data as expected from the training. As indicated by the higher RMS values of the selected sounds types, the variance for each sound type is high, showing that sounds were produced at varying frequencies or pitches. It also an indicates the ability of the system to detectboth maximum and minimum variance of the selected features (ZCR and FFT). **Table 2** presents a status summary of sounds predicted.

The result indicates that 123 audio files were correctly identified as emergency sound and 169 were correctly identified as normal sound. 27 emergency sounds were incorrectly identified as emergency sounds and 31 normal sounds were incorrectly identified as normal sounds. From the analysis, we compute the percentage of the error generated. The percentage error is calculated mathematically as:

$$\%\text{Error} = \frac{\text{Number of wrong predictions}}{\text{Total number of sample size}} \times 100 \tag{6}$$

$$\%\text{Error} = \frac{27+31}{150+200} \times 100$$
$$= 16.57\%$$

This indicates that an accuracy of 83.43% was achieved.

The changes in errors show that the initial training error was zero (due to overfitting) and spiked to 0.13 after the third sample. But as the training process progresses, the error reduces steadily to about 0.06. Considering the simulation, it can be seen that all the features are plotted and clearly shows the distinctness of one from the other. Moreover, the simulation, it resulted that the system was able to:

1) Detect automatically whether a particular sound is an emergency sound or not.

2) Show the maximum variation and the minimum variance of the selected features (ZCR and FFT).

3) Reached accuracy of 83% which shows that if more audio sounds were trained, an accuracy of about 100% will be achieved and this system will be very efficient.

As dipicted in **Figure 8**, to test the sound level of the monitoring equipment at the underground the signal was sent at the same time to know the sound level when the alarm is running till the time the alarm will be aspected or stop. The volume of the sound as it is running goes up at every point in time till the alarm is aspected or stop before the the volume of the sound start to go down and stop.

In **Figure 9**, a setpoint of 50 dB was used as the reference point so that any alarm that goes above the 50 dB the operator will see it as an emergency alarm and give information out for the workers and the authorities to respond to the kind of emergency and action that need to be taken.

**Table 2.** Results from training of audio data.

| Prediction | Emergency Sound | Normal Sound |
|---|---|---|
| Predicted Correctly | 123 | 169 |
| Predicted Wrongly | 27 | 31 |

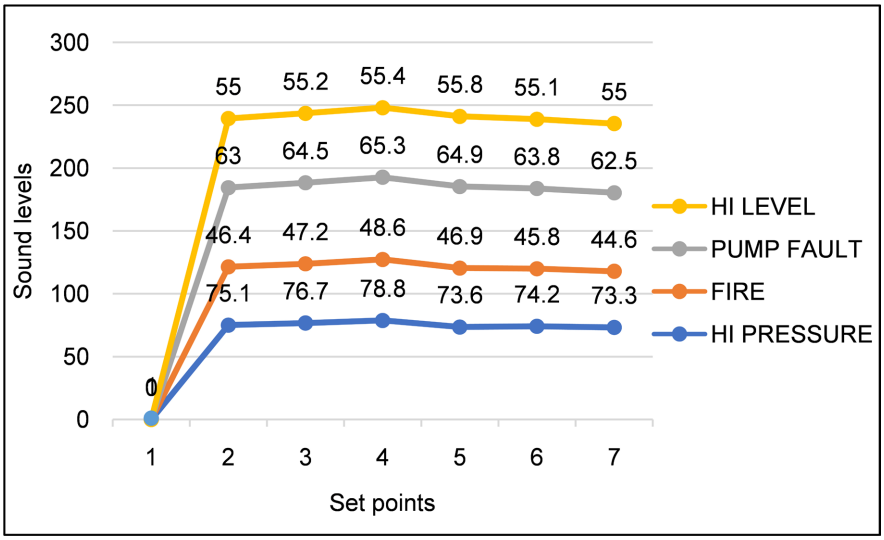


**Figure 8.** Tested sound levels.

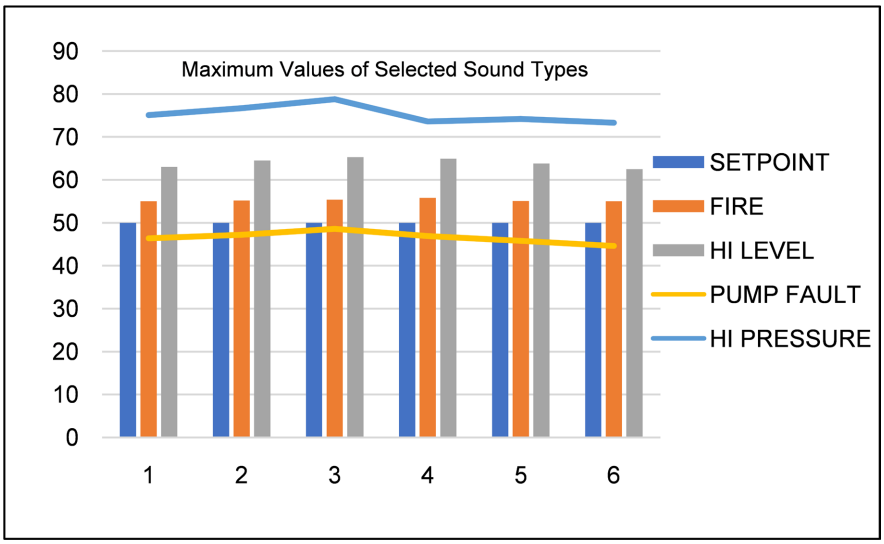


**Figure 9.** Results of the sound against the setpoint.

Figure 10 shows the functional block stages of the audio alarm system. The first stage talks about the fire level alarm, the second stage talks about the water level system. The third stage indicates the pump fault alarm systems, the fourth level depicts the pressure level, whiles the fifth level talks about the human emergency and the sixth or the last level indicates the running of the motor.

The human operators access the data through the computer which is a SCADA

system as illustrated in Figure 11. The software interprets and displays the data in an easy-to-understand form, so as to enable operators to quickly analyse and react to the alarm system. It also serves as records reference for future trouble-shooting.

With the location, it is displayed on the SCADA the location of the emergency and the condition of equipment at the emergency control room for action to be taken. The indication on the screen comes as red or green; when the system is in good condition, it shows green on the screen and when the system is not in good condition it shows red to alert the operator at the emergency control room to take necessary control measures. Again, when there is a fire at any level it will show on the screen.

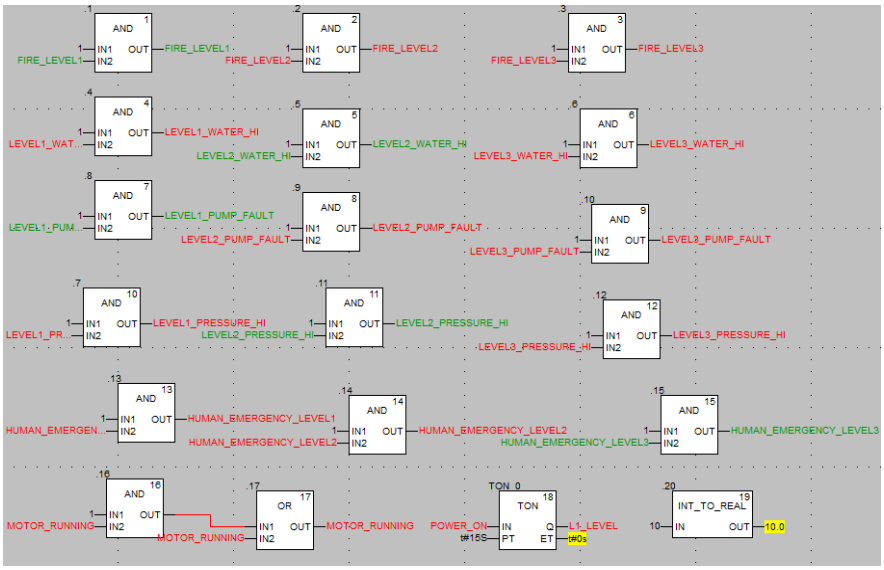


Figure 10. Function block diagram for audio alarm systems.

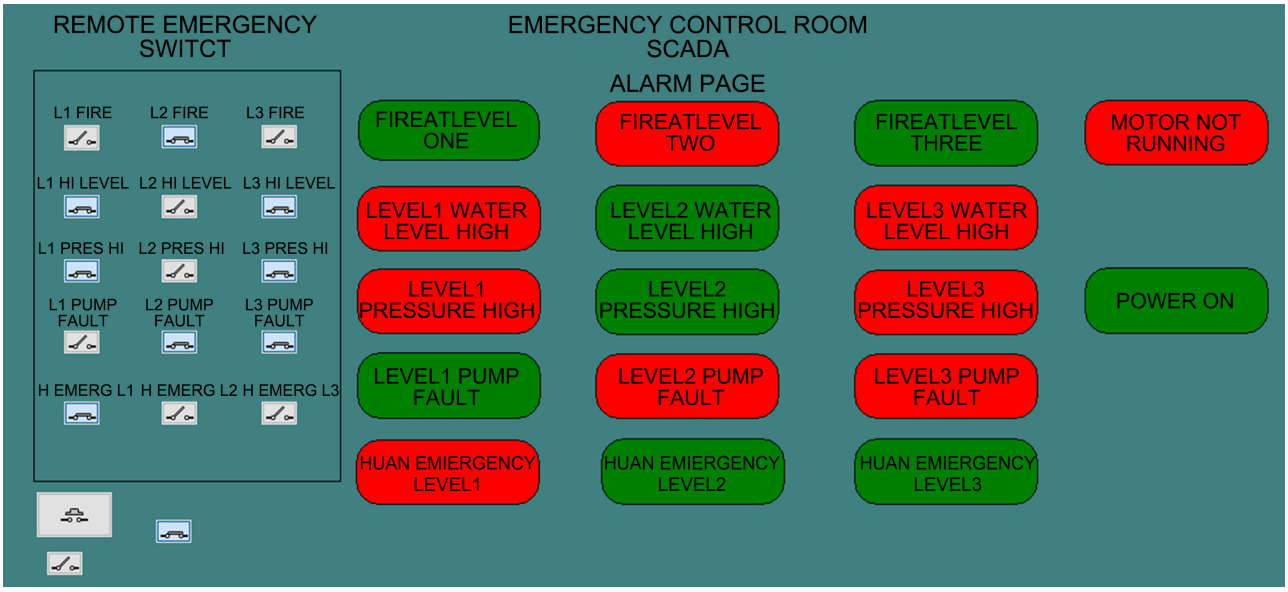


Figure 11. Simulation of the design in unity pro XL.

As depicted in Figure 12, the circuit design of the two pumps is to start one pump at a time. Sensors monitoring the safety interlock to give commands to the pumps to start. When the running pump is faulty, the second pump automatically takes over and runs after 10 seconds. Pressure and flow sensors are to monitor the flow and pressure to give information to the control room. This image is the graphic Representation of the circuit diagram above which shows the two pumps When running in automatic operation. With the graphic representation of the equipment on SCADA, it helps the operator to monitor and control without necessary someone being on the equipment to start it manually.

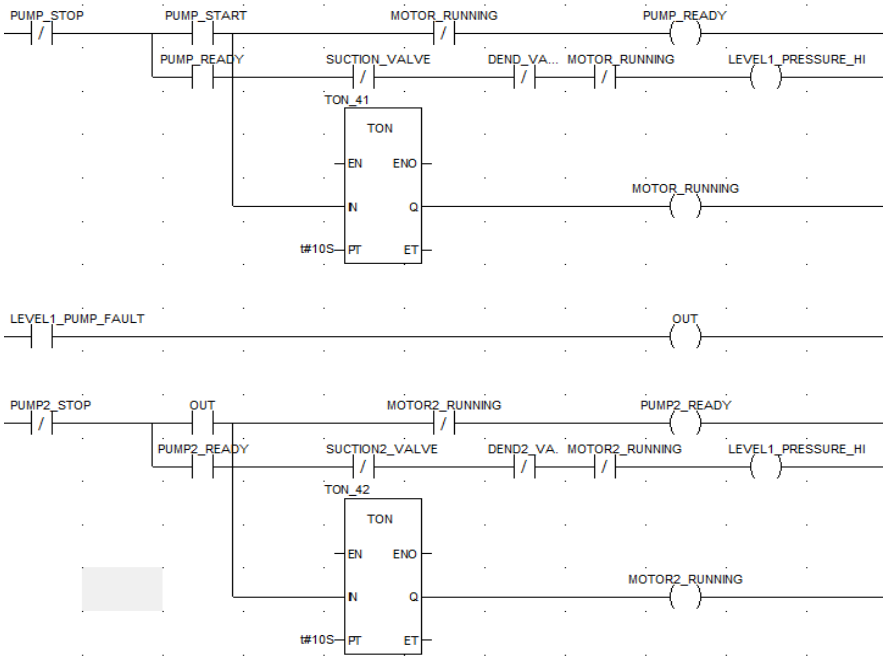


Figure 12. The circuit design of the system.

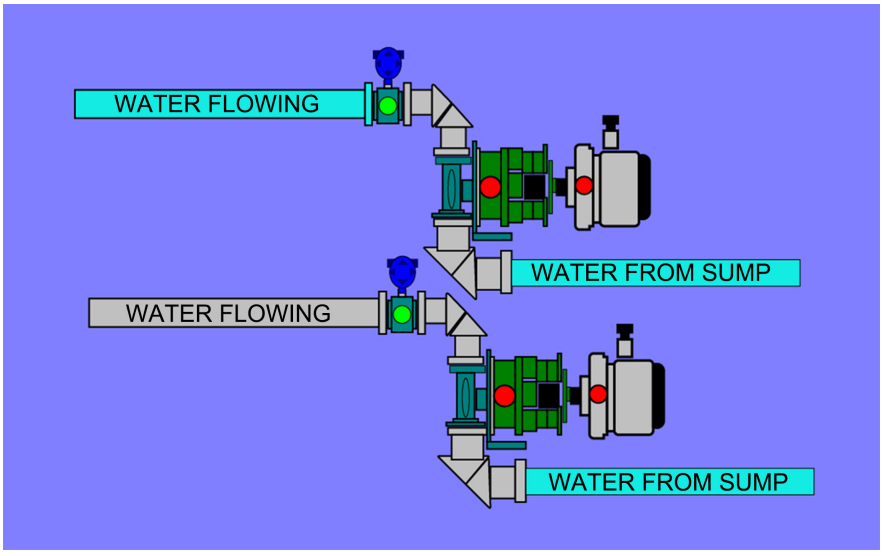


Figure 13. Automation and simulation of the automatic pumps.

With the current situation, the operation of the pump and communication to the emergency response is done manually by the operator being at the pump chamber 2 hours per shift. This put the operator in a very high-risk working environment. With the automation of the pumps (**Figure 13**), the pumps can be controlled automatically in the emergency control room and monitor the equipment by the sensor to automatically start and stop the motor in an emergency condition. A limitation to the application of the system emanates from the absence of a GSM module, with the sure implication that information cannot be directly channelled to the emergency response team upon the detection of an emergency. It is, therefore, recommended that future research in this direction should extend this paper by incorporating a GSM module so that the information can be channelled directly to the emergency response team immediately as an emergency is detected.

## 5. Conclusion

Considering the main aim of this paper, an automatic audio signal alerting and automatic control system has been designed for underground mine to improve communication and safety. The proposed system can automatically interpret audio signals, to classify them either as emergency or normal sounds. This information can be relayed to emergency authorities to facilitate fast emergency response. Sounds were detected with an error of approximately 17%, indicating the ability of the designed system to detect sounds with an accuracy of 83% approximately. Thus, with more data training, the system can detect sounds with an accuracy rate close to 100%. Conclusively, an automatic control system has been successfully designed and implemented to improve the safety of workers. The paper, therefore, has critical policy implications about communication, safety, and health for underground mine and other risky works. The emergency response team will know the type of incident and prepare for. The need for operator being on the equipment the whole of his shift to monitor and record every event will be done by the system itself at the control room which will reduce the labour cost for the company.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Can, A.B., *et al.* (2022) A New Method of Automatic Content Analysis in Disaster Management. 2022 10*th International Symposium on Digital Forensics and Security* (*ISDFS*), Istanbul, 6-7 June 2022, 1-5.

[2] Van Der Meer, T.G. (2016) Automated Content Analysis and Crisis Communication Research. *Public Relations Review*, **42**, 952-961. https://doi.org/10.1016/j.pubrev.2016.09.001

[3] Liang, J., *et al.* (2015) Detecting Semantic Concepts in Consumer Videos Using Au-

dio. 2015 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Vol. 2, 2279-2283.

[4] Zhang, J. (2021) Music Feature Extraction and Classification Algorithm Based on Deep Learning. *Scientific Programming*, **2021**, Article ID: 1651560. https://doi.org/10.1155/2021/1651560

[5] Neumann, P.G. (1994) Computer-Related Risks. Addison-Wesley Professional, Boston.

[6] Laurence, D. (2005) Safety Rules and Regulations on Mine Sites—The Problem and a Solution. *Journal of Safety Research*, **36**, 39-50. https://doi.org/10.1016/j.jsr.2004.11.004

[7] Topp, K., *et al.* (2018) Anatomical Society Summer Meeting in Galway.

[8] Sujyothi, P. (2014) Software Test Rig Development and Testing Sequences for Genset Paralleling System: A Research Report. Doctoral Dissertation, Vellore Institute of Technology, Vellore.

[9] Michael, I.A. (2022) Design of Lift Group Control Systems Based on PLC. *Journal of Engineering Research and Reports*, **22**, 31-44. https://doi.org/10.9734/jerr/2022/v22i317528

[10] Porter, E. (2005) Wearing and Using Personal Emergency. *Journal of Gerontological Nursing*, **31**, 26-33. https://doi.org/10.3928/0098-9134-20051001-07

[11] Sixsmith, A. (2004) A Smart Sensor to Detect the Falls of the Elderly. *IEEE Pervasive Computing*, **3**, 42-47. https://doi.org/10.1109/MPRV.2004.1316817

[12] Mann, W. (*2005*) Use of Personal Emergency Response Systems by Older Individuals with Disabilities. *Assistive Technology*, **17**, 82-88. https://doi.org/10.1080/10400435.2005.10132098

[13] Hamill, M. (2018) Development of an Automated Speech Recognition Interface for Personal Emergency Response Systems. *Journal of NeuroEngineering and Rehabilitation*, **6**, Article No. 26. https://doi.org/10.1186/1743-0003-6-26

[14] Govender, D. (2018) Investigating Audio Classification to Automate the Trimming of Recorded Lectures. University of Cape Town, Cape Town.

[15] Grosche (2018) Audio Content-Based Music Retrieval. Artificial Intelligence Research Institute (IIIA-CSIC), Campus UAB, Barcelona.

[16] Dong (2016) Design of Wireless Automatic Fire Alarm System. *Procedia Engineering*, **135**, 413-417. https://doi.org/10.1016/j.proeng.2016.01.149

[17] Lakshmana, P. (2018) Efficient Smart Emergency Response System for Fire Hazards Using IoT. *International Journal of Advanced Computer Science and Applications*, **9**, 314-320. https://doi.org/10.14569/IJACSA.2018.090143

[18] Félix-Brasdefer, J.C. (2010) Data Collection Methods in Speech Act Performance. In: Martínez-Flor, A. and Usó-Juan, E., Eds., *Speech Act Performance*: *Theoretical, Empirical and Methodological Issues*, John Benjamins Publishing Company, Amsterdam, 69-82. https://doi.org/10.1075/lllt.26.03fel

[19] Wang, Z., *et al.* (2020) Data Pre-Processing Methods for Electrical Impedance Tomography: A Review. *Physiological Measurement*, **41**, 2-9. https://doi.org/10.1088/1361-6579/abb142

[20] Majidi, M., *et al.* (2015) Improving Pattern Recognition Accuracy of Partial Discharges by New Data Pre-Processing Methods. *Electric Power Systems Research*, **119**, 100-110. https://doi.org/10.1016/j.epsr.2014.09.014

[21] Alasadi, S.A. and Bhaya, W.S. (2017) Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences*, **12**, 4102-4107.

[22] Danubianu, M. (2015) Step-by-Step Data Pre-Processing for Data Mining. A Case Study. *Proceedings of the International Conference on Information Technologies* (*InfoTech*-2015), Varna, 17-18 September 2020, 117-124.

[23] Chicco, D. (2017) Ten Quick Tips for Machine Learning in Computational Biology. *Biodata Mining*, **10**, Article No. 35. https://doi.org/10.1186/s13040-017-0155-3

[24] Choi, K., *et al.* (2018) A Comparison of Audio Signal Pre-Processing Methods for Deep Neural Networks on Music Tagging. 2018 26*th IEEE European Signal Processing Conference* (*EUSIPCO*), Rome, 3-7 September 2018, 1870-1874. https://doi.org/10.23919/EUSIPCO.2018.8553106

[25] McLoughlin, I.V. Speech and Audio Processing: A MATLAB-Based Approach. Cambridge University Press, Cambridge.

[26] Greener, J.G., *et al.* (2022) A Guide to Machine Learning for Biologists. *Nature Reviews Molecular Cell Biology*, **23**, 40-55. https://doi.org/10.1038/s41580-021-00407-0

[27] Zhou, Z.H. (2021) Machine Learning. Springer Nature, Berlin. https://doi.org/10.1007/978-981-15-1967-3

[28] Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H. (2018) State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon*, **4**, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

[29] Moffat, D., Ronan, D. and Reiss, J.D. (2015) An Evaluation of Audio Feature Extraction Toolboxes.

[30] Tate, M. (2013) Principles of Hearing and Audiology. 283.

[31] Piersol, A.G. and Paez, T.L. (2010) Harris' Shock and Vibration Handbook. McGraw-Hill Education, London.

[32] Addor, J.A., Wiah, E.N. and Alao, F.I. (2022) An Improved Two-States Cyclical Dynamic Model for Plastic Waste Management, *Asian Research Journal of Mathematics*, **18**, 52-68. https://doi.org/10.9734/arjom/2022/v18i530378

[33] Slabbekoorn, H., Bouton, N, van Opzeeland, I., Coers, A., ten Cate, C. and Popper, A.N. (2010) A Noisy Spring: The Impact of Globally Rising Underwater Sound Levels on Fish. *Trends in Ecology & Evolution*, **25**, 419-427. https://doi.org/10.1016/j.tree.2010.04.005

[34] Takahashi, D. (2019). Fast Fourier Transform Algorithms for Parallel Computers. Springer, Singapore. https://doi.org/10.1007/978-981-13-9965-7

[35] Joo, S., Choi, J., Kim, N. and Lee, M.C. (2021) Zero-Crossing Rate Method as an Efficient Tool for Combustion Instability Diagnosis. *Experimental Thermal and Fluid Science*, **123**, Article ID: 110340. https://doi.org/10.1016/j.expthermflusci.2020.110340

[36] Room, C. (2021) Audio Feature Extraction. *Machine Learning*, **16**, 51.

[37] Zaw, T.H. and War, N. (2017) The Combination of Spectral Entropy, Zero Crossing Rate, Short Time Energy and Linear Prediction Error for Voice Activity Detection. 2017 20*th IEEE International Conference of Computer and Information Technology* (*ICCIT*), Dhaka, 22-24 December 2017, 1-5. https://doi.org/10.1109/ICCITECHN.2017.8281794

[38] Sharm, A.G., Umapathy, K. and Krishnan, S. (2020) Trends in Audio Signal Feature Extraction Methods. *Applied Acoustics*, **158**, Article ID: 107020. https://doi.org/10.1016/j.apacoust.2019.107020

[39] Martin-Morato, I., Cobos, M. and Ferri, F.J. (2018) Adaptive Mid-Term Represen-

tations for Robust Audio Event Classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **26**, 2381-2392.
https://doi.org/10.1109/TASLP.2018.2865615

[40] Saddam, S.A.W. (2022) Wind Sounds Classification Using Different Audio Feature Extraction Techniques. *Informatica*, **45**, 57-65.
https://doi.org/10.31449/inf.v45i7.3739

[41] Wu, J., Xu, Y., Zhang, S.X., Chen, L.W., Yu, M., Xie, L. and Yu, D. (2019) Time Domain Audio Visual Speech Separation. 2019 *IEEE Automatic Speech Recognition and Understanding Workshop* (*ASRU*), Sentosa, 14-18 December 2019, 667-673.
https://doi.org/10.1109/ASRU46091.2019.9003983

[42] Bartusiak, E.R. and Delp, E.J. (2022) Frequency Domain-Based Detection of Generated Audio.

[43] Fang, Y., Liu, D., Jiang, Z. and Wang, H. (2023) Monitoring of Sleep Breathing States Based on Audio Sensor Utilizing Mel-Scale Features in Home Healthcare. *Journal of Healthcare Engineering*, **2023**, Article ID: 6197564.
https://doi.org/10.1155/2023/6197564

[44] Ayvaz, U., Gürüler, H., Khan, F., Ahmed, N., Whangbo, T. and Bobomirzaevich, A. (2022) Automatic Speaker Recognition Using Mel-Frequency Cepstral Coefficients through Machine Learning. *CMC-Computers Materials & Continua*, **71**, 5511-5521.
https://doi.org/10.32604/cmc.2022.023278