

**REAL TIME ABNORMAL SOUND DETECTION AND CLASSIFICATION FOR  
HOME ENVIRONMENT**

By

**SITI NUR AZIMAH BT MOHAMED AMIN**

Dissertation submitted in partial fulfillment of  
the requirements for the  
Bachelor of Engineering (Hons)  
(Electrical & Electronics Engineering)

SEPT 2011

Universiti Teknologi Petronas  
Bandar Seri Iskandar  
31750 Tronoh  
Perak Darul Ridzuan

**CERTIFICATION OF APPROVAL**

**Real Time Abnormal Sound Detection and Classification for Home Environment**

by

**Siti Nur Azimah bt Mohamed Amin**

A project dissertation submitted to the  
Chemical Engineering Programme  
Universiti Teknologi PETRONAS  
In partial fulfillment of the requirement for the  
BACHELOR OF ENGINEERING (Hons)  
(ELECTRICAL & ELECTRONIC ENGINEERING)

Approved by,



---

(Puan Zazilah bt May)

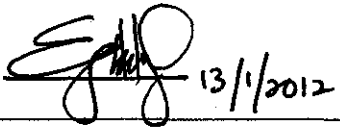
**UNIVERSITI TEKNOLOGI PETRONAS**

**TRONOH, PERAK**

**Sept 2011**

## CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

A handwritten signature in black ink, followed by the date '13/1/2012'. The signature is written over a horizontal line.

Siti Nur Azimah bt Mohamed Amin

## **ABSTRACT**

This paper is a final report for the final year project titled Real Time Abnormal Sound Detection and Classification for Home Environment. This project utilizes signal processing and signal analyzing techniques to classify any abnormal sounds in any house environment. The project will be applied specifically for surveillance system in the house so that the system could recognise and differentiate the abnormal sounds in house environment. There are various algorithms available to identify and classify the abnormal event sound. In this project, a system is designed to have abnormal sound detection and classification. For abnormal sound detection, mean signal approach being used while for classification process, there two main methods being used which are features extraction using Mel-Frequency Cepstral Coefficient (MFCC) and classifier using Gaussian Mixture Model (GMM). Mel scale frequencies in MFCC are distributed linearly in the low range but logarithmically in the high range. These characteristic corresponds to the physiological of the human ear. Therefore, MFCC has assurance of high accuracy in output. Gaussian Mixture Model (GMM) technique is suitable for usage in Matlab software. Signals from sensor will be detected and MFCC will extract the features which meet the requirement decided. The valuable features will be feed in to the classifier GMM. For this project, a microphone will be used as audio sensor, while all programming codes will be tested using Matlab software. As a result, the project would be able to classify the detected any abnormal event sound. The user would be able to employ the system as a supervision and precaution for any unexpected situation.

## **ACKNOWLEDGEMENT**

This project would not have been successful without the assistance and guidance of certain individuals and organization. First of all and most importantly, I would like to express my sincere thanks and utmost appreciation to my helpful project supervisor, Puan Zazilah bt May for her valuable input and guidance throughout the course of this project. I would also like to express my gratitude to Electrical and Electronic Engineering Department of Universiti Teknologi PETRONAS (UTP) for providing the chance for me to undertake this remarkable final year project. My knowledge had been put to a test after completing four years of electrical and electronic engineering course. A word of sincere gratitude to the Final Year Project Coordinator for arranging various seminars to assist the students. The seminars were indeed very helpful and insightful. Special thanks to all lecturers and technicians from Universiti Teknologi PETRONAS who had provided untiring guidance throughout the period of this project.

## Table of content

<b>Abstract</b> .....	4
<b>Chapter 1: Introduction</b>	
1.1 Background Study.....	9
1.2 Problem Statement.....	10
1.3 Objectives.....	11
1.4 Scope of Study.....	11
<b>Chapter 2: Literature Review</b>	
2.1 Comparison of Methods Available.....	12
2.1.1 Included System Structure.....	12
2.1.2 System Structure Not Included.....	14
2.2 Overview of Selected Methods.....	17
<b>Chapter 3: Methodology</b>	
3.1 Flowchart.....	18
3.2 Research.....	19
3.3 Project Overview.....	19
3.4 Data Acquisition.....	20
3.5 Audio Event Models.....	21
3.6 Audio Event Detection.....	21
3.7 Audio Event Classification.....	21
3.7.1 Features Extraction.....	22
3.7.2 Classification.....	22
3.8 Gantt Chart.....	24
3.9 Key Milestones.....	24
<b>Chapter 4: Result and Findings</b>	
4.1 Factor Analysis.....	26
4.2 Equipment Analysis.....	26
4.3 Abnormal Event Data Collection.....	27
4.4 Audio Event Detection.....	27
4.5 Features Extraction.....	29
4.6 Statistical Modelling.....	34
4.7 Testing.....	36
4.8 Challenges and Responses.....	38

<b>Chapter 5: Conclusion and Recommendation</b> .....	40
5.1 Conclusion.....	40
5.2 Recommendations.....	40
References.....	41
Appendices.....	44

## List of Figures

Figure 1	:	System Structure.....	13
Figure 2	:	MFCC.....	17
Figure 3	:	Project Flow Chart.....	19
Figure 4	:	System Flow Chart.....	19
Figure 5	:	Audio Event Detection System.....	21
Figure 6	:	Project Setup.....	22
Figure 7	:	Infant Cry Audio Signal.....	30
Figure 8	:	Infant Cry MFCC Spectrogram.....	30
Figure 9	:	Gunshot Audio Signal.....	31
Figure 10	:	Gunshot MFCC Spectrogram.....	31
Figure 11	:	Scream Audio Signal.....	32
Figure 12	:	Scream MFCC Spectrogram.....	32
Figure 13	:	Fire Alarm System Audio Signal.....	33
Figure 14	:	Fire Alarm System MFCC Spectrogram.....	33

## List of Tables

Table 1	:	Gantt Chart.....	25
Table 2	:	Key Milestones.....	25
Table 3	:	Factor Analysis.....	26
Table 4	:	Abnormal Event Sound.....	27
Table 5	:	Mean Comparison.....	29
Table 6	:	Real Time Audio Acquisition Result.....	29
Table 7	:	Number of Gaussian.....	34
Table 8	:	Euclidean Distance.....	34
Table 9	:	Standardized Euclidean Distance.....	34
Table 10	:	Minkowski Distance.....	34
Table 11	:	Chebychev Distance.....	36
Table 12	:	Mahalanobis Distance.....	36



# **Chapter 1**

## **Introduction**

### **1.1 Background Study**

The area of automatic surveillance systems is mostly focused on detecting abnormal events based on the acquired video and audio information. Current system in market typically involves cameras and microphones distributed in an area and connected to a controller. While the implementations give information needed, for this project, we concentrate on detecting irregular events by working only on the audio signal. Audio features are elements in a sound that allow experimentally recorded, observed and reproduced. Audio approach has low computational needs, low cost for system structure and the installation is simple. However, sound based system has some limitation because sound signals are dynamic and unpredictable.

#### **1.1.1 Abnormal event analysis**

Before a project being developed, it is a necessity to gain as much information as possible about the situation involve. For this project, the abnormal event detection system is specifically design for home environment; therefore an analysis being done for any type of abnormal events happened inside a home. There are several severe cases that involve household accident or abusive action in Malaysia. All the cases are avoidable if proper supervision or certain precautions taken in the house itself.

##### **1.1.1.1 Armed Robbery February 2008**

In Kota Tinggi, Johor [19], five masked robbers attacked oil palm worker's house when the family was asleep at night. The robbers however succeed to take cash, watches and mobile phone before they escaped. Father of the family fought one of the robbers and died before he could get to hospital. The robbers were armed with rifles. Until now, the case is still unsettled. The case would be able to solve if a system that could alert the police on the time the robbery happened. The gunshot sound could be detected as an abnormal event and a signal would be send to the police immediately.

### **1.1.1.2 Toddler Dies After Abused August 2008**

Kuala Lumpur [19], a toddler believed to have been a victim of child abuse died without waking up from coma. The girl was under supervision of her mother's friend while her mother went for work. The child's mother brought her to hospital when she saw her daughter unconscious. Sources revealed that the girl had suffered bleeding in the head probably due to a violent shake. This case can be avoided if the surveillance system was installed in the house. When the toddler was violently shake, the cry which can be classified as abnormal event would be detected by the system and alert the mom right away. The mom could go home immediately and stop the abuse or get help for the girl before the girl unconscious.

### **1.1.1.3 House on Fire April 2010**

In Tumpat, Kelantan [19], a house was caught in fire when a grandmother and a son were left at home while the mother went for work. Both of them are safe but the house was totally burnt. According to source, the fire was initiated by the son himself who played with a lighter inside the house while the grandmother was nearby the house. When the fire being seen by the grandmother, the system could detect any shout or scream from her as an abnormal event sound. The system could send a message to the mother at work and to fire station nearby promptly. The house or at least some of important items from the house could be save.

## **1.2 Problem Statement**

Both combined video and audio surveillance system is an establishment within the production industry proved by numerous companies that commercialize the system. The system predominately being utilized by business premises for security and supervision which involve a large area. Therefore, the system would not be efficient when it is being use for home environment. Most of the security department hired an officer to watch over the surveillance output from a monitor. The recorded data from the surveillance system will be used as one of the evidence by the authority when a crime had been done. In another word, the surveillance system is just a recorder device to capture visual or sound for an area.

### **1.3 Objectives**

This project has a few objectives that need to be achieved. The objectives are:

- a) To develop a set of algorithms needed for abnormal sound detection and classification.
- b) To build a system that could detect and classify any abnormal sound as a monitoring of unexpected event.

### **1.4 Scope of Project**

The project is utilizing signal processing and signal analyzing techniques that can be retrieved from previous courses completed in earlier years of study such as Digital Signal System, Digital System Design and Signal & System. The project is mainly consists of data processing and computer programming that involves Matlab software; therefore Matlab basic acquirement would be a necessity. Before any step taken for the project, basic information about audio sensor or microphone available in market also should be encountered. Fundamental process would be required by collecting reading materials related to the project. In order to have the best and most suitable methods for the project, step by step analysis process must be done utilizing all the materials gathered. The main subject for this project is the sound signals. Any regular everyday sounds need to be recorded as a normal event set along with abnormal sounds also. Sound detected by sensor has special characteristics which provide helpful information that can be extract and compare with the abnormal event set. The compared output will then classify the sound as abnormal or normal event. All the required equipments and materials are available in laboratory and within the reachable region such as internet and library; therefore the project is feasible for less than a year time allocation.

## **Chapter 2**

### **Literature review**

#### **2.1 Comparison of Methods Available**

There are two sections of review that being analyzed. One section consist a number of journals that mentioned the system structured or specific equipments used for their project. Meanwhile, another section did not provide any specific structure or equipments used. The section only emphasizes the description or comparison of methods used.

##### **2.1.1 Included system structure**

G. Várallyay Jr., Z. Benyó1, A. Illényi, Z. Farkas3, L. Kovács [9] produce a paper for acoustic analysis of the infant cry using classical and new methods. The cry signal characteristics being process in time and frequency domain. From this paper, authors implemented the role of hearing organ to the system so that the output can be as accurate as possible. Methods used are signal acquisition, Fundamental Frequency Detection using Smoothed Spectrum Method (SMM), dominant frequency detection with max value detection in the whole spectrum and visualisation using Classical Five Line method. The input signal acquisition required a microphone and recorder with 1 meter distance from the subject. SSM proves to be the most reliable method for fundamental frequency detecting in this paper.

Eric Castelli, Michel Vacher, Dan Istrate, Laurent Besacier and Jean-François Sérignat [8] present a habitat telemonitoring system based on the sound surveillance. The system specifically applied to elderly, convalescent persons or pregnant women. This paper emphasized that audio sensor is preferred than video monitoring by the consumers. The system presented applied the main concept of the proposed project whereby the system detected distress situation. The authors used input of 5 microphones and infrared sensors in 30m<sup>2</sup> area. Acoustical parameters used in speech recognition are MFCC, LFCC, LPC and non-classically like zero crossing rate, roll-off point, centroid. The authors choose Audio signal monitoring because it gives more privacy to consumer. Combination MFCC method with non-classically like zero crossing rate, roll-off point, centroid, lessen the false alarm and give higher accuracy to the output. Figure 1 shows the structure used by Eric Castelli, Michel Vacher, Dan Istrate, Laurent Besacier and Jean-François Sérignat [8]

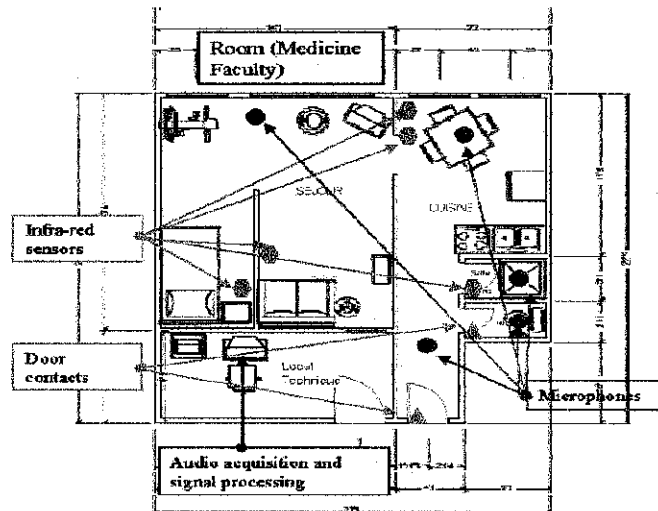


Figure 1 : System Structure

Stavros Ntalampiras, Ilyas Potamitis and Nikos Fakotakis [16] introduce a system that able to detect abnormal event that take place in noisy environment such as public place using acoustic surveillance. The authors consider typical situations like scream, gunshots and explosion. The same procedures involve such as extraction and classifier to produce low false alarm and high accuracy. This paper described that the project utilized one microphone for an area. It is stated that the authors used Mel-Frequency Cepstral Coefficients to compute the power of the short time Fourier transform for every frame and pass them through a triangular Mel scale filterbank so that signal components which play an important role are emphasized and Gaussian Mixture Models and Hidden Markov models as classifier. The main goal of this paper is to efficiently characterize the acoustic environment in terms of threatening/non-threatening conditions while using a single microphone. Acoustic surveillance system is low cost and relatively easy during setup, solution. The methods above are practical and suitable for real time operation.

### 2.1.2 System structure not included

Ing-Jr Ding [11] completed a project that capable to detect abnormal event based on audio signals by decision window using Fuzzy Logic Control. The authors brought up the point whereby the audio signals convey more useful and dominant clues or features. The project also implements the same basic flow of procedure that involve extraction and classification, but with additional decision window. For features extraction, LPC (Linear Prediction Coefficients), LPCC (Linear Prediction Cepstral Coefficients) and MFCC (Mel Frequency Cepstral Coefficients) being used. Result classification by Gaussian Mixture Model (GMM) and Decision Window (DW). The Fuzzy Logic Controller (FLC) that being used is able to stretch the window whiles the monitored environment for collecting more audio frames. FLC ensure the reliability and correctness of the detection.

C. Clavel, I. Vasilescu, L. Devillers, G. Richard, T. Ehrette [4] present an audio based surveillance system that able to recognise fear type emotion. The authors focus on emotion recognition during abnormal situation. Later in the paper mentioned that all the algorithms used being applied to audio surveillance system. By the end of this paper, the authors discovered a few challenges faced for improvising. In this paper, the authors tell that they utilized High-level features (Pitch, Intensity), Low-level features (Spectral and Cepstral) as feature extractor and Gaussian Mixture Models (GMM) as classifier. GMM has been thoroughly benchmarked in the speech community. Therefore, the expected output should more than the minimum requirement of accuracy.

The paper is a thesis of Computer Engineering by Riccardo Levorato [15] . The author mainly emphasized the classifier Gaussian Mixture Model as a method. The paper present the advantages of audio analysis process to be apply for surveillance system. The author used standard structures of system such as detection, extraction and classification to produce the output. All methods and algorithm used in this paper are design to be test on Matlab software. Based on the result and conclusion of the paper, the methods that the author chose and worked on, developed output that has accuracy more than 60%. The paper described a number of features extraction which are Teager Energy Operator (TEO), Zero Crossing Rate (ZCR), 30 Mel-Frequency Cepstral Coefficients (MFCCs), Spectral Flatness Measure (SFM), Spectral Centroid, Spectral Skewness, Spectral Slope, Spectral Decrease, Whole-Band Periodicity, Filtered-Band Periodicity, Correlation Slope and Correlation Decrease. As the classifier, Gaussian Mixture Model (GMM) is chosen. The author specifically highlighted the advantages of GMM. The classifier is simple and fast code

programming. Moreover, the classifier can work with simple audio signal access from file .wav (function wavread and wavwrite). GMM training algorithm also already implemented Statistic Toolbox in Matlab.

Apart from selected materials mentioned, there are other methods that being discovered along the analysis process. G. Tzanetakis, G. Essl and P. Cook, 2001 [10] implemented comparison of features extractor, Discrete Wavelet Transform (DWT), The short-time Fourier transform (STFT), Mel-Frequency Cepstral Coefficients (MFCC) and Gaussian Mixture Model (GMM) as classifier in the method. DWT specifically being chose because it is relatively recent and computationally efficient technique for extracting information about no stationary signals like audio. D. Neiberg, K. Elenius, K. Laskowski, 2006 [7], C. Clavel, T. Ehrette, G. Richard, 2005 [4], V. T. Vu, Q. C. Pham, J. L. Rouas, 2006 [18] and M. Vacher, D. Istrate, J. F. Serignat, 2004 [14] used Mel-Frequency Cepstral Coefficients (MFCC) for extraction and Gaussian mixture models (GMMs) for classification. The output for usage of MFCC yields 78% to 80% accuracy.

A. Rabaoui, M. Davy, S. Rossignol, and N. Ellouze, 2008 [2] chose different approach by compared One-class support vector machines (1-SVMs) with Hidden Markov Model (HMM) as classifier and compare a few features extractor like Time domain (Zero crossing rate), Frequency domain (Spectral centroid, Spectral rolloff point), Mel frequency cepstral coefficients (MFCCs) and Wavelet Transform. The best recognition accuracy (96.89%) is obtained when combining wavelet-based features, MFCCs, and individual temporal and frequency features. 1-SVM-based multiclass classification approach over performs the conventional Hidden Markov Model-based system in the experiments conducted, the improvement in the error rate can reach 50%. Cheung-Fat Chan and Eric W.M. Yu, 2010 [6] apply Weighted Average Delta Energy, LPC Spectrum Flatness, Fast Fourier Transform Spectrum Flatness, Zero Crossing Rate, Harmonicity, Mid-Level Crossing Rate, Peak and Valley Count Rate for the extraction and sliding window Hidden Markov Model (HMM) as their classifier. HMM produce only 5.5% error rate when applied trained training sounds set on human/non human sounds. T.L. Nwe, S.W. Foo and L. C. De Silva, 2003 [17] employed short time Log Frequency Power Coefficients (LFPC) to Hidden Markov Model (HMM) classifier. Results show that the proposed system yields an average accuracy of 78% and the best accuracy of 96% in the classification of six emotions. Results also revealed that LFPC is a better choice as feature parameters for emotion classification than the traditional feature parameters.

C. Doukas, L. Athanasiou, K. Fakos and I. Maglogiannis, 2008 [5] proposed Short-Time-Fourier Transformation (STFT) to a number of classifier for comparison. The classifier algorithms used are BayesNet, NaiveBayes, Logistic, MultiLayerPerception, SVM, IB1 (Nearest Neighbour), IBK (K Nearest), NNge, PART, NBTree, SimpleCart, AdaBoost, Regression, CVParameterSelection, RandomSubSpace, NestedDichotomies, Dagging and ThresholdSelector. From the 18 classification algorithms used, just over half were able to correctly identify an emergency situation with greater than 95% accuracy and 2 of the classification algorithms (NBTree and AdaBoost) were 100% accurate. M. F. McKinney and J. Breebaart, 2003 [12] practised Low level signal properties, Mel-frequency cepstral coefficients (MFCC), psychoacoustic features including roughness, loudness and sharpness, an auditory model representation of temporal envelope fluctuations as the features extractor and Gaussian-based quadratic discriminant analysis (QDA) as the classifier. The overall classification performance on general audio classes yields correctness by 93% and error  $\pm 2\%$  while with music genre classification yields 74% correctness with  $\pm 9\%$  error.

M. Cowling and R. Sitter, 2003 [13] compared a few classifiers; Learning vector quantization, Artificial neural networks, Dynamic time wrapping, Long term statistic and Gaussian Mixture Model (GMM) with two types of extractor; stationary ( Frequency extraction, Mel-frequency cepstral coefficients, Homomorphic cepstral coefficient) and non Stationary (Short time Fourier Transform, Fast wavelet transform, Continuous wavelet transform). From the reult, the highest accuracies are combination of continuous wavelet transform with dynamic time warping produces a classification rate of 70%. Combination of MFCCs with dynamic time warping also produced 70%. A. Temko and C. Nadeu, 2005 [3] achieve a 31.5% relative error reduction using Support Vector Machines (SVM) classifier for Perceptual-spectral coefficients, Cepstral-based spectral coefficients and FF-based spectral coefficients.



## 2.2 Overview of Selected Methods

### Feature Extractor

Extraction is one of the most important processes in the project. The feature vectors extracted will be used for classification step later. The chosen extractor is Mel-Frequency Cepstral Coefficient (MFCC). Mel scale is a perceptual scale of pitches whereby it is based on pitch comparisons. MFCCs can be obtained using several steps of process. In the mel frequency cepstrum MFC the frequency bands are equally spaced on the mel scale, which is similar to the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. In fact MFCC reflects the energy distribution over the basilar membrane of the cochlea. Hence, MFCCs are considered to carry a high amount of useful information related to a sound signal. Furthermore, they are commonly used to characterize a sound signal in such applications as automatic sound recognition.

Each signal is represented as a sequence of spectral vectors and speech signal is represented as a sequence of cepstral vectors. MFCC is basically spectrum that undergoes Mel-filters to become Mel-spectrum. Cepstral coefficients obtained for Mel-spectrum are called MFCC.

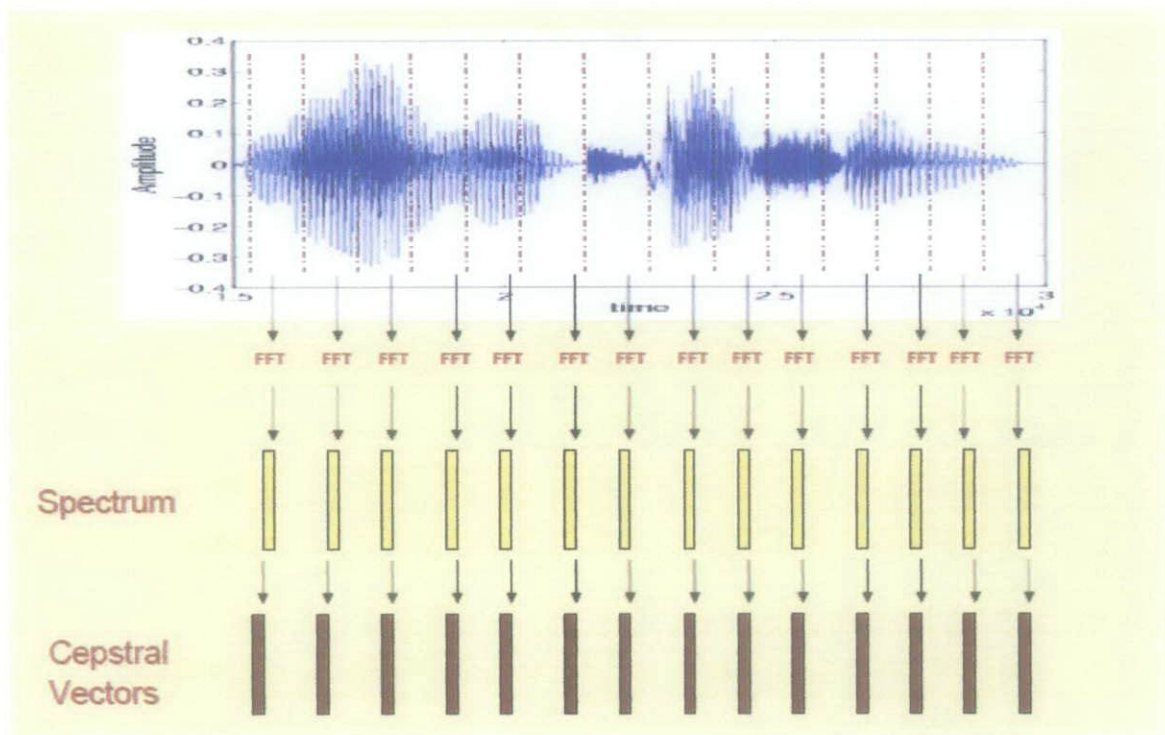


Figure 2 : MFCC

## *Classifier*

The classification in this project is based on Gaussian Mixture Models (GMM). GMM are among the most statistically mature methods for clustering. GMM consists of 3 overall steps. First, the characteristic parameters of abnormal event sound data will be calculated. Second, each GMM model with different Gaussian mixtures was trained. The GMM model was set with different Gaussian mixtures to obtain an optimal classification performance. Third, trained GMM models were used as a classifier to perform the classification tasks.

The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. GMMs are often used in biometric systems, most notably in speaker recognition systems, due to their capability of representing a large class of sample distributions. One of the powerful attributes of the GMM is its ability to form smooth approximations to arbitrarily shaped densities. The classical uni-modal Gaussian model represents feature distributions by a position (mean vector) and an elliptic shape (covariance matrix) and a vector quantize (VQ).

## Chapter 3 Methodology

### 3.1 Flowchart

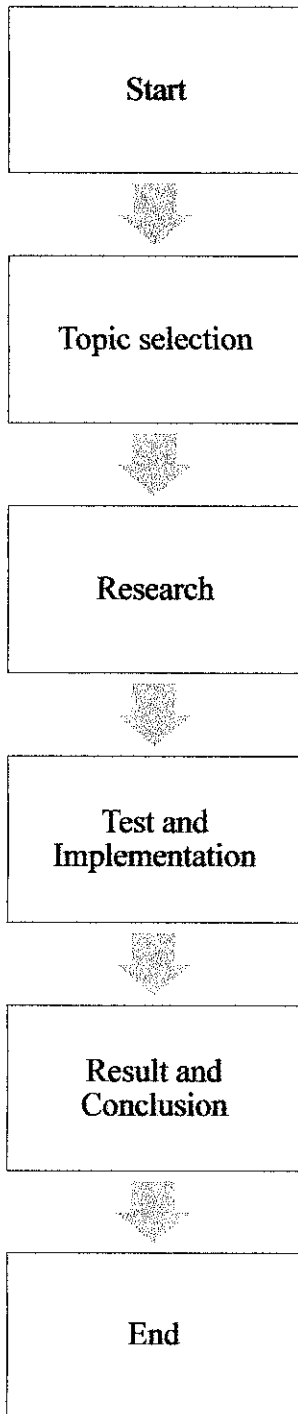


Figure 3 : Project Flow Chart

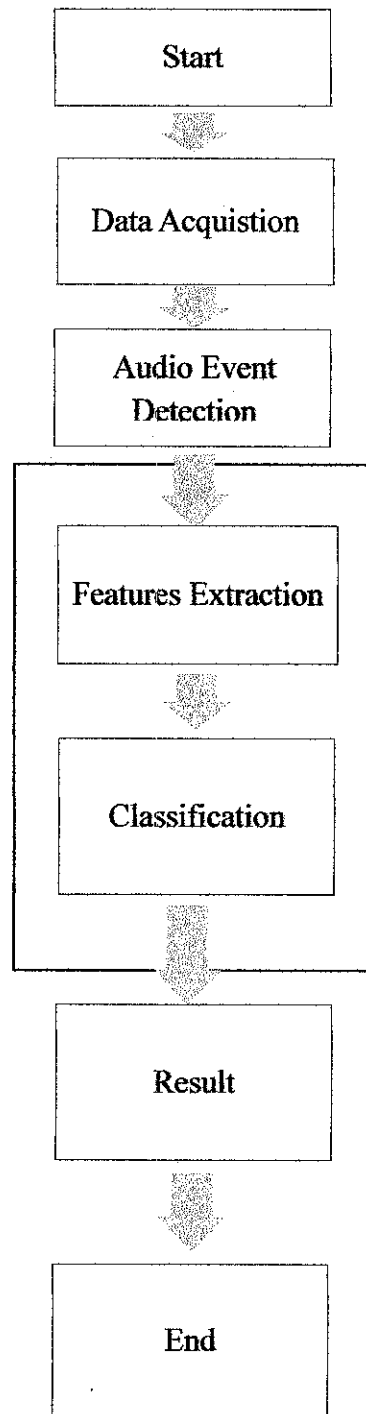


Figure 4 : System Flow Chart

### **3.2 Research**

Research for the project scope involves *quantitative* and *qualitative* research methods. Both research methods are aimed to meet the objective and requirement of the project. Keywords like surveillance system, audio analysis and abnormal detection system are being used to narrow down the searching process so that materials found are valid and reliable. Moreover, analysis and data collection about accidents involve in the house also became part of the project. Most common mediums such as internet, books, and notes from lecturer are being utilized for the research process. The research mainly conducted to gain familiarity or to achieve new insights in signal processing and signal analysis. A part from methods mentioned, the process also involved analytical research whereby facts or information being used, and analyzed to make a critical evaluation of the already available system. The method can help to improvise this project to be better. *Applied* research applied to aim at finding a solution for an immediate problem facing a society or an industrial or business organisation. This project is a design particularly for surveillance system that involves household; therefore, it is crucial for the project to be relevantly suitable for society.

### **3.3 Project Overview**

The audio event detection system consists of two phases; online training or event modelling and offline testing or event detection as shown in figure 3. The identification of the abnormal event sound is based on the algorithm chosen for extraction and classification. Both procedures are very important in order to determine whether the sound detected is cause by any abnormal event or not. Using Matlab programme, the algorithm will be implemented and tested to obtain result. Figure below shows the flow for each of processing level as well as the components used are stated and briefly discussed. The steps taken for latter approach includes audio acquisition, audio feature extraction and finally classification.

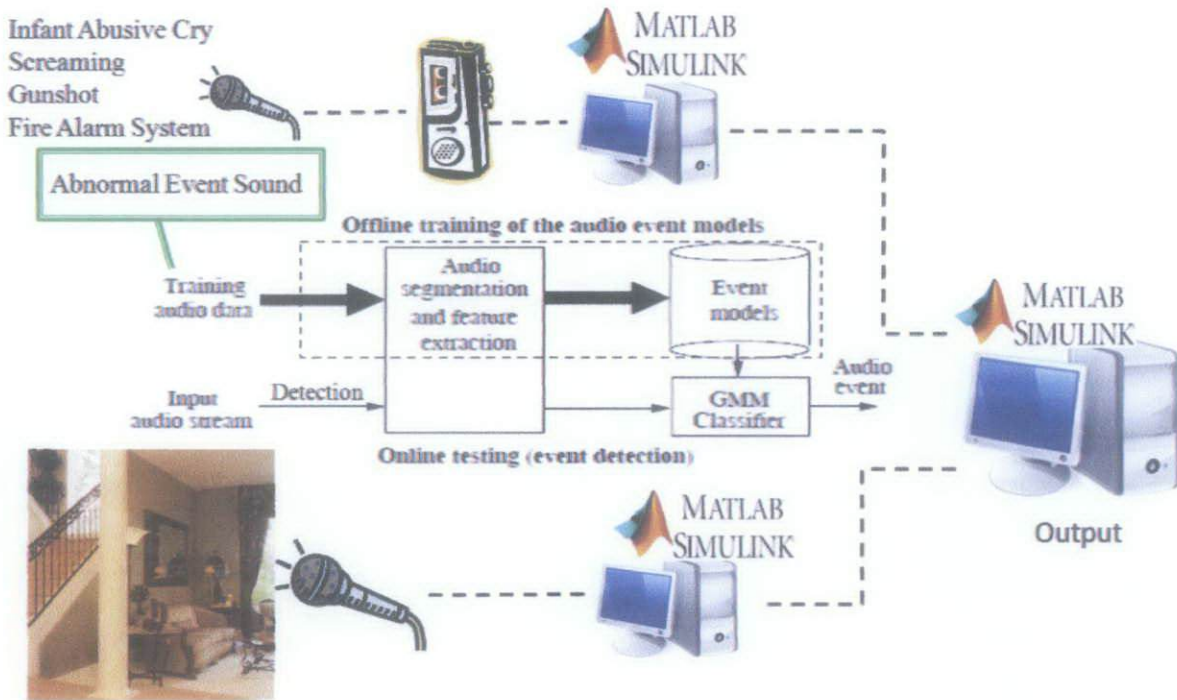


Figure 5: Audio Event Detection System

### 3.4 Data Acquisition

There are two phases of data acquisition process for this project. The first phase is training audio data where the abnormal and normal event sound for home environment will be recorded using recorder. The second stage is input audio stream for online testing where microphone will be use to capture real time event signal to be process. Simple set up would be required for the system as only a microphone is involved. The distance between the subject and the audio sensor is based on the microphone sensitivity. Approximately, for highest accuracy and quality possible, the distance apart should more or less than one metre.

All sample audio collected for abnormal model set being segmented into audio frames of 2 seconds each. The duration 2 seconds is chosen by experimentally observing the minimum length of audio frames which can capture any abnormal events. All recorded and real time audio will be assigned to waveform audio format (WAVE) or commonly use as WAV. The wav format is a Microsoft standard audio format which makes the audio bit stream easy to store on PCs.

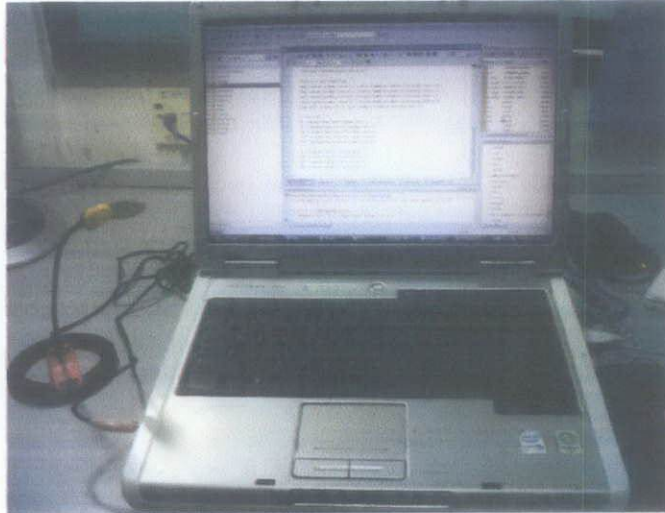


Figure 6: Project Setup

### 3.5 Audio Event Models

This stage on the project is the offline testing phase. This phase apply the recorded abnormal and normal sound event that mentioned in section 3.4. This project employs four abnormal sound events to be assigned as audio event models. The models are; abusive infant cry, screaming sound, gunshot and fire alarm system. They are obviously very synonym with home environment accidents that being analyzed in section 1.1 under background study. The audio event models are subjects to features extraction process as preparation for online testing phase.

### 3.6 Audio Event Detection

In section 3.4, I discussed on how the system acquire sound signals for both phases of the project. This section, audio event detection is the second step after real time audio acquisition. The sound captured by the microphone is updated every 2 seconds and each signal gained being detected to scan any abnormal sound. This process utilized differences of mean between abnormal and normal sound signals. A simple filter would filter out any 2 seconds sound signal that does not meet the minimum requirement set by the filter.

### 3.7 Audio Event Classification

When the sound signal being picked out from audio sound detection, the 2 seconds signal would undergo classification process. There are two phase involve in this stage which are; features extraction and classifier.

### 3.7.1 Features Extraction

As mentioned in 2.2, Mel-Frequency Cepstral Coefficient (MFCC) is being selected as extractor. A popular transformation between  $f$  Hertz and  $m$  mel as below equation:

$$m = 2595 \log_{10} \left( \frac{f}{700} + 1 \right) = 1127 \log_e \left( \frac{f}{700} + 1 \right)$$
$$f = 700(10^{m/2595} - 1) = 700(e^{m/1127} - 1)$$

The MFCCs are derived as follows:

- a) Take the Fourier transform of a frame of a signal.
- b) Map the power of the spectrum obtained onto the mel scale.
- c) Take the log of the powers at each of the mel frequencies.
- d) Take the discrete cosine transform of the list of mel log powers.
- e) The MFCCs are the amplitudes of the resulting spectrum.

### 3.7.2 Classification

Gaussian Mixture Models (GMM) is used in most of the pattern recognition projects. It is proven by analysis made with more than 30 journals and more than half of the authors chose to implement GMM as classifier. An interesting property of GMMs is that the training procedure is done independently for the classes by constructing a Gaussian mixture for each given class separately. Therefore, adding a new class to a classification problem does not require retraining the whole system and does not affect the topology of the classifier making it reliable for audio classification applications. As mentioned in 2.2, there are three main steps of GMM. There is a linear combination of multivariate Gaussian probability density function that can be express generally as:

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

$d$  is the number of features in the model;

$x$  is a  $d$ -component feature vector;

$\mu$  is the  $d$ -component vector containing the mean of each feature;

$\Sigma$  is the  $d$ -by- $d$  covariance matrix.  $|\Sigma|$  is the determinant.

All the data required will be train with the Expectation Maximization (EM) algorithm with a maximum likelihood criterion. EM is an iterative method which starts from a random distribution and alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimate for the latent variables, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step.

Next would be classification test with trained data using pairwise distance between two sets of observations which in this project is between the training audio data and testing audio data. For this stage, an extended analysis was required due to unsuccessful result from the previous approach. Pairwise methods evaluate all pairs of sequences and transform the differences into a distance. This essentially is a data reduction from a possibly many state difference to a single number. There several functions for Pairwise distance measure which are Euclidean distance, standardized Euclidean distance, city block metric, minkowski distance, Chebychev distance, Mahalanobis distance, cosine, correlation, Spearman, Hamming distance and Jaccard.



### 3.8 Gantt Chart

Activity	FYP 1				FYP 2			
	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
Early stage documentation								
Studies on automated surveillance system . (obj. 1)								
Studies on human behavioral and sound acquisition. (obj. 1)								
Studies on speech emotion recognition. (obj. 1)								
Conduct critical review on possible methods. (obj.2)								
Run selected method and analyze the result for improvement. (obj. 2)								
End stage documentation								

Table 1 : Gantt Chart

### 3.9 Key Milestones

Key Milestones	FYP 1				FYP 2			
	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
Completion of study on the information related								
Determine the method (obj. 1)								
Completion of construct the online audio monitoring system (obj.2)								

Table 2 : Key Milestones

## Chapter 4

### Result and findings

#### 4.1 Factor analysis

Based on analysis made from section 1.1, there are three different abnormal events that being analyzed. Therefore the project result is based on these three main abnormal events that occur inside the house. The cases have distinct factors that can be recognized immediately by the house owner.

Case	Factor
Armed Robbery	Gunshot
Child Abuse	Infant Cry
House on fire	Scream
House on fire	Alarm system

Table 3 : Factor Analysis

Based on above factors, there four main classes being construct so that any sound can be categorize and may fall under one of the main abnormal sound classes. The project utilized the factors as audio event models as mentioned in section 3.5.

#### 4.2 Equipment analysis

For the project, a microphone is needed to perform recording and input capture. Before a decision was made, there is a comparative analysis being done to several alternatives available in market. The comparison analysis had come to a conclusion where a microphone being selected based on a few characteristics which are very useful to increase accuracy of project output. The microphone is a compact and space saving microphone suitable for small area usage. The microphone draw maximum impedance 2.2 kilo Ohm. The frequency response range ability is sufficient for any room in home environment which being able to cover approximately area of  $5m^2$ . This microphone can also being use with a recorder or straight connection to PC. This model fit the project because the microphone with recorder is used in abnormal event data collection while microphone with computer is used for real time input acquisition. For the project, it is a necessity for the microphone to have high sensitivity because any sound input such as door opening or footstep must be captured as the

input. Noise cancelling feature is also an additional to the microphone and the project, therefore clear sound expected that could lessen the amount of tolerance due to noises.

### 4.3 Abnormal event data collection

Abnormal event sound recording is the first section of the project accomplishment. Based on section 1.1, there are four main factors; therefore the classes for classification process are based on the factors which are infant cry, gunshot, scream and fire alarm system. Several samples of abnormal event sounds being recorded for comparison with everyday event sounds. Using microphone and recorder, the sound data are successfully recorded in .WAV format for further procedures. Variety of data sounds being collected to produce high accuracy in final result.

Type of data sound	Number of file
a) Abusive infant cry	11
b) Gunshot	4
c) Scream	6
d) Fire alarm system	2

Table 4 : Abnormal Event Sound

### 4.4 Audio event detection

As mentioned in section 3.6, detection for abnormal sound is based on a simple filter which utilized the difference mean signal between abnormal and normal sound signal. Selections for cut-off mean value being done computationally. The cut-off mean value must be as accurate as possible so that result for detection shows zero false alerts. For this project, a comparison between normal sound event mean and abnormal sound mean being compute. The result yielded as below:

Normal Event Sound		Abnormal Event Sound	
Type	Mean	Type	Mean
1	0.00094296	Baby	0.00103443
2	0.00001566	Gunshot	0.00504557
3	0.00001391	Scream	0.01970305
4	0.00006607	Alarm	0.00132530

Table 5 : Mean Comparison

From above table, means from abnormal sound signal produced much higher value than normal event sound signal. However, due to real time data acquisition process, the mean difference values between two events above decrease the differences. Nevertheless, the mean for abnormal sound signal still higher than normal sound signal. Therefore, we took cut-off mean value at approximately from 0.00094 to 0.00095.

Cut-off mean value				
0.00091	0.00092	0.00093	0.00094	0.00095
normal	normal	abnormal	normal	Normal
abnormal	normal	normal	normal	Normal
normal	abnormal	normal	normal	Normal
normal	abnormal	abnormal	normal	Normal
abnormal	normal	abnormal	normal	Normal
abnormal	abnormal	normal	normal	Normal
normal	normal	normal	normal	Normal
normal	abnormal	normal	normal	Normal
normal	abnormal	normal	normal	Normal
normal	normal	normal	abnormal	Normal
abnormal	normal	abnormal	normal	Normal
abnormal	normal	normal	normal	abnormal
normal	normal	abnormal	normal	Normal
normal	abnormal	normal	normal	abnormal
abnormal	normal	normal	abnormal	Normal
normal	abnormal	normal	abnormal	Normal
abnormal	abnormal	abnormal	normal	Normal
normal	normal	abnormal	normal	Normal
normal	abnormal	normal	normal	Normal
abnormal	abnormal	normal	normal	Normal

Table 6 : Real Time Audio Acquisition Result

Table 6 is the result from real time audio acquisition taken with 20 set of testing audio signal which each consists of 2 seconds duration. All the results were taken in normal audio event condition. Above results produced 85% accuracy with 0.00094 cut-off mean value and 90% accuracy with 0.00095 cut-off mean value. Any mean value that higher than 0.00095 would yield 100% normal results however, the detection system did not response accurately for any weak abnormal event audio signal such as abusive infant cry or fire alarm system. Both abnormal event produce slightly lower values compared to other abnormal event audio signal. Therefore, any values that within 0.00094 until 0.00095 would give results that have acceptable error.

#### **4.5 Features Extraction**

For experimentation and testing purpose, the testing audio sample is one of the abnormal sound samples. Methods discussed in section 3.5.1 being transformed into codes of Matlab and run with four audio samples format .wav. Before Mel Cepstrum of each signals being produced, the signals must provide sample rate which is parameter needed for the procedures. Other parameters needed being fixed accordingly. The output of Mel Cepstrum for each signals are being represent as spectrogram with standard 12 coefficients. Spectrogram is a representation of spectral density that varies in time. Below are signals and results of features extracted from each abnormal audio signal.

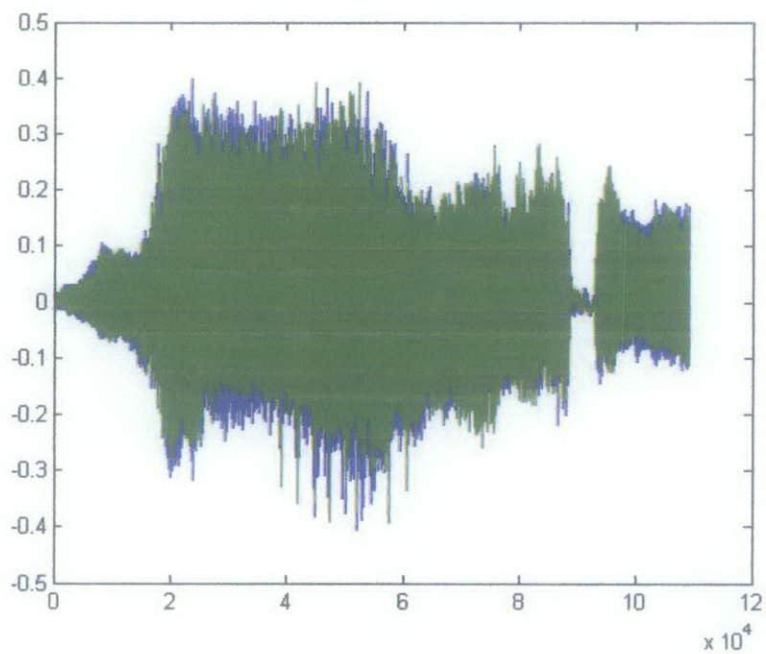


Figure 7 : Infant Cry Audio Signal

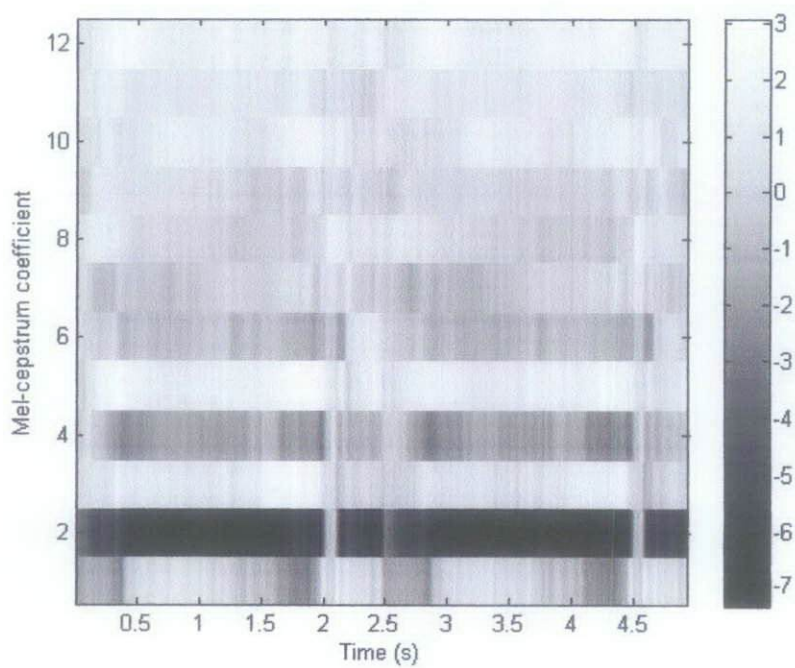


Figure 8 : Infant Cry MFCC Spectrogram

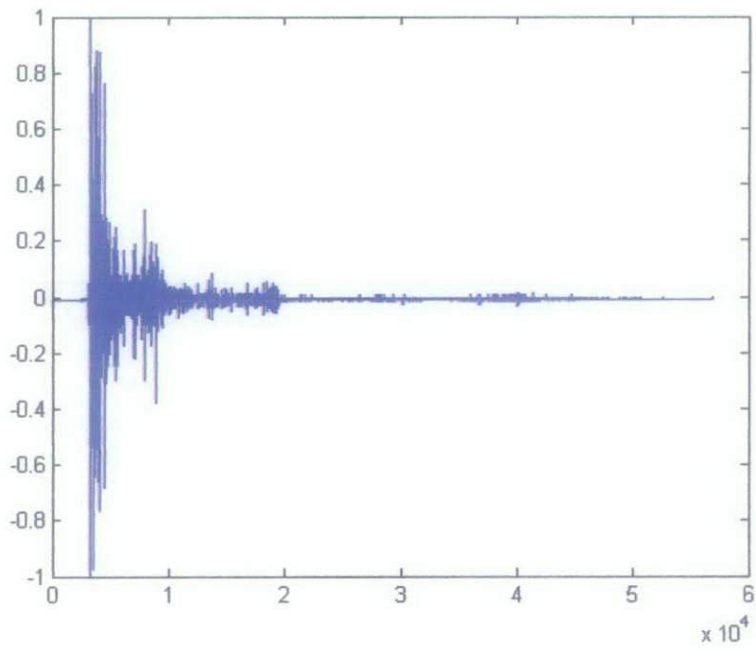


Figure 9 : Gunshot Audio Signal

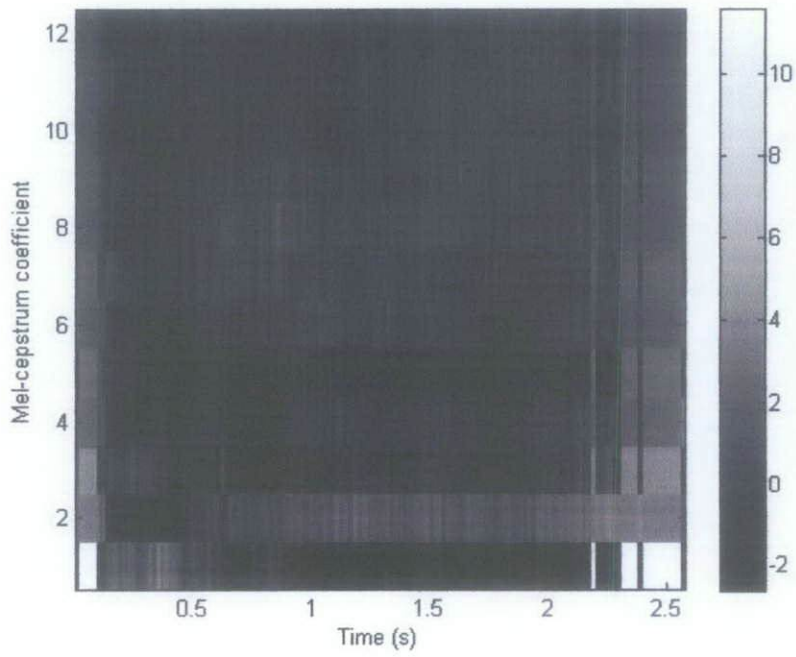


Figure 10 : Gunshot MFCC Spectrogram

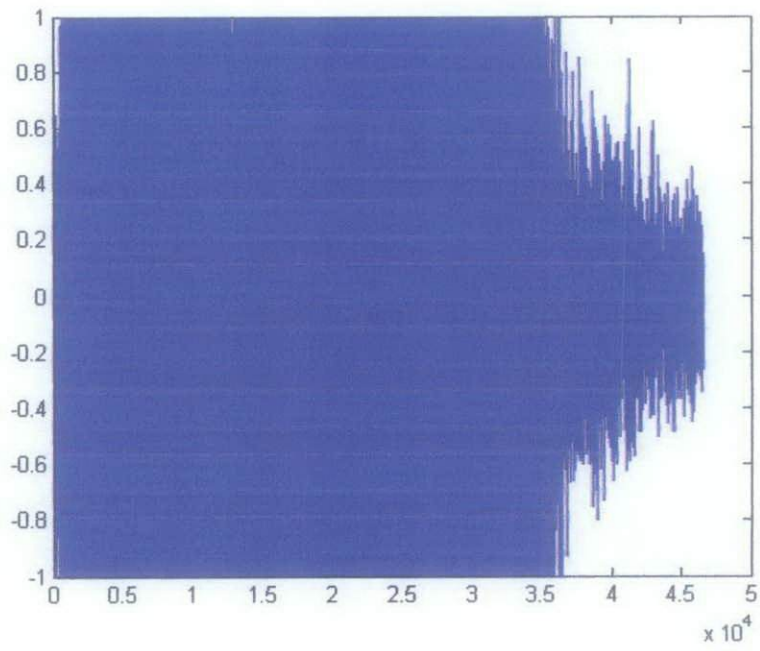


Figure 11 : Scream Audio Signal

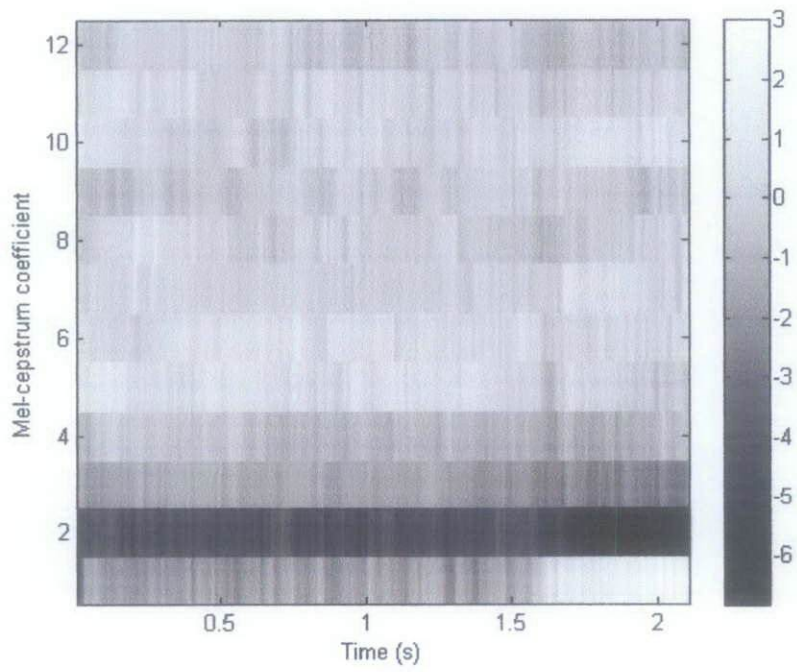


Figure 12 : Scream MFCC Spectrogram



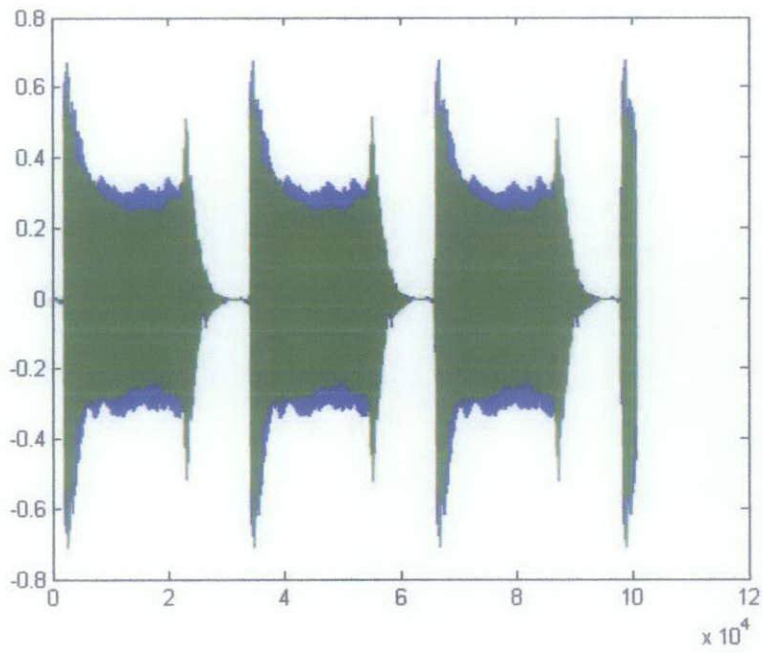


Figure 13 : Fire Alarm System Audio Signal

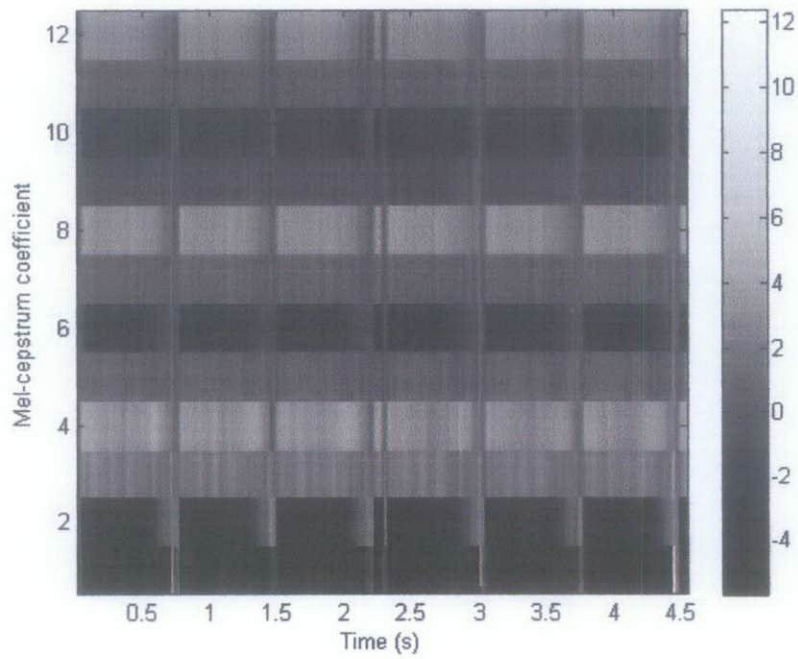


Figure 14 : Fire Alarm System MFCC Spectrogram

## 4.6 Statistical Modelling

Gaussian Mixture Model (GMM) code being used with features extracted in previous procedures as the input parameters. Besides the features from MFCC, number of Gaussian also required to produce adequate output for the next method. The outputs for GMM estimation process are means, covariance matrices and weight. Moreover, log likelihood values for each signal also appear as the output for the process. In order to determine a range of suitable number of Gaussian that acceptable and errors free, log likelihood values being use as observation point. Below are results recorded for each signal.

	Infant Cry	Gunshot	Scream	Fire Alarm System
Number of Gaussian : 1	log-likelihood : -424.230901 log-likelihood : -437.515585 converge	log-likelihood : -58.477178 log-likelihood : - 86.274708 converge	log-likelihood : -128.458143 log-likelihood : -123.694080 log-likelihood : -123.694080 converge	log-likelihood : -143.841638 log-likelihood : -434.635126 converge
Number of Gaussian : 2	log-likelihood : 603.180454 log-likelihood : - 437.515585 converge	log-likelihood : 1616.062100 log-likelihood : -86.274708 converge	log-likelihood : 416.666557 log-likelihood : 6675.985432 log-likelihood : 8543.737767 log-likelihood : 8855.567271 log-likelihood : 8556.143906 converge	log-likelihood : 1763.474644 log-likelihood : 16158.295639 log-likelihood : 20529.198593 log-likelihood : 32267.908613 log-likelihood : 32041.885375 converge
Number of Gaussian : 3	log-likelihood : 2474.456938 log-likelihood : 12177.830935	log-likelihood : 1501.200212 log-likelihood : 4932.129405	log-likelihood : 7266.360460 log-likelihood : 106060.801769	log-likelihood : 6783.317674 log-likelihood : NaN

	log-likelihood : 48801.771834	log-likelihood : 23473.058239	log-likelihood : 106060.801769	
	log-likelihood : 48901.895228	log-likelihood : NaN	converge	
	log-likelihood : 48801.771834			
	converge			
Number of Gaussian : 4	log-likelihood : 4678.302978	log-likelihood : 4212.178813	log-likelihood : 2352.906537	log-likelihood : 9245.486759
	log-likelihood : 54371.701618	log-likelihood : 130237.342302	log-likelihood : 24063.045800	log-likelihood : NaN
	log-likelihood : NaN	log-likelihood : 130032.524118	log-likelihood : 24102.915218	
		converge	log-likelihood : 23617.376016	
			converge	
Number of Gaussian : 5	log-likelihood : 10695.057339	log-likelihood : 11257.842247	log-likelihood : 20506.882824	log-likelihood : 23368.232102
	log-likelihood : NaN	log-likelihood : NaN	log-likelihood : NaN	log-likelihood : NaN

Table 7 : Number of Gaussian

From the table above, results of log likelihood produced are converging for all signals when number of Gaussian was set to 1 and 2. Meanwhile, log likelihood brings out result of Not a Number (NaN) for at least one signal number of Gaussian equals to 2. Therefore, number of Gaussian must be 1 or 2 in order to run the whole project.

## 4.7 Testing

First step of testing the detection result is convert covariance and sigma into a probability distribution using Multivariate Probability Density Function (MVNPDF). The method uses two parameters which are covariance and sigma. Both testing signal and training signal being distributed using MVNPDF. The result from MVNPDF was compared between testing signal distribution with every each of training signal distribution. The comparison method is Pairwise Distance as discussed previously in section 3.5.2. As mentioned before, the comparison method is consists of several function. All functions being test to encounter the most accurate result for the project. Below are the results of the testing process.

	Infant Cry	Gunshot	Scream	Fire Alarm		
Infant cry	0.00E+00	1.62E-06	3.86E-12	9.07E-13	0	√
Gunshot	1.77E-10	1.22E-09	3.84E-08	7.15E-11	7.15E-11	x
Scream	8.00E-10	3.84E-08	2.51E-09	1.02E-09	8E-10	x
Fire Alarm	1.65E-11	2.67E-10	1.96E-09	1.89E-47	1.89E-47	√

Table 8 : Euclidean distance

	Infant Cry	Gunshot	Scream	Fire Alarm		
Infant cry	0	6.65E-09	3.32E-08	6.84E-10	0	√
Gunshot	3.69E-09	0	5.33E-06	2.59E-19	0	√
Scream	8.91E-09	2.20E-05	3.45E-08	4.14E-07	8.91E-09	x
Fire Alarm	1.63E-09	5.59E-29	8.58E-10	4.74E-34	4.74E-34	√

Table 9 : Standardized Euclidean distance

	Infant Cry	Gunshot	Scream	Fire Alarm		
Infant cry	4.89E-13	2.50E-10	3.86E-12	1.65E-11	4.89E-13	√
Gunshot	6.83E-12	7.83E-10	3.63E-11	7.15E-11	6.83E-12	x
Scream	9.60E-10	1.95E-07	2.87E-10	1.04E-09	2.87E-10	√
Fire Alarm	1.65E-15	3.43E-19	3.65E-09	3.43E-19	3.43E-19	√

Table 10 : Minkowski distance

	Infant Cry	Gunshot	Scream	Fire Alarm		
Infant cry	9.48E-13	7.88E-14	8.82E-14	1.65E-15	1.65E-15	x
Gunshot	1.62E-10	1.14E-20	6.00E-13	1.61E-29	1.61E-29	x
Scream	1.23E-10	3.09E-07	4.75E-10	4.93E-09	1.23E-10	x
Fire Alarm	1.77E-30	2.05E-19	1.06E-09	4.23E-26	1.77E-30	x

Table 11 : Chebychev Distance

	Infant Cry	Gunshot	Scream	Fire Alarm		
Infant cry	1.72E-11	1.76E-16	4.36E-07	1.78E+17	1.76E-16	x
Gunshot	5.36E-10	2.39E-08	6.02E-06	1.01E-08	5.36E-10	x
Scream	5.70E-08	2.10E-05	0.00E+00	2.52E-06	0	√
Fire Alarm	3.45E-11	1.26E-17	1.08E-08	1.31E-17	1.26E-17	x

Table 12 : Mahalanobis Distance

Tables above are the result of comparison between testing distribution signal and training distribution signal. There several functions that gave errors and unable to compare the given distributions. Based on result above, there two functions that gave the highest accuracy. The result values are based on the minimum of distance which also gives the minimum range of value for each pair, which are Standardized Euclidean distance and Minkowski distance. Both methods yielded 75% accuracy result and they are the highest among other techniques.

## **4.8 Challenges and responses**

Throughout the process to find the most accepted result for the project, I faced a few challenges and difficulties. At the very early stage of the project, there was a problem discovered in audio event detection. Audio event detection stage is part of the real time audio acquisition process whereby all the project processes are in time loop. Initial design was involved filter that utilized frequency or amplitude of a signal; however the design was not executable in the time loop and produced error without any results. Extensive research being performed to open up other possibilities or enhancement; nevertheless extra time and knowledge required to have the success. Finally, a simplified filter method being discovered by utilized signals distinguishable mean value. The method is suitable for the project within time allocated and produces acceptable accuracy. Another problem encountered in GMM method used for the project. The last part of the method which required testing stage did not gave the desired results for the project and moreover the part went error for the input given to it. The part was then replaced by Multivariate Probability Density Function (MVNPDF) method whereby utilized both training and testing data to find match between them. The new method was discovered when another extended research being executed with keywords such as probability distance and log-likelihood analysis. Using the new method, results produced as expected however the accuracy varied. A few variations of method are being used to produce the most accurate results and finally, Standardized Euclidean distance with overall 75% accuracy being used in the project.

## **Chapter 5**

### **Conclusion and recommendations**

#### **5.1 Conclusion**

At the beginning of this report, there are two problems that initiate the construction of this project. The problems are; current system that only targeted for large and business premises and the usage of current surveillance system that majorly for evidence after crime done. The project used distinguishable mean value for real time audio detection stage and four abnormal event sound signals for classification stage to build a system that could detect and classify any abnormal event for home environment. As a conclusion, all the objectives set already accomplished. This project developed a set of algorithms needed for abnormal sound detection and classification. This project built a system that could detect and classify any abnormal sound as a monitoring of unexpected event.

#### **5.2 Recommendation**

This project could be improvised to gain better and more accurate results. Extensive study and research on detection method can produce more accurate result to the system. Instead of mean value of the signal approach, the detection stage could make use of frequency in the sound signal to be the subject of the difference between normal and abnormal. Involvement of Fourier transformation would be necessary in the recommended method. The result might gives less false alarm for the project.

## References

- [1] Andrey Temko and Climent Nadeu. “*Classification Of Meeting-Room Acoustic Events With Support Vector Machines And Variable-Feature-Set Clustering*”, ICASSP 2005, vol. V, pp. 505–508 (2005).
- [2] Andrey Temk1, Robert Malkin, Christian Zieger, Dusan Mach1, Climent Nade1, Maurizio Omologo. “*Acoustic Event Detection And Classification In Smart-Room Environments: Evaluation Of Chil Project Systems*”, Fourth Conference on Speech Technology, Saragossa, 2006.
- [3] Aki Harma, Martin F. McKinney and Janto Skowronek. “*Automatic Surveillance Of The Acoustic Activity In Our Living Environment*”, Multimedia and Expo, 2005. ICME 2005. IEEE International Conference, 2005.
- [4] Asma Rabaoui, Manuel Davy, Stéphane Rossignol, and Noureddine Ellouze. “*Using One-Class SVMs and Wavelets for Audio Surveillance*”, IEEE Transactions On Information Forensics And Security, December 2008.
- [5] Asma Rabaoui, Zied Lachiri, and Noureddine Ellouze. “*Using HMM-based Classifier Adapted to Background Noises with Improved Sounds Features for Audio Surveillance Application*”, World Academy of Science, Engineering and Technology 59, 2009.
- [6] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, T. Ehrette. “*Fear-type emotion recognition for future audio-based surveillance systems*”, Speech Communication, Elsevier, pp. 487-503, 2008.
- [7] C. Clavel, T. Ehrette and G. Richard. “*Events Detection For An Audio-Based Surveillance System*”, IEEE International Conference on Multimedia and Expo, Amsterdam, July 2005.
- [8] Charalampos Doukas, Lampros Athanasiou, Kostantinos Fakos and Ilias Maglogiannis. “*Advanced Sound and Distress Speech Expression Classification for Human Status Awareness in Assistive Environments*”, The Journal on Information Technology in Healthcare, 2008.
- [9] Cheung-Fat Chan and Eric W.M. Yu. “*An Abnormal Sound Detection And Classification System For Surveillance Applications*”, 18th European Signal Processing Conference (EUSIPCO-2010), 2010.
- [10] Cory McKay, Ichiro Fujinaga. “*Automatic Genre Classification Using Large High-Level Musical Feature Sets*”, Proc. 5th Int. Conf. Music Information Retrieval, 2004.



- [11] Daniel Neiberg, Kjell Elenius and Kornel Laskowski. “*Emotion Recognition in Spontaneous Speech Using GMMs*”, Proc. Int'l Conf. Spoken Language Processing (ICSLP '06), pp. 809-812., 2006.
- [12] Dong Zhao, Huadong Ma and Liang Liu. “*Event Classification For Living Environment Surveillance Using Audio Sensor Networks*”, Beijing IEEE/IET Electronic Library (IEL), VDE VERLAG Conference Proceedings, 2010.
- [13] Eric Castelli, Michel Vacher, Dan Istrate, Laurent Besacier and Jean-François Sérignat. “*Habitat Telemonitoring System Based On The Sound Surveillance*”, ICICHTH (International Conference on Information Communication Technologies in Health), 11-13 July 2003, Samos Island, Greece, 2003.
- [14] G. Várallyay Jr., Z. Benyó1, A. Illényi, Z. Farkas3, L. Kovács. “*Acoustic analysis of the infant cry: classical and new methods*”, Proceedings of the 26th Annual International Conference of the IEEE EMBS San Francisco, CA, USA • September 1-5, 2004.
- [15] George Tzanetakis, Georg Essl and Perry Cook. “*Audio Analysis using the Discrete Wavelet Transform*”, Proc. Conf. Acoustics and Music Theory Applications 2001.
- [16] Ing-Jr Ding. “*Events Detection For Audio Based Surveillance By Variable-Sized Decision Windows Using Fuzzy Logic Control*”, Tamkang Journal of Science and Engineering, Vol. 12, No. 3, pp. 299\_308, 2009.
- [17] Jianfeng Chen, Alvin Harvey Kam, Jianmin Zhang, Ning Liu, and Louis Shue. “*Bathroom Activity Monitoring Based on Sound*”, Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 47-61. Springer, Heidelberg 2005.
- [18] Jose Orozco Garcia, Carlos A. Reyes Garcia. “*Mel-Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry with Feed-forward Neural Networks*”, in Proceedings of the International Joint Conference 2003, pp. 3140 – 3145, 2003.
- [19] Kevin Kuo. “*Feature Extraction and Recognition of Infant Cries*”, IEEE Int. Conf. on Electro/Information Technology, Normal, Illinois, 2010.
- [20] Martin F. McKinney, Jeroen Breebaart. “*Features for Audio and Music Classification*”, Proceedings of the International Symposium on Music Information Retrieval. 151–8, 2003.

- [21] Michael Cowling and Renate Sitter. “*Comparison of techniques for environmental sound recognition*”, Pattern Recognition Letters 24 (2003) 2895–2907, 2003.
- [22] Michel Vacher, Dan Istrate and Jean-Francois Serignat. “*Sound Detection through Transient Models using Wavelet Coefficient Trees*”, Proc. CSIMTA, Cherbourg, France, 2004.
- [23] Mihail Popescu and Abhishek Mahnot. “*Acoustic Fall Detection Using One-Class Classifiers*”, 31st Annual International Conference of the IEEE EMBS, 2009.
- [24] Orion F. Reyes-Galaviz and Carlos Alberto Reyes-Garcia. “*A System for the Processing of Infant Cry to Recognize Pathologies in Recently Born Babies with Neural Networks*”, SPECOM’2004: 9th Conference Speech and Computer St. Petersburg, Russia September 20-22, 2004.
- [25] Pradeep K. Atrey, Namunu C. Maddage and Mohan S. Kankanhalli. “*Audio Based Event Detection For Multimedia Surveillance*”, in ICASSP06, 2006.
- [26] Riccardo Levorato. “*Gmm Classification Of Environmental Sounds For Surveillance Applications*”, Phd Thesis Riccardo Levorato, University Of Padova Department Of Information Engineering, Class 2009/2010.
- [27] Rolf Bardeli. “*Algorithmic Analysis of Complex Audio Scenes*”, Phd Thesis Rolf Bardeli, June 2008, <http://hss.ulb.uni-bonn.de/2008/1571/1571.pdf>, retrieved on 9<sup>th</sup> June 2011.
- [28] Sandra E. Barajas-Montiel, Carlos A. Reyes-García. “*Identifying Pain and Hunger in Infant Cry with Classifiers Ensembles*”, Proceeding Of The 2005 International Conference On Computational Intelligence For Modelling, Control And Automation, And International Conference On Intelligence Agents, Web Technologies And Internet Commences, vol 2, pp 770 – 775, 2005.
- [29] Stavros Ntalampiras, Ilyas Potamitis and Nikos Fakotakis. “*On Acoustic Surveillance Of Hazardous Situations*”, Proc. IEEE-ICASSP, 2009.
- [30] Tin Lay Nwe, Say Wei Foo and Liyanage C. De Silva. “*Speech emotion recognition using hidden Markov models*”, Speech Comm. 41 (2003), pp. 603–623, 2003.
- [31] Van-Thinh Vu, Quoc-Cuong Pham, Jean-Luc Rouas. “*Audio-Video Event Recognition System For Public Transport Security*”, Proceedings of IET Conference on Imaging for Crime Detection and Prevention, 414{419, London, UK, June 2006.

- [32] Xinyu Wu, Haitao Gong, Pei Chen, Zhi Zhong and Yangsheng Xu. “*Surveillance Robot Utilizing Video and Audio Information*”, *Journal of Intelligent and Robotic Systems*, 55(4), 403-421, 2009.
- [33] Xiaodan Zhuang, Xi Zhou, Thomas S. Huang and Mark Hasegawa-Johnson. “*Feature Analysis And Selection For Acoustic Event Detection*”, *Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp.17-20, 2008.
- [34] Yukang Guo and Mike Hazas. “*Localising Speech, Footsteps and Other Sounds using Resource-Constrained Devices*”, <http://fitlab.eu/resources/IPSN2011.pdf>, 2011, retrieved on 9th June 2011.
- [35] Yousra Abdulaziz, Sharrifah Mumtazah Syed Ahmad. “*Infant Cry Recognition System: Comparison Of System Performance Based On Mel Frequency And Linear Prediction Cepstral Coefficient*”, *International Retrieval & Knowledge Management, 2010 IEEE*, pages 260 -263, 2010.
- [36] BERNAMA, The Malaysian National News Agency, 2008 - 2010.

## Appendices

### Audio event detection program

```
%%%% User Parameters
duration = 2; % How many seconds of acquisition per plot refresh?
Fs = 44100; % Acquisition sample rate in Hz (try 8000)
% -----

%%%% Initialization & configuration of sound card
AI = analoginput('winsound');
addchannel(AI, 1);
set(AI, 'SampleRate', Fs);
set(AI, 'SamplesPerTrigger', duration*Fs);

% properly when "ctr+c" is hit
try
    count = 0; % count how many time the while was executed
    while 1
        % increment loop counter
        count = count + 1;
        % calculate elapsed time
        ET = duration * count;

        % start acquisition and retrieve data
        start(AI);
        data = getdata(AI);

        % Results
        plot(data)
        xlabel('Sample');
        ylabel('Signal (Volts)');

        if (abs(mean(data)) < 0.00095)
            disp('normal')
        else
            disp('abnormal')

            %Record
            y = wavrecord(2*Fs,Fs);
            filename = 'test.wav';
            wavwrite(y,Fs,filename);

            stop(AI);
            count = 0;
        end
    end

catch
    disp('--> coninuous loop was manually interrupted')
end

%%%% Termination
% disp('--> Deleting Analog Input Object')
% stop(AI)
% delete(AI)
```

## Audio event classification program

```
%Read files
[train1,fs1,nbits1]=wavread('baby.wav');
[nSamples1,nChannels1]=size(train1);
l1=nSamples1/fs1;

[train2,fs2,nbits2]=wavread('gunshot.WAV');
[nSamples2,nChannels2]=size(train2);
l2=nSamples2/fs2;

[train3,fs3,nbits3]=wavread('scream.wav');
[nSamples3,nChannels3]=size(train3);
l3=nSamples3/fs3;

[train4,fs4,nbits4]=wavread('alarm.wav');
[nSamples4,nChannels4]=size(train4);
l4=nSamples4/fs4;

[test,fs,nbits]=wavread('baby.wav');
[nSamples,nChannels]=size(test);
l=nSamples/fs;

%Features extraction
training_feat1=melcepst(train1,fs1);
training_feat2=melcepst(train2,fs2);
training_feat3=melcepst(train3,fs3);
training_feat4=melcepst(train4,fs4);
testing_feat=melcepst(test,fs);

%Statistical modeling
[mu_train1,sigma_train1,c_train1]=gmm_estimate(training_feat1,2);
[mu_train2,sigma_train2,c_train2]=gmm_estimate(training_feat2,2);
[mu_train3,sigma_train3,c_train3]=gmm_estimate(training_feat3,2);
[mu_train4,sigma_train4,c_train4]=gmm_estimate(training_feat4,2);
[mu_test,sigma_test,c_test]=gmm_estimate(testing_feat,2);

% Testing
X = mvnpdf(mu_test,sigma_test);
X1 = mvnpdf(mu_train1,sigma_train1);
X2 = mvnpdf(mu_train2,sigma_train2);
X3 = mvnpdf(mu_train3,sigma_train3);
X4 = mvnpdf(mu_train4,sigma_train4);

D1 = pdist2(X,X1,'seuclidean');
D2 = pdist2(X,X2,'seuclidean');
```

```

D3 = pdist2(X,X3,'seuclidean');
D4 = pdist2(X,X4,'seuclidean');

if min(min(D1))<min(min(D2))<min(min(D3))<min(min(D4))
    disp('In range of abusive infant cry sound')
else if min(min(D2))<min(min(D1))<min(min(D3))<min(min(D4))
    disp('In range of gunshot sound')
else if min(min(D3))<min(min(D1))<min(min(D2))<min(min(D4))
    disp('In range of screaming sound')
else if min(min(D4))<min(min(D1))<min(min(D2))<min(min(D3))
    disp('In range of fire alarm sound')
else
    disp ('Other abnormal sound')
end
end
end
end
end

```

## Sub program

### Features extraction

```
function c=melcepst(s,fs,w,nc,p,n,inc,fl,fh)
```

```
w='M';
```

```
nc=12;
```

```
p=floor(3*log(fs));
```

```
n=pow2(floor(log2(0.03*fs)));
```

```
fh=0.5;
```

```
fl=0;
```

```
inc=floor(n/2);
```

```
if any(w=='R')
```

```
    z=enframe(s,n,n/2);
```

```
elseif any (w=='N')
```

```
    z=enframe(s,hanning(n),n/2);
```

```
else
```

```
    z=enframe(s,hamming(n),n/2);
```

```
end
```

```
f=fft(z.);
```

```
[m,a,b]=melbankm(p,n,fs,fl,fh,w);
```

```
pw=f(a:b,:).*conj(f(a:b,:));
```

```
pth=max(pw(:))*1E-20;
```

```
if any(w=='p')
```

```
    y=log(max(m*pw,pth));
```

```
else
```

```
    ath=sqrt(pth);
```

```
    y=log(max(m*abs(f(a:b,:)),ath));
```

```
end
```

```
c=rdet(y).';
```

```
nf=size(c,1);
```

```
nc=nc+1;
```

```
if p>nc
```

```
    c(:,nc+1:end)=[];
```

```
elseif p<nc
```

```
    c=[c zeros(nf,nc-p)];
```

```
end
```

```
if ~any(w=='f')
```

```
    c(:,1)=[];
```

```
    nc=nc-1;
```

```
end
```

```
if any(w=='E')
```

```
    c=[log(sum(pw)).' c];
```

```
    nc=nc+1;
```

```

end

% calculate derivative

if any(w=='D')
    vf=(4:-1:-4)/60;
    af=(1:-1:-1)/2;
    ww=ones(5,1);
    cx=[c(ww,:); c; c(nf*ww,:)];
    vx=reshape(filter(vf,1,cx(:)),nf+10,nc);
    vx(1:8,:)=[];
    ax=reshape(filter(af,1,vx(:)),nf+2,nc);
    ax(1:2,:)=[];
    vx([1 nf+2],:)=[];
    if any(w=='d')
        c=[c vx ax];
    else
        c=[c ax];
    end
elseif any(w=='d')
    vf=(4:-1:-4)/60;
    ww=ones(4,1);
    cx=[c(ww,:); c; c(nf*ww,:)];
    vx=reshape(filter(vf,1,cx(:)),nf+8,nc);
    vx(1:8,:)=[];
    c=[c vx];
end

```

```

[nf,nc]=size(c);
t=((0:nf-1)*inc+(n-1)/2)/fs;
ci=(1:nc)-any(w=='0')-any(w=='E');
figure,imagesc(t,ci,c');
axis('xy');
xlabel('Time (s)');
ylabel('Mel-cepstrum coefficient');
map = (0:63)/63;
colormap([map map map]);
colorbar;

```



## Statistical modeling

```
function [mu,sigm,c]=gmm_estimate(X,M,iT,mu,sigm,c,Vm)
% [mu,sigma,c]=gmm_estimate(X,M,<iT,mu,sigm,c,Vm>)
%
% X : the column by column data matrix (LxT)
% M : number of gaussians
% iT : number of iterations, by default 10
% mu : initial means (LxM)
% sigm: initial diagonals for the diagonal covariance matrices (LxM)
% c : initial weights (Mx1)
% Vm : minimal variance factor, by defaut 4 ->minsig=var/(M^2Vm^2)

DEBUG=0;
GRAPH=0;

% *****
% GENERAL PARAMETERS
[L,T]=size(X); % data length
varL=var(X'); % variance for each row data;

min_diff_LLH=0.001; % convergence criteria

% DEFAULTS
if nargin<3 iT=10; end % number of iterations, by default 10
if nargin<4 mu=X(:,[fix((T-1).*rand(1,M))+1]); end % mu def: M rand vect.
if nargin<5 sigm=repmat(varL./(M.^2),[1,M]); end % sigm def: same variance
if nargin<6 c=ones(M,1)/M; end % c def: same weight
if nargin<7 Vm=4; end % minimum variance factor

min_sigm=repmat(varL./(Vm.^2*M.^2),[1,M]); % MINIMUM sigma!

if DEBUG sqrt(devs),sqrt(sigm),pause;end

% VARIABLES
lgam_m=zeros(T,M); % prob of each (X:,t) to belong to the kth mixture
lB=zeros(T,1); % log-likelihood
lBM=zeros(T,M); % log-likelihood for separate mixtures

old_LLH=-9e99; % initial log-likelihood

% START ITERATATIONS
for iter=1:iT
if GRAPH graph_gmm(X,mu,sigm,c),pause;end
if DEBUG disp(['***** ',num2str(iter),' *****']);end
```

```

% ESTIMATION STEP *****
[IBM,IB]=imultigauss(X,mu,sigm,c);
if DEBUG IB,B=exp(IB),pause; end
LLH=mean(IB);
if fprintf('log-likelihood : %f',LLH),end

lgam_m=IBM-repmat(IB,[1,M]); % logarithmic version
gam_m=exp(lgam_m); % linear version -Equation(1)

% MAXIMIZATION STEP *****
sgam_m=sum(gam_m); % sum of gam_m for all X(:,t)

% gaussian weights *****
new_c=mean(gam_m); % -Equation(4)

% means *****
% (convert gam_m and X to (L,M,T) and .* and then sum over T)
mu_numerator=sum(permute(repmat(gam_m,[1,1,L]),[3,2,1]).*...
    permute(repmat(X,[1,1,M]),[1,3,2]),3);
% convert sgam_m(L,M,N) -> (L,M,N) and then ./
new_mu=mu_numerator./repmat(sgam_m,[L,1]); % -Equation(2)

% variances *****
sig_numerator=sum(permute(repmat(gam_m,[1,1,L]),[3,2,1]).*...
    permute(repmat(X.*X,[1,1,M]),[1,3,2]),3);

new_sigm=sig_numerator./repmat(sgam_m,[L,1])-new_mu.^2; % -Equation(3)

% the variance is limited to a minimum
new_sigm=max(new_sigm,min_sigm);

% UPDATE
if old_LLH>=LLH-min_diff_LLH
    disp('converge');
    break;
else
    old_LLH=LLH;
    mu=new_mu;
    sigm=new_sigm;
    c=new_c;
end
end

```