

# Infant Cry Detection System with Automatic Soothing and Video Monitoring Functions

G.V.I.S. Silva, D.S. Wickramasinghe\*

Department of Electrical & Computer Engineering, The Open University of Sri Lanka,  
Nawala, Nugegoda, Sri Lanka

\*Corresponding Author: email: dswic@ou.ac.lk, Tele: +94718568047

---

**Abstract** - The aim of this research is to develop a portable, efficient and cost effective automatic infant's cry detector and self-soother with real time monitoring system for employed parents.

The cry detection algorithm has developed according to the crying signals and it is segmented using the short time energy function which is used as a voice activity detector to disable the operation of the algorithm when voice activity is not present. The features are extracted using MFCC (Mel Frequency Cepstrum Coefficients) and pitch frequency. Statistical properties are calculated for the extracted features of MFCC and pitch frequency. K-NN (K-Nearest Neighbour) algorithm classifier is used to classify the cry signal. The system can easily identify the infant cry and it is verified using K-NN with accurate results by proposed detection algorithm.

The combination of Pitch and MFCC gives more promising approach to cry detection than using only MFCC. The total average accuracy of MATLAB simulation is 80.8335% and on the device accuracy was 77.5% for cry detection.

Immediate cry detection and self-soothing system helps to increase baby's cognitive development process. This all in one module approach gives great benefits to the first-time parents, adoptive parents, caretakers, researchers or physicians by both economically and scientifically.

**Key words** – Cry detection, MFC, K-NN, Pitch detection, Soothing system

---

## 1 INTRODUCTION

Infant cry is the first verbal communication of new born baby with the world. The crying of the infant is a common phenomenon and probably one of the most difficult problems which parents have to face when taking care of a baby.

The cry of an infant is a biological siren to alert for the care giving environment about their needs to motivate the listener to respond. Most of the times, caretaker's advocates' follow strict routines to train the child for regular feeding, waking and sleeping pattern without considering their emotional and physical needs. Researchers have found babies whose cries are usually ignored will not develop healthy intellectual and social skills [15]. On the other hand, leaving a distressed baby to cry on a regular basis could damage the brain development.

Currently there are many types of baby monitoring systems with wearable option, android applications, wirelessly controlled camera systems etc. Most of these systems are

covered only home using Wi-Fi or Bluetooth. Due to this condition, employed parents (especially mothers) cannot ensure the safety of their babies because, they are unable to connect with the child when they're at working places. There are few products which have remote monitoring facilities. However, those are priced high-end products which are not affordable in developing countries like Sri Lanka. Further, these products are not easy to set up and by fixing near to the baby may cause health hazards due to electromagnetic radiation.

This product is designed for an affordable cost and it can be used from birth to 12 months of babies with ability to detect the infant's crying immediately and send notification to warn parents/caregiver while a soft sound and lights playing to sooth the baby. On the other hand, if the parent has any doubt about babies, they can connect to the home wireless network and check the baby using a mobile phone in real time at any time while ensuring the safety of the baby. The other benefit of this product is, it is important for hearing impaired parents because the parent can get notifications through android application which is password protected, and hearing-impaired parent can configure their phone into the vibration mode. Parents can watch live video stream from the baby's room when the notification received.

## **2 OBJECTIVES**

This research work was carried out to achieve the following objectives,

- (a) Study feature extraction methods in audio processing and develop a cry detection algorithm in different approaches of feature extraction and classification methods.
- (b) Design and implement infant's cry detection device, by applying the developed cry detection algorithm with self-soothing and video monitoring functions.

## **3 LITERATURE REVIEW**

### **3.1 Theoretical Background**

#### **Automatic Speech Recognition systems (ASR)**

Automatic Speech Recognition (ASR) is the process of converting a speech signal to a sequence of words, by means of an algorithm. Five modules can be identified to develop an ASR. [1]

- i. Speech Signal acquisition.
- ii. Feature Extraction.
- iii. Acoustic Modelling.
- iv. Language & Lexical Modelling.
- v. Recognition.

Feature extraction requires more attention in speech recognition because recognition performance depends heavily on this phase. The main goal of the feature extraction step is to compute a parsimonious sequence of feature vectors providing a compact

representation of the given input signal. The feature extraction is usually performed in three stages. The first stage is called the speech analysis or the acoustic front end. It performs Spectro temporal analysis of the signal and generates raw features describing the envelope of the power spectrum of short speech intervals. The second stage compiles an extended feature vector composed of static and dynamic features. Finally, the last stage transforms these extended feature vectors into more compact and robust vectors that are then supplied to the recognizer. [1]



There are various techniques used for feature extraction. Cepstral Analysis, Mel Cepstrum Analysis, Mel-Frequency Cepstrum Coefficients (MFCC), Linear Discriminant Analysis (LDA), Fusion MFCC, Linear Predictive Coding (LPC) Analysis and Perceptually Based Linear Predictive Analysis (PLP) are some of the techniques being used.



After feature extraction, the most important step is speech recognition. Basically, there are three approaches of speech recognition [1]. Those are Acoustic Phonetic Approach, Pattern Recognition Approach and Artificial Intelligence Approach.

### 3.2 Literature survey of similar products

Comparison of similar products is shown in table 1.

**Table 1: Comparison of similar products in the market for baby care with cry detection**

Product name	Methodology	Price	Drawbacks
<b>Why Cry - Baby Cry Analyzer Monitor</b> [7] 	<p>This sound sensitive device is programmed to recognize different pitches and then digitally analyses and transmit the baby's cry into one of five simple expressions - hungry, bored, annoyed, sleepy or stressed.</p>	<p>US\$ 52.34 (LKR ~7800)</p>	<ol style="list-style-type: none"> <li>1. No real time video monitoring option</li> <li>2. No self-soothing system included</li> <li>3. The alarm is sound based, need to be in the range to hear the alarm</li> </ol>
<b>Wireless Baby Crying Detector with Parental Alarm</b> [8] 	<p>This crying detector consisting with transmitter and receiver.</p> <p>Once the baby cry or other noise occurs, the receiver starts to make electronic crying noise.</p> <p>It's a low cost and compact design.</p>	<p>US\$ 10.70 (LKR ~1600)</p>	<ol style="list-style-type: none"> <li>1. Low sensitivity of cry detection</li> <li>2. Alarm is sound based and not useful for hearing impaired</li> <li>3. No self-soothing system included</li> <li>4. Limited Range (in between 50-60 m)</li> </ol>

<p><b>Cry translator</b></p> 	<p>A simple and lightweight. It is designed to be handled with one hand and attend to baby with the other. Notify the reason for crying. Can soothe baby with the sound of the beating of a heart or lullabies.</p>	<p>79.90 € (Euro) (LKR ~12800)</p>	<ol style="list-style-type: none"> <li>1. No alarming system to care giver when baby needs attention</li> <li>2. Expensive</li> <li>3. Limited range /No real time video monitoring option.</li> </ol>
<p><b>Wi-Fi, baby 2.0</b></p> 	<p>This monitors video and hear audio from the baby's room via an iPad, iPhone, PC, Mac, Android phone, or Android tablet. The video and audio streams live from a single Wi-Fi Baby unit set up in the child's bedroom.</p>	<p>US\$ 230 (LKR ~34500)</p>	<ol style="list-style-type: none"> <li>1. Do not detect the crying of a baby</li> <li>2. Always need to watch the baby video and pay attention to the audio.</li> <li>3. Only covers home wireless network</li> </ol>

#### 4 METHODOLOGY

The block diagram of the proposed system is shown in Figure 1. This system consists of 2 units; child unit and parental unit.

The child unit consists with a data processing unit, which interfaces with camera, speaker, microphone and self-soothing system with wireless connectivity (Wi-Fi). For the data processing, single board computer has been used. Parental unit is a smart phone, laptop or tablet with internet connectivity. Child unit connected with home internet connections via Wi-Fi or Ethernet. Parental unit and child unit are connected via internet. A notification will be given to the parent if baby is crying continuously. Then the parent can operate the system remotely according to his or her wish. It can be connected to the internet and can check his/her child online with a mobile phone in real time.

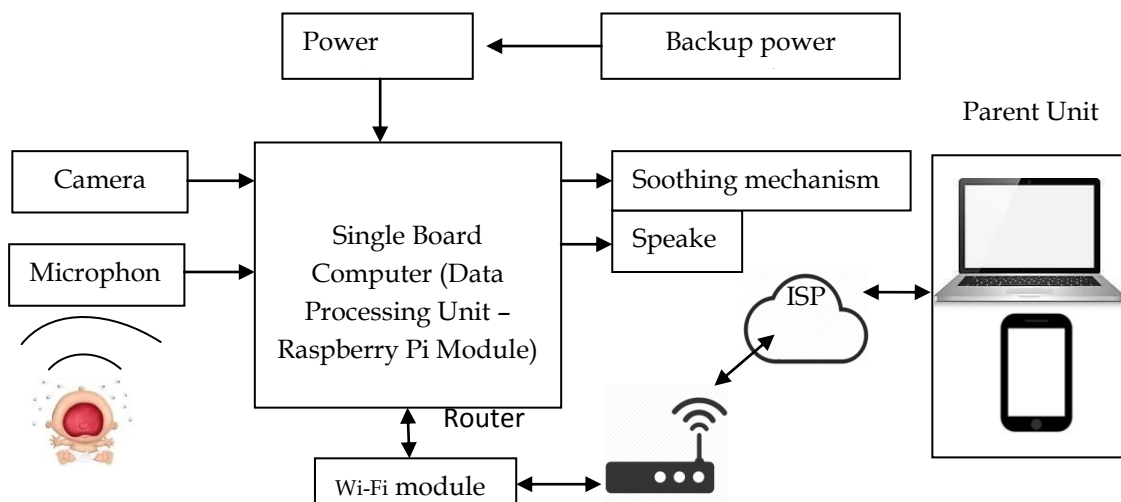


Figure 1: Block Diagram of the proposed infant cry detection and soothing system

### **Automatic cry detection system**

When the baby starts crying, at a predetermined time a microphone picks up a cry from the baby as an audio signal. At a certain sampling frequency, an A/D converter samples the audio signal received by the microphone and convert it. An audio Analyzer analyze the audio signal samples by the A/D converter and computes a characteristic quantity based on a frequency spectrum (Feature extraction).

After signal processing, data processing unit classifies the pattern from a recognition model by training it with infant cry samples (reference) and give the decision logic whether it is crying or not.

### **Self-Soothing system**

When the cry detection algorithm detects the sound input as cry at a predetermined time, a signal given to the two DC motors and 1<sup>st</sup>rotating lamp will be rotated continuously while the 2<sup>nd</sup> rotating lamp gives glow effect. Furthermore, the glowing effect can use the black perspective sheets with beautiful carvings which includes base to project different shapes on ceilings or walls, also according to different colours, these sheets can be changed manually, which will be very user friendly, and gives more realistic output to the baby at his developmental stage other than being bored by seeing the same light effect. The projection shapes and colours have been selected according to research papers [14] which are based on cognitive development of child at infancy period. In raspberry PI module (Single Board Computer), easily can operate the self-soother by using GPIO (General Purpose Input Output).

LED illusion mirror effect can be used for more glow effect and it affect to the babies' visual systems development, and at just 3 months of age, they have appeared perceive colours in a way that is analogous to adults. And simultaneously with the above process plays his favorite lullabies, or white noises soothe him, which are stored in raspberry PI as a sound library, after this sequence if baby continuously crying then notification will be sent to the parent.

From the proposed product, parents get instant notification when a cry detected. They can control (Playing music, call on their own voice) the soothing mechanism. Also, from live video streaming feature, parents can view their child real time. Further, crying sample can be saved for later statistical analysis.

The design of the system consists with two main subsystems.

- Cry detection algorithm
- Soothing mechanism and video streaming function

Overall integration of various modules also part of the design. Figure 2 shows the flow chart of the overall system.

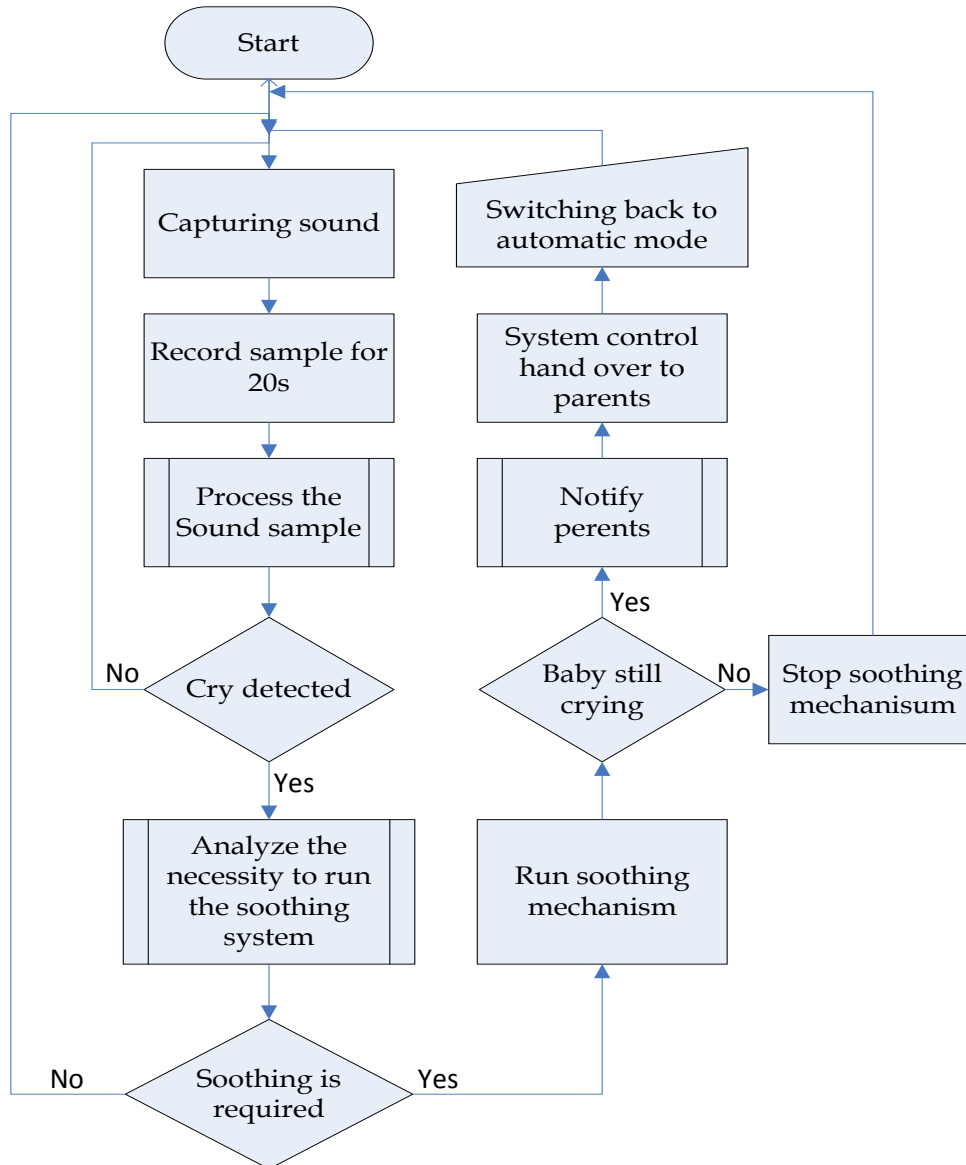


Figure 2: Top level flow chart of the system

#### 4.1 The Cry Detection Algorithm

The proposed algorithm is composed of three main stages:

- i) *Voice Activity Detector (VAD)* for detecting sections with sufficient audio activity. Short time energy is used as a VAD.

##### *Short-time energy:*

Where  $E_n$  is the energy of the sample 'n' of the audio signal,  $x[m]$  is the discrete time signal and  $w[m]$  is a rectangle window of size 'N'. The rectangle window function is defined by the expression below:  $w[n] = 1, 0 \leq n \leq N-1$ ; else  $w[n] = 0$

$$E_n = \frac{1}{2N} \sum (x[m].w[n-m].w[n+m])^2$$

## ii) Feature to be extracted

As mentioned in the section 3.1, there are various methods to extract the feature of the audio signal. In the proposed algorithm, two features have extracted from the cry audio signal.

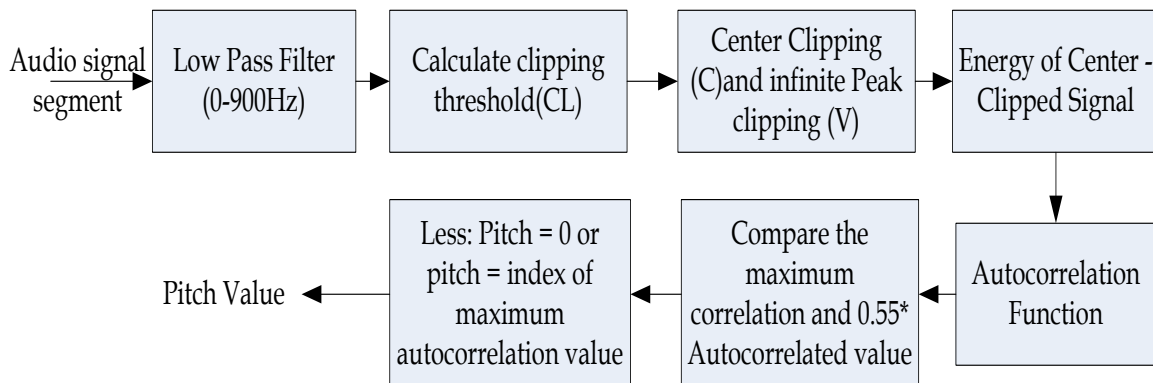
**Pitch:** It's a fundamental frequency of a periodic wave form. In humans, pitch is determined by the frequency of the vibration of vocal chords.

- Adult males average at 120 Hz, [85,155]
- Adult females average at 210Hz, [165, 255]
- Infants average at 450hz, [250, 700]

*Modified autocorrelation Method* used for detecting the fundamental frequency of the infant cry based on the center clipping method which gives more accurate results.

### Steps of Modified Auto Correlation Method for Pitch Frequency extraction <sup>[6]</sup>

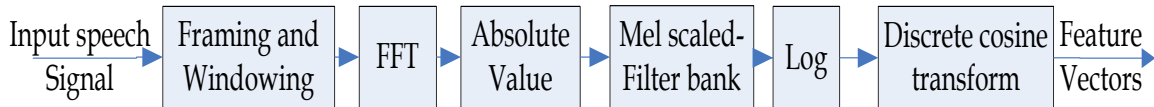
Figure 3 shows the block diagram of the pitch detection algorithm. The segmented audio signal is first required to be low pass filtered to 900Hz. Then the signal is digitized at a 10 kHz sampling rate and sectioned into overlapping 30ms (300 samples) sections for processing. Since the pitch period computation for all pitch detectors are performed 100 times/s, every 10ms, adjacent sections are overlapped by 20ms or 200 samples. The first stage of processing is the computation of a clipping threshold  $CL$  for the current 30-ms section of speech. The clipping level is set at a value which is 68% of the smaller of the peak absolute sample values in the first and last 10-ms portions of the section. Following the determination of the clipping level, the 30-ms section of the speech is center clipped, and then infinite peak clipped. Thereafter, clipping the autocorrelation function for the 30-ms section is computed over a range of lags from 20 samples to 160 samples (2-ms-20-ms period). Additionally, the autocorrelation at 0 delays is computed for voice/unvoiced determination. The autocorrelation function is then searched for its maximum value. If the maximum exceeds 0.55 of the autocorrelation values at 0 delays, the section is classified as voiced and the location of the maximum is the pitch period. Otherwise, the section is classified as unvoiced.



**Figure 3: Block diagram of pitch detection algorithm using the modified autocorrelation method**

## Mel-Frequency Cepstrum Coefficients (MFCC)

MFCC provides a representation of the short-term power spectrum of a signal. These coefficients are obtained by multiplying the short-time Fourier Transform (STFT) of each analysis frame by a series of  $M$  triangularly-shaped ideal band-pass filters, with their central frequencies and widths arranged according to a Mel -frequency scale. The total spectral energy  $E[i]$  contained in each filter is computed and a Discrete Cosine Transform (DCT) is performed to obtain the MFCC sequence.



**Figure 3: Steps to calculate MFC coefficients**

**Framing and Windowing:** -First split the signal up into several frames such that, analysing each frame in the short time instead of analysing the entire signal at once. At the range (10-30) ms, most part of speech signal is stationary. It is necessary to work with short term or frames of the signal. Windowing is performed to avoid unnatural discontinuities in the crying segment and distortion in the underlying spectrum. The choice of the windowing is a tradeoff between several factors. In speaker recognition, the most commonly used window shape is the hamming window <sup>[3]</sup>.The hamming window  $W_H(n)$ , defined as,<sup>[6]</sup>

$$W_H(n) = 0.54 - 0.46 \cos\left(\frac{2n\pi}{N-1}\right)$$

The hamming windows is used since, MFCC will be used which involves in the frequency domain. (Hamming windows will decrease the possibility of high frequency components in each frame due to such abrupt slicing of the signal.)

**Fast Fourier Transform (FFT):** To convert the signal from time domain to frequency domain preparing for the next stage (Mel frequency wrapping).Spectral analysis shows that cry signals have different timbres in speech signals correspond to the different energy distribution over frequencies. Therefore, usually perform FFT to obtain the magnitude, frequency response of each frame.

**Mel-scaled the filter bank:** The speech signal consists of tones with different frequencies. For each tone with an actual Frequency,  $f$ , measured in Hz, a subjective pitch is measured on the 'Mel' scale. The *Mel-frequency* scale is linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz.Using the following formula to compute the Mel( $f$ ) for a given frequency  $f$  in Hz:

$$\text{Mel}(f) = 2595 * \log_{10}(1 + f/700)$$

One approach to simulate the subjective spectrum is to use a filter bank; one filter for each desired Mel frequency component. The filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant Mel-frequency interval.



The reasons for using triangular band pass filters are twofold:

- Smooth the magnitude spectrum such that the harmonics are flattened to obtain the envelope of the spectrum with harmonics.
- Reduce the size of the features involved.

**Discrete cosine transforms, or DCT:** In this step, we apply DCT on the 20log energy  $E_k$  obtained from the triangular band pass filters to have L Mel-scale Cepstral coefficients. The formula for DCT is shown below.

$$C_m = \sum_{k=1}^N \cos \left[ \frac{m(k-0.5)\pi}{N} \right] E_k, m = 1, 2, \dots, L$$

Where N is the number of triangular band pass filters, L is the number of Mel-scale Cepstral coefficients. Usually we set N=20 and L=12. Since we have performed FFT, DCT transforms the frequency domain into a time-like domain called quefrequency domain. The obtained features are similar to cepstrum; thus, it is referred to as the Mel-scale Cepstral coefficients, or MFCC.

### *iii) Classification using k-nearest neighbours (k-NN) algorithm*

This algorithm operation is there to compare a given new record with training records and finding training records that are similar to it. It searches the space for the k training records that are nearest to the new record as the new record neighbours. In this algorithm nearest is defined in terms of a distance metric such as Euclidean distance. Euclidean distance between the two records (or two points in n-dimensional space) are defined by:

If  $x_1 = (x_{11}, x_{12}, \dots, x_{1n})$  and  $x_2 = (x_{21}, x_{22}, \dots, x_{2n})$

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^n (X_{1i} - X_{2i})^2}$$

Where  $x_1$  and  $x_2$  are two records with n attributes. This Formula measures the distance between two patterns  $x_1$  and  $x_2$ . The K-nearest neighbour classifier is a supervised learning algorithm where the result of a new instance query is classified based on the majority of the k-nearest neighbour category.

#### **4.1.1 Methodology of Infant Cry Detection**

The aim of the detection algorithm is to classify each incoming segment of a stream of input audio signals as 'cry' or 'no cry' The algorithm analyses the signal at various time-scales (segments of several seconds, sections of about 1 second, and frames of several tens of milliseconds). Figure 4 shows the audio processing algorithm to detect cry signal.

- A VAD is applied and the amount of activity is calculated for each segment.
- Each segment is further divided into sections of 1 second, with an overlap of 50%.
- If the activity duration of a given section is below a predefined threshold (30%) [4], the section is considered as having insufficient activity, and is classified as 'no cry' or '0'.

- If the activity is above the threshold, the section is divided into short-time frames (with duration of 32 msec and a hop size of 16 msec).<sup>[4]</sup>
- Each frame is classified either as 'cry' or as 'no cry', based on its extracted features using a k-NN classifier.
- For each section, if at least half of the frames are classified as 'cry', the whole section is considered as 'cry', Otherwise, it is considered as 'no cry'.
- Use K-NN classifier to classify the data *in* which each frame is classified either as a crying sound ('1'), if close enough to cry training samples, or as 'no cry' ('0'). The signal is divided into consecutive and overlapping segments, each of 10 seconds, with a step of 1 second.

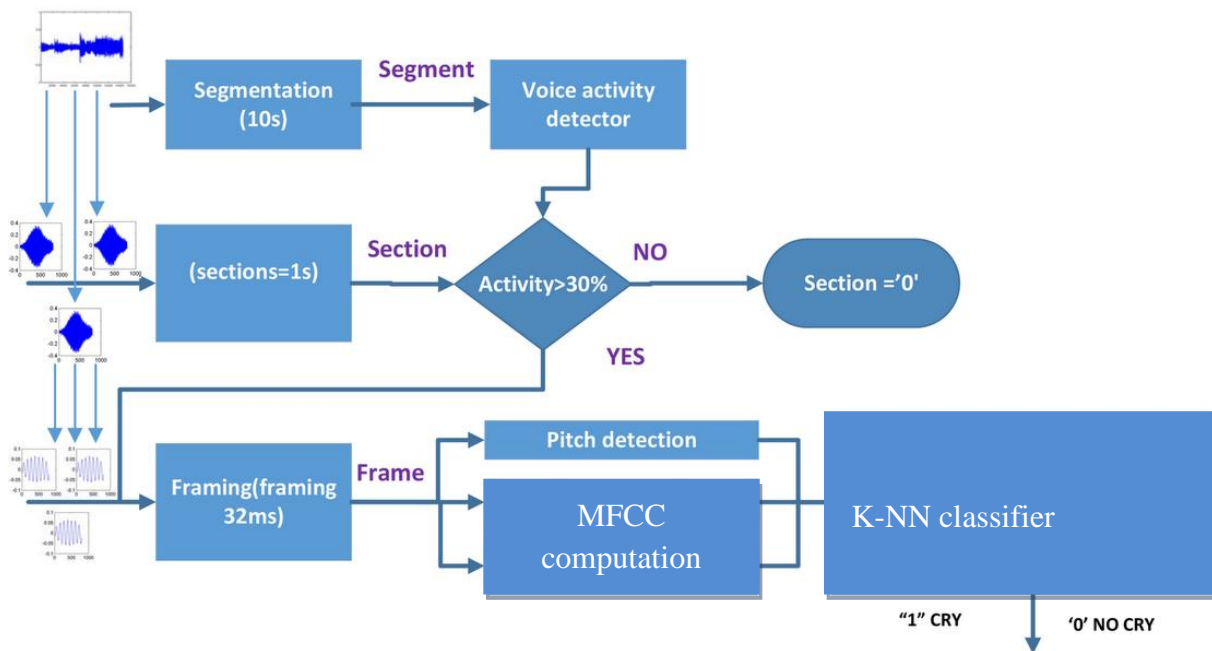


Figure 4: Infant cry detection algorithm –block scheme

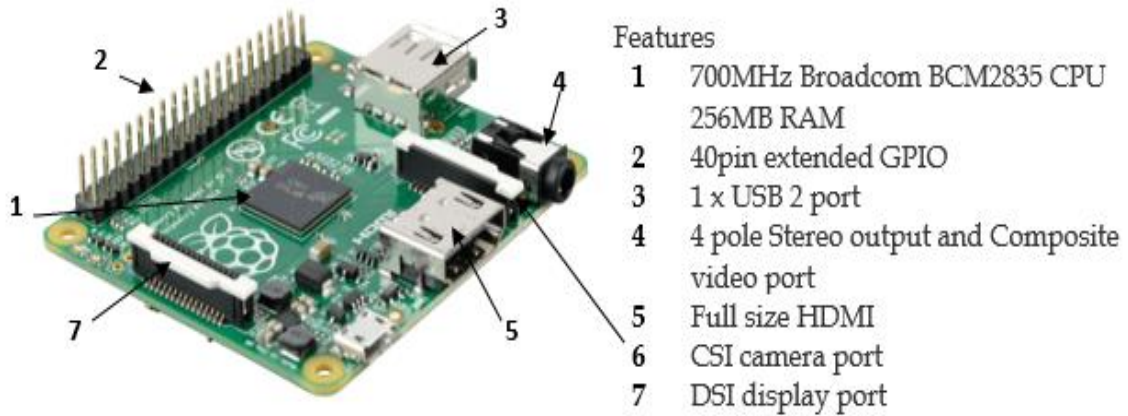
## 5 IMPLEMENTATION

Implementation has done in three steps.

- 1) Hardware implementation
- 2) Implementation of audio processing
- 3) Implementation of video streaming, parent monitoring and notification system

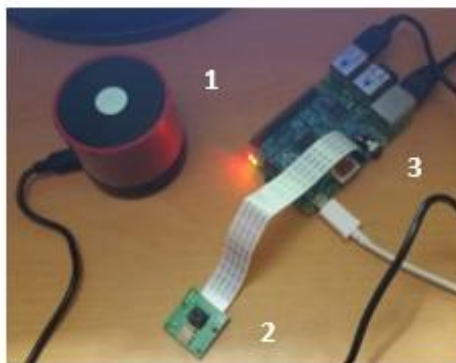
### 5.1 Hardware Implementation

For the prototype, Raspberry A+ Single board computer has been used. Main reasons for the selection is, its support multimedia and it has enough processing power and memory capacity to process audio processing algorithms introduced in section 4.



**Figure 5: Raspberry Pi A+ single board computer and its components**

For the video calling feature, Raspberry Pi compatible 5Mega pixel camera module has been used. This module capable of 1080p video and still image and it can connect to Raspberry Pi directly with CSI (Camera Serial Interface).Speaker system integrated via Raspberry Pi via USB hub. It has built in microphone and able to pay and store music files in MP3 format.



- Components**
- 1 USB compatible Speaker system
  - 2 Raspberry compatible 5MP camera
  - 3 Raspberry Pi B+ board (for testing purpose)

1

**Figure 6: Interconnection of Speaker, Camera and Raspberry PI board.**

**Implementation of the Soothing system**



**Figure 7: Implementation of Soothing system (Mechanical and Electronic circuit)**

Figure 7 shows the implementation of the mechanical arrangement of the soothing system. The right-hand side picture shows the fabrication of the Microcontroller based soothing control system.

## 5.2 Implementation of Audio Processing Algorithms

All the audio processing algorithms first implemented and tested in MATLAB 2014b environment. Later MATLAB Code converted to C code and uploaded to the Raspberry Pi board. In order to test the system, various crying samples are required.

### Data collection

The large number of baby cries has been recorded from daycare centers, neighbours and online databases, then fed to the computer. The signal is stored on the computer as a lossless WAV PCM file. Cry signals of babies ranging in age between 0-15 months. Collected data mainly has 3 voice domains; voice of baby, voice of the adult and mixture of baby & adult. It is assumed all the babies were healthy, Noise free environment (not engaged with engines, passersby, car horns, high hammering sounds, etc). Downloaded baby laugh/ splashes/ sneezing/ music and rattle sounds /giggles /happy vocals ranging between 0-15 months as negative samples. There are 150 training data, each of which represents the all sound forms, including cry and non-cry. 120 testing data are there respectively, including cry and non-cry.

### Testing the Algorithm in MATLAB

According to the algorithm,

- **Sample Data**-Sample data is the data that use instantly to check the results, according to the accuracy of the system simulations, we can change the sample size which we input to the system. The example MATLAB code is shown in Figure8.
- **Training Set**-It is the set that installed in the programmed memory and it used for distance calculation with sample data, 45 data set as the training set. With 37 babies cry samples and 08 non-cry samples and after that, gradually increased the training set 45 to 75 and 75 to 120 samples and 120 to 150. MATLAB implementation is shown in Figure 9.
- **Group Matrix**-This matrix defines the domains (baby cry domain, baby non-cry domain) of each and every data in training set.
- **Output** -Output is given according to the input sample size, each sample input data is calculated with each and every data in training set. All data in the training set are defined in the Group matrix as it belongs to cry or not. As an example, consider a one sample data, this sample is calculated with each and every data in the training set for obtaining the distance values and find out shortest distance given by which training data, then according to definition of group matrix algorithm identify the nearest neighbours of the sample data. MATLAB output results shown in Figure 10.

```

Command Window
>> % Matlab Function - knnclassify
% Syntax :
% Class = knnclassify(sample,training,group)

%Sample Matrix
A = [
    0.704,558.2278;
    0.683,525.2134;
    0.895,760.3448;
    0.702,490;
    0.955,958.6957;
    0.937,938.2979;

```

Figure 8: MATLAB code of Training matrix

```

%Training Matrix
B = [
    0.688,450.7042;
    0.595,324.2647;
    0.625,412.1495;
    0.655,428.1495;
    0.750,588;
    0.712,472.596;
    0.698,490;
    0.795,572.7273;
    0.760,604.1096;
    0.757,595.9459;
    0.775,612.5;
    0.785,630;
    0.735,537.8049;
    0.699,501.136;
    0.715,495.5056;
    0.698,479.3475;
    0.778,621.1268;
    0.812,711.2903;
    0.789,630;

```

Figure 9: MATLAB code of sample matrix

Figure 11 shows the MATLAB plot of variation Pitch Frequency against MFCC values. It confirms the other research finding, infant pitch frequency between 350-700Hz.

```

%Function
Class = knnclassify(A,B,G);

% Display result
disp('Result:');
disp(Class);
Warning: KNNCLASSIFY will be removed using ClassificationKNN.pred:
> In knnclassify at 80
Result:
'CRY'
'CRY'
'CRY'
'CRY'
'NOTCRY'
'NOTCRY'

```

Figure 10: Output of KNN algorithm

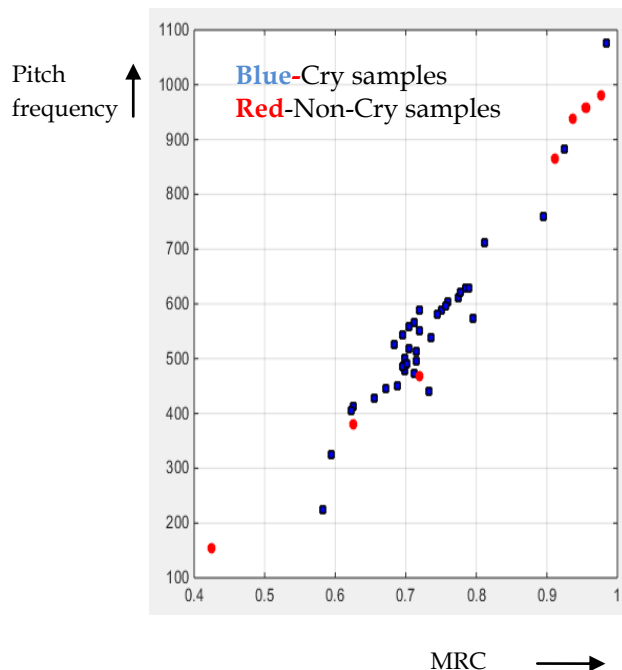


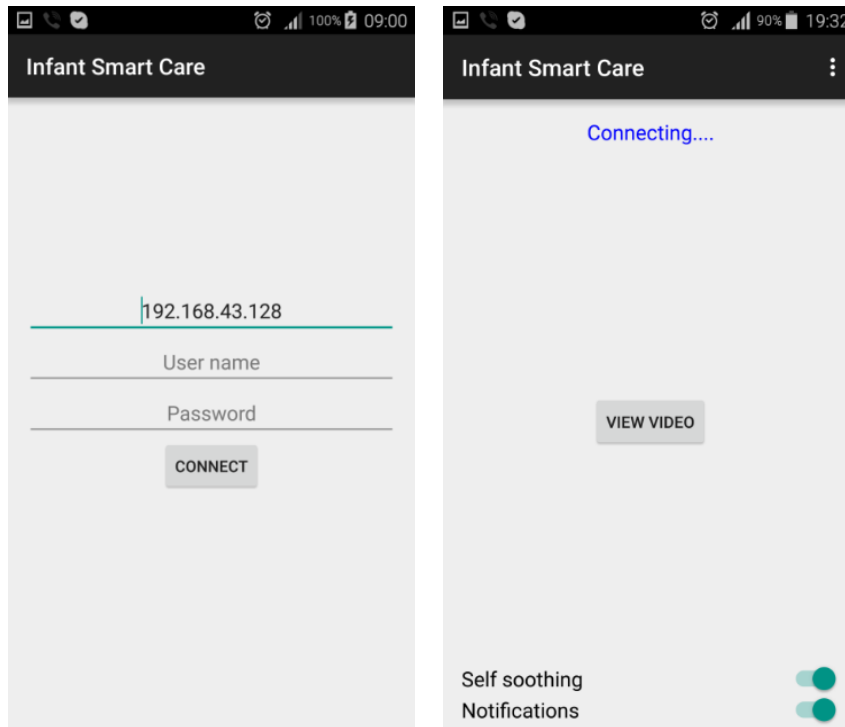
Figure 11: Graph of MFCCv'ss Pitch Frequency

### 5.3 Implementation of parent device application and video calling function

Parent device of the infant cry detection system is an android application which can be installed on a smart phone or a tablet computer. Once it linked to the home wireless network, users can install the application on the mobile device and have unlimited access to the video feed from the baby's room. The application is password protected to prevent outsiders from gaining access (Figure 12).

The login screen gives the IP address to connect the device after typing correct user name and pass word it enables to connect to the system. Then the application gives the option to on/off the self-soothing notification on the user's mobile device. By clicking on the top of the right side of the application (three dots) will give the user to disconnect or turn off the device.

Video calling feature implemented using available application called "VLC player". It has been installed and configured with the Raspberry Pi Operating system.



**Figure 12: Menus of the Android application**

## 6 RESULTS AND DISCUSSION

In this testing phase 150 training sample, and testing samples with 6 types of cry samples which have each 10 samples, as positive testing samples, and 60 testing negative samples to evaluate the performance of the algorithms. Testing was carried out in both MATLAB simulation and Raspberry Pi based prototype device.

When downloading and recording cry/laugh samples have been categorized mainly as above mentioned 12 categories, in this simulation it's detected only whether its cry or not cry. Table 2 shows the summary of testing results.

MATLAB simulation shows the cry detection accuracy 81.667% and on device shows the 78.33% accuracy which means low accuracy output than the MATLAB simulation results, this is because the MATLAB is a total theoretical simulation, but when testing it in practically there are some other noises can be added to the system, 60 negative samples shows 80% of non-cry detection accuracy and the device shows 76.667% non-cry detection accuracy. This means the average classification accuracy of the device is 77.5%.

**Table 2: Summary of Test results in MATLAB Simulation and on Device**

Positive	Test Samples	MATLAB simulation		Accuracy	On device		Accuracy
		Detected as cry	Detected as non-cry		Detected as cry	Detected as non-cry	
Pain	10	9	1	90%	9	1	90%
Screaming	10	8	2	80%	8	2	80%
Yell moan	10	9	1	90%	8	2	80%
frustrated	10	8	2	80%	7	3	70%
Whining	10	8	2	80%	7	3	70%
Upset	10	7	3	70%	8	2	80%
				81.667%			78.333%
Negative	Test samples	MATLAB simulation		Accuracy	On device		Accuracy
		Detected as non-cry	Detected as cry		Detected as non-cry	Detected as cry	
laugh	10	8	2	80%	8	2	80%
Happy vocal	10	7	3	70%	7	3	70%
giggles	10	8	2	80%	8	2	80%
gurgle	10	9	1	90%	8	2	80%
Adult baby speech mixture	10	7	3	70%	7	3	70%
Music play/rattle sounds	10	9	1	90%	8	2	80%
				80%			76.667%

Table 3 shows how is the effect of increasing number of training samples can increase the accuracy of the system.

**Table 3: Effect of the increasing of Number of samples**

No of training samples	MATLAB simulation	On Device
45	65.75%	60.22%
75	72.356%	68.003%
120	77.22%	73.58%
150	80.8335%	77.5%

Proposed infant cry detection algorithm is based on 2 decision levels in different time scales and classified either as 'cry' or 'non-cry' based on its spectral characteristics.

Multiple time scale analysis and detection levels are aimed in providing a classifier with high detection rate, whether the total average accuracy up to now is 80.8335% (Simulation) and 77.5% (On device). This can be improved by adding more training data

to the system as reference to the sample matrix. The detection rates fall in relatively broad range, whereas missed detection rate has a narrow range.

The main objective of this study is the development of a cry detection algorithm by applying different approaches in feature extraction and classification. Results show the developed algorithm working with acceptable accuracy. However, Soothing system needs to be tested with the real environment to test how babies responded to the soothing mechanism which is not done under this study. Also, feedback from parents about the overall product should take into consideration before moving to the next level.

## 7 CONCLUSION AND FUTURE WORK

With the help of the proposed cry detection algorithm, it can easily identify the infant's cry and verified it by using KNN with accurate results. Other than using only MFCC, the combination of Pitch and MFCC gives a more promising approach to cry detection.

Employed statistical model based voice activity detector in order to determine when is the cry detection algorithm should analyse the input signal. This leads to reduction of power consumption. All of these can improve the recognition accuracy.

Cry detection has been challenging because of the highly variable nature of input speech signals. Speech signals in training and testing sessions can be different due to many facts such as:

- Baby's voice change with time
- Health conditions. For example-deaf/asthma
- Speaking rates
- Variations in recording environments play a major role.

Therefore, increasing more training samples of different noises and speeches would give more accurate results.

### Future work:

- Improve the Accuracy of the cry detection algorithm by training with more samples.
- Test the soothing system in real environment and get parents' feedback regarding the overall product.
- Improve the audio processing algorithm to detect and notify the reason for baby cry.
- Improve the android application with more features.

## REFERENCES

1. Therese S. S. and Lingam, C. (2013). Review of Feature Extraction Techniques in Automatic Speech Recognition International Journal of Scientific Engineering and Technology [Online] Volume No.2, Issue No.6, June pp : 479-484 Available from: <http://ijset.com/ijset/publication/v2s6/paper606.pdf> [Accessed:24/02/2015]



2. Desai, N., Dhameliya, K. and Desai, V. (2013). Feature Extraction and Classification Techniques for Speech Recognition International Journal of Emerging Technology and Advanced Engineering[Online] Volume 3, Issue 12, December pp:367-371 Available from: [http://www.ijetae.com/files/Volume3Issue12/IJETAE\\_1213\\_64.pdf](http://www.ijetae.com/files/Volume3Issue12/IJETAE_1213_64.pdf)[accessed:24/02/2015]
3. Dhingra, S.D., Nijhawan,G. and Pandit, P. (2013). Isolated Speech Recognition Using MFCC And DTW [Online] Vol. 2, Issue 8, August Available from: [http://Www.Ijareeie.Com/Upload/2013/August/20P\\_ISOLATED.Pdf](http://Www.Ijareeie.Com/Upload/2013/August/20P_ISOLATED.Pdf) [Accessed:24/07/2015]
4. Cohen, R. and Lavner,Y. (2012). Infant Cry Analysis and Detection IEEE 27-th Convention of Electrical and Electronics Engineers in Israel [Online] November Available from: <http://spl.telhai.ac.il/speech/splnews/upload/Baby%20cry%20analysis.pdf>[Accessed:24/02/2015]
5. La Gasse, L. L., Neal, A. R. and Lester, B. M. (2005). Assessment Of Infant Cry: Acoustic Cry Analysis And Parental Perception Mental Retardation And Developmental Disabilities Research Reviews [Online] Volume11, December pp: 83-9 Available from:[http://homepage.psy.utexas.edu/homepage/group/Neallab/pubs/assessment\\_cry.pdf](http://homepage.psy.utexas.edu/homepage/group/Neallab/pubs/assessment_cry.pdf)[Accessed:24/02/2015]
6. Tan, L. and Karnjanadecha, M. (2003). Pitch Detection Algorithm: Autocorrelation Method and AMDF [online] Available from [http://five.dots.coe.psu.ac.th/~montri/Research/Publications/iscit2003\\_pda.pdf](http://five.dots.coe.psu.ac.th/~montri/Research/Publications/iscit2003_pda.pdf)[accessed:28/07/2015]
7. [www.amazon.com](http://www.amazon.com) (2015). Why Cry Mini Baby Cry Analyzer / Sleepy+ Stressed +Hungry[Online]Available from: [http://www.amazon.com/WhyCry-Analyzer-Sleepy-Stressed Annoyed/dp/B00IALBAD2](http://www.amazon.com/WhyCry-Analyzer-Sleepy-Stressed-Annoyed/dp/B00IALBAD2) [Accessed:24/02/2015]
8. [www.dx.com](http://www.dx.com)(2015). Wireless Baby Crying Detector with Parental Alarm [Online] Available from:[http://www.dx.com/p/wireless-baby-crying-detector-with-parental-alarm-13394?utm\\_source=dx&utm\\_medium=club&utm\\_campaign=club\\_skureview\\_page#.VPDvznyUeyI](http://www.dx.com/p/wireless-baby-crying-detector-with-parental-alarm-13394?utm_source=dx&utm_medium=club&utm_campaign=club_skureview_page#.VPDvznyUeyI)[Accessed:24/02/2015]
9. [www.amazon.com](http://www.amazon.com) (2015). Raspberry-Pi-Model-A [Online] Available from: [http://www.amazon.com/Raspberry-Pi-Model-A-256MB/dp/B00PEX5TO/ref=sr\\_1\\_fkmr1\\_1?s=electronics&ie=UTF8&qid=14258576&sr=1-1-fkmr1&key+words=raspi+a%2B](http://www.amazon.com/Raspberry-Pi-Model-A-256MB/dp/B00PEX5TO/ref=sr_1_fkmr1_1?s=electronics&ie=UTF8&qid=14258576&sr=1-1-fkmr1&key+words=raspi+a%2B)[Accessed: 24/02/2015]
10. [www.cpsc.gov/en/](http://www.cpsc.gov/en/) (2015). FAQs: Safety Standard for Children's Toys [online] Available from: <http://www.cpsc.gov/Business--Manufacturing/Business-Education/Toy-Safety/>[Accessed: 24/02/2015]
11. *Ian Poole* [www.radio-electronics.com](http://www.radio-electronics.com) (2015). IEEE 802.11n Standard [Online] Available from:<http://www.radio-electronics.com/info/wireless/wi-fi/ieee-802-11n.php> [Accessed: 24/02/2015]

12. wikipedia.org (2015). Sudden infant death syndrome [Online] Available from: [http://en.wikipedia.org/wiki/Sudden\\_infant\\_death\\_syndrome](http://en.wikipedia.org/wiki/Sudden_infant_death_syndrome) [Accessed: 24/02/2015]
13. www.adobe.com (2015). Live stream [Online] Available from: [http://www.adobe.com/devnet/adobe-media-server/articles/dynstream\\_live/popup.html](http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_live/popup.html) [Accessed: 24/02/2015]
14. www.scienceblogs.com (2009).Cognitive daily [online] Available from :<http://scienceblogs.com/cognitivedaily/2009/06/30/do-babies-like-color-if-so-whi/> [Accessed:10/06/2015]
15. <https://www.askdrsears.com/topics/health-concerns/fussy-baby/science-excessive-crying-harmful>[Accessed:10/06/2015]