International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5S3, July 2019

An Ann Based Real Time System for Classification of Normal and Abnormal Cries of Pre-Term and Neonates

Punith Kumar M B, T Shreekanth,Anupama M1, Sarsawath S

Abstract: ***Infants communicate with the external world through cry. Most of the problems in the infants can be explored through their cry within first year. Variations in cry can sometimes indicate the neurological disorders, genetic problems and many more. Classification of the infant cry as normal and abnormal at the early stages can reduce the course of action or any casualty. Hence this work proposes a computational approach for the early diagnosis of pre-term and neonates' infant cry. The previous works include various algorithms for classification, however the novelty in this work can be attributed to processing only voiced part of the cry signal. The cry signal is first preprocessed by decomposing it into three levels using db13 wavelet in order to remove any noise that has been inherited during signal acquisition. This signal is further processed to extract only voiced part of the speech by identifying the endpoints through Zero Crossing Rate and Energy. Then the MFCC features are extracted, as this kind of signal envelop is best estimated eventually using these kind of features and are used to train feed forward neural networks based on back propagation algorithm. In order to train the network 100 normal and 100 abnormal samples were used. The database has been obtained from the neonatal ward of JSS Hospital, Mysuru. The algorithm has been tested on the test dataset consisting of 50 samples. The performance of the proposed method has been evaluated on only voiced part of the cry signal using the diagnostic test measures and the efficiency is found to be 98%as compared to 90% efficiency if the same procedure is applied on the entire cry signal.***

Index Terms: Back Propagation, DCT, FFT, MFCC, STFT, Neural Network, Pre-term, Neonates, Hamming Window, Wavelet.

I. INTRODUCTION

It is an unquestionable that the infant communicates physiological and psychological distress to caregivers by crying. Crying accomplishes the purpose of continuous persistence by indicating to adult caregivers that the baby needs something [1]. Pre-term babies are those that are born before 37 weeks, and neonates refer to new born babies. Pre-term babies are usually susceptible to various complications growing up, and if not taken proper care, neonates may also develop complications. Premature baby's high pitched cry may be sign of something deeper.

Manuscript published on 30 July 2019.
***** Correspondence Author (s)
Punith Kumar M B, Associate Member of the Institution of Engineers (AMIE), Member of IEEE.

Thereby, the infant's cry analysis could soon become an important non-invasive complementary tool in early identification of infants at threat so that it will help in implementing prevention strategies and policies particularly important in the case of preterm neonates.Traditional studies of infant cry signals focus more on analyzing infant cry using features derived from fundamental frequency (F0) contour or pitch contour, energy of the cry signal in different frequency sub-bands and un-voicing present in the infant's cry [2]. This paper presents the development of an intelligent classification system for differentiating between normal and premature baby cry using wavelet packet transform (DWT) and machine learning techniques. This paper deals with some of the significant works on premature infant cry signal analysis. Ryuichi Kusaka et.al present result concerning with the design of intervention plans of care for premature infants with risks of high levels of crying that paved way for promotion of parental ability to recognize the neurobehavioral profile of the preterm infant [3].

Orlandi, Silvia & Reyes-Garcia, et al, proposed the exploitation of differences between full-term and preterm infant cry with robust automatic acoustical analysis and data mining techniques. Automatic infant cry recognition system for fast and proper identification of risk in preterm babies [4].

Whereas Alaie et al apply Gaussian mixture models to distinguish between healthy full term and premature infants and those with specific medical problems with true positive rate of 80.77% and true negative of 86.99 % [5].

Manfred et al performed robust tracking of main acoustic parameters on very short and time-varying signal frames of premature infant cry by making use of self-developed user-friendly software tool which helped in extracting very high fundamental frequency (*F*0) and resonance frequency (RFs) values, with abrupt changes and voiced/unvoiced features of very short duration in a single utterance used to deduce information on the state of health of preterm new-born babies [6].

Dhanashri U.S. Talauliker and Nayana Shenvi have analyzed the preterm infants cry by pre-processing to eliminate silenced region of cry signal and estimating the fundamental frequency (pitch) using time domain and frequency domain analysis as such parameters are of interest in exploring brain function at early stages of preterm infant development, for the timely diagnosis of neonatal disease and malformation [7].

Possible differences between full-term and preterm infants in their neurophysiologic maturity, and the subsequent impact on their speech development were analyzed with regard to the long-time average spectrum (LTAS) characteristics and preterm infants' crying behavior is believed to reveal different characteristics from that of full-term infants whose nervous systems are comparatively well-developed[12].

Johnston et al., in 1993 conducted study of premature infants which resulted in projecting that the premature infant has the basis for communicating pain via facial actions but that these are not well developed whereas full-term newborn is better equipped to interact with his caretakers and express his distress through specific facial actions [13].

Li-mei Chen et al., in 2014 used Long-time average spectrum (LTAS) to analyze the cry phonations of full-term and preterm infants and has shown significant differences in terms of spectral peak because of immature neurological development of preterm infants [14].

Shreekanth T et.al proposed an algorithm where in Discrete Wavelet Transform (DWT), Mel Frequency Cepstral Coefficients (MFCC), Vector

Quantization (VQ) and Euclidean Distance measure were employed to classify the cry signals [17].

The main goal of this current study is to find out how the normal and abnormal crying behavior between full-term and preterm infants differs from each other using wavelet packet transform. The detection might help in determining infants' health conditions. Furthermore, if the dissimilarity of the crying behavior can be systematically characterized, the measurements can be further applied to identify features in preterm infant cries. The limitations of all the techniques discussed above are that the entire cry signal is considered for feature extraction and for further classification of the signals. This work proposes the use of MFCC features and ANN classifier to classify the cry signal by processing only the voiced part of the cry and thereby leading to higher accuracy of classification. The rest of the paper is segmented as follows, Section II explains the various techniques used to develop the system. Section III explains the results obtained. The final section explains the inferences of the result obtained and improvements that can be made on the proposed system.

II. METHODOLOGY

A. Pre-Processing

Speech signals, especially those obtained in real-time are accompanied by noise. In order to remove the noise wavelet based de-noising is used. The main reason wavelets are used in de-noising is that they have perfect capability of reconstruction despite their irregular shape allowing them to morph into high order and linear polynomials. Hence wavelets can follow algebraic rules far better than other conventional filters that are based on Fourier transform design. This method hence, is capable of retaining the signal to the maximum extent removing only noise regardless of the frequency [Carl]. The steps for de-noising is as shown in Figure 1.

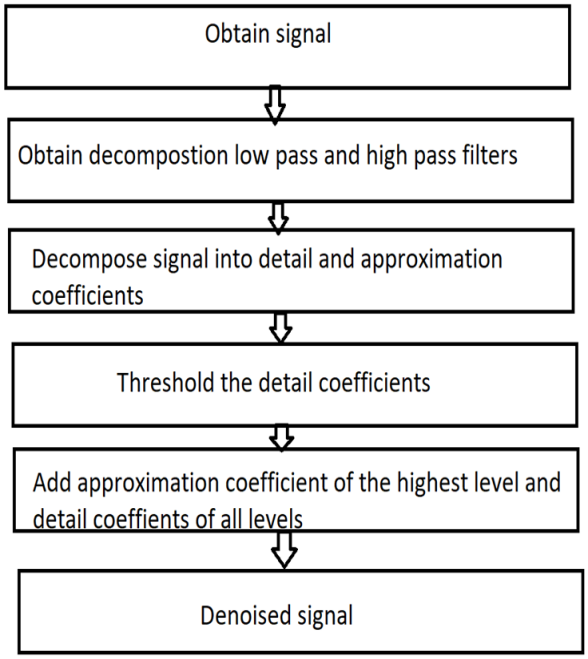
Assuming that the data X(t) is as represented in equation 1.

$$X(t) = S(t) + N(t) \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Where S(t) represents the signal, and N(t) represents additive noise. The aim is to remove the noise as much as possible.

The steps for decomposition are as follows:

a) The signal is subjected to decomposition into their coefficients, and dB13 consumes lesser time in processing, so it is used in this application [15]. Daubechies wavelets are used in a wide range of applications due to their asymmetrical nature and vanishing moments that help in analysis of the signal. The approximation and detail coefficients up to the third level are computed using the low-pass and high pass filters decomposition filters.



b) Thresholding is a very important step in wavelet de-noising. Hard thresholding can be described as the process of setting to zero the elements whose absolute values are lower than the threshold, and soft thresholding is called shrink or kill which is an extension of hard thresholding. It is based on first setting the elements with the absolute values lower than the threshold to zero, and then shrinking the other coefficients with different threshold selection rules [11].

There are various such thresholding methods and the one used in this method is heursure is used to obtain the threshold limit, as it provides a proper limit even if signal to for Heursure is represented in equation 2.

$$\lambda = \begin{cases} \lambda_1 & A < B \\ \min(\lambda_1, \lambda_2) & A \geq B \end{cases} \quad \ldots\ldots\ldots\ldots (2)$$

Where, A = $s-N/s$ and B =$(log2N)^{3/2}\sqrt{N}$, N is length of wavelet coefficient vector and s is the sum of squared wavelet coefficients [8].

c) The denoised signal is a sum of all the detail coefficients and the approximation coefficient of the last level, which helps in reconstructing the signal. The signal before and after de-noising is as shown in Figure 2.

d) The signal is then subjected to removal of any silence present in the cry signal, using methods of short time energy, and zero crossing rate. The voiced part of an audio signal had high energy due to periodicity, and the zero crossing rate is low for a signal that is unvoiced. Using these fundamentals as depicted in Figure.3, we can remove the silence in the audio signal containing the cry [16].

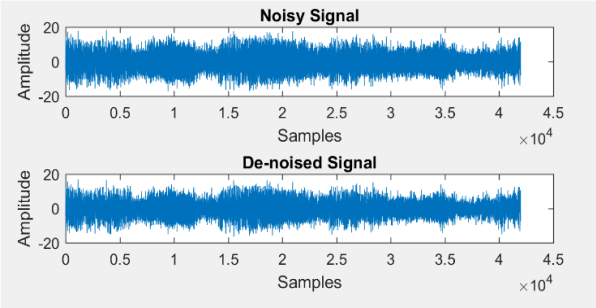


Figure 2: De-noised signal

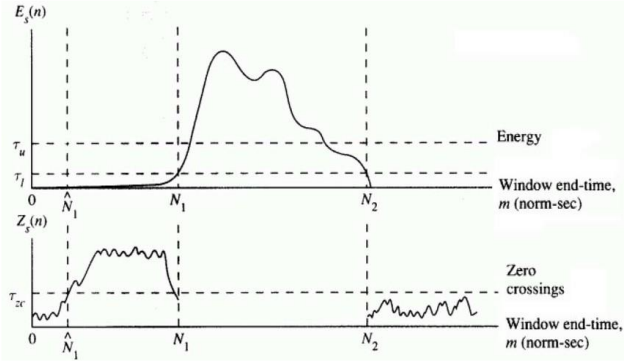


Figure 3: Method used to extract only voiced part of the cry

B. Feature Extraction and Training.

1. Mel-Frequency Cepstral Coefficients (MFCCs) Feature Extraction:

The first step to any audio processing is the extraction of features while discarding unwanted information. In order to accurately extract various useful information from speech, it is important to understand the mechanism used to generate sounds in humans. The vocal tract of a human being determines the type of sounds emitted and these sounds behave like an envelope of short time spectrums. The MFCC features are the best to represent this envelope and has hence been used in this system for feature extraction.

The algorithm for implementing MFCC features is as follows [9]:

a) The first 3.5 seconds of the signal is analyzed to determine if the cry is normal or abnormal due to presence of noise. The signal is sampled at 8000 Hz which is the sampling rate of the .wav file containing the cry.

b) Once the frames are obtained the high frequency components of the signal need to be amplified, hence a pre-emphasis filter is used. The filter is useful due to the following reasons:

i. Balance the frequency spectrum since high frequencies usually have smaller magnitudes compared to lower frequencies

ii. Avoid numerical problems during the Fourier transform operation

iii. May also improve the Signal-to-Noise Ratio (SNR). The first – order pre-emphasis filter is as given in equation 3.

$$y(t)=x(t)-\alpha x(t-1) \quad \ldots\ldots (3)$$

Where α is the filter co-efficient which is set to any value between 0.95 and 0.97 generally and is this case we set this value to 0.95.

c) An audio signal is constantly changing, so we need to obtain intervals where the signals do not change much, hence we need to break down the signal into smaller intervals so that it can be analyzed. This method is otherwise called as sampling. The intervals chosen are usually 20ms- 40ms. If the frame for analysis is any shorter than this then the spectral estimate cannot be obtained properly, and if it is longer than this then the signal varies too much to be analyzed.

d) We then apply a window function, in particular hamming to each frame. The equation representing the hamming window is as given in equation 4.

$$w[n]=0.54-0.46\cos(2\pi n/n-1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where, $0 \le n \le N-1$, and Nis the window length. Plotting this equation, the graph obtained is shown in Figure 4.

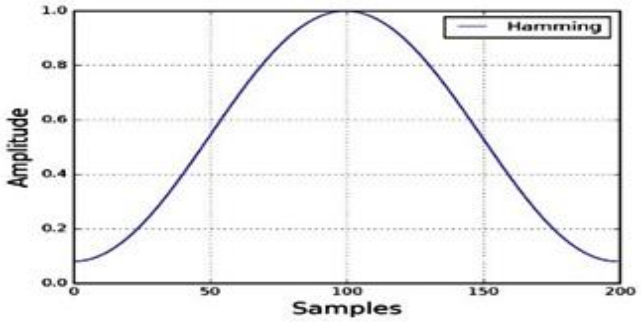


Figure 4: Hamming window

e) The N point FFT is calculated for each frame, this is also called short time Fourier transform (STFT), and then we compute the power spectrum.

f) The final step is applying triangular filters, which are usually forty in number on a Mel-scale to the power spectrum to obtain the frequency bands. The Mel-scale aims to mimic the non-linear human ear perception of sound, by being more discriminative at lower frequencies and less discriminative at higher frequencies. We can convert between Hertz (f) and Mel (m) using the following equations:

$$m = 2595 log\ 10(1+f/700) \quad \ldots\ldots\ldots (5)$$

$$f=700(10^{m/2595}-1) \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

The filter H is hence obtained using Hertz and Mel as in equation 7:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \le k < f(m) \\ 1 & k = f(m) \\ \dfrac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \le f(m+1) \\ 0 & k > f(m-1) \end{cases} \quad \ldots.(7)$$

g) However, the data obtained by the coefficients is highly correlated, which is problematic for machine learning techniques. Hence we use discrete cosine transform DCT to de-correlate the coefficients and obtain a compressed version of the signal.

2. Multi-Layer Perceptron

The very objective of this system is to estimate whether the given speech signal is a normal cry or an abnormal cry. Hence, we require logistic regression where a given data is classified into one of the two groups. The most elementary sections of a neural network are the input layer, a set of hidden layers and output layer. The signal is sent to the input layer and the output layer produces the result after classification. A perceptron is the most elementary part of a neural network that receives inputs, multiplies them by some weight, and then passes them into an activation function to produce an output. The activation function in this case is a logistic function given by equation 8.

$$F(x) = L/ (1+e^{-k(x-x0)}) \text{ ..... } (8)$$

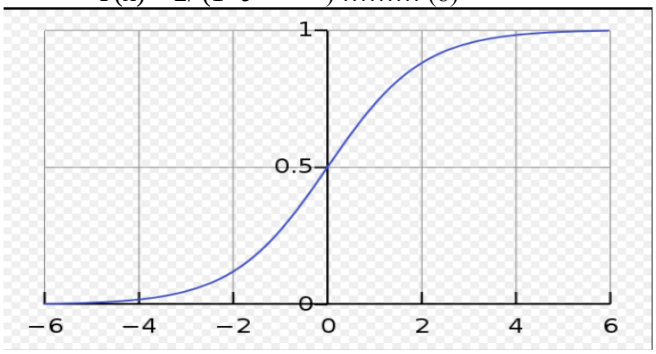


Figure 5: Sigmoid function

Multi-layer perceptron is often applied to supervised learning problems as it trains a set of input-output pairs and learns to model the correlation (or dependencies) between those inputs and outputs. Training involves adjusting the parameters, or the weights and biases, of the model in order to minimize error. Back propagation is used to make those weigh and bias adjustments relative to the error, and the error itself can be measured in a variety of ways, including by root mean squared error (RMSE).The amount of data that needs to be processed in machine learning is very high as every sample is multi-dimensional, and once the functions as well as the data set are in place, the function needs to be minimized to derive the result.

Multi-layer perceptron's ability to accurately represent a data set is dependent on the number of hidden layers, hidden units and an objective function [10]. There are two hidden layers, the first layer has ten hidden units and the second has five hidden units. The general method for choosing the number of hidden layers is by obtaining the sum between the numbers of inputs and outputs and a number in the range one to ten. In this case the minimum value is twenty-five, however reducing the number of the hidden units produces the same accuracy. Reducing the number of unit and layers below this leads to improper estimation of the model, and increasing beyond this value converts the model into a memory bank without allowing the untrained data to be properly estimated, while trained data to be estimated accurately. Sometimes one hidden layer is enough to produce a model that accurately represents the data-set, however since the model required for this particular application is complex more than one hidden layer is chosen and kept at a minimum of two hidden layers. The data is trained using the MFCC features of the training set, and the prediction is based on the MFCC features of the test set. The model was capable of representing the data enough to not over fit it, while producing accurate results in determining abnormality.

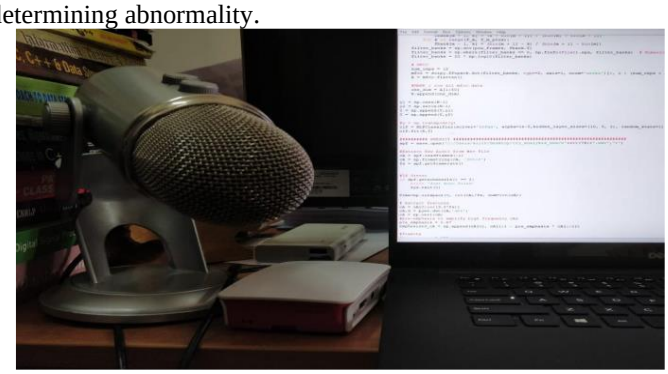


Figure 6: The system setup

III. RESULT

The system has been implemented using the Raspberry pi microprocessor as it is compact and powerful enough to run machine learning algorithms and provides a wide range of libraries and drivers that not only can run complicated sequence of operations with lower computational cost and time, but also allow for easier interfacing of various hardware. There are other microprocessors which perform at a higher efficiency but do not allow for cost reduction hence Raspberry pi is chosen as it offers cost reduction and reliability along with versatility. The Blue Yeti USB microphone is used to record the cry signals, where the signals are sampled using sampling frequency of 44.1 kHz. The real time system has been implemented using a microphone with a USB connection which can be connected to one of the four USB ports of the Raspberrypi. The microphone obtains the input and the processing is carried out on the Raspberrypi to determine if the cry is a normal or an abnormal cry ensuring portability. The system setup developed is as shown is Figure.6. The output of the system for the random signals under test is depicted in Figure 7 and Figure 8.

Table 1 and Table 2 depicts various results obtained through human judgment and automated interpretation by the system as subjected to the voiced part of the cry and the entire cry signal respectively. The total number of samples that were used to develop the system was two-hundred of which hundred were normal cries and the other hundred were abnormal cries. The training set consisted of 100 normal and 100 abnormal samples of cry. The 50 samples was used for testing. The sum of the columns indicates the total number of normal and abnormal cries as judged by a human. The sum of the rows indicates the total number of normal cries and abnormal cries as interpreted by the system. The system is said to perform correctly when the human and the system produce the same output. True positives are generated when the human and system agree that the cry is normal, and false negatives when both agree that the cry is abnormal. Hence the accuracy is measured using the equation (9)

$$\text{Accuracy} = (\text{true}-\text{positive} + \text{false}-\text{negative})/ (\text{total number of samples}) \text{ ..... } (9)$$

Table 1: Results of the cry analyzer: Considering only voiced part of the cry

Human/ System	Positive (normal)	Negative (abnormal)
True (normal)	24	0
False (abnormal)	1	25

Table 2: Results of the cry analyzer: Considering the entire cry signal

Human/ System	Positive (normal)	Negative (abnormal)
TRUE (normal)	24	1
FALSE (abnormal)	1	24

Table 3: Performance Comparison

System	Accuracy
[18]	96%
Proposed	98%

The number of true positives are twenty four, false-negatives are twenty five. Hence the accuracy is 98%. True-negatives indicate that the human has interpreted that the cry is abnormal, but the system classifies it as normal and false-positives indicates that the human has classified the cry as normal and system classifies it as abnormal. It is important to ensure that the true negatives are minimum as this indicates that the abnormality is ignored by the system. It is more crucial to identify the abnormality than it is to identify normality. It can be found that the system has achieved the task of identifying abnormality from Table 1. The performance of the proposed system is compared with the existing system in the literature as shown in Table.3 and the increase in accuracy of the proposed system is attributed to processing only voiced part of the signal.

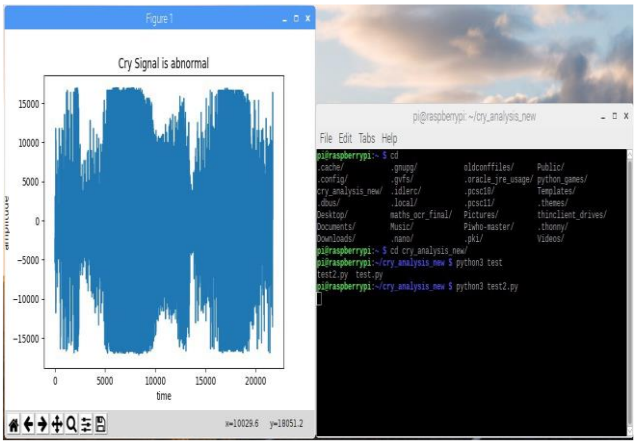


Figure 7: The waveform of the random signal: 1 under test and its classification.

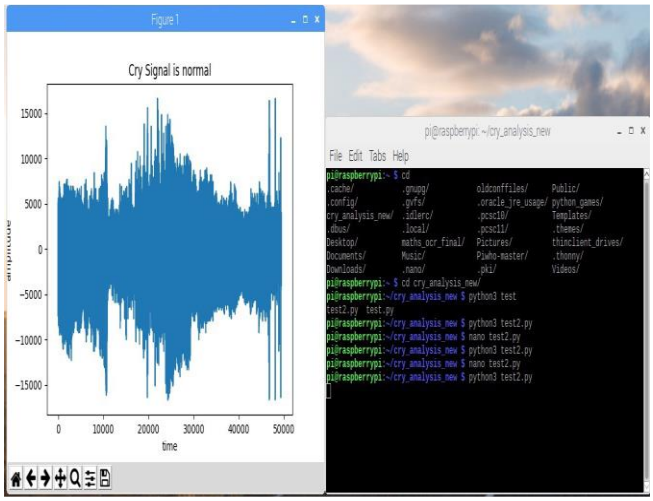


Figure 8: The waveform of the random signal: 2 under test and its classification.

VI. CONCLUSION

The system works in real time to identify the cries and classify as normal or abnormal. It is implemented using a cost effective microprocessor raspberry-pi, and a USB microphone. The implementation is simple as it requires the use of very few features that can be processed easily and quickly. The signal is decomposed using dB13 as it is takes very less processing time for the de-noising algorithm to convert the noisy signal into one that is capable of being processed. MFCC features are extracted from the voiced part of the cry signal to train and classify the signal using a multi-level perceptron. The data-set consists of two-hundred samples of which hundred are normal cries and the other hundred are of abnormal cries for training. The proposed method was evaluated on 50 data set, producing an accuracy of 98% when only voiced part of the signal is processed and 90% when the entire cry signal is processed. More sophisticated systems also include the capability of detecting the cry before classification which can be incorporated in this system. The entire system can be made more modular so that it is portable, all of which would provide the same efficiency but a better user-interface.

VII. ACKNOWLEDGEMENT

The author would like to thank Director **Dr. N P Nataraj,** JSS Institute of Speech and Hearing for his support in providing access to resources.

REFERENCES

1. Hofer M. A., “Unexplained infant crying: An evolutionary perspective”, ActaPaediatrica, 91, 491-496, 2002.
2. Chittora, Anshu and Patil, Hemant, “Newborn infant’s cry analysis”, International Journal of Speech Technology. 19. 10.1007/s10772-016-9379-8, 2016
3. Ryuichi Kusaka,, Shohei Ohgi, Kenta Shigemori,and Tetsuya Fujimoto, “Crying and Behavioral Characteristics in Premature Infants”, J Jpn Phys TherAssoc ,v.11(1); 2008
4. Orlandi, Silvia and Reyes-Garcia, Carlos Alberto & Bandini, Andrea & Donzelli, Gianpaolo & Manfredi Claudia, “Application of Pattern Recognition Techniques to the Classification of Full-Term and Preterm Infant Cry”. Journal of voice: official journal of the Voice , 2015

5. [Hesam Farsaie Alaie, Lina Abou-Abbas, Chakib Tadj, "Cry-based infant pathology classification using GMMs" , Speech Communication, Volume 77, March 2016, Pages 28-52
6. C.Manfredi, L.Bocchi, S.Orlandia, L.Spaccaterrab, G.P.Donzellib, "High-resolution cry analysis in preterm newborn infants",Medical Engineering and Physics ,Volume 31, Issue 5, June 2009, Pages 528-532
7. Dhanashri U.S. Talauliker, NayanaShenvi, "Analysis of Cry in New Born Infants", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, March 2015
8. Neema Verma, "Performance Analysis Of Wavelet Thresholding Methods In Denoising Of Audio Signals Of Some Indian Musical Instruments", International Journal of Engineering Science and Technology (IJEST), Vol. 4 No.05 May 2012.
9. Parwinder Pal Singh, Pushpa Rani, "An Approach to Extract Feature using MFCC", IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 08, August. 2014
10. Hassan Ramchoun, Mohammed Amine 11 JanatiIdrissi, Youssef Ghanou, Mohamed Ettaouil, " Multilayer Perceptron: Architecture Optimization and Training", International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No1,2016
11. Prashant Arora1, Kulwinder Singh, "Denoising of Speech Signals Using Wavelets", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue I, January 2017.
12. Goberman AM, Robb MP, "Acoustic examination of preterm and full-term infant cries: the long-time average spectrum", Journal of Speech Language and Hearing Research. 1999 Aug; 42(4):850-61.
13. Johnston, C., Stevens, B., Craig, K., & Grunau, R., "Developmental changes in pain expression in premature, full-term, two- and four-month-old infants", Pain, 52, 1993, 201-208.
14. Li-mei Chen, Yu-Hsuan Yang, Chyi-Her Lin, Yuh-Jyh Lin, Yung-Chieh Lin, "Spectrum Analysis of Cry Sounds in Preterm and Full-Term Infants", Conference on Computational Linguistics and Speech Processing ROCLING 2014, pp. 193-203
15. Jashanpreet Kaur, Seema, Sunil Kumar, "Audio Noise Reduction Using Discrete Wavelet Transform and Filters",International Journal of Advanced Research in Computer Science & Technology,Vol. 4, Issue 2, Apr. - Jun. 2016.
16. [Bachu R.G., Kopparthi S., Adapa B., Barkana B.D. "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal", Advanced Techniques in Computing Sciences and Software Engineering , by Elleithy, Khaled, ISBN 978-90-481-3659-9.
17. T. Shreekanth and S Saraswathi, " Acoustic analysis and classification of Infant Cry: A new approach", Journal on Digital Signal Processing, Vol.5, No.2, 2017, pp.1-7.

AUTHORS PROFILE

Dr. Punith Kumar M B was born in Mandya, India, in 1985. He received the B.E. Degree in Electronics and Communication Engineering from The National Institute of Engineering, Mysore in 2007, and the M.Tech in VLSI Design and Embedded Systems from PES College of Engineering, Mandya under the The Visvesvaraya Technological University (VTU) , Belgaum in 2010 and Ph.D. degrees in Electronics from the University of Mysore (UoM), Mysore, India, in 2017. In 2007, he joined the Department of Electronics and Communication Engineering as a Lecturer in JVIT, Bidadi, Bangalore. In 2010, he joined the Department of Electronics and Communication Engineering as a Assistant Professor in BGS Institute of Technology, BG Nagar, Mandya. In 2017, he joined the Department of Electronics and Communication Engineering as an Associate Professor in PES College of Engineering Mandya. His current research interests include image processing, video processing, video shot detection etc. Dr Punith Kumar M B is a Life Member of the Indian Society for Technical Education (ISTE) and Associate Member of the Institution of Engineers (AMIE) , Member of IEEE. He was the Judge, Chairperson and Review member for the National and International Conference.