# *A comparative Study Between Artificial Intelligence Techniques in an Automatic Infant's Pain Cry Identification System*

Yousra Abdulaziz Mohammed
College of Medicine
University of Fallujah
Baghdad , Iraq
yosraalhity@gmail.com

***Abstract*—This paper presents a comparative study between Continuous Density Hidden Markov Model (CDHMM) and Artificial Neural Network (ANN) on an automatic infant's cries classification system which main task is to classify and differentiate between pain and non-pain cries belonging to infants. In this study, Mel Frequency Cepstral Coefficient (MFCC) and Linear Prediction Cepstral Coefficients (LPCC) are extracted from the audio samples of infant's cries and are fed into the classification modules. Two well-known recognition engines, ANN and CDHMM, are conducted and compared. The ANN system (a feed-forward multilayer perceptron network with back-propagation using scaled conjugate gradient learning algorithm) is applied. The novel continuous Hidden Markov Model classification system is trained based on Baum –Welch algorithm on a pair of local feature vectors.**

**After optimizing system's parameters by performing some preliminary experiments, CDHMM gives the best identification rate at 96.1%, which is much better than 79% of ANN whereby in general the system that are based on MFCC features performs better than the one that utilizes LPCC features.**

***Keywords- Artificial Neural Networks, Continuos Density Hidden Markov Model; Mel Frequency Cepstral Coefficient, Linear Prediction Cepstral Coefficints, Infant Pain Cry Classification***

## I. INTRODUCTION

Infants often use cries as communication tool to express their physical, emotional and psychological states and needs [1]. An infant may cry for a variety of reasons, and many scientists believe that there are different types of cries which reflects different states and needs of infants, thus it is possible to analyze and classify infant cries for clinical diagnosis purposes.

A number of research work related to this line have been reported, whereby many of which are based on Artificial Neural Network (ANN) classification techniques. Petroni and Malowany [4] for example, have used three different varieties of supervised ANN technique which include a simple feed-forward, a recurrent neural network (RNN) and a time-delay neural network (TDNN) in their infant cry classification system. In their study, they have attempted to recognize and classify three categories of cry, namely 'pain', 'fear' and 'hunger' and the results demonstrated that the highest classification rate was achieved by using feed-forward neural network. Another research work carried out by Cano [8] used the Kohonen's self organizing maps (SOM) which is basically a variety of unsupervised ANN technique to classify different infant cries. Al-Azzawi [7] designed an automatic infant cry recognition system based on the fuzzy transform (F-transform) that classifies two different kinds of cries, which come from physiological status and medical disease, a supervised MLP scaled conjugate ANN was used and the classification accuracy obtained was 96%.

Apart from the traditional ANN approach, other infant cry classification technique studied is Support Vector Machine (SVM) which has been reported by Barajas and Reyes [2]. Here, a set of Mel Frequency Cepstral Coefficients (MFCC) was extracted from the audio samples as the input features. On the other hand, Orozco and Garcia [6], use the linear prediction technique to extract the acoustic features from the cry samples of which are then fed into a feed- forward neural network recognition module.

Hidden Markov Model is based on double stochastic processes, whereby the first process produces a set of observations which in turns can be used indirectly to reveal another hidden process that describes the states evolution [13]. This technique has been used extensively to analyze audio signals such as for biomedical signal processing [17] and speech recognition [18]. NN are defined as systems which has the capability to model highly complex nonlinear problems and composed of many simple processing elements, that operate in parallel and whose function is determined by the network's structure, the strength of its connections, and the processing carried out by the processing elements or nodes. The prime objective of this paper is to compare the performance of an automatic infant's cry classification system applying two different classification techniques, Artificial Neural Networks and continuous Hidden Markov Model.

Here, a series of observable feature vector is used to reveal the cry model hence assists in its classification. First, the paper describes the overall architecture of an automatic recognition system which main task is to differentiate between an infant 'pain' cries from 'non-pain' cries. The

978-1-5386-9188-5/18/$31.00 ©2018 IEEE
DOI 10.1109/AiCIS.2018.00019

performance of both systems is compared in terms of recognition accuracy, classification error rate and F-measure under the use of two different acoustic features, namely Mel Frequency Cepstral Coefficient (MFCC) and Linear Prediction Cepstral Coefficients (LPCC). Separate phases of system training and system testing are carried out on two different sample sets of infant cries recorded from a group of babies which ranges from newborns up to 12 months old.

The organization of the rest of this paper is as follows. Section 2 describes the nature of datasets used in the study. Section 3 presents an overall architecture of the recognition system, in terms of the features extracted and the classification techniques. On the other hand, Section 4 and Section 5, details out the descriptions of the experimental set-ups and the results of the findings respectively. Section 5 concludes the paper and highlights potential area for future works.

## II. Experimental Data Sets

The infant cry corpus collected is a set of 150 pain samples and 30 non-pain samples recorded from a random time interval. The babies selected for recording varies from newborns up to 12 month old, a mixture of both healthy males and females. The records are then sampled at 16000 Hertz. It is important to highlight here that the pain cry episodes are the result of the pain stimulus carried out during routine immunization at a local pediatric clinic, in Darnah, Libya. Recordings resulting from anger, or hunger were considered as non-pain utterances were recorded at quiet rooms at various infants' home. Recordings were made on a digital player at a sample rate of 8000 Hertz and 4-bit resolution with microphone placed between 10 to 30 centimeters away from the infant's mouth. The audio signals were then transferred for analysis to a sound editor and then re-sampled at 16000 with 16 bit resolution [3, 4, 5]. The final digital recordings were stored as WAV files.

## III. System Overview

In this paper, we consider two recognition engines for infant's cry classification system. The first is an artificial neural network (ANN) which is a very popular pattern matching in the field of speech recognition. Feedforward multilayer perceptron (MLP) network with a backpropagation learning algorithm is the most well known of ANN. The other technique is a continuous density hidden Markov model (CDHMM). The overall system is as depicted in Figure 1 below:

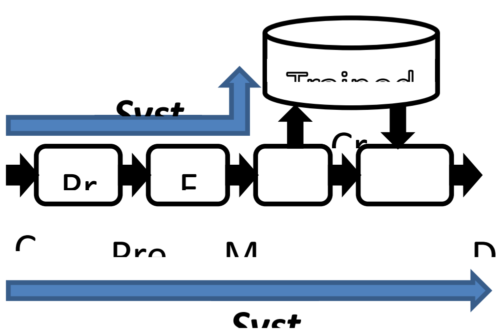

Figure 1: Infant Cry Recognition System Architecture

From Figure 1, we can say that infant cry recognition system implies three main tasks,

- Signal Preprocessing
- Feature Extraction
- Pattern Matching

### A. Pre-Processing

The first step in the proposed system is the pre-processing step which requires the removal of the 'silent' periods from the recorded sample. Recordings with cry units lasting at least 1 second from the moment of stimulus event were used for the study [4, 6]. The cry units are defined as the duration of the vocalization only during expiration [5].

The audio recordings were then divided further into segments of exactly 1 second length, where each represents a pre-processed cry segments as recommended by [2, 4, 6, 8]. Before these one second segments can be used for feature extraction, a process called *pre-emphasis* is applied. Pre-emphasis aims at reducing the high spectral dynamic range, and is accomplished by passing the signal through an FIR filter whose transfer function is given by.

$$F(z) = 1 - kz^{-1},\ (0 < k < 1) \qquad (1)$$

A typical value for the pre-emphasis parameter '$k$' is usually 0.97. Consequently the output is formed as follow:

$$y(n) = \square s(n)\ k.s(n-1)$$

(2)

where $s(n)$ is the input signal and $y(n)$ is the output signal from the first order FIR filter.

Every segment of 1 second is divided thereafter in frames of 50-milliseconds with successive frames overlapping by 50% from each other. The next step is to use a window function on each individual frame in order to minimize discontinuities at the beginning and end of each frame. Typically the window function used is the Hamming window and has the following form:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right),\quad 0 \le n \le N-1 \qquad (3)$$

Given the above window function and assuming that there are $N$ samples in each frame, we will obtain the following signal after windowing.

$$y(n) = x(n)w(n)\ ,\quad 0 \le n \le (N\text{-}1) \qquad (4)$$

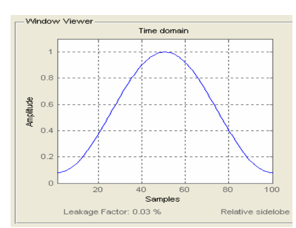


Figure 2: Hamming Window

Off the 150 pain and 30 non-pain recording samples, 625 and 256 one second cry segments were obtained respectively. Off these 881 cry segments, 700 were used for system training and 181 were used for system testing. It is important to use a separate set of cry segments for training and testing purposes in order to avoid obtaining biased testing results.

### B. Feature Extraction

*Mel-Frequency Cepstral Coefficients (MFCC)*

MFCCs are one of the more popular parameter used by researchers in the acoustic research domain. It has the benefit that it is capable of capturing the important characteristics of audio signals. Cepstral analysis calculates the inverse Fourier transform of the logarithm of the power spectrum of the cry signal, the calculation of the mel cepstral coefficients is illustrated in Figure 3.


Figure 3: Extraction of MFCC from Audio Signals

The cry signal must be divided into overlapping blocks first, (in our experiment the blocks are in Hamming windows) which is then transformed into its power spectrum. Because human perception of the frequency contents of sounds does not follow a linear scale. They are approximately linear with logarithmic frequency beyond about 1000 Hz. The mel frequency warping is most conveniently done by utilizing a filter bank with filters centered according to mel frequencies as shown in Figure 4.

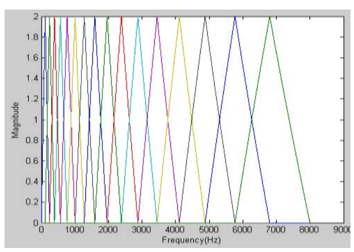


Figure 4: Mel Spaced Filter Banks

The mapping is usually done using an approximation (where *fmel* is the perceived frequency in mels),

$$f_{mel}(f) = 2595 \times \log(1 + \frac{f}{700}) \quad (5)$$

These vectors were normalized so that all the values within a given 1 second sample would lie between ±1 in order to decrease their dynamic range [4].

*Linear Prediction Cepstral Coefficients (LPCC)*

LPCC is Linear Predicted Coefficients (LPC) in the cepstrum domain. The basis of linear prediction analysis is that a given speech sample can be approximated with a linear combination of the past *p* speech samples [9]. This can be calculated either by the autocorrelation or covariance methods directly from the windowed portion of audio signal [10]. This method of linear prediction is known as an appropriate technique to process speech and some works also have been performed based on this technique [21]. In this study the LPC coefficients were calculated using the autocorrelation method that uses Levinson-Durbin recursion algorithm. Hence, LPCC can be derived from LPC using the recursion as follows:

$$c_0 = r(0) \quad (6)$$

where *r* is derived from the LPC autocorrelation matrix

$$c_m = a_m + \sum_{k=1}^{m-1}(\frac{k}{n})c_k a_{m-k} \text{ for } 1 < m < P \quad (7)$$

$$c_m = \sum_{k=m-p}^{m-1}(\frac{k}{n})c_k a_{m-k} \text{ for } m > P \quad (8)$$

where *p* is the so called prediction order, $a_m$ represents the $m^{th}$ LPC coefficient and $m$ is the number of LPCC's needed to be calculated. Whereas $c_m$ is the $m^{th}$ LPCC.

*DELTA coefficients*

It has been proved that system performance may be enhanced by adding time derivatives to the static parameters [10]. The first order derivatives are referred to as delta features and can be calculated as shown in formula 9,

$$d_t = \frac{\sum_{\theta=1}^{\Theta}\theta(c_{t+\theta}-c_{t-\theta})}{2\sum_{\theta=1}^{\Theta}\theta^2} \quad (9)$$

where *d,* is the delta coefficient at time ***t,*** computed in terms of the corresponding static coefficients $c_{t-\theta}$, to $c_{t+\theta}$ , and $\boldsymbol{\Theta}$ is the size of delta window.

### C. Patter Matching

*Artificial Neural Network (ANN) Approach*

Neural Networks are defined as systems which has the capability to model highly complex nonlinear problems and composed of many simple processing elements, that operate in parallel and whose function is determined by the network's structure, the strength of its connections, and the processing carried out by the processing elements or nodes.

The feedforward multilayer perceptron (MLP) network architecture using a backpropagation learning algorithm is one of the most popular neural networks. It consists of at least three layers of neurons: an input layer, one or more hidden layers and an output layer. Processing elements or neurons in the input layer only act as buffers for distributing the input signal xi to neurons in the hidden layer. The hidden and output layers have a non-linear activation function. A backpropagation is a supervised learning algorithm to calculate the change of weights in the network. In the forward pass, the weights are fixed and the input vector is propagated through the network to produce an output. An output error is calculated from the difference between actual output and the target. This is propagated backwards through the network to make changes to the weights [19].

For this work, a feed-forward multilayer perceptron using full connections between adjacent layers was trained and tested with input patterns described above in a supervised manner with scaled conjugate gradient back-propagation learning algorithm since it has shown good results in classifying infant's cries than other NN algorithms [3, 20]. The number of computations in each iteration is significantly reduced mainly because no line search is required. Different feature sets were used in order to determine the set that results in optimum recognition rate. Sets of 12 MFCC, 12MFCC+1ST derivative, 12 MFCC+1st & 2nd derivative, 20 MFCC, 16 LPCC and 16 LPCC+ 1st derivative were used. Two frame length was used, 50ms and 100 ms in order to determine the right choice that gives the best results. Two architectures were investigated in this study. First, with one hidden layer then with two hidden layers. The number of hidden neurons was varied to obtain optimum performance. The number of neurons in the input layer is decided by the number of elements in the feature vector. Output layer has two neurons each for one cry class The activation function used in all layers in this work is hyperbolic tangent sigmoid transfer function 'TANSIG'. Training stops when any of these conditions occur:

- The maximum number of epochs (repetitions) is reached. we established 500 epochs at maximum because above this value, the convergence line do not have any significant change.
- The networks were trained until the performance is minimized to the goal i.e. mean squared error is less than 0.00001.

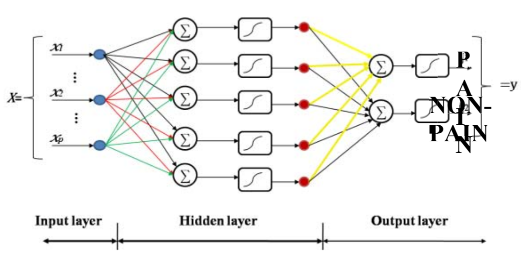


Figure 5: MLP Neural Network

*Hidden Markov Model (HMM) Approach*

The continuous HMM is chosen over the discrete counterparts since it avoids losing of critical signal information during discrete symbol quantization process and that it provides for better modeling of continuous signal representation such as the audio cry signal [12]. However, computational complexity when using the CHMMs is more than the computational complexity when using DHMMs. It normally takes more time in the training phase[11]

An HMM is specified by the following:

- $N$, the number of states in the HMM;
- $\pi_i = P(x_1 = s_i)$, the prior probability of state $s_i$ being the first state of a state sequence. The collection of $\pi_1$ forms the vector $\pi = \{ \pi_1, \ldots, \pi_N\}$;
- $a_{ij} = P(x_{t+1} = s_j | x_t = s_i)$ , the transition coefficients gives the probability of going from state $s_i$ immediately to state $s_j$. The collection of $a_{ij}$ forms the transition matrix A.
- Emission probability of a certain observation $o$, when the model is in state $s_i$ . The observation $o$ can be either discrete or continuous [11]. However, in this study a continuous HMM is applied whereby continuous observations $o \in \Re^D. b_i = P(o_t \mid x_t = s_i)$ indicates the probability density function (pdf) over the observation space for the model being in state $s_i$.

For the continuous HMM, the observations are continuous and the output likelihood of a HMM for a given observation vector, can be expressed as a weighted sum of $M$ multivariate Gaussian probability densities [15] as given by the equation

$$b_j(o_t) = \sum_{m=1}^{M} C_{jm} \aleph(o_t; \mu_{jm}, \Sigma_{jm}), \quad \text{for } 1 \leq j \leq N \quad (10)$$

where $o_t$ is a $d$-dimensional feature vector, $M$ is the number of Gaussian mixture components $m$ is the mixture index ($1 \leq m \leq M$), $C_{jm}$ is the mixture weight for the $m^{th}$ component, which satisfies the constraint

- $C_{jm} > 0$ and
- $\sum_{m=1}^{M} C_{jm} = 1$ , $1 \leq j \leq N$, $1 \leq m \leq M$

where $N$ is the number of states, $j$ is the state index, $\mu_{jm}$ is the mean vectors of the $m^{th}$ Gaussian probability densities function, and $\aleph$ is the most efficient density functions widely used without loss of generality[11, 15] which is defined as,

$$\aleph(o_t; \mu_{jm}, \Sigma_{jm}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left\{-\frac{1}{2}(o_t - \mu_{jm})^\tau \Sigma_{jm}^{-1}(o_t - \mu_{jm})\right\} \quad (11)$$

In the training phase, and in order to derive the models for both 'pain' and'non-pain' cries (i.e. to derive $\lambda_{pain}$ and $\lambda_{non-pain}$ models respectively) we first have to make a rough guess about the parameters of an HMM, and based on these initial parameters more accurate parameters can be found by applying the Baum-Welch algorithm. This mainly requires for the learning problem of HMM as highlighted by Rabiner in his HMM tutorial [13]. The re-estimation procedure is sensitive to the selection of initial parameters. The model topology is specified by an initial transition matrix. The state means and variances can be initialized by clustering the training data into as many clusters as there are states in the model with the K-means clustering algorithm and estimating the initial parameters from these clusters [16].

The basic idea behind the Baum-Welch algorithm (also known as Forward-Backward algorithm) is to iteratively re-estimate the parameters of a model, and to obtain a new model with a better set of parameters , which satisfies the following criterion for the observation sequence $O = (o_1, o_2, \ldots, o_t)$,

$$P(O|\bar{\lambda}) \geq P(O|\lambda) \quad (12)$$

where the given parameters are, $\lambda = (\pi, A, \mu_j, \Sigma_j)$ . By setting, $\lambda = \bar{\lambda}$ , at the end of every iteration and re-estimating a better parameter set, the probability of $P(O|\lambda)$ can be improved until some threshold is reached. The re-estimation procedure is guaranteed to find in a local optimum. The flow chart of the training procedure is shown in Figure 6,

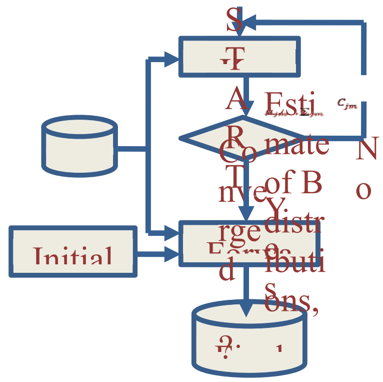

Figure 6: Training Flow Chart

Once the system training is completed, system testing is carried out to investigate the accuracy of the recognition system. The classification is done with a maximum-likelihood classifier, that is the model with the highest probability with respect to the observations sequence, i.e., the one that maximizes *P*(*Model* | *Observations*) will be the natural choice,

$$\hat{\lambda}_{ML} = \arg\max_{1 \le k \le K} P(O|\lambda_k) \quad (13)$$

This is called the *maximum likelihood* (*ML*) estimate. The best model in maximum likelihood sense is therefore the one that is most probable to *generate* the given observations.

In this study, separate untrained samples from each class where fed into the HMM classifier and were compared against the trained 'pain' and 'non-pain' model. This testing process mainly requires for evaluation problem of HMM as highlighted by Rabiner [13]. The classification follows the following algorithm:

If P(test_sample| $\lambda_{pain}$) > P(test_sample| $\lambda_{non\text{-}pain}$)

Then test_sample is classified *'pain'*

Else

test_sample is classified *'non_pain'* (14)

## I. EXPERIMENTAL RESULTS

A total of 700 cry segments of 1 second duration is used during system testing. Each feature vector is extracted at 50ms windows with 75% overlap between adjacent frames. The size of the input data frame used was 50ms and 100ms in order to determine which would yield the best results for this application. These settings were used in all experiments reported in this paper which compare between the performance of both types of infant cry recognition systems described above utilizing 12 MFCCs (with 26 filter banks) and 16LPCCs (with $12^{th}$ order LPC) acoustic features. The effect of the dynamic coefficients for both used features was also investigated.

After optimizing system's parameters by performing some preliminary experiments, it was found that a fully connected (an ergodic) HMM topology with five states and eight Gaussian mixture per state is the best choice to model the cry utterances because it scored the best identification rate. On the other hand, a fully connected feedforward MLP NN trained with backpropagation with Scaled Conjugate Gradient algorithm, with one hidden layer and five hidden neurons has given the best recognition rates. The best results for both systems were obtained using 50 ms frame length.

The results of training and testing datasets are evaluated according to the overall classification accuracy which is calculated by taking the percentage number of correctly classified test samples.

$$\text{System Accuracy (\%)} = \frac{x_{correct}}{T} \times 100\% \quad (15)$$

where $x_{correct}$ is the total number of correctly classified test samples and $T$ is the overall total number of test samples.

Both systems performed optimally with 20 MFCC's 50 ms window. For the NN based system the hierarchy of one hidden layer having 5 hidden nodes showed to be the best, while an ergodic HMM with 5 states and 8 Gaussians per state has resulted in the best recognition rates. The optimum recognition rates obtained were 96.1% for HMM trained with 20 MFCC, whereas for ANN the highest recognition rate was 79% using 20 MFCC also. For both systems trained with LPCC's, the best recognition rate obtained for HMM was 78.5% using 16 LPCC+DEL, 10 Gaussians , 5 states and 50 ms whereas for ANN was 70.2% using 16 LPCC, 5 states, 8 Gaussians per state and 50 ms window. Table I and II show the results obtained from both systems for different feature sets. Figure 7 summarizes a comparison between the performance of both systems for different feature sets used.

TABLE I. PERFORMANCE OF HMM SYSTEM

| Feature Set | Accuracy | HMM Topology &Window size |
|---|---|---|
| 20 MFCC | **96.1%** | 8Gaussians 5 states, 50 ms |
| 12 MFCC | 82.3% | 10Gaussians, 5 states, 100 ms |
| 12MFCC+ energy+DEL, | **92.3%** | 8Gaussians, 5 states, 50 ms |
| 12 MFCC+ energy + DEL+ DEL DEL | 91.2% | 8Gaussians, 5 states, 50 ms |
| 16 LPCC | 72.4% | 14 Gaussians, 5 states, 50 ms |
| 16 LPCC + DEL | 78.5% | 10 Gaussians 5 states,, 50 ms |

TABLE II. PERFORMANCE OF NN SYSTEM

| Feature Set | Accuracy | NN Structure & window size |
|---|---|---|
| 20MFCC | **79%** | 1 hidden layer, 5 nodes, 50ms |
| 12 MFCC | 73% | 1 hidden layer, 5 nodes, 100ms |
| 12MFCC+ energy+DEL | **76.2%** | 2 hidden layers, 5 nodes each, 50ms |
| 12 MFCC+ energy + DEL+ DEL DEL | 72.4% | 1 hidden layer, 5 nodes, 100ms |
| 16 LPCC | 70.2% | 1 hidden layer, 5 nodes,50ms |
| 16 LPCC+DEL | 66.3% | 2 hidden layers, 10 nodes, 50ms |

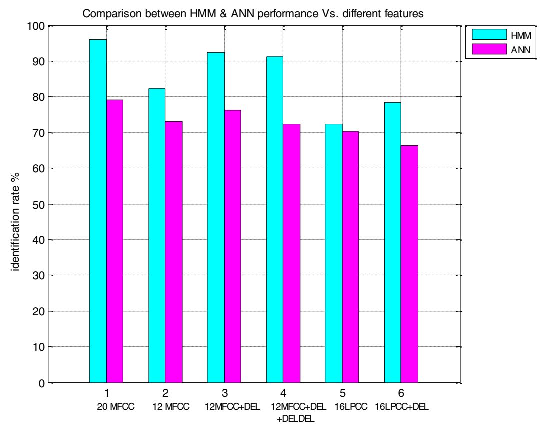


Figure 7: Identification Rate of ANN and CDHMM

## II. Conclusion and Future Work

In this paper we applied two artificial intelligence techniques to identify infant pain and non-pain cries . From the obtained results, it is clear that HMM has been proved to be the superior classification technique in the task of recognition and discrimination between infants' pain and non-pain cries with 96.1% classification rate with 20 MFCC's extracted at 50ms. From our experiments, the optimal parameters of CHMM developed for this work were 5 states , 8 Gaussians per state, whereas the ANN architecture that yielded the highest recognition rate of 79% was with one hidden layer and 5 hidden nodes. In general, both systems performed better with 50 ms frame length than with 100 ms. The results also show that the system accuracy performs better with MFCC's rather than with LPCC's features.

## References


**[1]** J.E. Drummond, M.L. McBride, "The Development of Mother's Understanding of Infant Crying,", *Clinical Nursing Research,* Vol 2, pp 396 – 441, 1993.

**[2]** Sandra E. Barajas-Montiel, Carlos A. Reyes-Garcia, "Identifying Pain and Hunger in Infant Cry with Classifiers Ensembles", *Proceedings of the 2005 International Conference on Computational Intelligence for Modeling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, Vol. 2, pp 770 – 775, 2005

**[3]** Yousra Abdulaziz, Sharrifah Mumtazah Syed Ahmad, "Infant Cry Recognition System: A Comparison of System Performance based on Mel Frequency and Linear Prediction Cepstral Coefficients", *Proceedings of International Conference on Information Retrieval and Knowledge Management (CAMP 10),* Selangor, Malaysia, PP 260-263, March, 2010.

**[4]** M. Petroni, A.S. Malowany, C.C. Johnston, B.J. Stevens, "A Comparison of Neural Network Architectures for the Classification of Three Types of Infant Cry Vocalizations", *Proceedings of the IEEE 17$^{th}$ Annual Conference in Engineering in Medicine and Biology Society*, Vol. 1, pp 821 – 822, 1995

**[5]** H.E. Baeck, M.N. Souza, "Study of Acoustic Features of Newborn Cries that Correlate with the context", *Proceedings of the IEEE 23$^{rd}$ Annual Conference in Engineering in Medicine and Biology ,* Vol. 3, pp 2174 – 2177, 2001.

**[6]** 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", *proceedings of international joint conference on neural Networks*, Vol. 4, pp 3140-3145, 2003**.**

**[7]** Nemir Ahmed Al-Azzawi**, "** Automatic Recognition System of Infant Cry based on F-Transform**"** *International Journal of Computer Applications (0975 – 8887)*, Volume 102– No.12, pp. 28-32, September 2014

**[8]** Suaste, I., Reyes, O.F., Diaz, A., Reyes, C.A., "Implementation of a Linguistic Fuzzy Relational Neural Network for Detecting Pathologies by Infant Cry Recognition," *Advances of Artificial Intelligence-IBERAMIA,*Vol. 3315, pp. 953-962, 2004

**[9]** Jose Orozco, Carlos A. Reyes-Garcia, "Implementation and Analysis of Training Algorithms for the Classification of Infant Cry with Feed-forward Neural Networks", *in IEEE International Symp. Intelligent Signal Processing,* pp. 271 – 276, Puebla, Mexico, 2003.

**[10]** Eddie Wong, Sridha Sridharan, "Comparison of Linear Prediction Cepstrum Coefficients and Mel-Frequency Cepstrum Coefficientsfor Language Identification", *in Proc. Int. Symp. Intelligent Multimedia, Video and Speech Processing*, Hong Kong May 24, 2001.

**[11]** Hesham Tolba," Comparative Experiments to Evaluate the Use of a CHMM-Based Speaker Identification Engine for Arabic Spontaneous Speech", *2nd IEEE International Conference on Computer Science and Information Technology, ICCSIT*, pp. 241 – 245, 2009.

**[12]** Joseph Picone, "Continuous Speech Recognition Using Hidden Markov Models", *IEEE ASSP magazine*, pp.26-41, 1991.

**[13]** L.R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition". *Proceedings of the IEEE*, Vol. 77 (2, pp. 257-286), 1989.

**[14]** Yang Liu, Elizabeth Shriberg, "Comparing Evaluation Metrics For Sentence Boundary Detection", *in Proc. IEEE Int. Conf.* Vol. 4, pp185-188, 2007.

**[15]** Jun Cai, Ghazi Bouselmi, Yves Laprie, Jean-Paul Haton, "Efficient likelihood evaluation and dynamic Gaussian selection for HMM-based speech recognition", *Computer Speech and Language,* vol.23, pp. 47–164, 2009.

**[16]** B. Resch, "Hidden Markov Models, A tutorial of the course computational intelligence", *Signal Processing and Speech Communication Laboratory.*

**[17]** Dror Lederman, Arnon Cohen, Ehud Zmora, "On the use of Hidden Markov Models in Infants' Cry Classification", *22$^{nd}$ IEEE Convention,* pp 350 – 352, 2002.

**[18]** Mohamad Adnan Al-Alaoui, Lina Al-Kanj, Jimmy Azar, Elias Yaacoub, “Speech Recognition Using Artificial Neural Networks And Hidden Markov Models”, *The 3rd International Conference on Mobile and Computer Aided Learning, IMCL Conference,* Amman, Jordan, 16-18 April 2008.

**[19]** Sawit Kasuriya, Chai Wutiwiwatchai, Chularat Tanprasert, “Comparative Study of Continuous Hidden Markov Models (CHMM) and Artificial Neural Network (ANN) on Speaker Identification System”, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)*, Vol. 9, Issue: 6, pp. 673-683, 2001.

**[20]** José Orozco, Carlos A. Reyes García, "Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks”, Proceedings of European Symposium on Artificial Neural Networks, pp 349 – 354, 2003.

**[21]** J. Saraswathy, M.Hariharan, Sazali Yaacob and Wan Khairunizam, "Automatic classification of infant cry: A review", International Conference on Biomedical Engineering (ICoBE),27-28 February 2012, Penang.