2009 Second International Conference on Computer and Electrical Engineering

A System for Detecting of Infants with Pain from Normal Infants Based on Multi-Band Spectral Entropy by Infant's Cry Analysis

Mahmoud Mansouri Jam
Technical and Engineering dept. of Shahed University
Tehran, Iran
E-mail: mansourijam@gmail.com

Hamed Sadjedi
Technical and Engineering dept. of Shahed University
Tehran, Iran
E-mail: sadjedi@shahed.ac.ir

***Abstract*—Infant cry is a multimodal behavior that contains a lot of information about the infant, particularly, information about the health of the infant. In this paper a new feature in infant cry analysis is presented for recognition two groups: infants with pain and normal infants, by Mel frequency multi-band entropy extraction from infant's cry. In signal processing stage we made pre-processing included silence elimination, filtering, pre-emphasizing. After taking Fourier transform, spectral entropy was computed as single feature of signal. In classifying stage, by training artificial neural network, correction rate of recognition was obtained 66.9%. In order to enhancement in results, we used Mel filter bank. Entropy of each sub-band constitutes elements of next feature vector. We used PCA analysis for reducing in dimension of the recent feature vector. After ANN training, correction rate improved to 88.5%. So multiband spectral entropy enhanced results in salient correction rate.**

Keywords- Infant cry; Spectral Entropy; Pain; Artifical Neural Network.

I. INTRODUCTION

Infant's cry is a multimodal and dynamic behaviour that carries a lot of information in its acoustic signal. An infant's cry contains a lot of information about the baby, as hunger, pain, sleepiness or boredom. Mothers and specialists in the area of child care can distinguish their infant from others by crying [1]. Such entity vividly clarifies an infant's physiological anatomy and psychological condition. Physiological quantities from the laryngeal configuration, i.e. the length of the vocal tract, in return exemplify the resonance and formant effects. Reason of infant's cry is the same reason of speech in adults i.e. to let others know about their needs or problems.

In this investigation, cry of infants with pain and cry of normal infants is classified. The pioneers of studying on infant's cry were Wasz-Hockert and et al in Scandinavia in the1960s [2, 3]. In previous papers some acoustic features for cry recognition has been studied, but in this paper we study spectral entropy features that before was not used in cry recognition. The production of infant's cry can be modeled similarly to the generation of adult speech, so infant's cry classification can consider a pattern recognition problem such as automatic speech recognition (ASR), thus we can use the same techniques in ASR. Based on the information that cry carries, we may detect disorders in infant as soon as possible.

II. INFANT'S CRY PRODUCTION

Infant's cry is in the most sensitive range of the human auditory sensation area, if any disorder occurs with the infant the cry may differ, this issue is main idea in recognition systems based on infant's cry [4, 5].

The first stage of the cry production mechanism is initiated in the infant's brain upon external or internal stimuli (hunger, pain, etc.). In the second stage the brain command is translated into series of commands through the nervous system to the speech and respiratory limbs which are responsible for the creation of acoustic signals at the physiological level. This process continues with the ejection of air from the lungs to the vocal tract. The vocal tract starts at the vocal cords and ends at the lips or nozzle. The tracts are constructed from the pharynx which interconnects between the mouth to the esophagus, the mouth cavity and the nasal cavity which starts at the velum and ends at the nozzle. At the sound utterance time the status of the vocal tract changes by changes the status and the position of its internal organs, in this way, different acoustic characteristics are formed. The larynx consists of the vocal cords which vibrate as a result of the air pressure, the muscle movements and the physical characteristics of the cords [6].

Several models of cry production have been theorized. The sound-filter theory underlies most acoustic analyses of cry sounds. As Fig. 1, the universal model of speech generation suggests that the waveform that impinges upon the listener's ear is a function of the characteristics of the source and its filters [4].

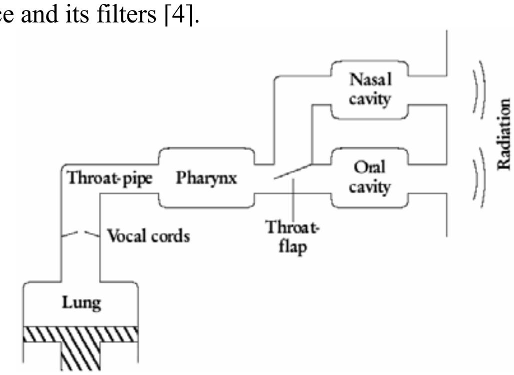

Figure 1. Universal schema for the voice production [4]

978-0-7695-3925-6/09 $26.00 © 2009 IEEE
DOI 10.1109/ICCEE.2009.164

Spectrogram of infant's cry exhibits some differences between normal cry and pathological cry. As Fig. 2 normal crying features constitute of raising-falling pitch contour, ascending-descending melody and high intensity seen from the spectrum. In pathological infant's crying spectral intensity will be lower than normal, rapid pitch shifts, weak phonation and silences during the crying as Fig. 3.

III. AUTOMATIC CRY RECOGNITION PROCESS

The infant's Automatic Cry Recognition (ACR) process is, in generally, a pattern recognition problem similar to Automatic Speech Recognition (ASR). The system takes the wave of infant's cry as the system input, and at the end obtains the kind of cry [7].

Generally, the process of ACR is done in two steps. The first step is known as signal processing, whereas the second is known as pattern classification. In the first step, the goal is feature extraction, the cry signal is firstly normalized and cleaned according to arguments in the next section of paper, and then it is analysed to extract the most important characteristics, that we know it "acoustical analysis". The set of obtained characteristics is represented by a vector, which can consider a pattern for second step.

In second step, the set of all vectors is used to train the classifier. Some classifiers in ACR system, in recently papers, have been Hidden Markov Model (HMM), Bayesian Classifier, Artificial Neural Networks (ANN), that in this paper we used ANN classifier. Later on, a set of unknown feature vector, the test patterns, is compared with the knowledge that the computer has to measure the classification output efficiency. Fig. 4 shows the different stages of the described recognition process. In next section we will explain each step with details.

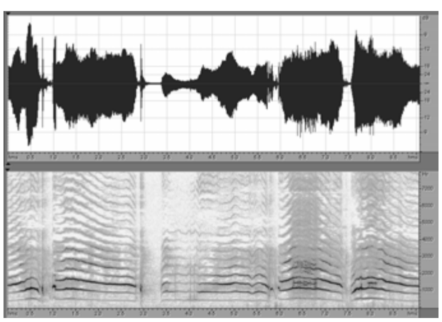

Figure 2. Waveform and spectrum of normal cry

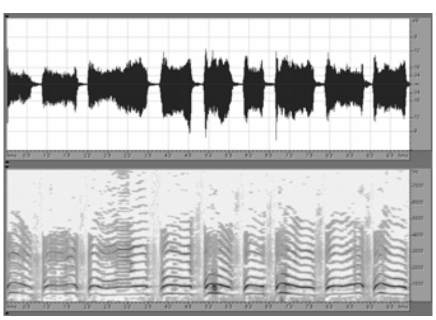

Figure 3. Waveform and spectrum of pathological cry

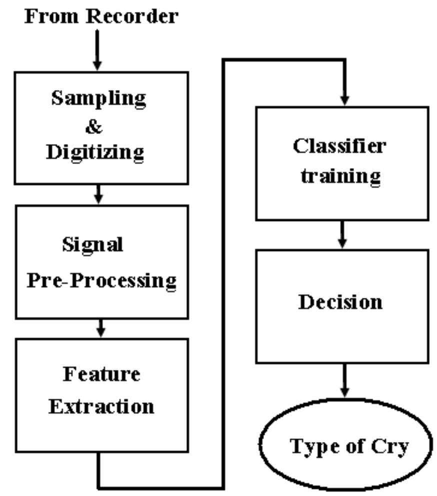

Figure 4. Automatic Cry Recognition (ACR) process

A. DataBase

For providing database of infant's cry, we collect some signals from private infants by parents contribution and some signals were used from the "Baby Chillanto" database from "Instituto Nacional de Astrofisica Optica y Electronica – CONACYT, Mexico". We made, uniformly, all signals by 16 KHz sampling rate and each sample in 16 bit was quantized. Then, each signal is labelled by its infant specifications.

B. Signal Pre-Processing

1) Segmentation with silence elimination: In order to have cry samples in sufficient numbers, we must make segments from cry signals. In segmentation, firstly it is necessary to distinguish more important parts of signal from others. Because in parts of signal with less important contents, there are not any valuable information and they only cause computational cost. As Fig. 5 we can see silence as a kind of low importance sections in cry signal, that we must eliminate them.

Automatic segmentation process needs to detect silence from main components of signal. We used threshold level on short time energy of cry signal. The short time energy is computed by equation (1).

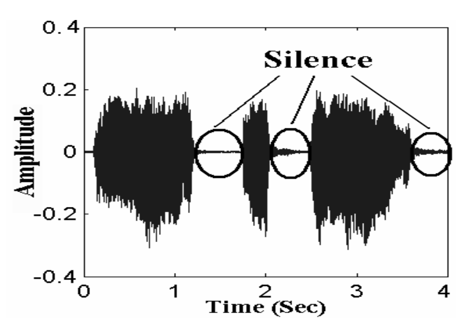

Figure 5. Cry signal before silence elimination

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

That N is number of samples in small time frame and x(m) is cry signal, and w(n) is rectangular window by:

$$w(n) = \begin{cases} 1, & 0 \le n \le N-1, \\ 0, & otherwise \end{cases} \qquad (2)$$

It provides a convenient representation of the magnitude variation over the time. It is fact that values of energy for the less importance section of cry i.e. silence and etc, are in general significantly smaller than those of the more importance i.e. the real crying. This approach can be used as the measurement to distinguish audible sounds from silence when the signal-to-noise (SNR) ratio is high [4].

2) Windowing: Infant's cry signal is a dynamic signal. In other hand, similar to speech signal, cry signal is non-stationary signal. It is fact that, position and shape of speech organ in adult or cry organ in infants, can change very fast. But infant's cry is more stationary than adult speech, because infants have not ability to control cry organ [8].

In non-stationary signals case, we can not use many tools such as Fourier Transform. For overcoming to this problem we use short time Fourier transform or STFT. In order to use STFT, we must make signal to short windows that we know them as frames. We used Hamming window over signal by short length, and over lapping. The hamming window equation is:

$$W(n) = 0.54 - 0.46\cos(2\pi n/(N-1)) \qquad (3)$$

Hamming window has a soft spectral with less distortion relative to rectangular window, that these properties are very important for cry acoustic analysis.

3) Filtering: The recorded signals may contain unwanted effects, as background noises, echo, etc. Some of these effects can disadvantageously affect the results of the analysis. In the frequency domain the lowest component of the infant's cry is not less than 250-300 Hz, thus using a high-pass filter, with a frequency cut-off at 250 Hz is a good solution to reduce most of the background noise [4].

Based on vocal channel model for speech generation, response of vocal tract is similar to low pass filter. Amplitude of high frequency of cry after passing from vocal tract is attenuate relative to low frequency. In the other side, high frequency components of infant's cry have more importance information, so we use pre-emphasize filter for overcoming to this problem. Usually, a simple high pass FIR filter is used for this purpose [9].

C. Feature extraction

1) Entropy: We can consider entropy as a measure for level of stochastic sense for non-stationary signals. Entropy can be used to find "peakiness" or "flatness" of a probability mass function (PMF). A PMF with sharp peakiness has low entropy and a PMF with flat distribution has high entropy [10].To compute entropy of a spectrum we converted the spectrum into a PMF like function by equation (4).

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i} \quad for\ i = 1\ to\ N \qquad (4)$$

Where x_i is the amplitude of i^{th} frequency component of the spectrum, $x=(x_1,... ,x_N)$ is the PMF of spectrum and N is the number of points in the spectrum of each frame. Entropy of each frame is computed from x by:

$$H = -\sum_{i=1}^{N} x_i \log_2 x_i \qquad (5)$$

Fig. 6 illustrates the entropy distribution graph after entropy analysis in normal infant's cry and infant's cry with pain. According to difference between these two distributions, it is possible that considers the spectral entropy as a feature for distinguishing between two group cries.

2) Mel Filter Bank: By inspiration from human hearing and this fact that human hearing in linear frequency range is not equally sensitive, Mel frequency scale filter bank is defined. Mel Filter Bank is similar to cochlea filter bank in ear. By equation (6) Mel scale frequency can obtain from linear scale frequency.

$$F_{Mel} = 1125\ln(0.0016 F_{Hz} + 1) \qquad (6)$$

This scale from 1 Hz to about 1 KHz changes linear, upper than 1 KHz changes logarithmic. In linear scale area of each bin is 1. A Mel filter bank consist of some successive filters (bins) in Mel scale that band width of each filter is equal and middle of each filter is end point for previous filter and start point for next filter [11].

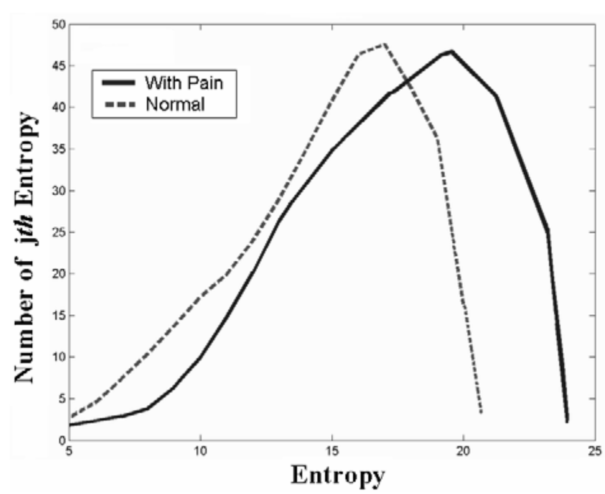

Figure 6. Spectral entropy distribution in normal infant's cry and infants' cry with pain

3) Multi-band Spectral Entropy: The single spectral entropy did not give good results in classification step, which is expressed in IV. section of this paper. In order to obtain better results, we used extraction of multi-band spectral entropy with Mel filter bank.

In fact, by single spectral entropy for total of spectrum we can indicate peakiness or flatness in spectrum, but we can not detect positions of peakiness or flatness in spectrum.

In multi-band spectral entropy analysis, we took FFT of each frame, then after passing from Mel filter bank, we computed spectral entropy of each sub-band (filter) in filter bank, separately, and now, we can detect position of peakiness or flatness in better approximation. In fact, by this method we can obtain entropy distribution in all of spectrum.

D. Artificial Neural Networks (ANN)

A neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements [12].

Connections are quantified by weights, which are dynamically adjusted during training. The required training can be done through the back-propagation technique. During training (or learning), a set of training instances is given. Each training instance, typically, is described by a feature vector. It should be associated with a desired output (a concept, a class), which is encoded as another vector that called the desired output vector.

We studied many learning algorithms and after implementation of them, confirmed the results of papers in [7, 13] and used scaled conjugate gradient (SCG) algorithm for ANN training. From an optimization point of view, learning in a neural network is equivalent to minimizing a global error function, which is a multivariate function that depends on the weights in the network. Many of the training algorithms are based on the gradient descent technique. Minimization is a local iterative process in which an approximation to the function, in a neighbourhood of the current point in the weight space, is minimized. Most of the optimization methods that are used to minimize functions are based on the same strategy. The SCG algorithm [15] denotes the quadratic approximation to the error E in a neighbourhood of a point w by:

$$E_{qw}(y) = E(w) + E^{'}(w)^{T} y + \frac{1}{2} y^{T} E^{''}(w) y \qquad (7)$$

In order to determine the minimum to $E_{qw}(y)$, the critical points for $E_{qw}(y)$ must be found. The critical points are the solution to the linear system defined by Moller in [14]:

$$E^{'}{}_{qw}(y) = E^{''}(w) y + E^{'}(w) = 0 \qquad (8)$$

SCG belongs to the class of Conjugate Gradient Methods, which show super linear convergence on most problems. By using a step size scaling mechanism SCG avoids a time consuming line-search per learning iteration, which makes the algorithm faster than other second order algorithms [13].

IV. RESULTS OF ACR EXPERIMENTS

We implemented our ACR system by cry signals from 23 infants with pain and 17 normal infants. Length of those signals was not equal, because of different situation in signal recording. After silence elimination by energy threshold approach, we segmented each signal to segments with 1 second length. We called each segment name as "cry sample", we obtained 260 cry samples for normal infants and 260 cry samples for infants with pain.

In next step we created 23 frames with 100msec length from each cry sample after windowing by Hamming window which these windows have had 60% of overlapping. These frames were passed from noise rejection filter and then pre-emphasize filter by frequency response H(z). Pre-emphasize function was $H(z)=1-0.95Z^{-1}$.

We took FFT of each frame and then computed spectral entropy of each frame. So we obtained a single feature for each frame of cry sample. Each cry sample has 23 frames, so the final feature vector for each cry sample has 23 elements. The used classifier is MLP neural network. We implemented our programs by MATLAB software [14]. Architecture of used MLP contains 23 input layers, 5 nodes in hidden layer and 2 nodes in output layer and scaled conjugate gradient method was used for network training.

In order to obtain valid results, we used permuted cross validation method in training process. In each time of training, randomly, 80% of data was dedicated to training phase and 20% of data for test phase, in separately. From data set of training phase, again, 80% of data was dedicated to training and 20% was dedicated to validation set. After 10 times training phase and correction rate calculations, we achieved to results in table I.

According to results in table I, it can observe the spectral entropy feature can be considered as convenient feature to classify infant's cry. We obtained 66.9% in mean of correction rate. In order to enhance in results, we took FFT of each frame, then after passing from Mel filter bank with 16 sub-bands, we computed spectral entropy of each sub-band (filter) in filter bank, separately. This coefficient makes a new feature vector that is called Multi band spectral entropy. So we obtained a feature vector with 16 coefficients for each frame of cry sample. Each cry sample has 23 frames, so the final feature vector for each cry sample has 368 elements.

TABLE I. CORRECTION RATE OF ACR SYSTEM BY SPECTRAL ENTROPY EXTRACTION AS SINGLE FEATURE

Cry type	Train Data Set	Test Data Set
With pain	82.7%	71.5%
Normal	69.5%	61.1%
Total Mean	76.8%	**66.9%**

The used classifier is MLP neural network, for training process we need to more cry sample with 368 elements in feature vector. In other side with this feature vector size, computational cost is high. For overcoming to this problem, we used principal components analysis (PCA). By applying the PCA analysis size of feature vector reduced from 368 to 30 elements, so the network can be trained in low computational cost, too.

Architecture of our used MLP has 30 input layers, 5 nodes in hidden layer and 2 nodes in output layer and scaled conjugate gradient method was used for network training. After 10 times training phase and correction rate calculations by permuted cross validation method, we achieved to results in table II. As results in table II, it can be observed, clearly, the multi-band spectral entropy has enhanced the results of single band spectral entropy. We achieved to 88.5% mean of correction rate in this system. So, we can claim the multi-band spectral entropy, are efficient features in classification of infant's cry.

However, researches on features for cry recognition, conversely speech recognition, is not very extended, so for insuring for goodness of this feature it is necessary to compare this feature with other features in future.

V. CONCLUSION AND FUTURE WORKS

Infant's cry is a multimodal and dynamic behaviour that carries a lot of information in its acoustic signal. In this paper, we could recognize infants with pain from normal infants with spectral entropy coefficients extraction. We use multi layer perceptron artificial neural network as classifier. We obtained correction rate (66.9%) in implementation of ACR system. In order to enhance in results, we calculated entropy of each sub-band (filter) in filter bank, separately. This coefficient makes a new feature vector. We achieved correction rate (88.5%) in implementation of ACR system.

The main reason of results enhancement can be this fact that single entropy only gives peakiness or flatness in total of spectrum in cry signal, but multi-band spectral entropy features give distribution of entropy in total spectrum i.e. we can obtain positions of peakiness or flatness in spectrum. In other hand with multi-band entropy we can be able to determine detail of entropy in all over of spectrum, as result the changes of information in spectrum of cry signal is increased.

According to these results and previous works in other papers, we purpose to compare multi-band spectral entropy with good features such as LPCCs, MFCCs, etc. in identical data base [16]. In other side, our results and results of other investigations have not acquired exactly in the same algorithm or software, for example pre-processing may be different in other papers, or databases are different. Finally, we will be able to select the best feature(s) for ACR system to detect infant with pain from normal infant.

TABLE II. CORRECTION RATE OF ACR SYSTEM BY MULTI-BAND SPECTRAL ENTROPY

Cry type	Train Data Set	Test Data Set
With pain	92.3%	86.9%
Normal	97.6%	90.4%
Total Mean	94.7%	**88.5%**

ACKNOWLEDGMENT

We like to thank Dr. Carlos A. Reyes-Garcia, Dr. Emilio Arch-Tirado and his INR-Mexico group, and Dr.Edgar M.Garcia-Tamayo for their dedication of the collection of the Infant's cry data base.

REFERENCES

[1] H. M. Truby and J. Lind, "Cry sounds of the newborn infant", Acta Paediatr. Scand. Suppl., vol. 163, pp. 8-59, 1965.

[2] G.Varallyay Jr., Z. Benyo, A. Illenyi, Z. Farkas, L. Kovacs," Acoustic analysis of the infant cry: classical and new methods", IEEE, EMBS, USA, 2004.

[3] O. Wasz-Hocket, J. Lind, V. Vuorenkoski, T. Partanen, and E. Valanne, The Infant Cry-A Spectrographic and Auditory Analysis, Spastics International Medical Publications in Association with William Heinemann Medical Books Ltd. , 1968.

[4] Gyorgy Varallyay Jr., "Future prospects of the application of the infant cry in the medicine", Periodica Polytechnica Ser. El. Eng. Vol. 50, No. 1–2, PP. 47–62, 2006.

[5] Purhonen, N.–Paakkonen, A.–Ypparila, H.–Lehtonen, J.–Karhu, J., Dynamic Behavior of the Auditory N100 Elicited by a Baby's Cry, Int. J. Psychophysiol., 41/3, pp. 271–278, 2001.

[6] Dror Lederman, "Automatic classification of infants' cry", Ben Gorion University, M.Sc. Thesis P: 1-11, 2002.

[7] Orion F. Reyes-Galaviz, Carlos Alberto Reyes-Garcia, "A System for the Processing of Infant Cry to Recognize Pathologies in Recently Born Babies with Neural Networks", SPECOM: 9thConference Speech and Computer, Russia, 2004.

[8] Kathiresan Manickam, Haizhou Li, "Complexity Analysis of Normal and Deaf Infant Cry Acoustic Waves", MAVEBA: 4th International Workshop on Model and Analysis of Vocal Emission for Biomedical Applications, Italy, 2005.

[9] Qiaobing Xie, Rabab K. Ward, and Charles A. Laszlo, "Automatic Assessment of Infant's Levels-of-Distress from the Cry Signals", IEEE transactions on speech and audio processing, Vol.4, No.4, 1996.

[10] Hemant Misra, Shajith Ikbal, Sunil Sivadas_, Herv´e Bourlard,"Multi-resolution spectral entropy feature for robust ASR", IDIAP Research Institute, Switzerland, IEEE, ICASSP 2005.

[11] Gold, B., Morgan, N. , Speech and Audio Signal Processing and perception of speech and music, John Wiley & Sons, Inc, 2000.

[12] DARPA Neural Network Study, AFCEA International Press, 1988, p.60.

[13] Jose Orozco, Carlos A. Reyes Garcia, "Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks", ESANN'2003 proceedings - European Symposium on Artificial Neural Networks, Belgium, pp. 349-354, 2003.

[14] MathWoks, Inc., Matlab, V. 7.8.

[15] Moller, A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, 6 (4), 1993, pp. 525-533.

[16] Orion F. Reyes-Galaviz, Sergio Daniel Cano-Ortiz, Carlos Alberto Reyes-García, "Evolutionary-Neural System to Classify Infant's cry Units for Pathologies Identification in Recently Born Babies", Proceedings of the Special Session MICAI, pp. 330-335, 2008.